

JAMA Guide to Statistics and Methods

Nonparametric Statistical Analysis

John M. Lachin, ScD

In this issue of JAMA, Baxter and colleagues¹ from the Non-Invasive Abdominal Aortic Aneurysm Clinical Trial (N-TA³CT) research group report findings from a clinical trial that evaluated the ef-



Related article [page 2029](#)

fect of doxycycline vs placebo on aneurysm growth among patients with small infrarenal abdominal aortic aneurysms. The primary outcome was the maximum transverse diameter of the aneurysm relative to the initial baseline value after 2 years of treatment. However, the investigators anticipated that some patients might die or experience rupture of the aneurysm requiring endovascular repair. In these patients, the maximum transverse diameter measurement at 2 years would be missing but not by chance alone because the missing data are informative about the status of these patients relative to those who completed the study and had 2-year measurement data available. To allow for this informatively missing data, the statistical analysis plan for the study prespecified that the primary efficacy analysis would be conducted using *worst ranks*, a nonparametric analytic method. In this JAMA Guide to Statistics and Methods, general nonparametric statistics are addressed.

What Are Nonparametric Statistics?

Many statistical methods start with a statistical assumption that the distribution of measured values can be summarized by relatively few parameters. For example, a normal or bell-shaped distribution is completely defined by 2 parameters, the mean and the standard deviation. The commonly used *t* test that compares the means of 2 groups assumes the data arise from 2 populations and that each has a normal distribution with the same standard deviation but with different means. This is called a *parametric* analysis because the assumed distribution (eg, normal) can be completely summarized by a few parameters (eg, the mean and standard deviation).

On the other hand, a rank-based nonparametric analysis provides an alternative approach that requires fewer assumptions. Rather than assume that the data have a specific parametric distribution, *nonparametric* methods assess whether the distributions between groups appear to differ, without assuming a specific shape for those distributions. The simplest such nonparametric test is the *Wilcoxon rank sum test*,² which examines the order of the observed values—their ranks—in the 2 groups. In this approach, the observed values are replaced by their ranks and the ranks are analyzed.

As a simple example, consider 2 groups (A and B) with the maximum transverse diameter values in **Table 1**. In a rank analysis, each value is replaced by its rank across all groups. This yields the ascending ranks (lowest to highest) in **Table 2**.

Under the null hypothesis of no difference in the distributions between groups, regardless of their underlying shape, each group should have a similar mixture of low, intermediate, and high ranks. The statistical test of significance can then be based on the sum of the ranks in either group. Consider group A that is the smaller of the 2 groups with 4 observations, for which the sum of the ranks is $S_a = 25$. The sum of all 9 ranks is $1 + 2 + \dots + 9 = 45$ and the average rank is $45/9 = 5$. Thus,

under the null hypothesis, the expected sum in group A by chance alone is the average rank multiplied the number of observations $(5)(4) = 20$ that is not substantially different from what was observed (25), and the resulting 2-sided *P* value is .29. The same *P* value is obtained if the test is based on the sum of the ranks in the other group B.

The Wilcoxon test is also equivalent to a Mann-Whitney analysis³ that provides an alternate derivation and computation. Thus, the test is commonly called the *Wilcoxon-Mann-Whitney test*. This test is also a member of a family of *linear rank statistics*^{4,5} that includes as special cases the Mantel-Haenszel or log-rank test⁶ for survival data among many others.

A linear rank test extends the concept of the sum of ranks to a sum of rank scores, for which the score is a specified function of the simple rank. If one hypothesizes that the observations in the population follow a given distribution, then it is possible to describe the rank scores that provide an optimal linear rank statistic. For example, for a logistic distribution, the simple ranks themselves provide an optimal test (ie, the Wilcoxon test). Likewise, for a normal distribution with equal standard deviations, a test using van der Waerden or *normal scores*⁷ is optimal. The power of the test is also nearly equivalent to the power of the *t* test.

Why Are Nonparametric Statistics Important?

The principal advantage of a nonparametric test is that it provides a test of a null hypothesis of equality of distributions that does not require specific parametric assumptions. For example, the *t* test provides an optimal and valid test of the equality of the distributions in the 2 groups if it is assumed that both distributions are normal (bell shaped) with the same mean and with the same standard deviation. Under these assumptions, the test of equality of means is equivalent to a test of the equality of the entire distributions, and the type I (false-positive) error probability does not exceed the specified significance level, such as .05. However, when the distributions in the 2 groups have the same means but have different standard deviations or even different shapes, the type I error probability of the *t* test can be increased, leading to more false-positive results than specified by the significance level (eg, more than .05). On the other hand, a nonparametric test will provide the desired type I error probability (eg, .05) regardless of whether the normal distribution assumptions apply.

