Journal of Statistical Research *https://doi.org/10.3329/jsr.v58i1.75411*
2024, Vol. 58, No. 1, pp. 33-48 *ISSN 0256 - 422 X*

RANDOMIZATION-BASED COVARIANCE ADJUSTMENT OF WIN RATIOS AND WIN ODDS FOR RANDOMIZED MULTI-VISIT STUDIES WITH ORDINAL OUTCOMES

ANN MARIE K. WEIDEMAN[†]
Department of Biostatistics, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7420, USA
Email: anndo1@umbc.edu

ELAINE K. KOWALEWSKI[⋆,†]
Department of Biostatistics, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7420, USA
Email: elainekkowalewski@gmail.com

GARY G. KOCH
Department of Biostatistics, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7420, USA
Email: bcl@bios.unc.edu

SUMMARY

In many randomized multi-visit studies, some response variables have an ordinal scale for outcomes. The win odds (accounts for ties) and the win ratio (ignores ties) are useful for treatment comparisons for ordinal outcomes. This paper discusses the application of randomization-based covariance and stratification adjustment of the win odds (and win ratio) to enable their more convenient use. Adjustment for strata is through the weighted average of within stratum two-sample U statistics for numerators and denominators for the stratified win odds (or win ratio). As randomization-based, invocation of covariance adjustment is through constraints to zeros for baseline covariate differences in the joint vector with logarithms of stratified win odds (or win ratios) for the respective visits. Such adjustment has no formal assumptions about the distributions of response variables or covariates or the relationships of covariates to response variables; but the resulting adjusted stratified win odds (or win ratios) have narrower confidence intervals than their unadjusted counterparts when covariates have at least moderately strong associations with response variables. There is illustration of such results for the stratified win odds (and win ratio) for a randomized multi-visit clinical trial with an ordinal outcome for a chronic respiratory disorder.

Keywords and phrases: Goodman-Kruskal $Gamma$; Mann-Whitney probability; multivariate two-sample U statistics; Somers' D; stratified rank analysis of covariance; van Elteren statistic

---

⋆ Corresponding author
† Shared first authorship

1 Introduction

Many randomized multi-visit studies collect response variables for participants on an ordinal, rather than interval, measurement scale. An example is the randomized multi-visit clinical trial discussed in Section 2 and Section 4 for comparing test treatment to control for an ordinal global rating as terrible, poor, fair, good, and excellent by participants with a chronic respiratory disorder (Koch et al., 1989; Stokes et al., 2012). This ordinal, global rating has assessments at baseline and four follow-up visits. Also, this example has centers as a factor for stratified randomization; and the baseline global rating, age, and sex are covariates of interest for their associations with the ordinal response variables at the four follow-up visits. As indicated in the subsequent paragraph, the win ratio and the win odds are useful for describing the extent to which participants with the test treatment have better outcomes for an ordinal response variable than those with the control. This paper discusses randomization based methods for stratification and covariance adjustment for win ratios and win odds for the respective visits of randomized multi-visit studies in Section 2 and Section 3; and it illustrates their application to the previously described chronic respiratory disorder example in Section 2 and Section 4.

As discussed in Chapters 4 and 7 of Stokes et al. (2012), Gasparyan et al. (2021b), and many other references, the Wilcoxon rank sum statistic enables a randomization-based comparison between two randomized treatment groups for an ordinal response variable at a single visit for a study. The Mann-Whitney probability (i.e., win proportion WP) is a corresponding measure for the difference between the two treatment groups; and Kawaguchi et al. (2011), and perhaps other references, note that $(WP–0.5) = (\overline{R}_T - \overline{R}_C)/n$, where $\overline{R}_T$ and $\overline{R}_C$ are the mean ranks for the test (T) and control (C) treatment groups, with sample sizes n_T and n_C and $n = n_T + n_C$. Moreover, Chapter 4 of Stokes et al. (2012), as well as Dong et al. (2020) and Gasparyan et al. (2021b), note that $WP = (P(T > C)+0.5P(T = C))$ for the probability of a better outcome for a random participant on test treatment (T), with ties managed as half wins, compared to a random participant on control treatment (C). Equivalently, $2(WP - 0.5) = (Somers'\ D) = (P(T > C) - P(C > T)) = WD$ expresses the win difference WD; and $WO = WP/(1-WP) = (1+WD)/(1-WD)$ is the win odds. A related measure is the win ratio $WR = P(T > C)/P(C > T)$ as in Pocock et al. (2012); and since $(WR - WO)$ is proportional to WD, $WR > WO > 1$ when $WD > 0$ and $WR < WO < 1$ when $WD < 0$; and $WR = WO = 1.0$ when $WD = 0$ or $WR = WO$ if no ties. Also, $WR = (1 + Gamma)/(1 - Gamma)$, where $Gamma$ is the Goodman-Kruskal (Goodman and Kruskal, 1963, 1972) version of the Kendall tau rank correlation coefficient. Relatedly, simulation studies in Carr et al. (1989) indicate that the Fisher (1925, 1992) Z transformation of a correlation coefficient as applied to $Gamma$ as $Z = 0.5\log((1 + Gamma)/(1 - Gamma)) = 0.5\log(WR)$ has better statistical properties than $Gamma$ for control of Type I error.

