

Exploiting the Link Between the Wilcoxon-Mann-Whitney Test and a Simple Odds Statistic

Ralph O'Brien¹ and John Castelleo²
 Cleveland Clinic Foundation¹ and SAS²

Abstract:

The Wilcoxon-Mann-Whitney (WMW) rank-sum test is often regarded as comparing two independent medians, but this is only true under conditions rarely met in practice. So, what does it test? Let Y_1 and Y_2 be sample values from two independent groups, and $\pi = \text{Prob}(Y_1 > Y_2) + \text{Prob}(Y_1 = Y_2)/2$. This is the nonparametric area under the ROC curve in diagnostic testing, a field that routinely estimates and forms confidence intervals (CIs) for π and tests $H_0 : \pi = 0.50$. Consider $\theta = \pi/(1-\pi)$. *With no ties*, this is a special case of the generalized odds ratio by Agresti (1980), who gave asymptotic standard errors for estimates of θ and $\ln(\theta)$, thus providing CIs and a test for $H_0 : \theta = 1.0$. Generalizing theta to consider ties gives us a meaningful parameter and CI to augment the ubiquitous WMW p value. This also gives us a new way to handle sample-size analyses, competing with Kolassa's (1995) solution. We assess these methods using Monte Carlo studies and illustrate them with a two-arm clinical trial involving a seven-point Likert measure. Finally, we show perplexities that can result when the WMW test is applied to ordered categorical data, thus illustrating that it does not necessarily compare central tendencies.

KEY WORDS: Wilcoxon rank-sum test, Mann-Whitney U test, effect size, generalized odds ratio, power, sample-size analysis

1. Introduction

How should you interpret results from the Wilcoxon-Mann-Whitney (WMW) two-group test? Some software, including PROC NPAR1WAY in SAS/STAT, provides the mean ranks for each group, but these have no interpretation outside the study, and their difference increases as the sample size increases. A misconception commonly found in research reports and articles (and even in some statistics textbooks) is that the WMW test compares the two medians. This is not true except in the rare case in which the population distributions of the two groups are merely shifted versions of each other (i.e., differing only in location, and not shape or scale). In fact, a

WMW statistic can have a p value near 0.00 even when the two groups have identical sample medians. Here, we present a simple and highly useful way to better understand, interpret, and present results from the Wilcoxon-Mann-Whitney test by estimating a simple odds parameter and computing its confidence interval. The methodology also appears to provide to a sound approximation for computing statistical powers for the WMW test.

2. A Simple Odds Parameter for the Wilcoxon-Mann-Whitney (WMW) Test

Consider a hypothetical clinical trial to treat interstitial cystitis (IC), a painful, chronic inflammatory condition of the bladder with no known cause, though it most commonly affects women. Two treatments will be compared: lidocaine alone ("lidocaine") versus lidocaine plus a fictitious experimental drug called mironel ("Mir+lido"). The design is balanced, randomized, double-blind, and female-only. The primary outcome is a measure of overall improvement at week 4 on the study, measured on a seven-point Likert scale:

"Compared to when I started this study, my condition is:"

much worse	-3
worse	-2
slightly worse	-1
the same	0
slightly better	+1
better	+2
much better	+3

Suppose the counts for this trial are:

	Response						
	-3	-2	-1	0	+1	+2	+3
lidocaine	3	8	19	78	29	7	3
Mir+lido	1	5	22	52	42	16	10

The relevant results from the NPAR1WAY procedure in SAS/STAT software are:

Group	N	Mean Score
lidocaine	147	134.1
Mir+lido	148	161.8

Wilcoxon Two-Sample Test	
Z	-2.95
Two-Sided Pr > Z	0.0032

The higher mean rank for Mir+lido (161 vs. 134) along with the low p-value supports the conclusion that the Mir+lido patients reported better outcomes. But how much better?

