2014

# Confidence intervals for ranks of age-adjusted rates across states or counties

Shunpu Zhang
*University of Nebraska - Lincoln*, szhang3@unl.edu

Jun Luo
*Information Management Services, Inc.*

Li Zhu
*National Cancer Institute*

David G. Stinchcomb
*Westat, Inc.*

Dave Campbell
*Information Management Services, Inc.*

***See next page for additional authors***

Research Article

Statistics in Medicine

Received 4 February 2013, Accepted 22 November 2013 Published online 13 January 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6071

# Confidence intervals for ranks of age-adjusted rates across states or counties

**Shunpu Zhang,[a*†] Jun Luo,[b] Li Zhu,[c] David G. Stinchcomb,[d] Dave Campbell,[b] Ginger Carter,[b] Scott Gilkeson[e] and Eric J. Feuer[c]**

**Health indices provide information to the general public on the health condition of the community. They can also be used to inform the government's policy making, to evaluate the effect of a current policy or healthcare program, or for program planning and priority setting. It is a common practice that the health indices across different geographic units are ranked and the ranks are reported as fixed values. We argue that the ranks should be viewed as random and hence should be accompanied by an indication of precision (i.e., the confidence intervals). A technical difficulty in doing so is how to account for the dependence among the ranks in the construction of confidence intervals. In this paper, we propose a novel Monte Carlo method for constructing the individual and simultaneous confidence intervals of ranks for age-adjusted rates. The proposed method uses as input age-specific counts (of cases of disease or deaths) and their associated populations. We have further extended it to the case in which only the age-adjusted rates and confidence intervals are available. Finally, we demonstrate the proposed method to analyze US age-adjusted cancer incidence rates and mortality rates for cancer and other diseases by states and counties within a state using a website that will be publicly available. The results show that for rare or relatively rare disease (especially at the county level), ranks are essentially meaningless because of their large variability, while for more common disease in larger geographic units, ranks can be effectively utilized. Copyright © 2014 John Wiley & Sons, Ltd.**

**Keywords:** age-adjusted rate; rank; simultaneous confidence interval

## 1. Introduction

Health indices such as age-adjusted rates are summary measures of health and of the factors that affect health for a given population. It is often useful to monitor and compare health indices across different geographic units, population sub-groups, socioeconomic groups, and others. Such information can be used to provide information to the general public on the health condition of the community, to help the government's policy making, to evaluate the effect of a current policy or healthcare program, or for program planning and priority setting.

A commonly used method to extract the information from health indices is ranking the health indices based on the observed data. It is well recognized in the research community that the estimated health indices are random variables with a measure of uncertainty often expressed as a standard error or a confidence interval. However, the ranks obtained from the estimated health indices are often perceived as fixed in the existing scientific literatures, and conclusions are drawn from them on the basis of fixed values. In recent years, it has gradually come to researchers' attention that it is necessary to treat the

[a]*Department of Statistics, University of Nebraska Lincoln, Lincoln, NE 68583, U.S.A.*
[b]*Information Management Services, Inc, 3901 Calverton Blvd, Suite 200, Calverton, MD 20705, U.S.A.*
[c]*Statistical Methodology and Applications Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD 20892, U.S.A.*
[d]*Westat, Inc., Rockville, MD 20850, U.S.A.*
[e]*Independent consultant, Takoma Park, MD 20912, U.S.A.*
**Correspondence to: Shunpu Zhang, Department of Statistics, University of Nebraska Lincoln, Lincoln, NE 68583, U.S.A.*
[†]*E-mail: szhang3@unl.edu*

ranks as random variables and that they should be accompanied by a measure of uncertainty if they are to be useful to the public, patients, or the government. In this regard, Marshall and Spiegelhalter [1] proposed to incorporate the uncertainty associated with the ranks by constructing the 95% confidence intervals for the ranks of the success rates of in vitro fertilization clinics. They proposed in their paper a Monte Carlo simulation method to construct the confidence intervals for the proportions obtained from simple surveys. More recently, Lemmers *et al.* [2] proposed to describe the uncertainty in the ranks by using so-called best-case and worst-case scenarios. The best-case and worst-case scenarios for a geographic unit are obtained by counting the number of other units whose performance is better and worse than this unit, in terms of whether the upper and lower bounds of the confidence intervals of their corresponding odds ratios are greater or less than one. Gerzoff and Williamson [3] examined the uncertainty associated with ranks based on three common methods used to derive public health indicators: age adjustment, calculations based on census estimates, and calculations based on survey data.

In this paper, we propose a Monte Carlo method to construct individual and simultaneous confidence intervals for the ranks of age-adjusted rates of different geographic units. One challenge for constructing the confidence intervals for the ranks of age-adjusted rates is the zero-count problem for a specific age group, which researchers may experience when the populations of some age groups are small or the characteristic of interest is rare. The zero-count problem may cause the Monte Carlo method to produce simulated ranks with variation smaller than the true variation and hence lead to incorrect conclusions. Another challenge is the construction of the simultaneous confidence intervals for all the ranks. Marshall and Spiegelhalter [1] only provided a method for constructing the individual confidence interval for each rank. Although the individual confidence intervals for the ranks can provide information on what the range of the rank for a specific geographic unit might be, it is more typical that one is interested in comparing all the ranks simultaneously. Hence, simultaneous confidence intervals seem to be a more generally applicable choice than the individual confidence intervals. Finally, in some instances, only the age-adjusted rate and confidence interval are available, rather than all of the age-specific rates. We propose an alternative Monte Carlo method when this situation occurs.

We organize the paper as follows. In Section 2, we first propose a general Monte Carlo method for generating the simulated age-adjusted rate distribution and its corresponding rank distribution. Then, we propose a Bayes method to handle the zero-count problem. Further in this section, we present a novel method for constructing the simultaneous rectangular confidence intervals for the ranks. We choose to construct the simultaneous rectangular confidence intervals over the traditional simultaneous confidence region (such as the confidence ellipsoid based on the Hotelling's $T$-square statistic) because the rectangular confidence intervals have the advantage of easier interpretation and graphical representation. Finally, we provide a modified Monte Carlo method when one only has access to the age-adjusted rate and confidence interval. In Section 3, we apply the proposed method to age-adjusted cancer mortality rates and incidence rates by state and county. We demonstrate that more features can be extracted from the simulated rank distribution, which can provide us further insight into the data. We provide conclusions and discussions in Section 4.