For studies with stratified randomization and no other relevant baseline covariates, the van Elteren (1960) extension of the Wilcoxon rank sum statistic enables a randomization-based stratified comparison between two treatment groups; and references such as pages 220-222 in Lehmann (1975) note how it has a locally most powerful property. As noted in Kawaguchi et al. (2011), a weighted mean of the within-stratum $(WP_h - 0.5)$, where WP_h is the win proportion for the h-th stratum with $h = 1, 2, \ldots, q$ for q strata, is the corresponding measure in the numerator of the van Elteren test statistic for the difference between two treatment groups. A convenient expression for

this weighted mean is $(\overline{WP} - 0.5) = \sum_{h=1}^{q} w_h(WP_h - 0.5)$ where

$$w_h = \left(\frac{n_{Th}n_{Ch}}{n_{Th} + n_{Ch} + 1}\right) / \left(\sum_{h'=1}^{q} \frac{n_{Th'}n_{Ch'}}{n_{Th'} + n_{Ch'} + 1}\right)$$

is the van Elteren weight for the h-th stratum with n_{Th} and n_{Ch} as the sample sizes for the test and control treatment groups within the h-th stratum. Relatedly, $\overline{WD} = 2(\overline{WP} - 0.5) = \sum_{h=1}^{q} w_h WD_h$ and $\overline{WO} = (1 + \overline{WD})/(1 - \overline{WD})$ respectively are the win difference and the win odds that correspond to the van Elteren test statistic for the combined strata. Also, Kawaguchi et al. (2011) indicate that the logit transformation $\log\left(\frac{\overline{WP}}{1-\overline{WP}}\right) = \log(\overline{WO})$ can have better statistical properties than $\overline{WP}$ in terms of Type I error control and coverage of confidence intervals, particularly when $\overline{WP}$ is further from its null value of 0.5 (for no difference between treatments for their distributions of a response variable in all strata). In Carr et al. (1989), there is consideration of extensions of the Goodman-Kruskal $Gamma$ for the combined strata and related statistical tests for the homogeneity of $Gamma$ among the strata; and they are indirectly applicable to the win ratio via the previously noted Fisher Z transformation of $Gamma$ for a study with two treatment groups. More directly, Dong et al. (2018) provide methods for the stratified win ratio and a related test of homogeneity; and Section 14.6 of Stokes et al. (2012) illustrates Wald tests of homogeneity for the Mann-Whitney probability (i.e., WP).

The appendices in Koch et al. (1998) describe the general nature of randomization-based covariance adjustment, and their scope briefly includes the Mann-Whitney probability (i.e., WP). For this method, there is invocation of constraints to zeros for measures of treatment differences for baseline covariates that are in a joint vector that includes measures of treatment differences for response variables on the basis of randomization. Accordingly, this method has no formal assumptions about the distributions of response variables or baseline covariates or the relationships of covariates with response variables. As explained in Section 3, the resulting adjusted measures of treatment differences for the response variables have smaller estimates for standard errors than their unadjusted counterparts without such covariance adjustment when there are at least moderately strong associations between the response variables and the baseline covariates, and so they can provide narrower confidence intervals and better power for comparisons between treatments. Kawaguchi et al. (2011) more fully addresses randomization-based adjustment of the Mann-Whitney probability (i.e., WP) for randomized multi-visit clinical trials, and it describes applications to three examples.

Gasparyan et al. (2021a) discuss methods for stratification and numeric covariate adjustment for the win proportion (and thereby the win odds by transformation), although their discussion refers to the win ratio. Related clarification of terminology concerning the distinction between the win odds and the win ratio is in Dong et al. (2020), Brunner et al. (2021) and Gasparyan et al. (2021b), where the win odds is noted to be a win ratio with ties as half wins. Gasparyan et al. (2021b) also provide power and sample size calculation formulas for the win odds.

This paper summarizes in Section 3 and the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 how the methods in the Appendices of Koch et al. (1998) and Kawaguchi et al. (2011) enable randomization-based stratification and baseline covariate adjustment for the win ratio and win odds estimates for the visits in a randomized multi-visit clinical trial to compare a test treatment to a

control treatment. Methods for two-sample U statistics, together with linear Taylor series approximations, are used to estimate the corresponding covariance matrices of these estimates (with these methods being slightly more appropriate than the methods for one sample U statistics in Carr et al. (1989) and Kawaguchi et al. (2011)). The previously noted example with respect to a chronic respiratory disorder has illustration in Section 2 of some basic methods for comparing two treatments for ordinal response variables for studies with stratified randomization. Additionally, it has illustration of randomization-based methods for both stratification adjustment and covariance adjustment in Section 4 so as to enable their more convenient use. In Section 5, there is a brief discussion that summarizes the methods and provides comments for related topics.

2 Example with Results Adjusted for Strata

As indicated in Section 1, the chronic respiratory disorder example for describing the applications of the methods in this paper is a clinical trial with stratified randomization to test treatment or control at two centers for participants with a chronic respiratory disorder. The illustrated response variable is an ordinal global rating as terrible, poor, fair, good, or excellent at baseline and four follow-up visits; and age and sex are other baseline covariates of interest. The distributions of this ordinal global rating at follow-up visit 1 are shown in Table 1 for the participants with control and test treatment at the two centers.