Consider the following way to characterize the WMW test. Let Y_1 and Y_2 be observations drawn independently from two distributions. Then the null hypothesis being tested is:

$$H_0 : \pi = 0.5, \text{ where}$$

$$\pi = \text{Prob}(Y_1 < Y_2) + \frac{1}{2}\text{Prob}(Y_1 = Y_2)$$

with alternative depending on the sidedness of the test:

$$H_1 : \begin{cases} \pi > 0.5, & \text{upper 1-sided} \\ \pi < 0.5, & \text{lower 1-sided} \\ \pi \neq 0.5, & \text{2-sided} \end{cases}$$

In other words, H_0 asserts that it is equally likely for Y_1 to be less than or greater than Y_2 , splitting ties evenly.

Now, converting π to an odds measure, we define

$$WMW_{\text{odds}} = \frac{\pi}{1 - \pi}.$$

For example, if $WMW_{\text{odds}} = 2.0$, the odds are 2:1 that Y_1 is less than Y_2 , splitting ties evenly.

Expressed in terms of WMW_{odds} , the hypothesis for the MWM test is:

$$H_0 : WMW_{\text{odds}} = 1$$

$$H_1 : \begin{cases} WMW_{\text{odds}} > 1, & \text{upper 1-sided} \\ WMW_{\text{odds}} < 1, & \text{lower 1-sided} \\ WMW_{\text{odds}} \neq 1, & \text{2-sided} \end{cases}$$

Estimating WMW_{odds} just involves counting properly, and it provides a clear way to quantify how much the two distributions differ in the manner examined by the WMW test. For the interstitial cystitis data, there are 9456 Y_1, Y_2 pairs having $Y_1 > Y_2$, 5808 pairs having $Y_1 < Y_2$, and 5736 pairs having $Y_1 = Y_2$. Thus $WMW_{\text{odds}} = (9985 + 5877/2)/(5894 + 5877/2) = 1.46$. But what is its sampling distribution? To get this, we can exploit the fact that WMW_{odds} is an extension of the *generalized odds ratio* (*genOR*) developed by Agresti (1980), the extension being the accommodation of ties.

3. Extending Agresti's *genOR* to WMW_{odds}

A measure quite similar to WMW_{odds} is Agresti's (1980) generalized odds ratio:

$$genOR = \frac{\text{Prob}(Y_1 < Y_2)}{\text{Prob}(Y_1 > Y_2)}.$$

When used in a 2x2 table, it is identical to the usual odds ratio, hence the term "generalized odds ratio." But it fails to be a suitable effect size measure for the WMW test, because it ignores the ties rather than split them evenly, in effect "overstating" the group difference: $genOR \geq WMW_{\text{odds}}$, with equality holding when there are no ties. For the interstitial cystitis data, $genOR = 9985/5894 = 1.69$ versus 1.46 for WMW_{odds} .

Fortunately, Agresti's formulas for *genOR* can be readily applied to WMW_{odds} . *genOR* is a simple transformation of the Goodman-Kruskall gamma (γ) statistic,

$$genOR = \frac{1 + \gamma}{1 - \gamma}$$

Accordingly, Agresti used the delta method to take known results for γ and $SE(\hat{\gamma})$ and derive estimates for *genOR*, $SE(\widehat{genOR})$, $\ln(genOR)$, and $SE(\ln(\widehat{genOR}))$. The approximate normality of $\ln(\widehat{genOR})$ provides confidence limits and p-values for *genOR*.

We can extend *genOR* to WMW_{odds} by modifying the data in a way that leaves WMW_{odds} unaffected but destroys all ties between the two groups. Thus, $\widehat{genOR} = WMW_{\text{odds}}$ and Agresti's results apply immediately.

