

A nonparametric approach for determining significance of county cancer rates compared to the overall state rate: illustrated with Minnesota data

Alan P. Bender · Allan N. Williams ·
John Soler · Margee Brown

Received: 7 June 2011 / Accepted: 12 February 2012 / Published online: 11 April 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract

Background The study of the geographical distribution of disease has expanded greatly with GIS technology and its application to increasingly available public health data. The emergence of this technology has increased the challenges for public health practitioners to provide meaningful interpretations for county-based state cancer maps.

Methods One of these challenges—spurious inferences about the significance of differences between county and overall state cancer rates—can be addressed through a non-parametric statistical method. The Wilcoxon's signed rank test (WSRT) has a practical application for determining the significance of county cancer rates compared to the statewide rate. This extension of the WSRT, developed by John Tukey, forms the basis for constructing a single confidence interval for all differences in county and state directly age-adjusted cancer rates. Empirical evaluation of this WSRT application was conducted using Minnesota cancer incidence data.

Results The WSRT procedure reduced the impact of statistical artifacts that are frequently encountered with standard normal significance testing of the difference between directly age-adjusted county and the overall state cancer rates.

Conclusion Although further assessment of its performance is required, the WSRT procedure appears to be a useful complement for mapping directly age-adjusted state cancer rates by county.

Keywords Cancer · Counties · Significance · Surveillance

Introduction

The county as a geographic basis for describing state cancer incidence

For many years, investigators have evaluated geographic patterns of cancer occurrence by county to generate etiologic hypotheses [1–3]. For example, observations of higher lung cancer and mesothelioma rates in the shipbuilding areas on the eastern seaboard of the United States contributed to an understanding of the carcinogenic potential of asbestos [4]. Cancer rates in counties with nuclear power plants have been compared to similar counties without plants to address public concerns [5]. In Minnesota, the observation of a large and persistent increase in mesothelioma incidence in the north-eastern region of the state led to the discovery that taconite miners were at high risk of the disease [6].

Evaluation of cancer rates by county for etiologic hypothesis development and testing has limitations [7, 8]. These evaluations are ecologic analyses [9] that are based on aggregated exposure and demographic data for larger areas and frequently diverse populations. The emergence of the analytic capabilities of geographic information systems (GIS) has led many investigators to study geographical variations of cancer occurrence for geographic units smaller than the county [10].

Displaying cancer incidence by county remains a valuable tool for describing the occurrence of cancer to the general public. County cancer rates are also important to community health assessments and allocation of public health resources [11]. County administrators and local public health officials are keenly interested in knowing how their county's cancer rates compare to the state and whether any differences are significant. A wide range of audiences identify with county-based cancer rates, and for

A. P. Bender (✉) · A. N. Williams · J. Soler · M. Brown
Minnesota Department of Health, 85 East 7th Place,
PO Box 64882, St. Paul, MN 55164-0882, USA
e-mail: alan.bender@state.mn.us

the purpose of communicating with these audiences, the county is often the geographic unit of choice [12, 13].

Challenges of disease mapping

In 1976, the eminent statistician Tukey [14] described the future challenge for disease mapping as the requirement to inform and not mislead through the creation of what he called a “propaganda device.” Today, the development of GIS has expanded the capacity to map large databases of health outcomes at a high level of spatial specificity. These data can now be linked to other environmental information and population risk factors [15]. The challenges articulated by Tukey 35 years ago continue to apply to the spatial measurements and statistics of GIS analysis today [16].

The appropriate methodology to ensure that maps of health statistics inform and not mislead is the subject of much discussion and research [17, 18]. Statistical inaccuracies often arise from mapping differences between county and the overall state directly age-adjusted cancer rates. Inaccurate specification of significance when multiple counties and the state are compared, and inaccuracy of estimates used in calculating standard normal tests of significance can lead to spurious inferences about the number of statistically significant differences between county and overall state cancer rates.

Multiple comparisons

Developing a map representing differences between county rates and the overall state rate involves many different geographic comparisons (eighty-seven in Minnesota). A table of eighty-seven differences provides the opportunity for simultaneously evaluating eighty-seven separate hypotheses. If there is a 5% chance that each of the individual confidence intervals does not contain the true difference in rates, the probability that all eighty-seven confidence intervals contain the true difference is considerably less than 95%. The distinction between a single statistical test of hypothesis that the county rate and state rate are not different and a test of all differences is referred to as the multiple inference or multiple comparison problem [19].

When the number of comparisons is large, the potential for error in the inferences increases. In Minnesota, the probability that at least one of the differences will be found to be statistically significant (at the 0.05 level) is $1 - (0.95)^{87} = 0.99$ [20]. Comparing many county rates creates a markedly greater probability of generating positive findings due to chance than does the stated 0.05 level.

The problem of multiple comparisons applies to cancer mapping as well. The graphical representation is based on the tabular values of differences between county and

overall state rates. If the comparisons are mapped based on a single comparison test of statistical significance, then eighty-seven different comparisons are implied but not distinctly stated by the map itself. The high likelihood that one or more of the “statistically” different county rates is due to chance is integrated into the global representation of the map.

Errors in estimation

The standard normal statistical test for the difference between two directly age-standardized cancer rates can be approximated by a formula that assumes that the covariance of the two rates is zero, that is, the two rates are independent. The estimate for the variance of the difference between two directly age-adjusted cancer rates is the sum of each cancer rate’s variance estimate [21]. If the independence assumption is violated, a more accurate representation for the variance of the difference in rates is the sum of the two variances minus twice their covariance [22]. As the number of tests of significance increases, the possibility of major violations of the underlying assumption of independence also increases.

The assumption of independence is frequently violated when one of the rates is for a populous county. Cancer rates for populous counties are often positively correlated with the overall state rate since the county has a relatively large impact on the state rate. Assuming that the covariance is zero, when it is numerically positive, creates a loss of statistical power (the confidence intervals are spuriously large). Removing the county before calculating the overall state rate would mitigate this dependence to some degree but would result in a different overall state rate for each county—an undesirable outcome.

A negative covariance can occur for sparsely populated rural counties. For example, if populous urban counties had a higher smoking prevalence (compared to some non-urban counties) as in Minnesota, then the overall state rate for smoking-related cancers would tend to be higher and the cancer rates for counties with lower smoking prevalence would be relatively lower for these cancers. The result of assuming that the numerically negative covariance is zero would be an over specification of significance (spuriously narrow confidence intervals).

This paper describes a method to evaluate the significance of county cancer incidence rates compared to the overall state rate that is more resistant to the artifacts created by these statistical inaccuracies than are the frequently used standard normal tests of significance. This method is complementary to the graphical presentation of county cancer rates employed by states and larger governmental units when specification of significance is also desired.

Methods

Tukey’s modification of the Wilcoxon’s signed rank test (WSRT) is a nonparametric procedure. A detailed description of the theoretical basis of the large sample approximation for the distribution-free confidence interval employed here is given in the text [23]. This text is still considered by many statisticians as a standard for applied nonparametric statistics [24].

The WSRT procedure as applied here uses a measure of the average of all possible boundaries of the m (number of counties) 95% standard normal confidence intervals. Six steps based on Tukey’s extension of the WSRT procedure are proposed to calculate a joint confidence interval for all differences between directly age-adjusted county cancer rates and the overall state rate.

1. Form m 95% standard normal confidence intervals using the Keyfitz formula [25] for the standard errors of the directly age-adjusted rates creating the m lower and upper confidence limits: (CIL_i, CIU_i) ; $i = 1, m$.

Let $diff_i = c \text{ rate}_i - s \text{ rate}$; $i = 1, m$ (c is the county, s is the state), and $se_i =$ standard error of $c \text{ rate}_i$, is calculated as follows, for a given sex and 18 age groups, $se_i^2 = \sum_{j=1}^{18} (N_j/P_j^2)(W_j)^2$; where N_j is the number of specific cancers (e.g., lung) in the i th county, P_j is the population of the j th age group for the given sex of the i th county, W_j is the US 2000 standard population weight for the j th age group, $se =$ standard error of $s \text{ rate}$, (calculated with total state data for N_j and P_j) and, the standard error of the difference between the county and state rate, $se(diff_i) = (se_c^2 + se_s^2)^{1/2}$, with the 95% standard normal confidence interval = $diff_i \pm 1.96 se(diff_i) = (CIL_i, CIU_i)$ [21].

2. Form the $M = \frac{m(m+1)}{2}$ averages of the lower and upper bounds of the m 95% standard normal confidence intervals [23]. These are called Walsh (W) averages.