Illustration is possible for basic methods for the win difference (WD), the win proportion (WP), the win odds (WO), and the win ratio (WR) through the use of the MEASURES option in the SAS FREQ Procedure (SAS Institute Inc., 2018) to produce Somers' D (as the win difference (WD)) and Goodman-Kruskal $Gamma$ for the comparisons between test treatment and control within the two centers; and similar capabilities for basic methods are available for other software. Accordingly, the first row of Table 2 provides estimates of Somers' D (and their standard errors) for Center 1 and Center 2 for the example; and the 5th row correspondingly provides estimates of Goodman-Kruskal $Gamma$ (and their standard errors). As noted in Section 1, $WP = (WD + 1)/2$, $WO = WP/(1 - WP)$, and $WR = (1 + Gamma)/(1 - Gamma)$; and they are also shown in Table 2 for Center 1 and Center 2 with the estimates for their corresponding standard errors (SE). In this regard, the determination of the estimates for their standard errors (SE) is possible though the straightforward use of linear Taylor Series methods for their relationships, as well as those for $\log(WR)$, to WD and $Gamma$. More specifically,

$$\begin{aligned} SE(WP) &= SE(WD)/2, SE(WO) = SE(WP)/(1-WP)^2, \\ SE(\log(WR)) &= 2\,SE(Gamma)/(1-Gamma^2)\ and \\ SE(WR) &= 2\,SE(Gamma)/(1-Gamma)^2. \end{aligned}$$

In accordance with Section 1, stratification adjusted comparisons between the test and control treatments are possible with weighted averages, $\overline{d} = \sum_{h=1}^{2} w_h d_h$ of within stratum comparisons d_h for the centers $h = 1, 2$ with van Elteren weights, $w_h = 0.504, 0.496$ so as to be compatible with the van Elteren test statistic. For the win difference, $\overline{WD}$, and the win probability, $\overline{WP}$, for the combined centers, the d_h are their within stratum counterparts, WD_h and $(WP_h - 0.5)$, respectively;

and for the win odds, $\overline{WO}$, for the combined centers, the transformation $\overline{WO} = (\overline{WP}/(1 - \overline{WP}))$ is applicable. The d_h for the win ratio are the $\log(WR_h)$, and the reverse transformation, $\widetilde{WR} = \exp(\overline{d})$, provides a win ratio for the combined strata. However, when the within stratum sample sizes for the two treatment groups are moderate rather than large (e.g., $15 \leq n_{Th}, n_{Ch} \leq 50$), the $\overline{WR}$ version of the win ratio from Expressions (A.1) and (A.2) in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 has more robust statistical behavior; and so it is the stratified win ratio for the methods for randomization-based covariance adjustment in Section 3 and their application in Section 4. In this regard, $\overline{WR}$ in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 is a ratio of weighted means across the strata for two-sample U statistics within strata; whereas $\widetilde{WR}$ is a geometric mean of within stratum win ratios WR_h based on one-sample U statistics for the corresponding Goodman-Kruskal $Gammas$.

Standard errors for $\overline{d}$ are produced for $\overline{WD}$, $\overline{WP}$, and $\log(\widetilde{WR})$ via $SE(\overline{d})$ $= \sqrt{\sum_{h=1}^{2}(w_h SE(d_h))^2}$; and they are shown in Table 2 for the example, with two-sided 95% confidence intervals, $\overline{d} \pm 1.96 \times SE(\overline{d})$. The confidence interval for $\widetilde{WR}$ is $\exp(\log(\widetilde{WR}) \pm 1.96 \times SE(\log(\widetilde{WR})))$; and that for $\overline{WO}$ is $\exp(\log(\overline{WO}) \pm 1.96 \times SE(\log(\overline{WO})))$ where $SE(\log(\overline{WO})) = SE(\overline{WP})/(\overline{WP}(1 - \overline{WP}))$ via linear Taylor series methods. Also, the $(\widetilde{WR} - 1)/(\widetilde{WR} + 1)$ transformation of the confidence interval for $\widetilde{WR}$ provides the confidence interval for $Gamma$ for the combined strata. Although the lower limits of the confidence intervals in Table 2 for $\overline{WD}$, $\overline{WP}$, and $\widetilde{WR}$ for the combined strata suggest that better global ratings are more likely for test treatment than control, this interpretation needs some caution since the van Elteren test statistic from SAS PROC FREQ (with center*treatment*response/cmh scores=modridits) has two-sided $p = 0.0524$.

In Section 3, methods for randomization-based adjustment for baseline covariates are described for the stratification adjusted win odds, $\overline{WO}$, and win ratio, $\overline{WR}$, for which more formal definitions are in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3. For the example in this section, the application of randomization based adjustment for baseline covariates is then illustrated in Section 4 with the baseline global rating, age, and sex as the covariates for the global ratings at the four follow-up visits.

3 Methods for Randomization-based Covariance Adjustment for the Win Ratio and the Win Odds

Let $\bar{\boldsymbol{x}}_{Th}$ and $\bar{\boldsymbol{x}}_{Ch}$ denote means for vectors of s numeric covariates (with no missing values) for test treatment and control for participants in the h-th stratum; and let $\boldsymbol{g} = \sum_{h=1}^{q} w_h(\bar{\boldsymbol{x}}_{Th} - \bar{\boldsymbol{x}}_{Ch})$ denote the vector of stratified differences between treatments for means of covariates with respect to van Elteren weights for the q strata of a stratification factor (as specified in Section 1 and illustrated for centers in Section 2). Also, in accordance with Section 2 and the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3, let $\boldsymbol{f} = (f_0, f_1, \ldots, f_r)'$ denote the vector of logarithms for the stratification adjusted win odds $\overline{WO}_j$ (or the win ratios $\overline{WR}_j$ as defined with Expression (A.1) and Expression (A.2) in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 where (A.3) and