Using the interstitial cystitis example, one can see that the trick is rather simple. Recall the original table of counts:

	Response						
	-3	-2	-1	0	+1	+2	+3
lidocaine	3	8	19	78	29	7	3
Mir+lido	1	5	22	52	44	16	10

Now add and subtract some arbitrarily small quantity, $\epsilon > 0$, from each of the values in the second group, as so:

	Response									
	-3.1	-3.0	-2.9	-2.1	-2.0	-1.9	-1.1	-1.0	-0.9	...
lidocaine	-	3	-	-	8	-	-	19	-	...
Mir+lido	0.5	-	0.5	2.5	-	2.5	11	-	11	...

Here, $\epsilon = 0.1$, but it can be any positive value small enough to preserve the ordering of the counts across the categories.

This 2 x 21 table of counts produces the same $\widehat{WMW}_{\text{odds}} = 1.46$ value as we obtained from the original 2 x 7 table. On the other hand, \widehat{genOR} is reduced from 1.69 to 1.46. Using Agresti's formulas for \widehat{genOR} on the modified table yields a 95% confidence interval for WMW_{odds} of [1.11, 1.93], with $p = 0.005$ (corresponding to $Z = 2.78$). For comparison, the 95% confidence interval for \widehat{genOR} is [1.19, 2.41], with $p = 0.002$ (corresponding to $Z = 3.07$). Since there are 5877 ties, $\widehat{WMW}_{\text{odds}}$ is closer to 1.0, and its corresponding Z statistic is smaller. Operating somewhat like odds ratios and hazard ratios, $\widehat{WMW}_{\text{odds}}$ and its confidence interval give us a simple effect size that can be compared with those obtained from other ordinal responses in this or any other study.

Explicit formulas for WMW_{odds} , $SE(\widehat{WMW}_{\text{odds}})$, $\ln(\widehat{WMW}_{\text{odds}})$, and $SE(\ln(\widehat{WMW}_{\text{odds}}))$ are merely the sample analogues of those given for the population values given in Section 4..

4. Power Analysis Based on WMW_{odds}

The asymptotic distribution of $\widehat{WMW}_{\text{odds}}$ can be obtained by applying Agresti's (1980) distributional results for \widehat{genOR} to the contingency table trivially modified (as demonstrated in Section 3.) to break ties. This strategy turns out to be identical to using extended versions of Agresti's (1980) equations (with correction terms to split ties evenly) on the *unmodified* table. These extended equations for use with the unmodified table are developed in this section. Furthermore, the use of WMW_{odds} is generalized to comparisons of any two distributions with ordered values, not just ordered categorical distributions.

While the power computation method is most directly applicable to the test using $\widehat{WMW}_{\text{odds}}$ (being derived from its asymptotic distribution, after all), it also appears to serve as a sound power approximation for the traditional flavors of the WMW test (such as those implemented in PROC NPAR1WAY in SAS/STAT). This is not surprising considering the simple relationship between WMW_{odds} and the traditional WMW test, with the former based on the odds parameter constructed from the latter.

4.1 Power Formula

Let Y_1 and Y_2 be independent observations from any two distributions that we wish to compare using a WMW test. For purposes of deriving the asymptotic distribution of $\widehat{WMW}_{\text{odds}}$ (and consequently

the power computation as well), these distributions must be formulated as ordered categorical ("ordinal") distributions. In addition, all of the conditional probabilities (of response given group membership) must be specified, along with the usual power analysis ingredients such as alpha, sample size per group, and sidedness of test.

If a distribution is not ordinal, it can be discretized using a large number of categories with negligible loss of accuracy. Our suggested discretization method is to break each non-ordinal distribution into b categories (where the choice of b depends upon computational feasibility and desired accuracy), with breakpoints evenly spaced on the probability scale. That is, each bin contains an equal probability $1/b$ for that distribution. Then the breakpoints across both distributions are pooled to form a collection of C bins (heretofore called "categories"), and the probabilities of bin membership for each distribution are re-calculated. The motivation for this method of binning is to avoid degenerate representations of the distributions (i.e., small handfuls of large probabilities among mostly empty bins), as may be caused by something like an evenly spaced grid across *quantiles* rather than probabilities.