## 2. Methods

### *2.1. Age-adjusted rates*

Public health data often need adjustment for confounding factors such as age or other demographic factors. The most common adjustment for public health data is age adjustment because the occurrence of many health conditions is related to age. An age-adjusted rate is a weighted average of the age-specific rates (the number of events divided by the total number of individuals in the population for a specific age group), where the weights are the proportions of persons in the corresponding age group of a standard population. The age-adjusted rates provide a standard way of comparing rates for different population groups. Age adjustment minimizes the effect of a difference in age distributions when comparing rates to allow comparisons between these population groups independent of their age structure. Both direct and indirect methods for age adjustment are possible. Direct age adjustment applies age-specific rates from each area or group to a standard population age profile. Indirect age adjustment is used when only age-specific populations from each area are available, but not age-specific rates. For this paper, we use direct age adjustment, assuming either age-specific rates from each area (Section 2.2) or at least the directly standardized age-specific rate and its confidence interval (Section 2.5) are available.

More specifically, assume that there are $I$ geographic units and $J$ age groups in the study, and the data available are the count $x_{ij}$ of health events and the count $n_{ij}$ of the population size from geographic unit $i$ and age-specific group $j$. The age-specific (crude) rate is the number of health events occurring in a specified population during a year, usually expressed as the number of health events per 100,000 people. It is calculated using the following formula:

$$\text{Age-specific rate of region } i \text{ and age group } j = \frac{x_{ij}}{n_{ij}} \times 100,000. \tag{2.1}$$

Consequently, the age-adjusted rate per 100,000 population for region $i$ is

$$aarate_i = \frac{1}{\sum_{j=1}^{J} stdmil_j} \sum_{j=1}^{J} \left[ \left( \frac{x_{ij}}{n_{ij}} \right) \times 100,000 \times stdmil_j \right] \tag{2.2}$$

where $stdmil_j$ is the population for age group $j$ in a standard population, $j = 1, \ldots, J$. Standard populations can be found at http://seer.cancer.gov/stdpopulations/index.html.

### 2.2. *The Monte Carlo method for generating simulated age-adjusted rates and ranks*

The purpose of the current paper is to propose methods for comparing the ranks of the age-adjusted rates of the $I$ regions to see if the ranks of some regions are significantly different from others. To do so, we need to have the knowledge of the joint probability distribution (the joint rank distribution) of the ranks. If one is also interested in the distribution of the rank of a specific region, it is also necessary to know the marginal distribution of the rank (the marginal rank distribution) of that region. Theoretical derivation of the joint distribution and marginal distribution of the ranks is difficult because of the dependence among the ranks. In this section, we propose to simulate the joint and marginal rank distributions using the Monte Carlo method.

In applications, the observed count $x_{ij}$ is often viewed as a realization from the random variable $X_{ij}$, which has the Poisson distribution $Poisson(\lambda_{ij})$, where $\lambda_{ij}$ is the age-specific rate. Because $X_{ij}$ is the maximum likelihood estimator (as well as an unbiased estimator) of $\lambda_{ij}$ under the Poisson model, we can generate the simulated counts $x_{ij}^k$ for geographic unit $i$ and age-specific group $j$ from $Poisson(x_{ij})$, where $1 \leqslant k \leqslant K$ is the index of simulations. We will discuss the choice of $K$ in Section 2.4. With $x_{ij}^k$ being available, the $k$-th simulated age-adjusted rate for region $i$ can be calculated by using the weighted average method (2.2) with $x_{ij}$ being replaced by $x_{ij}^k$. Denote the $k$-th simulated age-adjusted rate for region $i$ by $aarate_i^k$ and its rank relative to $aarate_1^k, \ldots, aarate_I^k$ by $rank_i^k$, we have the simulated joint distribution of the age-adjusted rates

$$aarate_i^k, \quad i = 1, \ldots, I; k = 1, \ldots, K, \tag{2.3}$$

and its corresponding joint rank distribution

$$rank_i^k, \quad i = 1, \ldots, I; k = 1, \ldots, K. \tag{2.4}$$

For each fixed $i$, $aarate_i^k$, $k = 1, \ldots, K$, provide a sampling distribution of the estimated age-adjusted rate of geographic unit $i$. Therefore, the $100(1-\alpha)\%$ confidence interval for the true age-adjusted rate of unit $i$ can be easily calculated from the $100(\alpha/2)$-th and $100(1-\alpha/2)$-th percentiles of the simulated rates $aarate_i^k$, $k = 1, \ldots, K$. Similarly, the $100(1-\alpha)\%$ confidence interval of the ranks of geographic unit $i$ can be easily obtained by finding the $100(\alpha/2)$-th and $100(1-\alpha/2)$-th percentiles of $rank_i^k$, $k = 1, \ldots, K$.

### 2.3. *Adjusting for zero count using the Bayes method*

A potential problem in the described Monte Carlo method is the use of $x_{ij}$ as the estimator of $\lambda_{ij}$. When the age-specific population is small or the characteristic of interest is rare, the count $x_{ij}$ may be zero, and therefore, $\hat{\lambda}_{ij}$ (the estimator of $\lambda_{ij}$) is also zero, which leads to a degenerate Poisson distribution. This will cause the simulated count for this age-specific group to be zero for every Monte Carlo iteration. Thus, the contribution of this age-specific group to the overall variability in the simulated age-adjusted

rates for this geographic unit will be zero. It is obvious that a count of zero does not necessarily mean that the true age-specific rate is zero, and hence, the existence of the zero count for certain age-specific populations may cause underestimation of the age-adjusted rate, as well as its variability. We call this the zero-count problem. The zero-count problem is worst when all the age-specific rates are zero and therefore the age-adjusted rate is zero as well.

The preceding discussion shows that it is necessary to adjust for the zero-count problem before the Monte Carlo method is conducted. Ideally, we expect that a new method with the zero-count problem corrected satisfies the following two properties: (1) $\hat{\lambda}_{ij}$ should be bounded away from zero when the observed count is zero, and (2) it should reflect the fact that zero count observed from a large age-specific population implies stronger evidence that the true age-specific rate is closer to zero. A natural way that can achieve both properties is to use a Bayes estimator. For developing this Bayes estimator, instead of the Poisson model, it is more convenient to use the binomial model $B(n_{ij}, p_{ij})$, where $p_{ij}$ is the probability of success (a person exhibits the characteristic of interest) for geographic unit $i$ and age-specific group $j$, and satisfies $\lambda_{ij} = n_{ij}\, p_{ij}$.