(A.4) pertain to the $\overline{WO}_j$) for a randomized clinical trial with $(r+1)$ visits, where $j=0$ denotes the baseline visit and $j=1,2,\ldots,r$ denote r post-baseline follow-up visits; i.e., $f_j=\log(\overline{WO}_j)$ for the stratification adjusted win odds (or $f_j=\log(\overline{WR}_j)$ for the stratification adjusted win ratios). Let $\boldsymbol{F}=(\boldsymbol{g}',\boldsymbol{f}')'=(\boldsymbol{g}',f_0,\boldsymbol{f}_*')'$ denote the vector that jointly pertains to differences between treatments for both all pertinent baseline covariates as $(\boldsymbol{g}',f_0)'$ and logarithms of stratified win odds (or win ratios) at the r post-baseline follow-up visits as $\boldsymbol{f}_*$. For the example in Section 2, $r=4$ for the four post-baseline visits, and $\boldsymbol{F}=(g_1,g_2,f_0,f_1,f_2,f_3,f_4)'$, where g_1 corresponds to age and g_2 corresponds to sex (as 1 if male and 0 if female). More generally, f_0 can include additional ordinal covariates that pertain to the severity of the baseline status of a participant, and $\boldsymbol{g}$ can include categorical covariates as 0 or 1 indicators.

Let $\boldsymbol{V}_{\boldsymbol{F}}$ denote a consistent estimate of the covariance matrix for $\boldsymbol{F}$ from the methods in the Estimate for Covariance Matrix $\mathbf{V}_{\mathbf{F}}$ for $\mathbf{F}$ in Section 3. On the basis of stratified randomization of test and control treatments to the participants, constraints to zeros are applicable to $\boldsymbol{g}$ and f_0. Accordingly, the invocation of these constraints for $\boldsymbol{F}$ by weighted least squares with weights from $\boldsymbol{V}_{\boldsymbol{F}}^{-1}$ and $\boldsymbol{X}=[\mathbf{0}_{r,(s+1)},\boldsymbol{I}_r]'$ produces the randomization-based adjusted estimates $\boldsymbol{b}=(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{F}$ for the logarithms of the stratification adjusted win odds (or win ratios) for the r post-baseline follow-up visits; see the Appendices in Koch et al. (1998) for related discussion. Moreover, how $\boldsymbol{b}$ represents the covariance adjustment for $\boldsymbol{f}_*$ is more clearly evident with its simplified expression as $\boldsymbol{b}=(\boldsymbol{f}_*-\boldsymbol{V}_{\boldsymbol{F},12}'\boldsymbol{V}_{\boldsymbol{F},11}^{-1}[\boldsymbol{g}',f_0]')$ where $\boldsymbol{V}_{\boldsymbol{F},11}$ is the $(s+1)\times(s+1)$ upper left block of $\boldsymbol{V}_{\boldsymbol{F}}$ and $\boldsymbol{V}_{\boldsymbol{F},12}$ is the $(s+1)\times r$ upper right block of $\boldsymbol{V}_{\boldsymbol{F}}$.

A consistent estimator for the $(r\times r)$ covariance matrix $\boldsymbol{V}_{\boldsymbol{b}}$ for the covariance adjusted estimates for the logarithms of the stratification adjusted win odds (or the win ratios) is $\boldsymbol{V}_{\boldsymbol{b}}=(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{X})^{-1}=(\boldsymbol{V}_{\boldsymbol{f}_*}-\boldsymbol{V}_{\boldsymbol{F},12}'\boldsymbol{V}_{\boldsymbol{F},11}^{-1}\boldsymbol{V}_{\boldsymbol{F},12})$ where $\boldsymbol{V}_{\boldsymbol{f}_*}$ is the $(r\times r)$ lower right block of $\boldsymbol{V}_{\boldsymbol{F}}$ for the covariance matrix that corresponds to the logarithms of stratified win odds (or win ratios) for the r post-baseline visits. In this regard, the structure of $\boldsymbol{V}_{\boldsymbol{b}}$ expresses how linear functions $\boldsymbol{c}'\boldsymbol{b}$ have smaller estimated variance $\boldsymbol{c}'\boldsymbol{V}_{\boldsymbol{b}}\boldsymbol{c}$ than their unadjusted counterparts $\boldsymbol{c}'\boldsymbol{f}_*$ for which $\boldsymbol{c}'\boldsymbol{V}_{\boldsymbol{f}_*}\boldsymbol{c}$ is the estimated variance. This useful property of $\boldsymbol{b}$ applies without any assumptions for the distributions of the response variables or the covariates or the relationships of the covariates to the response variables; and so it is an important statistical property of randomization-based covariance adjustment.

As indicated in the Estimate for Covariance Matrix $\mathbf{V}_{\mathbf{F}}$ for $\mathbf{F}$ in Section 3, $\boldsymbol{F}$ has an approximately multivariate normal distribution on the basis of central limit theory when the within stratum sample sizes n_{Th} and n_{Ch} and their totals $n_T=\sum_{h=1}^q n_{Th}$ and $n_C=\sum_{h=1}^q n_{Ch}$ are sufficiently large. Accordingly, $\boldsymbol{b}$ has an approximately multivariate normal distribution (see Koch et al. (1977), Koch and Wiener (2017), and Chapter 14 of Stokes et al. (2012)). Thus, with v_{b_j} as the estimated variance of b_j from the j-th diagonal element of $\boldsymbol{V}_{\boldsymbol{b}}$ and $z_{\alpha/2}$ as the $100(1-\frac{\alpha}{2})$ quantile of the standard normal distribution with mean 0 and variance 1, $\exp\left(b_j\pm z_{\frac{\alpha}{2}}\sqrt{v_{b_j}}\right)$ provides a two-sided $100(1-\alpha)\%$ confidence interval from randomization-based covariance adjustment for the stratification adjusted win odds (or win ratio) at the j-th visit. Moreover, for $\boldsymbol{C}$ as a specified $(c\times r)$ matrix with full rank $c\leq r$, $Q_{\boldsymbol{Cb}}=\boldsymbol{b}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{V}_{\boldsymbol{b}}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{b}$ approximately has the chi-squared distribution with c degrees of freedom (d.f.) under the null hypothesis that $\boldsymbol{C}\boldsymbol{b}$ consistently estimates $\mathbf{0}_c$ as the $(c\times 1)$ vector of zeros. In this regard, with $\boldsymbol{C}=[\boldsymbol{I}_{(r-1)},\ -\mathbf{1}_{(r-1)}]$ where $\boldsymbol{I}_{(r-1)}$ and $\mathbf{1}_{(r-1)}$ respec-