For $m = 87$ (Minnesota), $M = \frac{m(m+1)}{2} = 3,828$,

$$W_L^k = (CIL_i + CIL_j)/2,$$

$$W_U^k = (CIU_i + CIU_j)/2; \quad i \leq j, \quad k = 1, 2, \dots, 3,828.$$

3. Rank (sort) both lower and upper sets of Walsh averages, so that:

$$W_L^1 \leq W_L^2 \leq \dots \leq W_L^M,$$

$$W_U^1 \leq W_U^2 \leq \dots \leq W_U^M.$$

4. The $1 - \alpha$ confidence interval for the lower limit is $(W_L^{C\alpha}, W_L^{(M+1-C\alpha)})$, and the $1 - \alpha$ interval for the

upper limit is $(W_U^{C\alpha}, W_U^{(M+1-C\alpha)})$, where $C\alpha = \frac{m(m+1)}{4} - Z_{(\alpha/2)} \left[\frac{m(m+1)(2m+1)}{24} \right]^{1/2}$.

($C\alpha$ rounded to the nearest integer determines the two elements of the ordered arrays of Walsh averages that serve as end points of the joint confidence interval, α is usually 0.05).

5. For $i = 1, m$ use $(W_L^{C\alpha}, W_U^{(M+1-C\alpha)})$ as the joint confidence interval for all the $diff_i$.
 - 6a. If the (CIL_i, CIU_i) does not contain zero, the significance of the $diff_i$ is then determined by the joint confidence interval derived in 5).
 - 6b. If the $diff_i$ lies outside the joint confidence interval, it is concluded that the county rate is different than the state rate. For $m = 87$ counties and $Z_{(\alpha/2)} = 1.96$, $C\alpha = 1,451$, $M + 1 - C\alpha = 2,378$; and 5) becomes (W_L^{1451}, W_U^{2378}) .

A detailed example of the calculations required for the application of this nonparametric procedure described by steps (1)–(6b) is given in “Appendix 1.” This algorithm produces a joint (single) confidence interval for all differences between the county and state rates, for a given sex and cancer. The 95% standard normal confidence intervals, (CIL_i, CIU_i) , are specific to each county for the null hypothesis H_0 : $diff_i$ (county rate $_i$ —state rate; for a given sex and cancer) = 0.

The purpose of the analysis is to identify differences in cancer rates that lie outside (CIL_i, CIU_i) whose statistical significance is less likely due to multiple comparisons and possible lack of the independence of the two rates. Therefore, the application of step 6b) requires that step 6a) first be satisfied. Any difference where zero lies inside the standard normal 95% confidence interval is judged as nonsignificant and is not compared to the joint interval $(W_L^{C\alpha}, W_U^{(M+1-C\alpha)})$.

The number of years, age groups, number, and types of cancer were varied, and the performance of the WSRT procedure was evaluated for 300 separate analyses of different combinations of these variables for Minnesota and its eighty-seven counties. The performance was evaluated by the empirical method of comparing inferences derived from the WSRT procedure to those derived from the 95 and 99% standard normal tests and the 90% Bonferroni test of significance (the Bonferroni test is based on the standard normal test of significance with the significance level adjusted to reduce the number of false positives) [26]. Specificity of the WSRT procedure was assessed by determining if an elevated number of significant differences were reduced. Sensitivity was evaluated by determining which of the discrepant inferences derived from the

three tests and the WSRT procedure were valid, based on a detailed examination of additional data.

The data used for these comparisons were from the Minnesota Cancer Surveillance System [27]. The MCSS is a population-based statewide cancer registry that has been in operation since 1988. Several examples of these analyses are provided to illustrate the performance of the WSRT procedure to this application. Results of the WSRT procedure are labeled in the parlance suggested by Tukey as, “unusually low,” “not unusual,” and “unusually high” [14].

Results

Figure 1 is a map of the 95% standard normal significance tests of the difference between the directly age-standardized,

county-specific female lung cancer incidence rates for 1988–2007 and the corresponding overall Minnesota rate [21, 25]. The number of counties with significantly low lung cancer rates is the dominant feature of this map.

Figure 2 contains the WSRT procedure alternative to the map provided by Fig. 1. The difference between Figs. 1 and 2 is the method used to determine which of the county rates were different than the overall state rate. The WSRT procedure determination of significance generated a map (Fig. 2) in which the number of counties with significantly different rates was substantially reduced.

The National Cancer Institute’s SEER*Stat software provides additional parametric methods, based on rate ratios, for determining statistical significance of county age-adjusted cancer rates compared to the overall state rate [28]. The results of the SEER*Stat determinations of

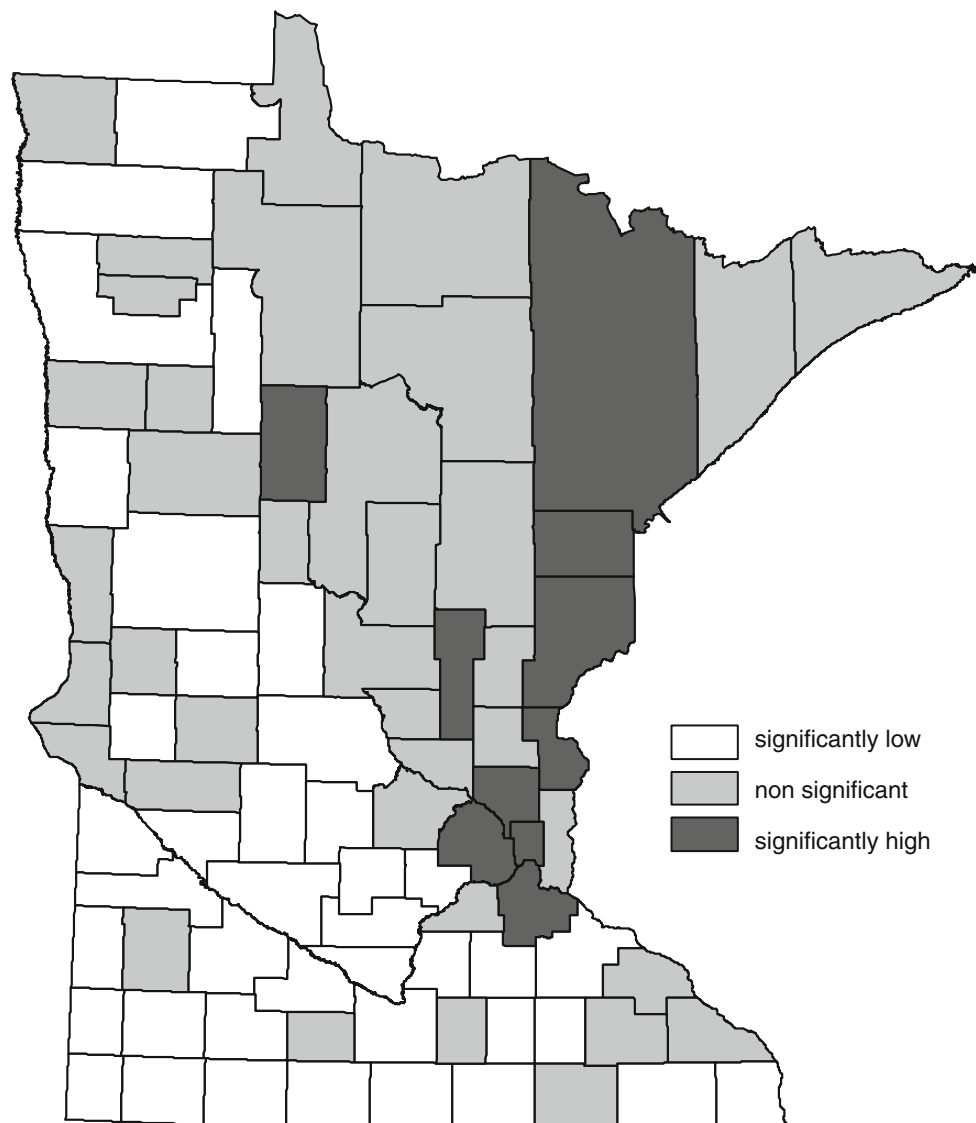


Fig. 1 Age-adjusted female lung cancer incidence compared to overall state female age-adjusted rate by county (1988–2007). Significance determined by 95% standard normal confidence interval for difference in the two rates