Note that the maximum likelihood estimator (MLE) of $p_{ij}$ under the binomial model is

$$\hat{p}_{ij} = X_{ij}/n_{ij} \tag{2.5}$$

and the zero-count problem is reflected by the fact that $\hat{p}_{ij} = 0$ when $X_{ij} = 0$. Assume that $p_{ij}$ has the beta prior $Beta(\alpha, \beta)$, that is,

$$\pi(p_{ij}|\alpha, \beta) = \frac{p_{ij}^{\alpha-1}(1-p_{ij})^{\beta-1}}{B(\alpha, \beta)},$$

where $B(\alpha, \beta)$ is the beta function with $\alpha > 0, \beta > 0$. Then, the posterior distribution is

$$f(p_{ij}|x_{ij}) \propto Beta(x + \alpha, n_{ij} - x_{ij} - \beta),$$

and the Bayes estimator of $p_{ij}$ when $x_{ij} = 0$ is

$$\hat{p}_{ij}^{B} = E(p_{ij}|x_{ij} = 0) = \frac{\alpha}{n_{ij} + \alpha + \beta}. \tag{2.6}$$

It can be seen that $\hat{p}_{ij}^{B}$ possesses the two desired properties we discussed previously by noticing the fact that (1) $\hat{p}_{ij}^{B}$ $\left(\text{hence, } \hat{\lambda}_{ij} = n_{ij}\,\hat{p}_{ij}\right)$ is always larger than zero; and (2) $\hat{p}_{ij}^{B}$ satisfies $\hat{p}_{ij}^{B} \downarrow$ as $n_{ij} \uparrow$, that is, zero count from a large population indicates stronger evidence that the true $p_{ij}$ is closer to zero. In contrast, the crude rate (the MLE) $\hat{p}_{ij} = x_{ij}/n_{ij}$ is always zero when $x_{ij} = 0$ regardless of the population size.

The remaining question about (2.6) is how to choose $\alpha$ and $\beta$. One method is to choose $\alpha = \beta = 1$. This is equivalent to the uniform [0, 1] distribution. The uniform prior is sometimes interpreted as 'non-informative' because it implies that we are 'ignorant' of $p_{ij}$, except that it can be anywhere between 0 and 1. Unfortunately, the uniform prior is not invariant under re-parameterization. Let $\gamma_{ij} = \log(p_{ij})$. It was shown in [4] that $\gamma_{ij}$ has the exponential distribution; hence, the uniform prior is not as 'non-informative' as it seems. Jeffreys [5] proposed an alternative prior that is invariant under re-parameterization. The Jeffreys' prior for the binomial model is $Beta(\alpha = 0.5, \beta = 0.5)$. It can be seen from (2.6) that the Bayes estimator of $p_{ij}$ with respect to the uniform prior and the Jeffreys' prior are

$$\hat{p}_{ij}^{1} = E(p_{ij}|x_{ij} = 0) = \frac{1}{n_{ij} + 2} \tag{2.7}$$

and

$$\hat{p}_{ij}^{2} = E(p_{ij}|x_{ij} = 0) = \frac{0.5}{n_{ij} + 1}, \tag{2.8}$$

respectively.

Because the observed count is zero, we do have some prior information on $p_{ij}$: it should be close to zero and is most likely less than $1/n_{ij}$ for geographic unit $i$ and age group $j$, that is, the majority of the probability mass of the prior distribution of $p_{ij}$ is within $(0, 1/n_{ij})$. Otherwise, because

$\hat{p}_{ij} \to p_{ij}$ when $n_{ij}$ goes to $\infty$, the observed count $X_{ij}(= n_{ij}\hat{p}_{ij})$ will be greater than 1, that is, one would mostly likely have observed a count of one or more. A distribution with such property is $Beta(\alpha = 1/n_{ij}, \beta = 1 - 1/n_{ij})$. It can be shown that for $n_{ij} \geqslant 100$, at least 95% of the probability mass of $Beta(\alpha = 1/n_{ij}, \beta = 1 - 1/n_{ij})$ is concentrated between $(0, 1/n_{ij})$. With this as the prior distribution, the Bayes estimator of $p_{ij}$ is

$$\hat{p}_{ij}^{B} = E(p_{ij}|x_{ij} = 0) = \frac{1/n_{ij}}{n_{ij} + 1}. \tag{2.9}$$

We propose this method for use when the observed count is zero for any specific age group. For convenience, we will simply call $\hat{p}_{ij}^{B}$ the Bayes estimator in the remainder of the paper.

### *2.4. Construction of simultaneous confidence intervals for ranks*

As discussed in Section 2.1, a 95% confidence interval for the rank of a specific geographic unit can easily be constructed from the simulated marginal distribution (2.4). Such intervals can be used to see where a specific geographic unit stands relative to other geographic units included in the study. However, most users of ranks are not interested in one specific unit, but rather in looking at the ranks and the associated confidence intervals of all the units. One cannot examine a large set of 95% individual confidence intervals and still maintain that level of coverage over the entire set of units examined. This brings up the issue on how to construct the confidence intervals for the ranks of all the regions simultaneously.

One immediate answer to this is to use the Hotelling's $T^2$ test. However, the Hotelling's $T^2$ test only provides a simultaneous confidence region (ellipsoid), which is often difficult for non-statisticians to comprehend. Another drawback of the Hotelling's $T^2$ test is that it requires the independence of the test statistics being compared. An alternative method that does not require the independence assumption and has easy interpretation is the Bonferroni method. The Bonferroni method provides a confidence interval for each individual rank and therefore can be interpreted as the usual confidence interval in one-dimensional case. The confidence region formed by these individual confidence intervals is called a rectangular confidence region. However, the drawback of the Bonferroni method is that it can be very conservative, especially when the number of comparisons is large. Other attempts on developing rectangular confidence region have been focused on multivariate normal distribution and can be traced back to Dunn [6, 7]. Šidàk [8] showed that a confidence region constructed for the case of independent normal random variables is always conservative for any case of dependent multivariate normal variables. More recent attempts on constructing simultaneous rectangular confidence region include the Simes test [9], the Simes–Hochberg test [10], and the Hommel test [11], among others. Nevertheless, these methods require that the test statistics are independent of each other, which is not true for our case because the ranks are dependent on each other.

In the following, we propose a Monte Carlo method to construct the simultaneous rectangular confidence region at a prescribed level, which does not require the assumption of independence. To construct the simultaneous rectangular region, basically we need to find $[r_i, s_i]$ such that

$$P(r_1 \leqslant rank_1 \leqslant s_1, \ldots, r_I \leqslant rank_I \leqslant s_I) = 1 - \alpha, \tag{2.10}$$

where $i = 1, \ldots, I$. Equation (2.10) has infinite many solutions if there are no constraints imposed on the individual confidence intervals. We propose to fix each confidence interval $[r_i, s_i]$ at the same confidence level, say $\beta$, that is, all pairs of $(r_i, s_i)$ should satisfy

$$P(rank_i < r_i) = P(rank_i > s_i) = (1 - \beta)/2, \tag{2.11}$$

for $i = 1, \ldots, I$. Equation (2.11) implies that the probability of the individual event $r_i \leqslant rank_i \leqslant s_i$ is $\beta$. On the basis of (2.10), the joint probability of $r_i \leqslant rank_i \leqslant s_i$ for all $i$'s is $1 - \alpha$. Because the probability of joint (intersection) events is no larger than the individual events by definition, we need to search for $\beta$ in the range $1 - \alpha$ to 1. The value of $\beta$ is numerically searched on the basis of the simulated joint rank distribution (2.4) so that (2.10) holds. Because of the discrete nature of ranks, it is in general impossible to find $\beta$ satisfying (2.10) exactly. Instead, we will start with the two endpoints $1 - \alpha$ and 1 as possible values for $\beta$ and search inwards using a bisection algorithm for $\beta$ and its corresponding $[r_i, s_i]$'s so that the left-hand side of (2.10) is as close as possible to $1 - \alpha$.