tively are the $(r-1)\times(r-1)$ identity matrix and the $(r-1)\times 1$ vector of ones, $Q_{\boldsymbol{Cb}}$ provides a test statistic for homogeneity of the stratification adjusted win odds (or win ratios) across the r visits in the sense of no treatment*visit interaction. When homogeneity is reasonably applicable $Q_{\boldsymbol{Cb}}$ with $\boldsymbol{C} = \mathbf{1}_r'$ provides a test statistic for the geometric mean of the win odds (or win ratios) across the r visits equalling 1.0.

4 Results with Randomization-Based Covariance Adjustment for the Example

In accordance with the application of the methods in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 and Section 3 to the chronic respiratory disorder example in Section 2, Table 3 provides the estimates (and estimated standard errors) for the logarithms of the stratification adjusted win odds and win ratios via the f_j for the four post-baseline visits 1, 2, 3, 4; and it provides corresponding results which additionally have randomization-based covariance adjustment via the b_j (from constraints to zeros for $\boldsymbol{g}$ and f_0 in $\boldsymbol{F} = (\boldsymbol{g}', \boldsymbol{f}')'$). In this regard, the estimated standard errors $\sqrt{v_{f_j}}$ for the f_j (or $\sqrt{v_{b_j}}$ for the b_j) are square roots of the corresponding diagonal elements of $\boldsymbol{V_F}$ in Expression A.9 in the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 (or $\boldsymbol{V_b}$ in Section 3).

From the structure of the estimated covariance matrix $\boldsymbol{V_b}$ for $\boldsymbol{b}$ in Section 3, the estimated standard errors in Table 3 for the logarithms of the stratification adjusted win odds and win ratios for the four post-baseline visits are about 10% smaller with randomization-based covariance adjustment than without it. Moreover, this statistical property applies even though the estimated logarithms of the stratification adjusted win odds and win ratios for the four post-baseline visits from randomization-based covariance analysis via $\boldsymbol{b}$ are similar to those without it via $\boldsymbol{f}$. In this regard, randomization implies that $(\boldsymbol{b} - \boldsymbol{f}_*)$ is a consistent estimator of $\mathbf{0}_r$ as the $r \times 1$ vector of zeros via $(\boldsymbol{g}', f_0)'$ being a consistent estimator of $\mathbf{0}_{(s+1)}$. The previously noted interpretations are also applicable to the two-sided 95% confidence intervals in Figure 1 from $\exp(f_j \pm 1.96\sqrt{v_{f_j}})$ for the stratification adjusted win odds and win ratios without covariance adjustment and from $\exp(b_j \pm 1.96\sqrt{v_{b_j}})$ for those with covariance adjustment.

In Figure 1, the stratified win odds and the stratified win ratio are noticeably larger for Visit 2 than the other three visits. For this interpretation, there is support from corresponding test statistics $Q_{\boldsymbol{Cb}}$ with $\boldsymbol{C} = [\boldsymbol{I}_3,\ -\mathbf{1}_3]$ for treatment*visit interaction. Their results are $Q_{\boldsymbol{Cb}} = 9.12$ for the stratified win odds with covariance adjustment and $Q_{\boldsymbol{Cb}} = 8.18$ for the stratified win ratio with covariance adjustment, and so their corresponding p-values are 0.0277 and 0.0425, respectively, relative to the chi-squared distribution with d.f. = 3.

For both the f_j and the b_j in Table 3, those for the logarithms of the win ratios are consistently larger than their counterparts for the win odds. However, the estimated standard errors that pertain to the logarithms of the win ratios are correspondingly larger than those for the win odds in the sense that $z_{f_j} = f_j/\sqrt{v_{f_j}}$ (or $z_{b_j} = b_j/\sqrt{v_{b_j}}$) for the logarithms of the win ratios are reasonably similar to their counterparts for the win odds. In this regard, the (win ratio/win odds) ratios for the z_{f_j} are 1.001, 0.994, 1.001, 0.996 for the post-baseline visits 1, 2, 3, 4, respectively; and those

for the z_{b_j} are 1.013, 1.015, 1.017, 1.023 for the post-baseline visits 1, 2, 3, 4, respectively. Thus, for the comparisons between the test treatment and the control for the post-baseline visits 1, 2, 3, 4, the interpretation of the z_{f_j} (or the z_{b_j}) for the stratified win ratios would be similar to those for the stratified win odds; and such interpretations can be compatible with the extent to which the corresponding confidence intervals have lower limits above 1.0 (or upper limits below 1.0).