After the discretization process just mentioned above, we now have two ordinal distributions, each with a set of probabilities across a common set of C ordered categories. For simplicity of notation, we assume (without loss of generality) the response values to be $1, \dots, C$. Represent the conditional probabilities as

$$\begin{aligned} \tilde{p}_{ij} &= \text{Prob}(Y_i = j \mid \text{group} = i), \\ i &\in \{1, 2\} \quad \text{and} \quad j \in \{1, \dots, C\} \end{aligned}$$

and the group allocation weights as

$$w_i = \frac{n_i}{N} = \text{Prob}(\text{group} = i), \quad i \in \{1, 2\}$$

The joint probabilities can then be calculated simply as

$$\begin{aligned} p_{ij} &= \text{Prob}(\text{group} = i, Y_i = j) = w_i \tilde{p}_{ij}, \\ i &\in \{1, 2\} \quad \text{and} \quad j \in \{1, \dots, C\} \end{aligned}$$

The next step in the power computation is to compute the probabilities that a randomly chosen pair of observations from the two groups is concordant, discordant, or tied. It is useful to define these probabilities as functions of the terms $R_{s_{ij}}$ and $R_{d_{ij}}$, defined as follows, where Y is a random observation

drawn from the joint distribution across groups and categories:

$$\begin{aligned}
 Rs_{ij} &= \text{Prob}(Y \text{ is concordant with cell}(i, j)) + \\
 &\quad \frac{1}{2} \text{Prob}(Y \text{ is tied with cell}(i, j)) \\
 &= \text{Prob}((\text{group} < i \text{ and } Y < j) \text{ or} \\
 &\quad (\text{group} > i \text{ and } Y > j)) + \\
 &\quad \frac{1}{2} \text{Prob}(\text{group} \neq i \text{ and } Y = j) \\
 &= \sum_{g=1}^2 \sum_{c=1}^C w_g \tilde{p}_{gc} \left[\mathbb{I}_{(g-i)(c-j) > 0} + \frac{1}{2} \mathbb{I}_{g \neq i, c=j} \right]
 \end{aligned}$$

and

$$\begin{aligned}
 Rd_{ij} &= \text{Prob}(Y \text{ is discordant with cell}(i, j)) + \\
 &\quad \frac{1}{2} \text{Prob}(Y \text{ is tied with cell}(i, j)) \\
 &= \text{Prob}((\text{group} < i \text{ and } Y > j) \text{ or} \\
 &\quad (\text{group} > i \text{ and } Y < j)) + \\
 &\quad \frac{1}{2} \text{Prob}(\text{group} \neq i \text{ and } Y = j) \\
 &= \sum_{g=1}^2 \sum_{c=1}^C w_g \tilde{p}_{gc} \left[\mathbb{I}_{(g-i)(c-j) < 0} + \frac{1}{2} \mathbb{I}_{g \neq i, c=j} \right]
 \end{aligned}$$

For an independent random draw Y_1, Y_2 from the two distributions, we have

$$\begin{aligned}
 P_c &= \text{Prob}(Y_1, Y_2 \text{ concordant}) + \\
 &\quad \frac{1}{2} \text{Prob}(Y_1, Y_2 \text{ tied}) \\
 &= \sum_{i=1}^2 \sum_{j=1}^C w_i \tilde{p}_{ij} Rs_{ij}
 \end{aligned}$$

and

$$\begin{aligned}
 P_d &= \text{Prob}(Y_1, Y_2 \text{ discordant}) + \\
 &\quad \frac{1}{2} \text{Prob}(Y_1, Y_2 \text{ tied}) \\
 &= \sum_{i=1}^2 \sum_{j=1}^C w_i \tilde{p}_{ij} Rd_{ij}
 \end{aligned}$$