## The algorithm

To search for $\beta$ in the range $1-\alpha$ to 1, we start with choosing an initial pair of $\beta$ values $\beta_1 = 1, \beta_2 = 1-\alpha$ for the probability of individual event defined in (2.11). For $\beta_1 = 1$, we first find $(r_i^1, s_i^1)$ from the simulated rank distribution (2.4) that satisfies the condition for the individual probability (2.11)

$$P(rank_i < r_i^1) = P\left(rank_i > s_i^1\right) = (1-\beta_1)/2, \quad i = 1, \ldots, I,$$

and

$$P\left(r_1^1 \leqslant rank_1 \leqslant s_1^1, \ldots, r_I^1 \leqslant rank_I \leqslant s_I^1\right) = \beta_1^* \geqslant 1-\alpha$$

When $\beta_1 = 1, \quad (1-\beta_1)/2 = 0$, then $\left(r_i^1, s_i^1\right)$ is just the range of all the simulated ranks. Hence, the joint probability $P\left(r_1^1 \leqslant rank_1 \leqslant s_1^1, \ldots, r_I^1 \leqslant rank_I \leqslant s_I^1\right) = 1$.

Similarly, when $\beta_2 = 1-\alpha$, $(1-\beta_2)/2 = \alpha/2$. We know it is true that

$$P\left(rank_i < r_i^2\right) = P\left(rank_i > s_i^2\right) = (1-\beta_2)/2 = \alpha/2, i = 1, \ldots, I,$$

and

$$P\left(r_1^2 \leqslant rank_1 \leqslant s_1^2, \ldots, r_I^2 \leqslant rank_I \leqslant s_I^2\right) = \beta_2 \leqslant 1-\alpha,$$

because the joint confidence level is lower than the individual confidence level.

The aforementioned choices of $\beta_1$ and $\beta_2$ guarantee that the value of $\beta$, which satisfies (2.10), is between $\beta_1$ and $\beta_2$. With these initial values available, we can find the desired value of $\beta$ using the following algorithm:

Step 1. Calculate $\beta_0 = (\beta_1 + \beta_2)/2$ and find its corresponding $r_i^0, s_i^0$ such that

$$P\left(r_i^0 \leqslant rank_i \leqslant s_i^0\right) = \beta_0 \quad \text{for} \quad i = 1, \ldots, I.$$

Then, use the simulated joint rank distribution (2.4) to find the probability

$$P\left(r_1^0 \leqslant rank_1 \leqslant s_1^0, \ldots, r_I^0 \leqslant rank_I \leqslant s_I^0\right) \hat{=} \beta_0^*.$$

Step 2. If $\beta_0^* < 1-\alpha$, set the new pair of $\beta$ values $(\beta_1, \beta_2) = (\beta_1, \beta_0)$; otherwise, set $(\beta_1, \beta_2) = (\beta_0, \beta_2)$.

Repeat steps 1 and 2 until no improvement can be made. In practice, the algorithm is often terminated when the improvement is less than a certain pre-specified threshold. In our algorithm, the threshold is set at 1/number of simulation.

Three considerations must be kept in mind when using this algorithm:

(1) The aforementioned algorithm tries to find rectangular intervals that have confidence levels as close as possible to the pre-specified level $1-\alpha$. Hence, the resulting confidence intervals may have simultaneous levels higher than $1-\alpha$ if the algorithm is stopped at $\beta_1$ and lower than $1-\alpha$ if the algorithm is stopped at $\beta_2$. If one wishes to only have simultaneous levels to be at least $1-\alpha$, he or she should make sure the algorithm is stopped at $\beta_1$.

(2) The success of the aforementioned algorithm also depends on the number of simulations, $K$. A too small value of $K$ does not provide simulated ranks fine enough for the algorithm to find distinguishable $[r_i, s_i], i = 1, \ldots, I$, while a too large value of $K$ may make the algorithm computationally infeasible. In the Appendix, we show that, under the independence assumption of ranks, a lower bound of $K$ as a function of $I$ is $2/(1-(1-\alpha)^{1/I})$. Because the ranks are dependent on each other, this lower bound may not accurately reflect the true number of simulations needed to achieve the prescribed confidence level and hence can only be used as a rule of thumb for deciding $K$, the number of simulations.

(3) The same algorithm can be applied to the simulated joint age-adjusted rate distribution provided by (2.3) to obtain the simultaneous confidence intervals for the age-adjusted rates. However, in practice, it is not necessary to use simulations to obtain the simultaneous confidence intervals for the rates. Because the age-adjusted rates from different regions are assumed to be independent, one can directly use the Šidàk method to construct the simultaneous confidence intervals.

### 2.5. *Using a Gamma approximation when age-specific rates are not available*

In Equation (2.2), an age-adjusted rate is a weighted average of the age-specific rates, where the weights are the proportions of persons in the corresponding age groups of a standard population. The method in Section 2.2 assumes a binomial or Poisson distribution for deaths (or incidence) in each age group and applies the Monte Carlo method to estimate the confidence intervals of ranks. There are situations when age-specific rates are not available to calculate age-adjusted rates. Public health organizations often only provide disease surveillance data in the form of point and interval estimates of age-adjusted incidence rates, and data on age-specific incidence rates are often not available. For example, see the State Cancer Profiles website (http://statecancerprofiles.cancer.gov/). This section develops a method to estimate confidence intervals for ranks of age-adjusted rates based on the point and interval estimates of age-adjusted rates.

Assume the disease counts in a geographic region and the age groups are independent Poisson random variables. Then, the age-adjusted rate in that region is a weighted average of these Poisson random variables. Fay and Feuer [12] developed a confidence interval of an age-adjusted rate based on the gamma approximation. Specifically, suppose that the age-adjusted rate is composed of age groups 1 to $J$, and for each geographic region $i$, define

$$w_{ij} = \frac{stdmil_j}{n_{ij} \times \sum_{j=1}^{J} stdmil_j} \times 100,000$$

where $n_{ij}$ is the age-specific population of geographic region $i$ and age group $j$, and $stdmil_j$ is the standard million of the US population, $j = 1, \ldots, J$.