For clinical trials like the chronic respiratory disorder example for which the total sample size for the two treatment groups are in a moderate range, such as $40 \leq n_T,\ n_C \leq 100$, there may need to be some caution for the interpretation of the confidence intervals in Figure 1 (or the corresponding z_{f_j} (or the z_{b_j})) to evaluate whether better outcomes are more likely for the test treatment than control. For this purpose, a more rigorous method is stratified rank analysis of covariance as the extension of the van Elteren test statistic to have randomization-based covariance adjustment; and it is applicable with either the NParCov3 (Zink and Koch, 2012) or the NParCov4 (Zink et al., 2017) SAS macros. This method addresses $(\overline{WP} - 0.5)$ with the applicable covariance matrices determined under the strong null hypothesis for randomization-based methods (i.e., each participant has the same outcome regardless of their randomly assigned treatment). Specifications for the use of the NParCov4 macro include the RANK Procedure in SAS (SAS Institute Inc., 2018) with the NPLUS1 option to produce ranks within strata for the pooled treatment groups divided by the corresponding total sample size plus 1, HYPOTH=NULL, and the option COMBINE=FIRST for randomization-based covariance adjustment after stratification adjustment. The respective counterparts to the z_{b_j} from this method are 2.301, 4.313, 3.387, 2.561 for the four post-baseline visits. Thus, the respective z-values from stratified rank analysis of covariance are somewhat smaller than those from the z_{b_j}=2.362, 4.595, 3.630, 2.680 respectively for the win odds and the z_{b_j}=2.393, 4.663, 3.692, 2.742 respectively for the win ratio for the four post-baseline visits; and such tendencies can be helpful for understanding the extent to which the interpretation of the z_{b_j} may need some caution. Moreover, essentially exact randomization-based assessments of the z-values from stratified rank analysis of covariance are possible from their corresponding randomization distributions with respect to all possible stratified randomizations of participants to the test treatment or control. Nevertheless, the stratified win odds (or win ratio), together with their confidence intervals with randomization-based covariance adjustment, can be very helpful descriptively for interpreting the comparisons between the test treatment and the control for ordinal response variables in a randomized multi-visit study. In summary, a reasonable analysis strategy for ordinal response variables can be the parallel use of stratified rank analysis of covariance to evaluate rigorously whether better outcomes are more likely for test treatment than control, together with the description of such tendencies with the stratified win odds (or win ratio) by their confidence intervals from randomization-based covariance adjustment.

Table 1: Frequencies (and percentages) for outcomes of ordinal global rating according to treatment and center.

Center	Treatment	Terrible	Poor	Fair	Good	Excellent	Total
1	Control	4 (13.8%)	4 (13.8%)	9 (31.0%)	5 (17.2%)	7 (24.1%)	29
1	Test	1 (3.7%)	2 (7.4%)	10 (37.0%)	10 (37.0%)	4 (14.8%)	27
2	Control	0 (0%)	2 (7.1%)	10 (35.7%)	7 (25.0%)	9 (32.1%)	28
2	Test	0 (0%)	1 (3.7%)	3 (11.1%)	9 (33.3)	14 (51.9%)	27

Table 2: Estimates (and standard errors) for measures of difference between test and control treatment for Center 1, Center 2, and the combined centers and corresponding 95% confidence intervals for the combined centers.

	Center 1	Center 2	Combined Center	
Measure	Estimate (SE)	Estimate (SE)	Estimate (SE)	Confidence Interval
Somers' D (WD)	0.114 (0.150)	0.298 (0.140)	0.205 (0.102)	(0.004, 0.406)
Win Proportion (WP)	0.557 (0.075)	0.649 (0.070)	0.603 (0.051)	(0.502, 0.703)
Win Odds (WO)	1.257 (0.381)	1.847 (0.566)	1.515 (0.324)	(0.997, 2.304)
Goodman-Kruskal Gamma	0.148 (0.194)	0.421 (0.185)	0.289 (0.137)	(0.005, 0.530)
log(WR)	0.297 (0.396)	0.897 (0.448)	0.595 (0.299)	(0.009, 1.180)
Win Ratio (WR)	1.346 (0.533)	2.452 (1.099)	1.812 (0.541)	(1.009, 3.254)

Table 3: Estimates (and standard errors) for logarithms of the stratification adjusted win odds and win ratios for the combined centers in multi-visit clinical trial.

	Without Covariance Adjustment		With Covariance Adjustment	
Visit	Log(win odds)	Log(win ratio)	Log(win odds)	Log(win ratio)
1	0.416 (0.218)	0.569 (0.298)	0.437 (0.185)	0.603 (0.252)
2	0.931 (0.232)	1.256 (0.315)	0.965 (0.210)	1.315 (0.282)
3	0.675 (0.223)	0.903 (0.298)	0.726 (0.200)	0.982 (0.266)
4	0.494 (0.214)	0.692 (0.301)	0.528 (0.197)	0.754 (0.275)

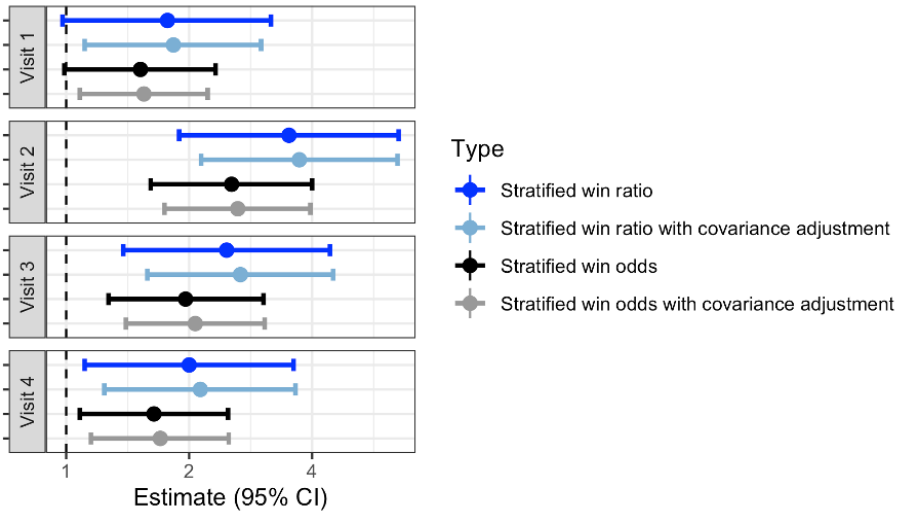

Figure 1: 95% Confidence Intervals (CIs) for the Stratified Win Ratio and Stratified Win Odds without and with Covariate Adjustment.