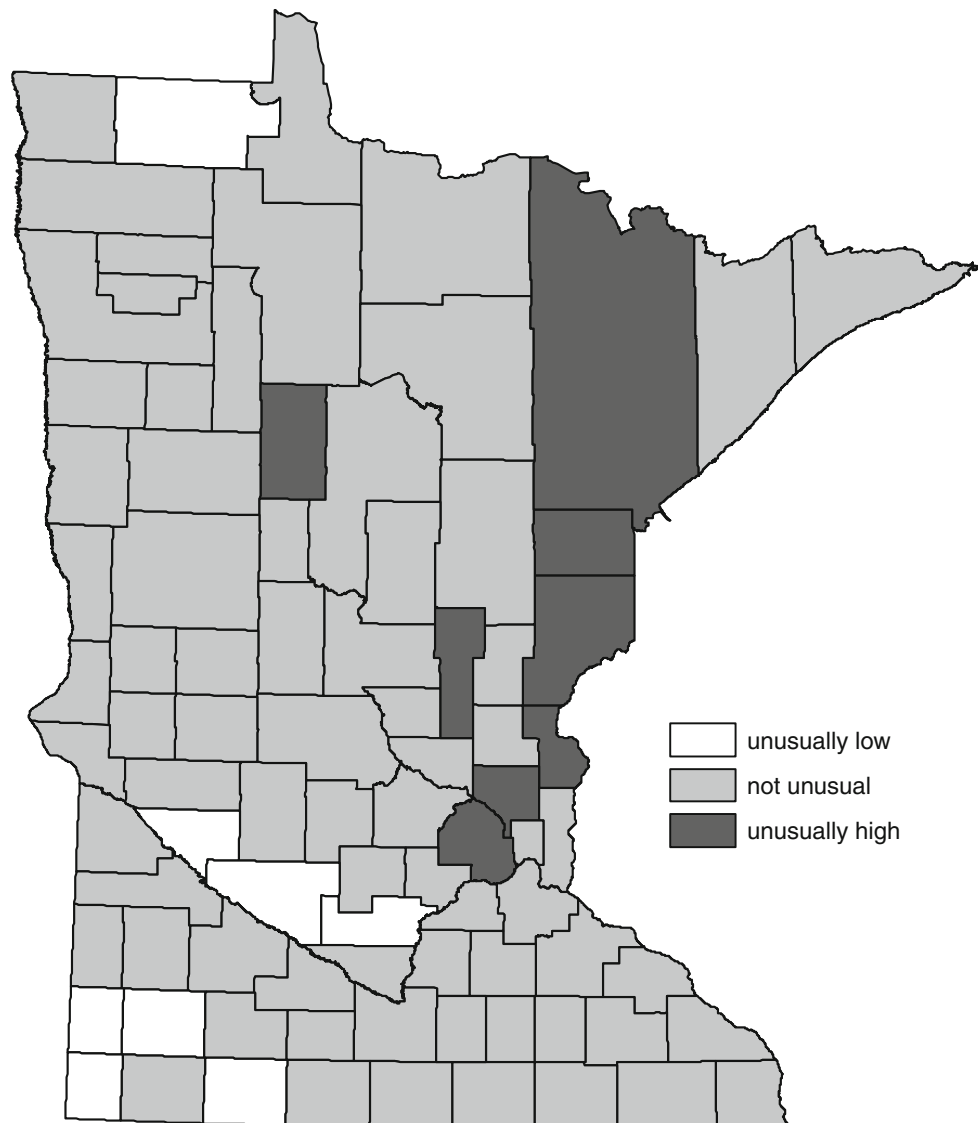


Fig. 2 Age-adjusted female lung cancer incidence compared to overall state female age-adjusted rate by county (1988–2007). Significance determined by Wilcoxon’s signed rank test (WSRT) procedure described in text

significance, and the standard normal methods evaluated in this report, for the data depicted in Figs. 1 and 2 are given in Table 1. Results of the standard normal tests and the SEER*Stat rate ratio tests of significance produced similar results. For relatively large counts, this similarity is expected [29, 30] and these methods would result in a map similar to Fig. 1.

Table 2 contains a summary for all MCSS data, 32 cancer types (“Appendix 2”), 20 years (1988–2007), and all age groups that represented 429,794 cancers. Biologically, inconsistent cancer and gender combinations were excluded. There were five of these. The number of total comparisons for the 1988–2007 aggregated data was 5,133 ($27 \times 87 \times 2 + 5 \times 87$).

The 95% standard normal test resulted in 882 (17.2% of 5,133) differences between county and overall state rates

being classified as significantly lower than zero and 170 (3.3%) as significantly larger than zero. In other words, there were 882 county rates significantly lower than the overall state rate and 170 significantly higher. This large proportion of significantly low rates is a measure of the spuriously significant differences that can result when comparing many county rates to the overall state rate and when the independence of the county rates and the overall state rate are an issue. Another indication of this artifact is provided by the 170 comparisons that were significantly high. Only 128 (2.5%) of the rates were expected to be significantly high.

Results of the 99% standard normal test demonstrated the expected overall reduced numbers of statistically significant results. However, there was still evidence of spuriously significant differences, most notably for low rates

Table 1 Comparison of significance assessments of age-adjusted female lung cancer incidence compared to overall state female age-adjusted rate by county (Minnesota, 1988–2007)

Significantly low						Significantly high						Nonsignificant					
SN95	SN99	B90	T95	F95	WSRT	SN95	SN99	B90	T95	F95	WSRT	SN95	SN99	B90	T95	F95	WSRT
<i>Summary counts for the 87 counties</i>																	
40	25	18	35	34	8	10	7	4	10	10	8	37	55	65	42	43	71

Standard Normal 95% CI (SN95), Standard Normal 99% CI (SN99), Bonferroni 90% CI (B90), Tiwari 95% CI (T95) [30], Fay 95% CI (F95) [29], and WSRT 95% CI (WSRT) [Wilcoxon's signed rank test—see text]

Table 2 Number (%) of statistically significant county rates for four tests of significance—32 cancers, 87 counties, and 2 sexes

Standard normal (95%)		Standard normal (99%)		Bonferroni (90%)		WSRT ^a (95%)	
Low	High	Low	High	Low	High	Low	High
882 (17.2)	170 (3.3)	616 (12.0)	69 (1.3)	488 (9.5)	34 (0.7)	130 (2.5)	76 (1.5)

Total comparisons = 5,133. Biologically inconsistent cancer and gender combinations excluded

MCSS 1988–2007 [All ages ($n = 429,794$ total cancers)]

^a Wilcoxon's signed rank test—see text

(12% significant). The 90% Bonferroni test, designed to reduce the number of false positives, was more conservative, in that, it severely reduced the number of significantly elevated rates to 0.7%. The Bonferroni test was not as effective in mitigating the impact of low rates; 488 (9.5%) remained significant.

Results from the WSRT procedure provided a different perspective. Only 2.5% of the county rates were significantly low and 1.5% were high. The number of significantly low rates was reduced, and the sensitivity for detecting elevated county rates did not appear to be greatly impaired. The 1.5% of the rates considered significantly high exceeded that estimate for both the 99% standard normal and the 90% Bonferroni tests. The only result that was larger (3.3%) was for the 95% standard normal test and that number was likely skewed too large by the multiple comparisons (5,133) that were made. For these data, the WSRT procedure greatly reduced the number of apparently spurious results generated by the three standard normal tests of significance.

Figures 3 and 4 represent the comparison of the female age-adjusted all-cancer incidence rate to the state rate for the years 1988–2007 by county. Determination of significance in Fig. 3 was by the 95% standard normal test and in Fig. 4 by the WSRT procedure. The difference in impressions of significance created by eliminating statistical artifact afforded by the WSRT procedure compared to the standard normal tests is seen again in these two figures. Throughout this evaluation, reduction in the number of counties with cancer rates statistically different than the overall state rate was a prominent feature of the WSRT procedure.

Large numbers

Precision of the estimates for age-adjusted rates mapped in Figs. 3 and 4 was relatively high. These rates included all female cancers, for all years, which incorporated the largest numbers available. The problem of estimating rates from a small number of occurrences, or small populations, is well documented [31]. Large numbers represent another problem [32] that frequently results in specification of statistical significance for very small differences in rates.

Figures 5 and 6 represent comparisons of the incidence of a common male cancer (colon/rectum) to the overall state rate for the years 1988–2007 by county. Results of the 95% standard normal significance determinations yielded 6 (6.9%) counties with low rates and 14 (16.1%) with significantly higher age-adjusted rates than the overall Minnesota rate. The WSRT analysis yielded 2 (2.3%) and 5 (5.7%) of the counties with unusually lower and higher rates, respectively, than the cancer rate for the state as a whole. For this example, the WSRT procedure reduced the number of significant outcomes by nearly two-thirds. The WSRT procedure consistently reduced the number of significant inferences for differences between county and overall state rates when the comparisons involved relatively common cancers (large numbers).