Then

$$WMW_{\text{odds}} = \frac{P_c}{P_d}$$

Proceeding to compute the theoretical standard error associated with WMW_{odds} (that is, the population analogue to the sample-based estimated standard error estimate), we have

$$\begin{aligned}
 SE(WMW_{\text{odds}}) &= \\
 &\frac{2}{P_d} \left[\sum_{i=1}^2 \sum_{j=1}^C w_i \tilde{p}_{ij} (WMW_{\text{odds}} Rd_{ij} - Rs_{ij})^2 \right]^{\frac{1}{2}}
 \end{aligned}$$

Converting to the log scale using the delta method,

$$SE(\ln(WMW_{\text{odds}})) = \frac{SE(WMW_{\text{odds}})}{WMW_{\text{odds}}}$$

The next step is to produce a “smoothed” version of the $2 \times C$ cell probabilities that conforms to the null hypothesis $WMW_{\text{odds}} = 1$ (in other words, independence in the $2 \times C$ contingency table of probabilities). Let $SE_{H_0}(\ln(WMW_{\text{odds}}))$ denote the theoretical standard error of $\ln(WMW_{\text{odds}})$ assuming H_0 .

Finally we have all of the terms needed to compute the power, using the noncentral Chi-square and normal distributions:

$$\text{power} = \begin{cases} P \left(Z \geq \frac{SE_{H_0}(\ln(WMW_{\text{odds}}))}{SE(\ln(WMW_{\text{odds}}))} z_{1-\alpha} - \delta^* N^{\frac{1}{2}} \right), & \text{upper 1-sided} \\ P \left(Z \leq \frac{SE_{H_0}(\ln(WMW_{\text{odds}}))}{SE(\ln(WMW_{\text{odds}}))} z_{\alpha} - \delta^* N^{\frac{1}{2}} \right), & \text{lower 1-sided} \\ P \left(\chi^2(1, (\delta^*)^2 N) \geq \left[\frac{SE_{H_0}(\ln(WMW_{\text{odds}}))}{SE(\ln(WMW_{\text{odds}}))} \right]^2 \chi^2_{1-\alpha}(1) \right), & \text{2-sided} \end{cases}$$

where

$$\delta^* = \frac{\ln(WMW_{\text{odds}})}{N^{\frac{1}{2}} SE(\ln(WMW_{\text{odds}}))}$$

and Z is a standard normal random variable, $\chi^2(df, nc)$ is a noncentral χ^2 random variable with degrees of freedom df and noncentrality nc , and N is the total sample size.

4.2 Power Analysis for Clinical Trial Example

Consider the clinical trial example introduced in Section 2.. Suppose we wish to compute the power of a test we are planning to compare the self-reports for “lidocaine” and “Mir+lido” using the WMW test (or the *genOR*-based or WMW_{odds} -based test), with a total sample size of 150 and $\alpha=0.01$.

Discretization of the underlying distributions is not necessary here, because the distributions are already ordinal, based on a seven-point Likert scale. But we do need to specify the planned sample size allocation weights and conjecture the collection of 14 underlying true probabilities of the 7 self-report categories given membership in each of the two groups. Suppose that we plan a balanced design, assigning “lidocaine” and “Mir+lido” to equal numbers of patients. We make scientifically educated guesses about the conditional probabilities as shown in Table 1.

	Response						
	-3	-2	-1	0	+1	+2	+3
lidocaine	.01	.04	.20	.50	.20	.04	.01
Mir+lido	.01	.03	.15	.35	.30	.10	.06

Table 1: Conjectured conditional probabilities $\{\tilde{p}_{ij}\}$.

The theoretical values of $genOR$ and WMW_{odds} are obtained from Table 1 by simply adding the relevant probabilities and computing ratios:

$$\begin{aligned} \text{Prob}[Y_{\text{lidocaine}} < Y_{\text{Mir+lido}}] &= .475 \\ \text{Prob}[Y_{\text{lidocaine}} = Y_{\text{Mir+lido}}] &= .271 \\ genOR &= \frac{.475}{1 - .475 - .271} = 1.87 \\ WMW_{odds} &= \frac{.475 + .135}{1 - .475 - .135} = 1.57 \end{aligned}$$

The computed power to detect a WMW_{odds} of 1.57, using the power formula in Section 4.1, is 0.826.