Fay and Feuer [12] specified that if

$$w_{im} = \max(w_{ij})$$

$$z_i = w_{im}^2$$

and

$$v_i = \sum_{j=1}^{J} \left(w_{ij}^2 \times x_{ij}\right)$$

where $x_{ij}$ is the count of disease or death in geographic region $i$ and age group $j$, then the age-adjusted rate of region $i$ $\quad aarate_i$ can be approximated by a gamma distribution with mean $aarate_i$ and variance $v_i$. Tiwari *et al.* [13] further modified the gamma approximation and developed more efficient estimates of confidence intervals, and these modified gamma confidence intervals have been taken as the default confidence interval estimates of age-adjusted rates in SEER*Stat (the program used to compute cancer rates for the US Surveillance Epidemiology and End Results (SEER) cancer registries). The modification is specified as

$$w_{im} = average(w_{ij})$$

$$z_i = average\left(w_{ij}^2\right)$$

and the lower and upper endpoints of a $(1-\alpha) \times 100\%$ confidence interval are calculated as

$$\begin{aligned} CI_{low,i} &= \left(\frac{v_i}{2 \times aarate_i}\right) \times \left(ChiInv\left(\frac{\alpha}{2}, \frac{(2 \times aarate_i^2}{v_i}\right)\right) \times 100,000 \\ CI_{high,i} &= \left(\frac{v_i + z_i}{2(aarate_i + w_{im}}\right) \times \left(ChiInv\left(1 - \frac{\alpha}{2}, \frac{2(aarate_i + w_{im})^2}{v_i + w_{im}^2}\right)\right) \times 100,000 \end{aligned} \tag{2.12}$$

where $aarate_i$ is the age-adjusted rate in region $i$, and $ChiInv$ denotes the cumulative distribution function of Chi-square distribution at probability $1 - \frac{\alpha}{2}$ and degrees of freedom $\frac{2(aarate_i + w_{im})^2}{v_i + w_{im}^2}$.

Age-adjusted rate $aarate_i$ can be viewed approximately as a gamma random variable with mean $aarate_i$ and variance $v_i$. When only the age-adjusted rate, $aarate_i$ and its confidence interval $CI_{low,i}$ and $CI_{high,i}$ are available, we can derive $v_i$ from (2.12). Because $w_{ij}$ can be calculated from the standard millions and the population in region $i$ and age group $j$, $w_{im}$ and $z_i$ are functions of $w_{ij}$ and can be calculated too. The only missing piece in calculating the lower and upper boundaries is $v_i$. Because $v_i$ does not have a closed-form solution from (2.12), we need to solve for it using an iterative algorithm in the following steps:

Step 1: Pick an arbitrary pair of numbers $v'$ and $v''$ such that $v' < v''$.
Step 2: Calculate $v$ as the average of $v'$ and $v''$.
Step 3: Using $v$ in (2.12) as $v_i$, calculate $CI_{low,i}$. If the calculated $CI_{low,i}$ is larger than the observed lower limit of the 95% confidence interval, then set $v' = v$, otherwise $v'' = v$.
Step 4: Repeat steps 2 and 3 until the calculated and observed values of $CI_{low,i}$ become equal to each other. This value of $v$ is identified as the optimal value of $v_i$ based on $CI_{low,i}$.

Similarly, we can solve for $v_i$ on the basis of the upper limit of the confidence interval, $CI_{high,i}$. The two solutions of $v_i$ will be slightly different, but the difference does not have an impact on the simulation of $aarate_i$ and hence its rank calculation, so either could be used. Now, with both $aarate_i$ and $v_i$ available, we can simulate age-adjusted rate $aarate_i$ as a gamma random variable with mean $aarate_i$ and variance $v_i$. Confidence intervals for the ranking of age-adjusted rates can be generated with the same Monte Carlo method as described in Sections 2.2–2.4.

Defining the length as the difference between upper and lower limits of the 95% confidence intervals, Figure 1 compares the length of 95% confidence intervals of ranks for age-adjusted rates of all cancer mortality in Maryland counties for the period of 2004–2008, simulated for 10,000 times using the Gamma method with the length of the 95% confidence intervals of the same ranks using the gold standard method in Sections 2.2–2.4. Because the gold standard is based on the assumption that the age-specific counts are realizations from the Poisson distributions, we refer to it as the Poisson method in this section. It can be seen that the lengths of the 95% confidence intervals of ranks generated from the two methods are identical for most counties in this example, with the exception of two counties whose lengths of the 95% confidence intervals are differed by 1 for the two methods. We also observed in the simulation results that the counties with the same length of the 95% confidence intervals have the same endpoints for the two methods. This figure demonstrates that the Gamma method can be used as a satisfactory approximation to the Poisson method when age-specific rates are not available.

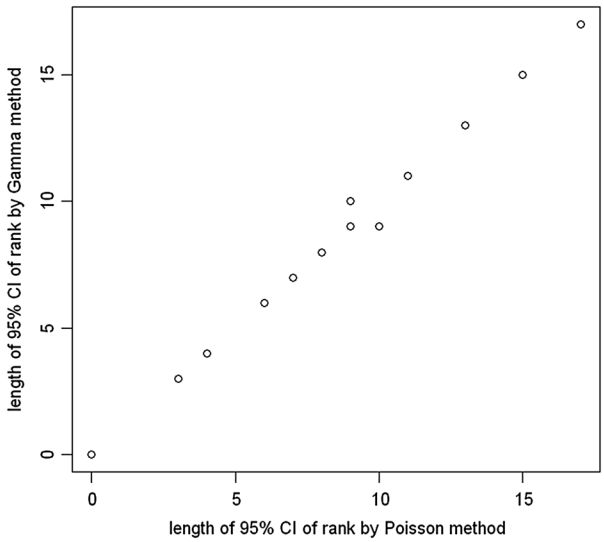


**Figure 1.** Length (upper limit minus lower limit) of the 95% confidence intervals of ranks for all cancer age-adjusted mortality rates of Maryland counties, 2004–2008, comparing results for the Gamma method versus the Poisson method.

## 3. Examples

In this section, we will demonstrate the proposed methods using US age-adjusted cancer incidence and mortality rates. US mortality data, collected and maintained by the National Center for Health Statistics (NCHS) from the National Vital Statistics System, are a fundamental source of demographic and geographic details of cause-of-death information. Current confidentiality restrictions for NCHS vital statistics data require suppression of counts of less than 10 deaths or births. At the national and state level, single-year data are available as long as the death counts are no less than 10. At the county level, NCHS requires a minimum of 3 years of data are aggregated to be published. We generally recommend aggregating a minimum of 5 years at the county level to provide more stability. The age-adjusted mortality rate provides a standard way of comparing cancer rates for different cancers and in different geographic regions. Such rates are often used in public health activities to describe cancer trends and to characterize the cancer burden in a standardized manner.