Small numbers

For less common cancers, there were many zero occurrences in a county during the time period (1988–2007). Of the 882 (Table 2) county rates significantly lower than the corresponding state rate, 358 of these were for zero rates.

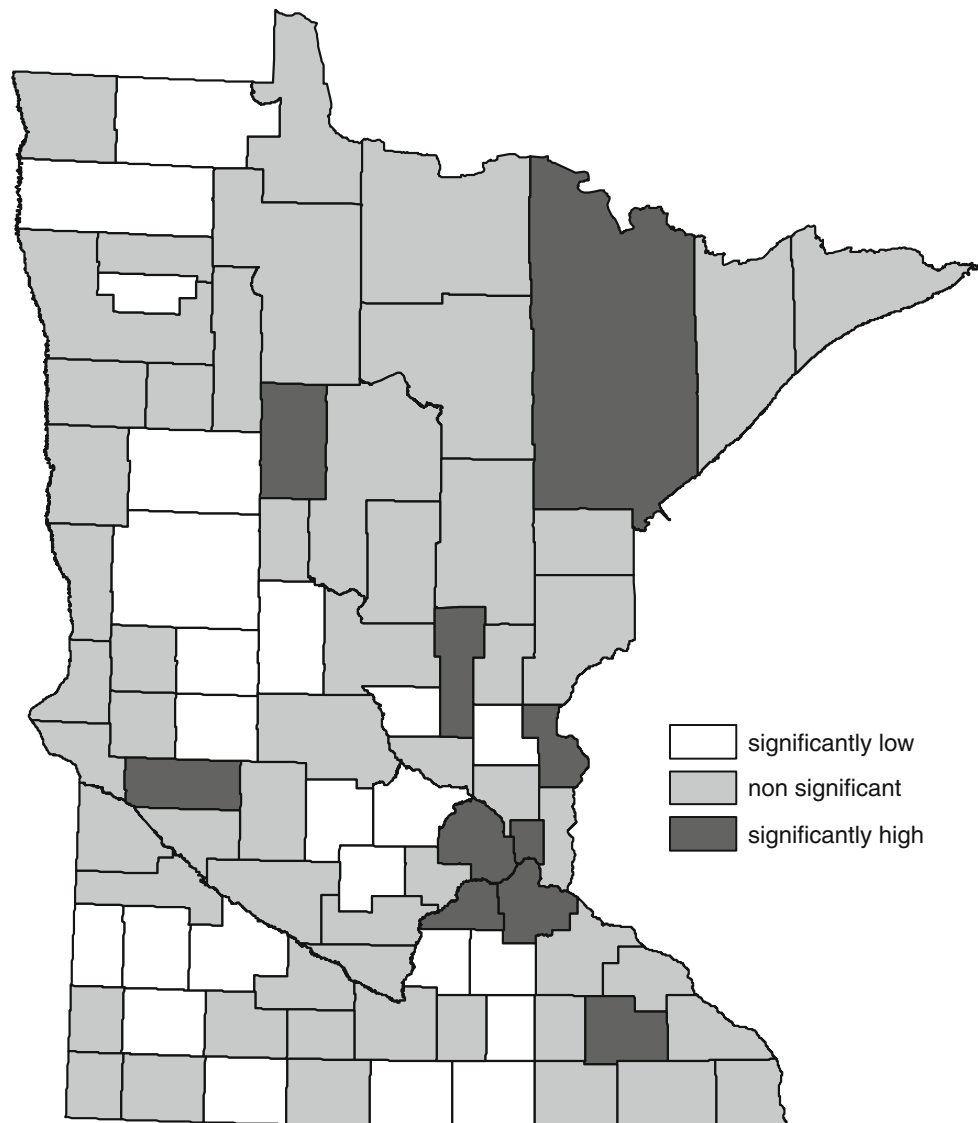


Fig. 3 Age-adjusted female all-cancer incidence compared to overall state female age-adjusted rate by county (1988–2007). Significance determined by 95% standard normal confidence interval for difference in the two rates

Of these 358, 339 were considered nonsignificant deviations from the overall state rate by the WSRT procedure; only 19 of 130 low rates were derived from zero county rates. The relative insensitivity to the impact of zero occurrences (and very small cell size in general) is a known behavior of the WSRT procedure [23] that is important to reducing small number artifact that complicates comparison of individual county cancer rates to the overall state rate.

Table 3 contains similar data as Table 2 except the years included were 2003–2007. The total number of cancers was reduced from 429,794 to 121,908 and the number of zero occurrences in the counties increased. The effect of this increase was to sharply elevate the number of significantly low rates derived from the standard normal

tests. The number of significantly low rates identified by the WSRT procedure decreased from 130 (2.5%) to 82 (1.6%). This decrease was a desirable outcome. As the precision of the estimates decreased (with less nonzero data), the number of occurrences considered significantly low should also decrease. As seen with the more numerous data from Table 2, the sensitivity of the WSRT procedure for elevated rates, determined by the percentage of comparisons that were significantly high, was between that of the 95% standard normal test and the other two methods.

Table 4 contains data for childhood (ages 0–14 years) cancers for all years combined (1988–2007). This analysis was based on a substantially smaller number of cancers (3,188) than Table 3. Since all 32 cancer types were analyzed separately, there were a very large number of zero

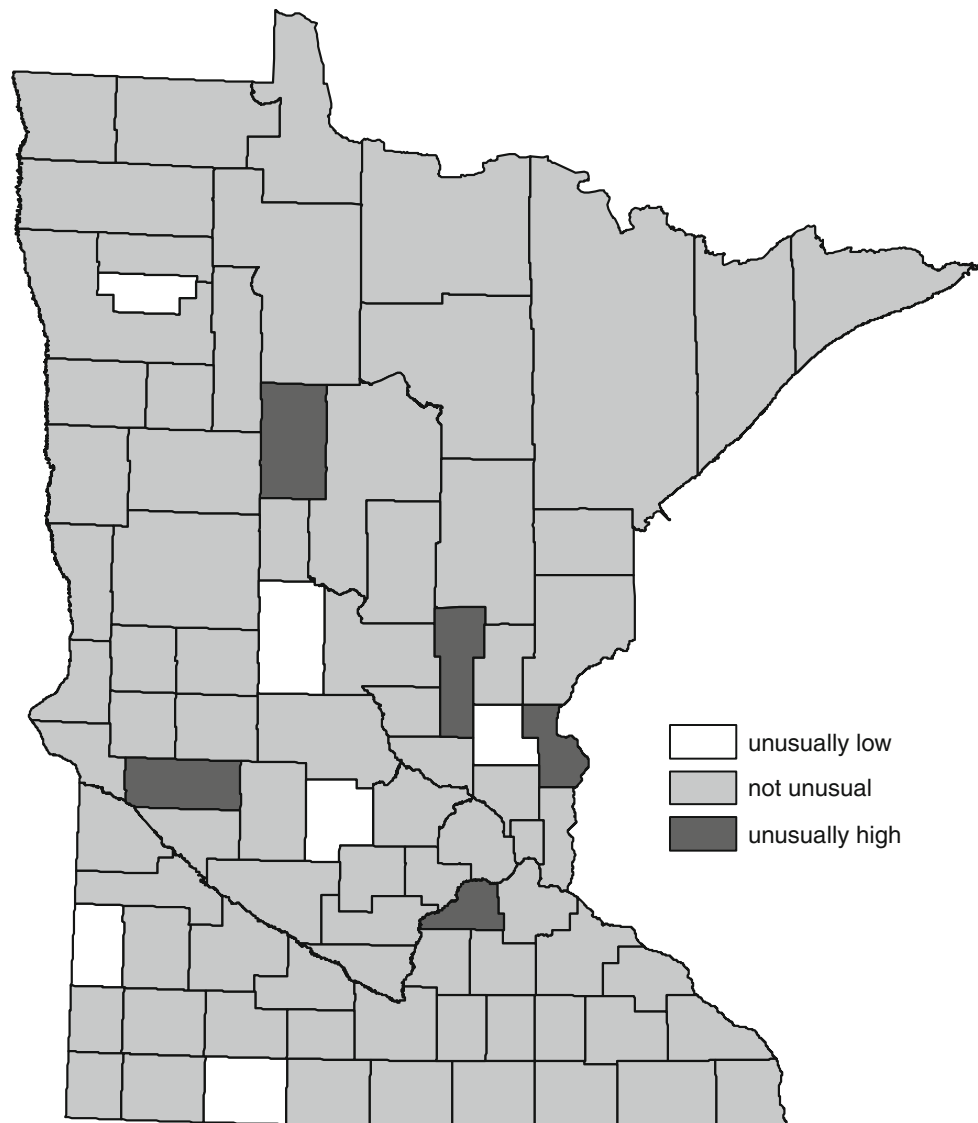


Fig. 4 Age-adjusted female all-cancer incidence compared to overall state female age-adjusted rate by county (1988–2007). Significance determined by Wilcoxon’s signed rank test (WSRT) procedure described in text

occurrences even for larger counties. This created an extreme situation for data populated with zero or a very small number of occurrences. The performance of the WSRT procedure in reducing the number of spuriously low rates was even more evident in this extreme situation.