5. Power Comparisons and Simulation Results

Tables 2 through 4 show power estimates using the WMW_{odds} power approximation developed in Section 4. for the scenarios presented in Kolassa (1995). For comparison, powers computed in Whitehead (1993) and Kolassa (1995) are also shown, along with empirical power estimates from Monte Carlo simulation.

Since Kolassa’s method applies only to the cumulative proportional odds model, all of the 18 scenarios in Table 2 feature constant cumulative proportional odds. While the individual cell probabilities are not shown, they can be derived from the assumed common log odds ratio (denoted θ) and the marginal column probabilities. The value of θ in each scenario was chosen to attain nominal powers of either 0.8 or 0.95 in Whitehead’s (1993) power approximation, since Kolassa’s (1995) paper uses Whitehead (1993) as a basis for comparison. All scenarios assume a one-sided test with $\alpha = 0.025$.

The “Simulated Power” uses the traditional z -based test reported in PROC NPAR1WAY of SAS/STAT software and is followed by 95% confidence limits. Each estimate is obtained from 20,000 samples, computed as the proportion of one-sided tests rejected at the 0.025 level in 20,000 data sets simulated from the distribution defined by the marginal column probabilities and θ .

Scenario	N per Group	Column Probs	θ
1	30	0.289, 0.486, 0.153, 0.072	1.3732
2	15	0.289, 0.486, 0.153, 0.072	1.9739
3	5	0.289, 0.486, 0.153, 0.072	3.6394
4	30	0.25, 0.25, 0.25, 0.25	1.3145
5	15	0.25, 0.25, 0.25, 0.25	1.8895
6	5	0.25, 0.25, 0.25, 0.25	3.4839
7	30	0.1, 0.2, 0.3, 0.4	1.3408
8	15	0.1, 0.2, 0.3, 0.4	1.9273
9	5	0.1, 0.2, 0.3, 0.4	3.5535
10	30	0.289, 0.486, 0.153, 0.072	1.7669
11	15	0.289, 0.486, 0.153, 0.072	2.5398
12	5	0.289, 0.486, 0.153, 0.072	4.6828
13	30	0.25, 0.25, 0.25, 0.25	1.6914
14	15	0.25, 0.25, 0.25, 0.25	2.4313
15	5	0.25, 0.25, 0.25, 0.25	4.4828
16	30	0.1, 0.2, 0.3, 0.4	1.7252
17	15	0.1, 0.2, 0.3, 0.4	2.4798
18	5	0.1, 0.2, 0.3, 0.4	4.5723

Table 2: Description of scenarios.

It should be noted that Kolassa’s method is developed for use with the score test presented in Whitehead (1993), rather than the WMW test. However, because of the similarities between these two tests, Kolassa’s method is often used as an approximation for power of the traditional WMW test. Another well-known power computation for the WMW test is that of Hilton and Mehta (1993), but it is excluded from this comparison because it is designed for the exact conditional form of the WMW test, whereas our scope here is limited to asymptotic tests.

6. Does the WMW Test the Difference Between Two Medians?

As mentioned in the introduction, the WMW test is a test of the difference between medians *only* under the shift hypothesis.

Below is a case with equal medians but significant p-values:

	-1	0	1	
group=1	3	6	1	median=0
group=2	0	6	4	median=0

Equal medians, insignificant WMW and WMW_{odds} .