Age-adjusted rates and confidence intervals for incidence data at the state and county level were obtained from the State Cancer Profiles website (http://statecancerprofiles.cancer.gov). Age-adjusted rates are suppressed if the counts are less than 16 over a 5-year period. The underlying data for these rates come from the Centers for Disease Control and Prevention's National Program of Cancer Registries Cancer Surveillance System and from the National Cancer Institute's SEER Program (http://www.SEER.cancer.gov).

Figure 2 displays the histograms of the ranks (the rank histograms obtained from the simulated rank distribution (2.4)) of the mortality rate of black female breast cancer for Macon County, Alabama, 2003–2007. This histogram can take up to 66 values, because there are 66 counties in Alabama. The top panel was obtained on the basis of the MLE with no zero-count correction (2.5), and the bottom panel used the Bayes estimator for zero-count correction (2.9). Macon County is a relatively small county in Alabama with a population of 10,500 so the counts from many of the age groups are zero. The blue lines (the individual 95% confidence intervals) on the two panels show that the confidence interval obtained based on the MLE (without adjustment for zero count) is narrower (the top panel) than that based on the Bayes estimator (with the adjustment for zero count, the bottom panel).

The four panels in Figures 3 and 4 show examples of state and county simultaneous confidence intervals of the ranks of mortalities and incidence rates based on the Bayes estimator (2.9). Figure 3a shows the 2007 mortality rates of all causes of death at the state level. It shows a clear difference among the state mortality rates with relatively narrow confidence intervals. West Virginia has the highest ranked

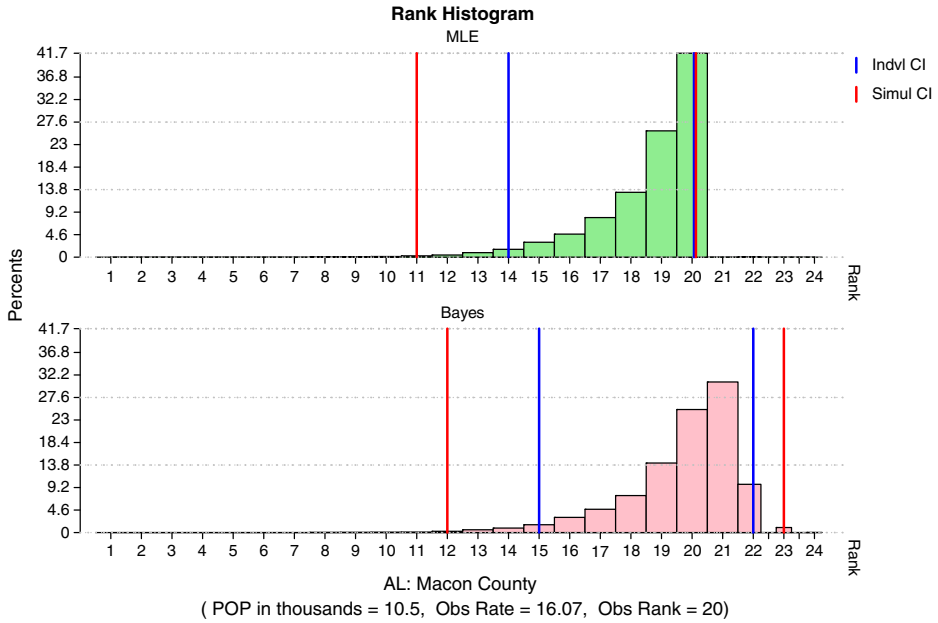


**Figure 2.** Individual and simultaneous confidence intervals based on the maximum likelihood estimator (MLE; top) and the Bayes method (bottom) for black female breast cancer mortality in Macon County, Alabama, 2003–2007.

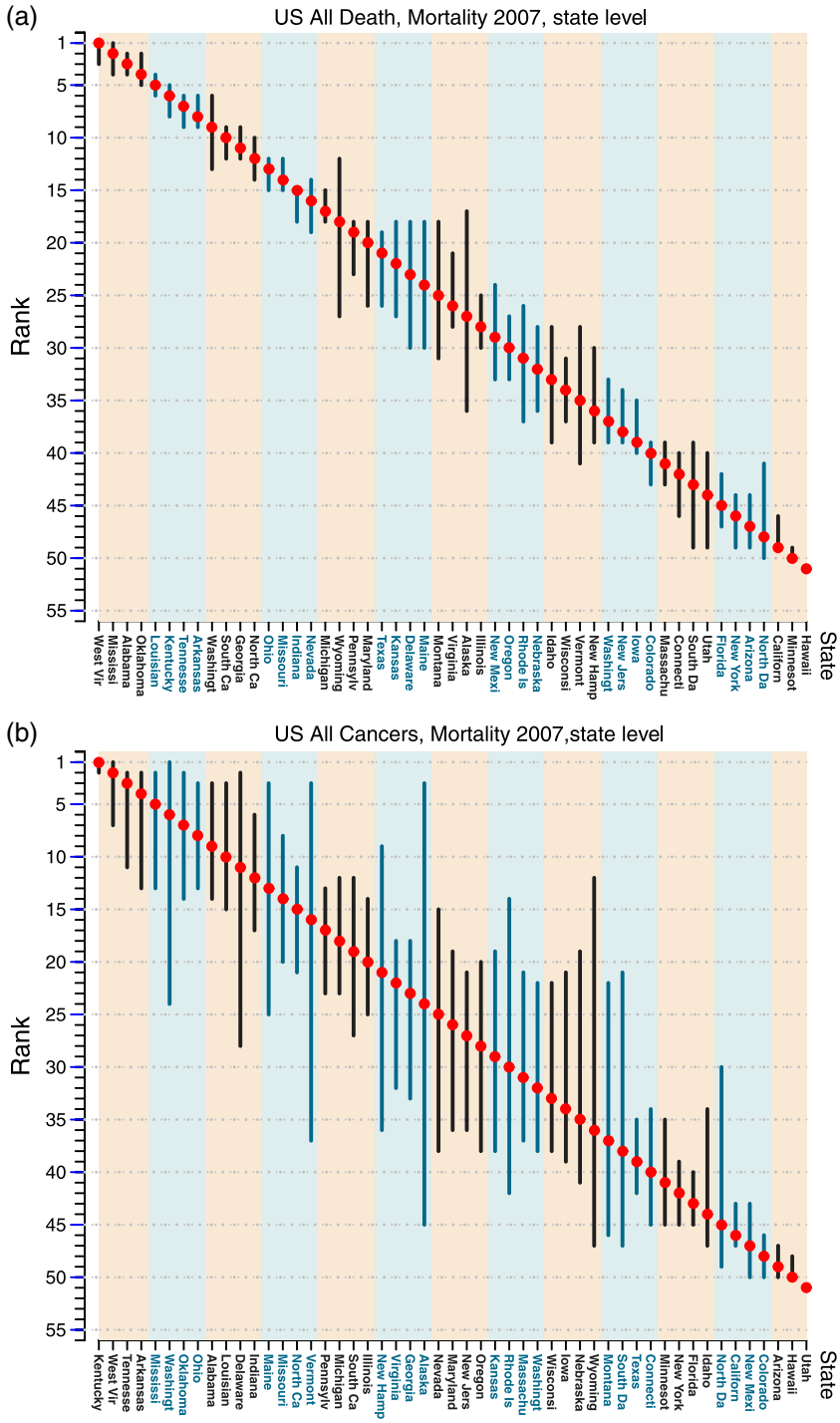


**Figure 3.** Examples of the simultaneous 95% confidence intervals of the ranks (using Bayes adjustment for zero counts) of mortality rates at the state level.