Example: bone cancer

For the period 2003–2007, there were 168 bone cancers diagnosed in males and 113 in females of all ages in Minnesota. The overall state male standard normal 95% confidence interval (CI) for the incidence rate per 100,000 per year was (1.1, 1.6) and the female 95% CI was (0.7, 1.0). By this measure, any county with a zero occurrence had a statistically significant low bone cancer incidence.

A zero occurrence resulted in a zero estimate for the county standard error that reduced the 95% CI for the difference between the county and the overall state rate to that of the 95% CI for just the state rate.

There were 31 counties with zero occurrences for males and 50 counties with zero occurrences for females. Thus, 81 (46.6%) of the 174 comparisons were considered as significantly low by the 95% standard normal test of significance. None of the counties had significantly elevated rates. (Another 59 counties had one or two occurrences, all of them not statistically different than the overall state rate). The results for the 99% standard normal and the 90% Bonferroni significance tests were identical to the 95% standard normal test; 81 (46.6%) were significantly low, and none were significantly elevated.

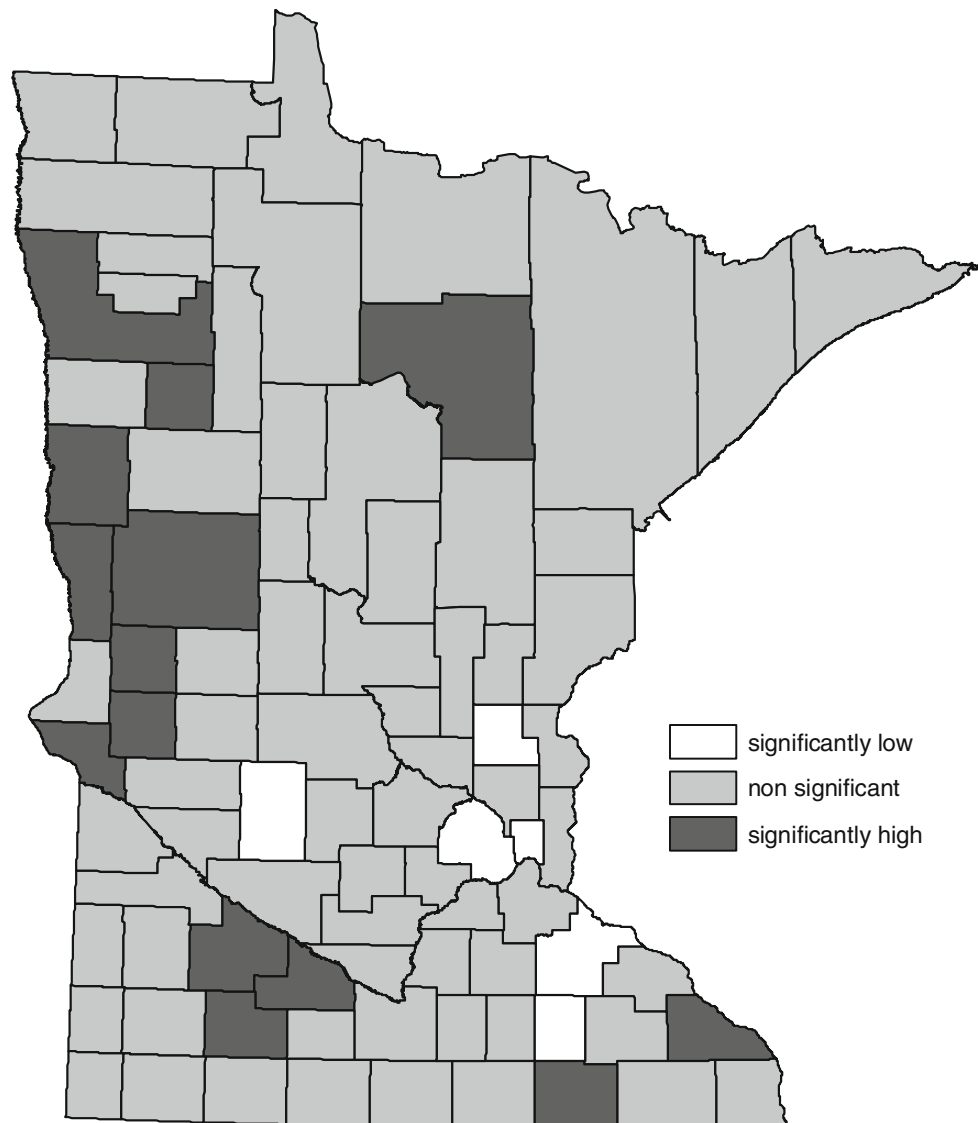


Fig. 5 Age-adjusted male colon/rectum cancer incidence compared to overall state male age-adjusted rate by county (1988–2007). Significance determined by 95% standard normal confidence interval for difference in the two rates

These evaluations could be used in two ways to create a cancer map. Maps (for males and females separately) portraying almost half of the state's counties as having significantly low rates of bone cancer could be created. Alternatively, a decision could be made that, except for a small number of larger counties, the rates were too unstable to warrant mapping cancer rates by county.

Both of these approaches would be problematic. The first set of maps would create a misrepresentation of significance. The second approach would not take advantage of the county-specific data on cancer incidence and the composite information available from these observations.

The WSRT procedure identified none of the county bone cancer incidence rates as unusually low or unusually high. Utilizing all the data on bone cancer incidence, there was

no evidence of an unusual rate of this cancer for the period 2003–2007 in any of the counties. Two maps could be created that would provide the impression that there were no unusual occurrences of male or female bone cancer incidence for the period 2003–2007 in any of the counties. This would be the most appropriate representation of the MCSS data and the one supported by the WSRT procedure.

Sensitivity

Sensitivity for detecting unusual occurrences of cancer, especially for childhood cancers, is important. In Table 4, there are three occurrences that were statistically elevated based on the 95% standard normal significance test, but