	Whitehead Power	Kolassa Power	WMW_{odds} Power	Simulated Power
1	0.80	0.781	0.786	0.779
2	0.80	0.755	0.771	0.756
3	0.80	0.615	0.691	0.506
4	0.80	0.795	0.799	0.786
5	0.80	0.785	0.797	0.773
6	0.80	0.728	0.779	0.658
7	0.80	0.793	0.797	0.789
8	0.80	0.783	0.794	0.773
9	0.80	0.754	0.772	0.638
10	0.95	0.933	0.930	0.933
11	0.95	0.909	0.906	0.909
12	0.95	0.706	0.773	0.608
13	0.95	0.945	0.939	0.944
14	0.95	0.937	0.926	0.934
15	0.95	0.895	0.869	0.830
16	0.95	0.944	0.938	0.939
17	0.95	0.936	0.924	0.930
18	0.95	0.900	0.867	0.818

Table 3: Estimated and simulated power.

	WMW_{odds} Power	Simulated Power
1	0.786	0.779 (0.773, 0.785)
2	0.771	0.756 (0.750, 0.762)
3	0.691	0.506 (0.499, 0.512)
4	0.799	0.786 (0.780, 0.791)
5	0.797	0.773 (0.767, 0.779)
6	0.779	0.658 (0.651, 0.664)
7	0.797	0.789 (0.783, 0.795)
8	0.794	0.773 (0.769, 0.779)
9	0.772	0.638 (0.632, 0.645)
10	0.930	0.933 (0.930, 0.936)
11	0.906	0.909 (0.905, 0.913)
12	0.773	0.608 (0.601, 0.615)
13	0.939	0.944 (0.941, 0.947)
14	0.926	0.934 (0.930, 0.937)
15	0.869	0.830 (0.828, 0.835)
16	0.938	0.939 (0.936, 0.943)
17	0.924	0.930 (0.926, 0.933)
18	0.867	0.818 (0.812, 0.823)

Table 4: Estimated and simulated power, showing 95% simulation confidence limits.

$$\begin{aligned}
 WMW \text{ p-value} &= 0.038 \\
 WMW_{\text{odds}} \text{ p-value} &= 0.024 \\
 WMW_{\text{odds}} \text{ confidence limits} &= [0.87, 9.2]
 \end{aligned}$$

Below is another case with equal medians, where the WMW p-value is again significant, but WMW_{odds} is significant due to accomodation of ties, which provide most of the information here:

	-1	0	1	
group=1	3	996	1	median=0
group=2	0	996	4	median=0

Equal medians, insignificant WMW, significant WMW_{odds} .

$$\begin{aligned}
 WMW \text{ p-value} &= 0.034 \\
 WMW_{\text{odds}} \text{ p-value} &= 0.85 \\
 WMW_{\text{odds}} \text{ confidence limits} &= [0.89, 1.15]
 \end{aligned}$$

7. Conclusion

We have demonstrated the utility of the WMW_{odds} parameter in providing interpretable estimates in

the nonparametric comparison of two distributions with ordered values. The tests and confidence intervals formulated using WMW_{odds} are a viable alternative to the traditional WMW analysis, which lacks meaningful and generalizable summary statistics. In addition, because of the similarity between the hypotheses of the two tests (both splitting ties evenly), the asymptotic distribution of WMW_{odds} provides a promising power approximation not only for its own test statistic, but also for the WMW test.

This paper is a preliminary communication of work in progress. Feedback is welcome.

References

- [1] Agresti, A. (1980), "Generalized Odds Ratios for Ordinal Data," *Biometrics*, 36, 59–67.
- [2] Hilton, J.F. and Mehta, C.R. (1993), "Power and Sample Size Calculations for Exact Conditional Tests with Ordered Categorical Data," *Biometrics*, 49, 609–616.
- [3] Kolassa, J.E. (1995), "A Comparison of Size and Power Calculations for the Wilcoxon Statistic for Ordered Categorical Data," *Statistics in Medicine*, 14, 1577–1581.

- [4] Whitehead, J. (1993), "Sample Size Calculations for Ordered Categorical Data," *Statistics in Medicine*, 12, 2257–2271.