In general, a nonparametric test such as the Wilcoxon test also has good power relative to a *t* test or other parametric tests. The relative power of 2 possible tests is assessed by the *relative efficiency*, often expressed as the ratio of sample sizes needed to provide the same power.⁵ When the actual distributions are normal but there is a difference in means, the Wilcoxon test has a relative efficiency of 0.955 vs the *t* test, meaning that if the *t* test required a sample size of 100 to achieve a given level of power, the Wilcoxon test would require a sample size of $100/0.955 = 105$ to provide the same power, a small difference. Furthermore, the Wilcoxon test is the optimal or most efficient test when the actual distributions are logistic, meaning a bell shape but with a higher fraction of the observations closer

Table 1. Groups A and B: Maximum Transverse Diameter Values

Group	Maximum transverse diameter, cm				
A	4.437	4.436	4.290	4.349	
B	4.045	4.091	4.479	3.985	4.295

Table 2. Groups A and B: Ranks of the Maximum Transverse Diameter Values

Group	Rank				
A	8	7	4	6	
B	2	3	9	1	5

to the mean than the normal distribution, and a slightly smaller fraction in the shoulders of the distribution. Then the relative efficiency of the Wilcoxon test to the t test is 1.1, meaning that the t test would need a 10% higher sample size to provide the same power as the Wilcoxon test. Similarly the normal scores test is highly efficient over a range of possible distributions, ie, is also robust.

Ranks (or rank scores) can also be used as the dependent variable in a regression model that adjusts for other covariates, termed a *rank transformation* analysis,^{8,9} in which the regression model-based test (eg, a t test) is approximately equal to the large-sample rank-based test. For example, in a 2-group comparative study with before and after treatment observations, a common approach would be to use an *analysis of covariance* (ANCOVA), in which the difference between posttreatment group means is adjusted for the pretreatment or baseline values in a simple regression model to increase study power. A similar nonparametric analysis can be conducted by using the ranks of both the baseline values and the post-treatment values. This approach has advantages^{8,9} over the analysis of the raw values because it avoids the substantial inflation of the type I (false-positive) error rate or loss of power that may occur if there are deviations from the normality assumptions required by the usual ANCOVA analyses.

Limitations and Alternatives to Using Nonparametric Statistics

With a parametric analysis such as a t test, the analysis often starts with estimates of the assumed parameters, such as the means in each group, the standard deviation, the mean difference and its standard error. The t test is then computed as a function of these

parameter estimates so that there is a 1:1 correspondence between the test and the confidence limits; ie, if $P < .05$, the confidence limits on the mean difference do not include zero. Conversely, a nonparametric test is not based on estimates of population parameters, in which case its results may conflict with parameter estimates. On the other hand, a nonparametric test can be based on summary quantities that can be interpreted without the need for any distributional assumptions, such as the Mann-Whitney difference mentioned above.

How Were Nonparametric Statistics Used in This Study?

Nonparametric methods were used in the N-TA³CT study analysis of the growth of abdominal aortic aneurysms. The authors used a rank transformation ANCOVA in which the changes in the maximum transverse diameter rank scores from baseline to 2 years were compared between groups when also adjusting for the rank score of the baseline measure, and sex. Van der Waerden or normal rank scores were used.

In this study, measures of the aortic diameter would be missing if the aneurysm required repair or had ruptured during the course of the study. When analyzing such study results, it is important to account for missing data.¹⁰ When missing data are related to the outcome of interest, nonparametric methods can also be used to account for nonrandomly missing data that are related to and provide information about the outcome. Baxter et al¹ used worst ranks methods to analyze the outcomes of their study. This method will be reviewed in a subsequent JAMA Guide to Statistics and Methods article.

ARTICLE INFORMATION

Author Affiliation: The Biostatistics Center, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Rockville, Maryland.

Corresponding Author: John M. Lachin, ScD, Biostatistics Center, George Washington University, 6110 Executive Blvd, Rockville, MD 20852 (jml@bsc.gwu.edu).

Section Editors: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, JAMA.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Baxter BT, Matsumura J, Curci JA, et al. Effect of doxycycline on aneurysm growth among patients

with small infrarenal abdominal aortic aneurysms: a randomized clinical trial. *JAMA*. Published May 26, 2020. doi:10.1001/jama.2020.5230

2. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bull*. 1945;1(6):80-83. doi:10.2307/3001968

3. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18(1):50-60. doi:10.1214/aoms/1177730491

4. Randles RH, Wolfe DA. *Introduction to the Theory of Nonparametric Statistics*. John Wiley & Sons; 1979.

5. Tolles J, Lewis RJ. Time-to-event analysis. *JAMA*. 2016;315(10):1046-1047. doi:10.1001/jama.2016.1825

6. van der Waerden BL. Order tests for the two-sample problem and their power. *Indag Math*. 1952;14:453-458. doi:10.1016/S1385-7258(52)50063-5

7. Conover WJ, Iman RL. Analysis of covariance using the rank transformation. *Biometrics*. 1982;38(3):715-724. doi:10.2307/2530051

8. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. *Am Stat*. 1981;35(3):124-133. doi:10.2307/2683975

9. Newgard CD, Lewis RJ. Missing data: how to best account for what is not known. *JAMA*. 2015;314(9):940-941. doi:10.1001/jama.2015.10516

10. Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials*. 1999;20(5):408-422. doi:10.1016/S0197-2456(99)00022-7