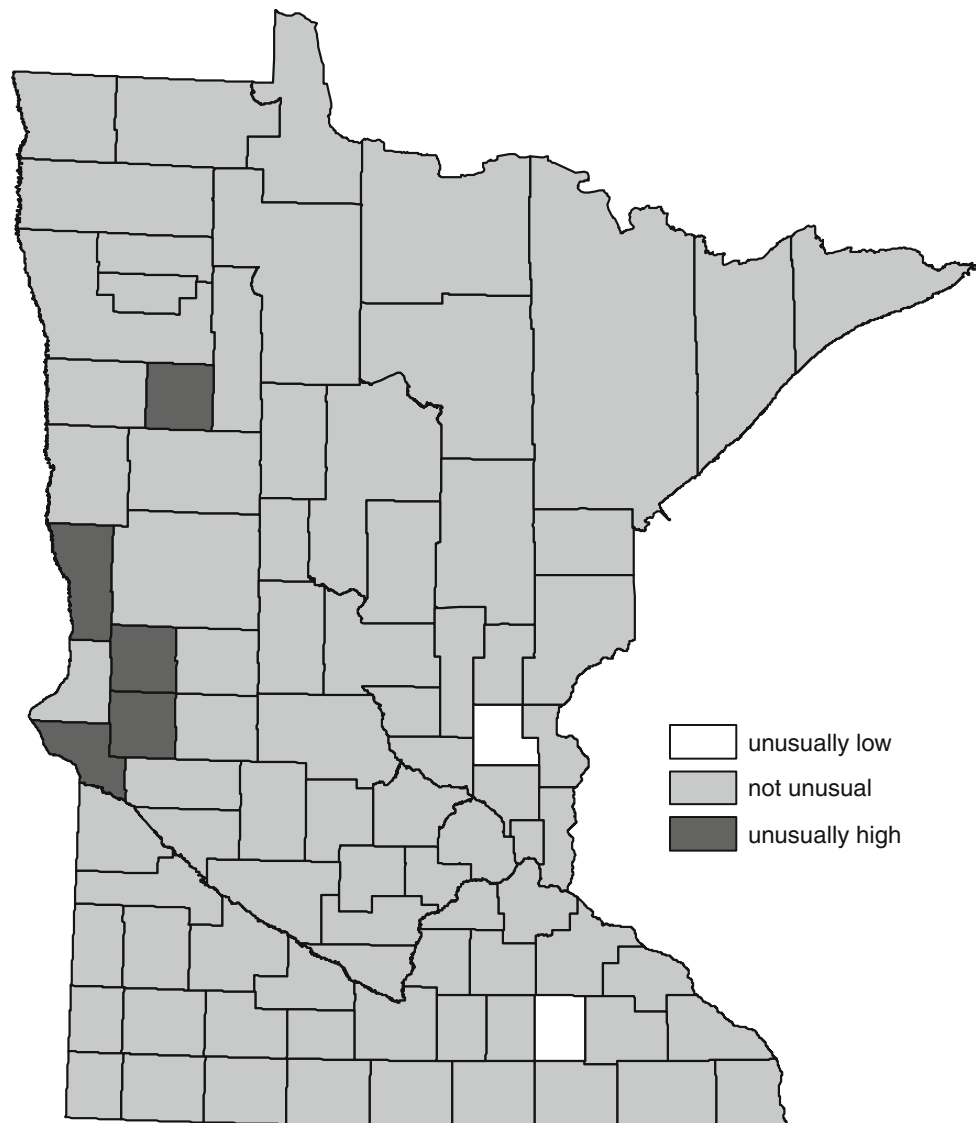


Fig. 6 Age-adjusted male colon/rectum cancer incidence compared to overall state male age-adjusted rate by county (1988–2007). Significance determined by Wilcoxon’s signed rank test (WSRT) procedure described in text

Table 3 Number (%) of statistically significant county rates for four tests of significance—32 cancers, 87 counties, and 2 sexes

Standard normal (95%)		Standard normal (99%)		Bonferroni (90%)		WSRT ^a (95%)	
Low	High	Low	High	Low	High	Low	High
1,256 (24.5)	85 (1.7)	1,094 (21.3)	27 (0.5)	943 (18.4)	7 (0.1)	82 (1.6)	38 (0.7)

Total comparisons = 5,133. Biologically inconsistent cancer and gender combinations excluded

MCSS 2003–2007 [All ages ($n = 121,908$ total cancers)]

^a Wilcoxon’s signed rank test—see text

were nonsignificant based on the WSRT procedure. These occurred in Stearns County (a large county located just northwest of the Minneapolis—St. Paul area). The three occurrences were due to the high childhood female acute lymphoblastic leukemia (ALL) rates. The high rate of childhood female ALL increased both the categories of “all

leukemia” and “all cancer” rates for females to bring the total elevated to 3.

The ALL data for Stearns County were examined by age groups (0–4, 5–9, and 10–14) to evaluate the consistency of this elevated rate. None of the rates for males were statistically significant based on the 95% standard normal test.

Table 4 Number (%) of statistically significant county rates for four tests of significance—32 cancers, 87 counties, and 2 sexes

Standard normal (95%)		Standard normal (99%)		Bonferroni (90%)		WSRT ^a (95%)	
Low	High	Low	High	Low	High	Low	High
1,990 (38.8)	3 (0.01)	1,813 (35.3)	0 (0.0)	1,728 (33.7)	0 (0.0)	9 (0.2)	0 (0.0)

Total comparisons = 5,133. Biologically inconsistent cancer and gender combinations excluded

MCSS 1988–2007 [Ages 0–14 ($n = 3,188$ total cancers)]

^a Wilcoxon's signed rank test—see text

ALL rates for females were consistently higher than the state rate. For the period 1998–2002, the difference reached statistical significance: 12.2 per 100,000 per year vs. 2.8 per 100,000 for the entire state. The importance of an observation of elevated ALL in young females in Stearns County without a corresponding elevation in young males is hard to interpret. Nonetheless, from a public health surveillance perspective of identifying unusual occurrences of cancer through disease mapping, the Stearns County observation was an example of decreased sensitivity for the WSRT procedure. This was the only such example identified during the evaluation of the WSRT procedure.

Discussion

There have been several methods developed to account for the loss of specificity due to multiple statistical comparisons [33]. Conventional approaches, such as the Bonferroni test of significance, are generally viewed as inappropriate in epidemiology as they diminish the ability to identify meaningful differences [26, 34, 35]. In the case of cancer mapping, there needs to be a middle ground due to the influence that it can have on public policy. Mapping of state cancer rates by county involves a large number of implied comparisons, and one of the goals of mapping these rates is the identification of unusual cancer occurrences that, when found, may have important public policy implications [15, 16, 36–38]. A recently established federal program, Environmental Public Health Tracking [39], is attempting to advance disease mapping and environmental correlations. A stated goal of this program is to recognize disease clusters in order to “understand the possible associations between the environment and the adverse health effects.”

This application of the WSRT procedure is another approach for assessing the significance of the difference between directly age-adjusted county cancer rates and the overall state age-adjusted rate. Traditional presentation of the pattern of county cancer rates occurs in several ways. Two of the most common are to portray the county rates as above or below the state average and to distribute the county rates by quartiles. These approaches have become a standard of practice [40]. The WSRT procedure does not

replace these methods. The WSRT procedure is complementary to them, serving the objective of identifying meaningful significant differences between county and overall state rates without a major loss of sensitivity.

The large amount of random variation associated with the analyses of rates based on a small number of cancers raises concern over the precision of the estimates [41]. Several strategies are employed to address this problem. Indirect standardization as an age-adjustment method has been recommended for small counties where the age-specific rates are often quite variable and unreliable [42]. While this approach is useful in comparing a single county to the overall statewide rate, the results of indirect standardization may not be as useful when comparing multiple counties if there are large differences in the age structures of the populations. The indirectly standardized rate is weighted to the specific age distribution of the population of interest [43]. Since mapping county cancer rates results in comparisons among the counties, indirectly age-adjusting cancer rates are not recommended as a method for creating state cancer (and other rate) maps [44].

The most common strategy to address the small number problem is to suppress analyses that are based on a number of occurrences judged to be too low to yield valid results. This number is not consistently defined. The Kentucky Cancer Registry uses 15 [45] and the National Center for Health Statistics recommends 25 as the minimum [46]. The National Program of Cancer Registries does not provide rates if the cell size is smaller than 16 or the population smaller than 50,000 people [47]. Important results from identifying and studying populations with low cancer incidence are well documented [48, 49]. Suppressing analyses of low cancer rates solely on the basis that they represent numbers that are considered too low to publish may result in missing useful insights [31]. Mapping important differences between low county cancer rates compared to the overall state rate can be facilitated by the WSRT procedure due to its decreased sensitivity to low rate instability.

The evaluation of the performance of the WSRT procedure was empirical and based on cancer incidence data from Minnesota. Population-based cancer incidence data in the United States are collected under similar protocols [50], and it is likely that the WSRT procedure would perform

similarly in other states. However, determination of whether the performance of the WSRT procedure found in this evaluation can be generalized for other applications and for use in all states requires further investigation.

The computational algorithm for the WSRT procedure used in this evaluation was coded in FORTRAN—a mathematically based programming language. The translation of this extension of the Wilcoxon's signed rank test into an executable algorithm was straightforward and can be incorporated as a FORTRAN (available from the first author upon request) or C subroutine, or it can be created within a SAS program. The simplicity of coding required for calculating the WSRT significance is an appealing attribute of this procedure.

As the number of observations increase, the precision of the estimates increase, often resulting in small differences becoming statistically significant. The WSRT procedure was not as greatly impacted by the large number effect but it is not immune to it. Identification and interpretation of important differences in cancer rates will always require skill and judgment. The WSRT procedure is an ally in this process as it reduces the number of cancer rates that need to be evaluated.

As illustrated throughout the evaluation of the WSRT procedure, the level of significance for differences in directly age-standardized rates, such as 0.05 or 0.01 is theoretical; each application will likely yield a computed significance that is different than the desired theoretical level of significance. For the purposes of cancer mapping, it is not the level of significance that is important, but the fact that the significantly different rates represent truly unusual occurrences. John Tukey recommended that for disease mapping, significant differences be described as “unusual occurrences.” The WSRT procedure provides a useful method to portray the differences in directly age-adjusted cancer rates occurring at the county level compared to the overall state rate as “unusually low,” “not unusual,” or “unusually high.”