observed value (i.e., has the highest mortality rate), and Hawaii has the lowest ranked observed value (i.e., has the lowest mortality rate). Figure 3b shows the 2007 mortality rates from all malignant cancers at the state level. This panel has wider confidence intervals and shows that Kentucky has the highest ranked observed value and Utah has the lowest ranked observed value because of their respective high and low smoking rates, while West Virginia and Hawaii have the second to the highest and the second to the lowest ranked observed values, respectively. The wider rank confidence intervals in Figure 3b than in Figure 3a shows that there is less variation between the states in the mortality rates for all causes than the mortality rates for all cancers.

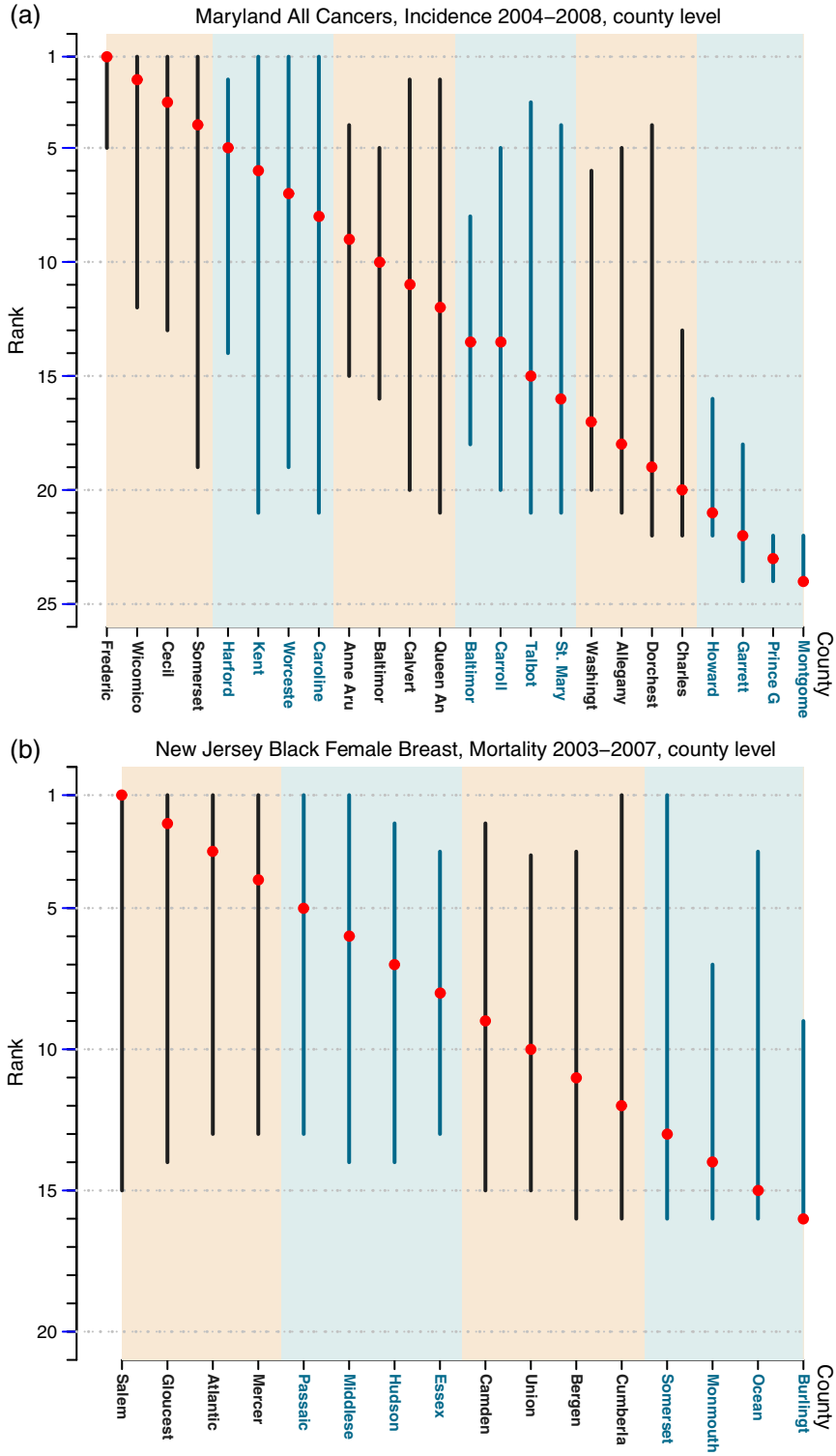


**Figure 4.** Examples of the simultaneous 95% confidence intervals of the ranks (using Bayes adjustment for zero counts) of incidence and mortality rates at the county level.

Figure 4a shows the 2004–2008 all cancer incidence rates of the counties in Maryland. This panel shows that Frederick County has the highest ranked observed cancer incidence rate in the state. While there has been concern expressing that Frederick County is the highest ranked county in Maryland in terms of its cancer incidence rates, Figure 4a indicates that it could have a rank between 1 and 5 with 95% of confidence. Figure 4b displays the breast cancer mortality rate for black women in

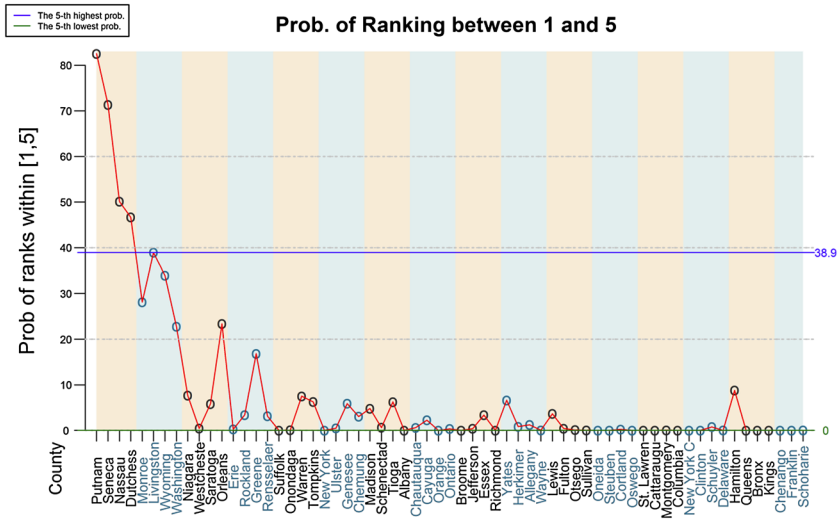


**Figure 5.** Probability of New York counties being ranked top 5 for female breast cancer incidence rates, 2005–2009.

New Jersey. This panel shows that because of the variability of the underlying rates, none of the counties in New Jersey is significantly different from others in breast cancer mortality rate for black women, and therefore, ranking these rates cannot provide meaningful results.