5 Discussion

This paper discusses randomization-based methods for stratification and baseline covariate adjustment for win odds and win ratio estimates for randomized multi-visit clinical trials with ordinal response variables and stratified randomization. As summarized in Section 3 and the Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3, the computations for the more widely applicable version of these methods (i.e., within-stratum sample sizes $n_{Th}, n_{Ch} \geq 15$) have four steps. The Estimate for Covariance Matrix $\mathbf{V_F}$ for $\mathbf{F}$ in Section 3 provides the first three steps, with the first being the construction of the vector for multivariate two-sample U statistics for the comparisons between the

test treatment and control within each stratum for both the numeric covariates and the ordinal response variables (as they pertain to the numerators and denominators for the win ratio and the win odds), together with the estimation of their corresponding covariance matrices. The second step is the determination of the weighted average across the strata for the vector of within-stratum two-sample U statistics with the use of van Elteren weights so as to correspond to the numerator of the van Elteren test statistic. The third step is the construction of the joint vector of both the differences between treatments for the means of covariates and the natural logarithms of the win ratios (or win odds) at each visit, together with the estimation of their covariance matrices. The last step is in Section 3, where randomization-based covariance adjustment has invocation through the use of weighted least squares methods to constrain treatment differences pertaining to baseline covariates to zeros. The chronic respiratory disorder example introduced in Section 2 has illustration of these methods in Section 4.

As related clarification, the stratification adjusted win ratio as $\exp(f_j)$ for the j-th visit is a consistent estimator for a population parameter which is $\theta_{WR_j} = (\sum_{h=1}^{q} w_h \theta_{1hj} / \sum_{h=1}^{q} w_h \theta_{2hj})$. For θ_{WR_j}, $\theta_{1hj} = P(y_{Thj} > y_{Chj})$ and $\theta_{2hj} = P(y_{Thj} < y_{Chj})$ for the ordinal response variables y_{Thj} and y_{Chj} at the j-th visit of a multi-visit clinical trial with $j = 0, 1, 2, \ldots, r$ by random participants with test treatment T and control C in the h-th stratum with $h = 1, 2, \ldots, q$; and the w_h are the van Elteren weights defined in Section 1. The population parameter for which the stratification adjusted win odds is a consistent estimator is $\theta_{WO_j} = (\sum_{h=1}^{q} w_h(\theta_{1hj} + 0.5\theta_{3hj}) / \sum_{h=1}^{q} w_h(\theta_{2hj} + 0.5\theta_{3hj}))$ where $\theta_{3hj} = P(y_{Thj} = y_{Chj})$. Moreover, the $\exp(b_j)$ for the win ratio and the win odds (as the randomization-based covariance adjusted counterparts of the $\exp(f_j)$) in Section 3 are correspondingly consistent estimators for the θ_{WR_j} and θ_{WO_j} (since $(\mathbf{b} - \mathbf{f}_*)$ is a consistent estimator of $\mathbf{0}_r$ as indicated in Section 4); and they can provide narrower confidence intervals than those without such covariance adjustment (as based on the f_j since $v_{b_j} \leq v_{f_j}$) when the covariates have at least moderately strong associations with the response variables.

In Section 4 for the chronic respiratory disorder example, the 95% confidence intervals for the covariance adjusted estimates are narrower compared to those that only have adjustment for strata. Also, the win odds estimates are closer to 1.0 than the win ratio estimates for the respective post-baseline visits, although the corresponding Z values for the comparisons between treatments are nearly identical.

Two SAS (SAS Institute Inc., 2018) macros, with one for the win ratio and one for the win odds, are conveniently available at github.com/elaineek/adj-wrwo for the methods provided in this paper for stratification adjustment and randomization-based covariance adjustment; see Kowalewski et al. (2023). An R package (R Core Team, 2021) with functions that utilize similar arguments to the SAS macro is also available at github.com/annweideman/winr. Both repositories contain the chronic respiratory disorder dataset used in Section 4 to demonstrate the methodology in Section 3 of this paper; and they also contain a dataset for a randomized multi-visit dermatology clinical trial as an additional example that has attention in both Kowalewski et al. (2023) and the documentation within the R package.

As additional considerations for the methods in this paper, Kawaguchi and Koch (2015) provide an R package R Core Team (2021), *sanon*, for the computation of stratified Mann-Whitney win prob-

abilities (WP) with randomization-based covariance adjustment; and this package provides several methods for managing missing data. They include managing missing data as tied with all observed data as well as for carried forward kernel functions for the within-stratum pairwise comparisons of each test treatment participant with each control participant. Sun et al. (2017) provide adaptions for the Mann-Whitney win probability (WP) as in (Kawaguchi et al., 2011) to multiple endpoints that are related to one another, and they also illustrate the use of such methods with multiple imputation for missing data. The previously noted methods of managing missing data in *sanon* (Kawaguchi and Koch, 2015), as well as multiple imputation, could be additions to the SAS macros and R package for the methods in this paper.