Acknowledgments The authors acknowledge the assistance of Eric Hanson for the preparation of the graphics in this report and Dr. Carin Perkins for assistance in its development.

Conflicts of interest The authors have no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix 1

Example calculations for the nonparametric procedure (WSRT)

Consider the data in Table 5 which contains age-standardized male cancer rates for all cancers combined for

Table 5 Age-standardized rates and their standard errors for males, all cancers combined

County/region	Standardized rate	Standard error
Entire State	555.62	1.18
Aitkin	546.56	16.10
Anoka	583.60	6.51
Becker	585.57	13.48
Beltrami	569.83	13.55
Big Stone	598.71	27.58
Blue Earth	528.55	11.07
Cass	566.25	13.16
Faribault	534.10	16.23
Fillmore	526.72	14.54
Goodhue	520.73	10.99
Kittson	466.45	25.96
Le Sueur	524.68	14.53
Lincoln	483.79	23.02
McLeod	546.21	13.11
Olmsted	628.29	8.34
Scott	564.36	11.58
St. Louis	554.88	5.21
Washington	541.64	7.27
Wright	553.24	9.72
Yellow medicine	516.53	19.23

MCSS 1988–2007

twenty counties and the entire state (per 100,000 per year). These data are from the MCSS for the period 1988–2007. The standard error for each of the counties (se_i) and the state (se) is also provided in Table 5. The twenty counties were a sample (small, medium, and large) of all 87 Minnesota counties used to illustrate the WSRT procedure. The steps are identical to those described in the text.

Step 1 Create the 20, 95% standard normal confidence intervals (CIL_i , CIU_i) for the difference between the county rate and the state rate ($diff_i$) from the formula: $diff_i \pm 1.96 (se_i^2 + se^2)^{1/2}$. For example, Aitkin County: $diff_1 = 546.56 - 555.62 = -9.06$; $(se_1^2 + se^2)^{1/2} = (259.21 + 1.3924)^{1/2} = 16.14$.

The (CIL_1, CIU_1) is $-9.06 \pm 31.64 = (-40.70, 22.58)$. The $diff_i$ and their corresponding (CIL_i, CIU_i) for all twenty counties are given in Table 6. Note that there are 11 occurrences that fall outside their respective 95% standard normal confidence intervals, that is, the confidence interval for the difference does not contain zero.

Step 2 For twenty counties, $m = 20$ and $M = m(m + 1)/2 = 210$. From the 210 Walsh averages for the upper and lower confidence limits. For example, for $i = 1$ and $j = 2$, $W_L^k = (CIL_i + CIL_j)/2 = (-40.70 + 15.01)/2 = -12.84$ (rounded to two decimal places) and $W_U^k = (CIU_i + CIU_j)/2 = (22.58 + 40.95)/2 = 31.76$.

Table 6 Differences between county and state rate (diff_{*i*}), lower and upper 95% standard normal confidence limits (CL), and significance (Sig)

County	diff _{<i>i</i>}	Lower CL	Upper CL	Sig
Aitkin	-9.06	-40.70	22.58	2
Anoka	27.98	15.01	40.95	3
Becker	29.95	3.43	56.47	3
Beltrami	14.21	-12.45	40.87	2
Big Stone	43.09	-11.02	97.20	2
Blue Earth	-27.07	-48.89	-5.25	1
Cass	10.63	-15.27	36.53	2
Faribault	-21.52	-53.41	10.37	2
Fillmore	-28.90	-57.49	-0.31	1
Goodhue	-34.89	-56.55	-13.23	1
Kittson	-89.17	-140.10	-38.24	1
Le Sueur	-30.94	-59.51	-2.37	1
Lincoln	-71.83	-117.01	-26.65	1
McLeod	-9.41	-35.21	16.39	2
Olmsted	72.67	56.16	89.18	3
Scott	-10.74	-21.21	-0.27	1
St. Louis	8.74	-14.07	31.55	2
Washington	-13.98	-28.42	0.46	2
Wright	-2.38	-21.57	16.81	2
Yellow medicine	-39.09	-76.85	-1.33	1

1 low, 2 not significant, 3 high

These entries can be found in the partial tables of the Walsh lower limit averages (Table 7) and Walsh upper limit averages (Table 8).

Step 3 Rank the lower and upper 210 Walsh averages in numeric order.

Step 4 Calculate $C\alpha = \frac{m(m+1)}{4} - Z_{(\alpha/2)} \left[\frac{m(m+1)(2m+1)}{24} \right]^{1/2}$.

For $m = 20$ and $Z_{(\alpha/2)} = 1.96$, $C\alpha = 52$, rounded to the nearest integer and $M + 1 - C\alpha = 159$; $M = 210$.

Table 7 Walsh lower confidence limits averages ($M = 210$)

	-40.70	-12.84	-18.64	-26.57	-25.86	-44.80	-27.98
	-47.06	-49.10	-48.63	-90.40	-50.11	-78.85	-37.96
	7.73	-27.39	-30.96	-34.56	-31.14	-58.78	15.01
	9.22	1.28	2.00	-16.94	-0.13	-19.20	-21.24
	-20.77	-62.55	-22.25	-51.00	-10.10	35.59	0.47
First 42 of the 210 Walsh lower limit averages	-3.10	-6.70	-3.28	-30.92	3.43	-4.51	-3.79

Table 8 Walsh upper confidence limits averages ($M = 210$)

	22.58	31.76	39.53	31.72	59.89	8.67	29.55
	16.48	11.14	4.68	-7.83	10.11	-2.04	19.49
	55.88	27.07	11.16	11.52	19.70	10.63	40.95
	48.71	40.91	69.07	17.85	38.74	25.66	20.32
	13.86	1.36	19.29	7.15	28.67	65.06	36.25
First 42 of the 210 Walsh upper limit averages	20.34	20.70	28.88	19.81	56.47	48.67	76.83

Step 5 The joint confidence interval is (W_L^{52}, W_U^{159}) . The 52nd ordered Walsh average is -55.45 and the 159th ordered Walsh average is 35.27. Thus, $(-55.45, 35.27)$ is the joint confidence interval for the diff_{*i*} when zero lies outside the (CIL_i, CIU_i) , $i = 1, 20$.

Steps 6a and b Compare each (CIL_i, CIU_i) in Table 6 to zero. If the (CIL_i, CIU_i) contains zero, the difference is considered nonsignificant. For Aitkin County, this interval contains zero. No further evaluation is required and comparison to the WSRT confidence interval, (W_L^{52}, W_U^{159}) , is not needed. The WSRT evaluation of the statistical significance of the difference is also taken as not unusual.

For Anoka County, the difference in age-standardized rates was 27.98. The corresponding (CIL_i, CIU_i) does not contain zero, and comparison of the Anoka County/State difference in rates to the WSRT confidence interval is appropriate. Since 27.98 is contained in the interval $(-55.45, 35.27)$, the WSRT evaluation of significance is that the Anoka County rate was not unusual.

For Olmsted County (Rochester), diff_{*i*} = 72.67 and the 95% standard normal confidence interval, (56.16, 89.18), does not contain zero. This difference is also larger than the upper limit of $(-55.45, 35.27)$ and the WSRT significance test indicates an elevated (unusually high) rate for Olmsted County.

Table 9 contains a summary of the 95% standard normal level of significance and the significance determined by the WSRT procedure. Eleven (8 low, 3 high) of the county rates were statistically different than the state rate as zero lies outside their respective 95% standard normal confidence intervals (Table 6). Only three (2 low, 1 high) were statistically unusual based on the WSRT procedure. This reduction in the number of significant differences is a salient feature of the WSRT procedure.

Table 9 95% standard normal (SN) and WSRT determination of statistical significance for cancer rate data in Table 5

County	95% SN ^a	WSRT ^b
Aitkin	2	2
Anoka	3	2
Becker	3	2
Beltrami	2	2
Big Stone	2	2
Blue Earth	1	2
Cass	2	2
Faribault	2	2
Fillmore	1	2
Goodhue	1	2
Kittson	1	1
Le Sueur	1	2
Lincoln	1	1
McLeod	2	2
Olmsted	3	3
Scott	2	2
St. Louis	1	2
Washington	2	2
Wright	2	2
Yellow medicine	1	2

^a 1 Low, 2 Not significant, 3 High

^b 1 Unusually low, 2 Not unusual, 3 Unusually high

Appendix 2

See Table 10.