In addition to the individual and simultaneous confidence intervals, the joint rank distribution (2.4) also provides the probabilities that a region takes different ranks. Such probabilities can be displayed by using a form of a probability histogram (or a rank histogram), as we have seen in Figure 2. The rank histogram can provide information such as what the most likely rank of a region is or which group of regions has the highest probability to be ranked in the top. The latter may have important implications in government policy making. For example, it is often desirable to identify the regions with the highest cancer incidence or mortality rates in order to allocate scarce resources, or to focus public health improvement efforts. Traditionally, this is carried out by finding the regions with the highest observed ranks. As we have seen, such a method ignores the inherent variability in the ranks, and the conclusion may not reflect the true ranks of the regions. Instead, we propose the concept of the probability of being ranked in the top for each region. Figure 5 shows the probability of being in the top 5 ranks for female breast cancer incidence rates for the counties of New York during 2005–2009. This figure shows that the fifth county by the observed rank is Monroe. However, in the five counties with the highest probability of being in the top 5, Monroe is replaced by Livingston. This type of graph helps to identify these issues for a given cut point and also helps visualize possible alternative cut points for further evaluation.

## 4. Conclusions and discussions

In this paper, we have proposed a Monte Carlo method for comparing the ranks of age-adjusted mortality rates of different geographic units based on the availability of the age-specific counts. We have also demonstrated the necessity of comparing the ranks simultaneously, and an efficient algorithm for constructing the simultaneous confidence intervals of ranks is proposed. We have further extended the method to the case in which only the age-adjusted rates and their associated confidence intervals on the rates are available. An online Web application called CI*Rank for performing these calculations is available to the public at http://surveillance.cancer.gov/cirank/.

In some instances, especially in states with many counties with small populations, the confidence intervals for the ranks will be very wide. In this instance, the counties in a state might be combined to produce more stable estimates. For example, the 120 counties in Kentucky are sometimes combined into 15 area development districts. In other instances, the group of counties to be compared will cross state boundaries. For example, one might be interested in ranking mortality or cancer incidence rates from counties across 13 states that comprise Appalachia (http://www.arc.gov/counties). At the suggestion of users, we have added the ability to suggest special groupings of counties to be added to the online

Web application. After they are added, these special groups suggested by our user base will be available to all users.

There are several possible extensions to the proposed methodology. First, to facilitate the application of the proposed method, we are considering the possibility to add an online tool on the website where users can upload and analyze their own data. Second, the age-adjusted rates are calculated assuming the regions to be ranked are independent from each other. A natural extension of our proposed method is to incorporate spatial autocorrelation in the estimation of age-adjusted rates as well as in the calculation of the confidence intervals using the well-known spatial adjacency network method [14]. Lastly, although our proposed method is only demonstrated for the age-adjusted rates, it can be modified for other health indices such as the proportion of smokers from a national survey such as the Behavioral Risk Factor Surveillance System (http://www.cdc.gov/brfss/) or from a small area estimation model (http://sae.cancer.gov/). In the case of direct estimates from a survey, the Monte Carlo simulation would probably use a normal approximation for the distribution of the statistic, with the standard error coming from the complex survey. Another similar problem to ranking incidence or mortality rates across geographic areas is ranking the different diseases rates within a single geographic area. For example, one might be interested in knowing the ranking of causes of death in California. The methods proposed here are not applicable when comparing the age-adjusted rates within the same geographic unit due to the dependence between the age-adjusted rates because they share the same denominator. In such a case, a simulation method could be developed so it can generate simulated age-adjusted rates under the same dependence structure as that of the original data before our method for constructing the simultaneous rectangular confidence intervals can be used.

## Appendix

Under the assumption that all ranks are independent of each other, direct application of the results of Šidàk [8] shows that the individual confidence level $1-\beta$ on the rank of a region for simultaneous confidence level $1-\alpha$ satisfies the following equation:

$$P(r_1 \leqslant rank_1 \leqslant s_1, \ldots, r_I \leqslant rank_I \leqslant s_I) = (1-\beta)^I = 1-\alpha.$$

Hence, $\beta = 1-(1-\alpha)^{1/I}$. Given that $K$ simulated ranks are generated for each region, this means that $K$ has to be at least $2/\beta = 2/(1-(1-\alpha)^{1/I})$ in order to delete at least one simulated rank from each end of the simulated rank distribution of a region. Hence, the minimal number of simulations is $2/(1-(1-\alpha)^{1/I})$.

## Acknowledgements

The authors would like to thank the associate editor and two referees for their valuable comments and suggestions, which led to significant improvement of the paper.

## References

1. Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 1998; **316**:1701–1705.
2. Lemmers O, Broeders M, Verbeek A, Heeten G, Holland R, Borm GF. League tables of breast cancer screening units: worst-case and best-case scenario ratings helped in exposing real differences between performance ratings. *Journal of Medical Screening* 2009; **16**(2):67–72.
3. Gerzoff RB, Williamson GD. Who's number one? The impact of variability on rankings based on public health indicators. *Public Health Reports* 2001; **116**(2):158–64.
4. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*, 3$^{rd}$ edn. CRC/Chapman and Hall: Boca Raton, 2009.
5. Jeffreys H. *Theory of Probability*, 3$^{rd}$ edn. University Press: Oxford, 1961.
6. Dunn OJ. Estimation of the means of dependent variables. *Annals of Mathematical Statistics* 1958; **29**:1095–1111.
7. Dunn OJ. Multiple comparisons among means. *Journal of the American Statistical Association* 1961; **56**:52–64.
8. Šidàk Z. Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association* 1967; **62**:626–633.
9. Simes R. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986; **73**:751–754.

10. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; **75**:800–802.
11. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988; **75**:383–386.
12. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine* 1997; **16**:791–801.
13. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Statistical Methods in Medical Research* 2006; **15**(6):547–69.
14. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics (AISM)* 1991; **43**:1–59.