Table 10 List of 32 cancers used in the evaluation of the WSRT procedure

All types combined	Liver
Bone	Melanoma of the skin
Brain and nervous system	Mesothelioma (all sites)
Breast	Myeloma
Cervix	Non-Hodgkin lymphoma
Colon and rectum	Oral cavity and pharynx
Hodgkin lymphoma	Ovary
Kaposi sarcoma (all sites)	Pancreas
Kidney	Prostate
Larynx	Soft tissues (including heart)
Leukemia (all types)	Stomach
Acute lymphoid leukemia	Testis
Chronic lymphoid leukemia	Thyroid
Acute myeloid leukemia	Urinary bladder
Chronic myeloid leukemia	Uterus

References

- Fraumeni JF Jr (1988) Etiologic insights from cancer mapping. In: Miller RW et al (eds) Unusual occurrences as clues to cancer etiology. Japan Sci Soc Press, Tokyo, pp 13–25
- Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni Jr JF (1980) Introduction. In: Atlas of cancer mortality for U.S. Counties 1950–1969. NIH Publ 75-780, p v
- Pickle LW, Mason TJ, Howard N, Hoover R, Fraumeni Jr JF (1987) Abstract. In: Atlas of U.S. cancer mortality among whites: 1950–1980. NIH Publ 87-2900, p iii
- Boffetta P, Stayner LT (2006) Pleural and peritoneal neoplasms. In: Schottenfeld D, Fraumeni JF Jr (eds) Cancer epidemiology and prevention, 3rd edn. Oxford University Press, Oxford, pp 659–673
- Jablon S, Hrubec Z, Boice JD Jr (1991) Cancer in populations living near nuclear facilities. A survey of mortality nationwide and incidence in two states. JAMA 265:1403–1408
- Brunner WM, Williams AN, Bender AP (2008) Investigation of exposures to commercial asbestos in northeastern Minnesota iron miners who developed mesothelioma. Regul Toxicol Pharmacol 52:S116–S120
- Zhu Y, Zhang J (2011) The spatial heterogeneity and correlation of population distribution in Xinjiang Uygur Autonomous Region. In: Xiang L, Shuming B (eds) 19th International Conference on Geoinformatics. IEEE Publ CFP1132H-PRT, pp 32–36
- Goldsmith JR (1999) The residential radon-lung cancer association in U. S. counties: a commentary. Health Phys 76:553–557
- Friedman GD (1980) Primer of epidemiology. McGraw-Hill Book Company, New York, pp 175–176
- Rushton G (2003) Public health, GIS, and spatial analytic tools. Annu Rev Public Health 24:43–56
- Hao Y, Ward EM, Jemal A, Pickle LW, Thun MJ (2006) U.S. congressional district death rates. Int J Health Geogr 5:28–40
- Gelman A, Price PN (1999) All maps of parameter estimates are misleading. Stat Med 18:3221–3234
- Boscoe FP, Pickle LW (2003) Choosing geographic units for choropleth rate maps, with an emphasis on public health applications. Cartogr Geogr Inf Sci 30:237–248
- Tukey JW (1979) Statistical mapping: what should and should not be plotted. In: Proceedings of the 1976 workshop on automated cartography and epidemiology. Natl Center Health Stat, DHEW (PHS) 79-1254, pp 18–26
- Cromley EK, McLafferty SL (2002) GIS and public health. The Guilford Press, New York, pp 5–10
- Mitchell A (2005) The ESRI guide to GIS analysis. Volume 2: spatial measurements and statistics. ESRI Press, Redlands, pp 2–5
- Melnick AL, Fleming DW (1999) Modern geographic information systems. J Public Health Manag Pract 5:viii–x
- Bell BS, Hoskins RE, Pickle LW, Wartenberg D (2006) Current practices in spatial analysis of cancer data: mapping health statistics to inform policy makers and the public. Int J Health Geogr 5:49–63
- Brown WB Jr, Hollander M (1977) Statistics a biomedical introduction. Wiley, New York, pp 231–234
- Savitz DA, Olshan AF (1995) Multiple comparisons and related issues of epidemiologic data. Am J Epidemiol 142:904–908
- Boyle P, Parkin DM (1991) Chapter 11. Statistical methods for registries. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet CS (eds) Cancer registration principles and methods. IARC Publ 95, [<http://www.iarc.fr/en/publications/pdfs-online/epi/sp95/sp95chap11.pdf>], pp 135–138
- Snedecor GW, Cochran WG (1973) Statistical methods: sixth edition. The Iowa State University Press, Ames, p 101

23. Hollander M, Wolfe DA (1973) Nonparametric statistical methods. Wiley, New York, pp 26–54
24. Pappas PA, DePuy V (2004) An overview of non-parametric tests in SAS: when, why, and how. Paper TU04. Duke Clinical Research Institute, Durham, pp 1–5
25. Keyfitz N (1966) Sampling variance of standardized mortality rates. *Hum Biol* 38:309–317
26. Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* 310:170–180
27. Minnesota Cancer Surveillance System (2008) <http://www.health.state.mn.us/divs/hpcd/cdee/mcss>
28. SEER*Stat (2010) Surveillance, epidemiology, and end results (SEER), SEER*Stat software, National Cancer Institute, Washington, DC, <http://seer.cancer.gov/seerstat>, version 6.6.2
29. Fay MP, Feuer EJ (1997) Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 16:791–801
30. Tiwari RC, Clegg LX, Zou Z (2006) Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Res* 15:547–569
31. Kennedy-Kalafatis S (1995) Reliability-adjusted disease maps. *Soc Sci Med* 41:1273–1287
32. Kalinowski P, Fidler F (2010) Interpreting significance: the differences between statistical significance, effect, size, and practical importance. *Newborn Infant Nurs Rev* 10:50–54
33. Neter J, Wasserman W (1974) Applied linear statistical models. Regression, analysis of variance, and experimental designs. Richard D Irwin, Inc., Homewood, pp 146–147
34. Rothman KJ, Greenland S (1998) Modern epidemiology, 2nd edn. Lippincott-Raven, Philadelphia, pp 227–228
35. Rothman KJ (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–46
36. Pike MC, Henderson BE (1980) “Clustering” of cancer cases. In: Levin DL (ed) Cancer epidemiology in the USA and USSR. NIH Publ 80-2044, pp 47–50
37. The Pew Environmental Health Commission (2000) America’s Environmental Health Gap: Why the Country Needs a Nationwide Tracking Network. Companion Report. Johns Hopkins School of Hygiene and Public Health, Baltimore [<http://healthyamericans.org/reports/files/healthgap.pdf>], pp 4–8
38. Jackson R (2002) Chemicals and chromosomes, children and cancer, clusters and cohorts in a new century. In: Wilson S, Jones L, Coussens C, Hanna K (eds) Cancer and the environment, gene-environment interaction. National Academy Press, Washington, pp 92–93
39. Centers for Disease Control (2005) CDC’s strategy for the National Environmental Health Tracking Program. Department of Health and Human Services, p 2
40. Centers for Disease Control (2009) CDC WONDER on-line database. Compressed Mortality File 1999–2006, Series 20, No. 2L, National Center for Health Statistics, Atlanta
41. Boscoe FP, O’Brien DK, McLaughlin CC (1997) Geographic information systems. In: Menck HR, Gress DM, Griffin A et al (eds) Cancer registry management principles and practices for hospitals and central registries, 3rd edn. Kendall Hunt, Dubuque, pp 413–431
42. Fleiss JL (1981) Statistical methods for rates and proportions, 2nd edn. Wiley, New York, pp 240–244
43. Rothman KJ (1986) Modern epidemiology. Little, Brown and Company, Boston, pp 45–46
44. Pickle LW, White AA (1995) Effects of the choice of age-adjustment method on maps of death rates. *Stat Med* 14:615–627
45. Kentucky Cancer Registry (2008) Annual report. Cancer Incidence and Mortality in Kentucky, 2001–2005. University of Kentucky, Lexington, p 1
46. Curtin LR, Klein RJ (1995) Direct Standardization (Age-Adjusted Death Rates). *Natl Center Health Stat, DHHS Publ* 95–1237, p 3
47. US Cancer Statistics Working Group (2007) United States Cancer Statistics: 2004 Incidence and Mortality. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, p 10
48. Lyon JL, Klauber MR, Gardner JW, Smart CR (1976) Cancer incidence in Mormons and non-Mormons in Utah, 1966–1970. *N Engl J Med* 294:129–133
49. Enstrom JE (1981) Reassessment of the role of dietary fat in cancer etiology. *Cancer Res* 41:3722–3723
50. National Program of Cancer Registries (2010) About the program. <http://www.cdc.gov/cancer/npcr/about.htm>