As indicated in Section 1, many randomized multi-visit clinical trials for the comparison of two treatments have response variables with ordinal outcomes. The discussion in this paper is potentially useful for such clinical trials through enabling the win odds and the win ratio for comparisons between a test treatment and control to have randomization-based adjustment for stratification factors and baseline covariates; and such adjustments have no assumptions for the distributions of response variables or covariates or the relationships of covariates to the response variables. Moreover, as indicated in Figure 1 in Section 4 for the chronic respiratory disorder example, the confidence intervals for the win odds and the win ratio from randomization-based covariance adjustment are narrower than those without such adjustment when the covariates have at least moderately strong associations with the response variables; and they are straightforward to interpret through the win odds or the win ratio.

A Appendix

A.1 Estimate for Covariance Matrix $\mathbf{V_F}$ for F in Section 3

In Section 3, the estimators $\overline{WR}_j$ for the win ratios and $\overline{WO}_j$ for the win odds for the respective visits $j = 0, 1, 2, \ldots, r$ of a multi-visit clinical trial are ratios of weighted means (across the strata with respect to the van Elteren weights in Section 1 of two sample U statistics (for which Puri and Sen (1971) is a reference). Accordingly, for $\boldsymbol{f}$ as the vector of their logarithms, $f_j = \log(\sum_{h=1}^{q} w_h U_{1hj} / \sum_{h=1}^{q} w_h U_{2hj})$. For the win ratio,

$$\begin{aligned} U_{1hj} &= \sum_{k=1}^{n_{Th}} \sum_{k'=1}^{n_{Ch}} [I(y_{Thjk} > y_{Chjk'})/n_{Th}n_{Ch}] \\ &= \sum_{k=1}^{n_{Th}} \sum_{k'=1}^{n_{Ch}} U_{1hjkk'}/n_{Th}n_{Ch} \end{aligned} \tag{A.1}$$

$$\begin{aligned} U_{2hj} &= \sum_{k=1}^{n_{Th}} \sum_{k'=1}^{n_{Ch}} [I(y_{Thjk} < y_{Chjk'})/n_{Th}n_{Ch}] \\ &= \sum_{k=1}^{n_{Th}} \sum_{k'=1}^{n_{Ch}} U_{2hjkk'}/n_{Th}n_{Ch} \end{aligned} \tag{A.2}$$

where y_{Thjk} and $y_{Chjk'}$ are random outcomes for the ordinal response variable at the j-th visit for the k-th test treatment participant and the k'-th control participant in the h-th stratum; also, $I(S)$ is the indicator function which equals 1 when S applies and equals 0 otherwise. Relatedly, (U_{1hj}/U_{2hj}) is the win ratio for the j-th visit for the h-th stratum. Alternatively, (U_{1hj}/U_{2hj}) becomes the win odds for the j-th visit for the h-th stratum by modifying (A.1) and (A.2) to replace the kernel function $U_{1hjkk'}$ and $U_{2hjkk'}$ by $U^*_{1hjkk'}$ and $U^*_{2hjkk'}$ in (A.3) and (A.4) and redefining the $\boldsymbol{U}_{1hj}$ and $\boldsymbol{U}_{2hj}$ accordingly.

$$U^*_{1hjkk'} = U_{1hjkk'} + 0.5I(y_{Thjk} = y_{Chjk'}) \tag{A.3}$$

$$U^*_{2hjkk'} = U_{2hjkk'} + 0.5I(y_{Thjk} = y_{Chjk'}) \tag{A.4}$$

With $\boldsymbol{U}_{\boldsymbol{x}hkk'} = (\boldsymbol{x}_{Thk} - \boldsymbol{x}_{Chk'})$ as the vector of differences between baseline covariates for the k-th test treatment participant and the k'-th control participant in the h-th stratum,

$$\begin{aligned} \boldsymbol{g} &= \sum_{h=1}^{q} w_h(\bar{\boldsymbol{x}}_{Th} - \bar{\boldsymbol{x}}_{Ch}) \\ &= \sum_{h=1}^{q} w_h \sum_{k=1}^{n_{Th}} \sum_{k'=1}^{n_{Ch}} \boldsymbol{U}_{\boldsymbol{x}hkk'}/n_{Th}n_{Ch} \\ &= \sum_{h=1}^{q} w_h \boldsymbol{U}_{\boldsymbol{x}h} \end{aligned} \tag{A.5}$$

in Section 3 is a weighted mean across the strata for two sample U statistics within the strata.

For the win ratio, let $\boldsymbol{U}_{1hkk'} = (U_{1h0kk'}, U_{1h1kk'}, \ldots, U_{1hrkk'})'$ and $\boldsymbol{U}_{2hkk'} = (U_{2h0kk'}, U_{2h1kk'}, \ldots, U_{2hrkk'})'$ with $\boldsymbol{U}_{hkk'} = (\boldsymbol{U}'_{\boldsymbol{x}hkk'}, \boldsymbol{U}'_{1hkk'}, \boldsymbol{U}'_{2hkk'})'$, then

$$\boldsymbol{U}_h = \sum_{k=1}^{n_T h} \sum_{k'=1}^{n_C h} \frac{\boldsymbol{U}_{hkk'}}{n_{Th}n_{Ch}} = (\boldsymbol{U}'_{\boldsymbol{x}h}, \boldsymbol{U}'_{1h}, \boldsymbol{U}'_{2h})', \tag{A.6}$$

where $\boldsymbol{U}_{1h} = (U_{1h0}, U_{1h1}, \ldots, U_{1hr})'$ and $\boldsymbol{U}_{2h} = (U_{2h0}, U_{2h1}, \ldots, U_{2hr})'$. Since $\boldsymbol{U}_h$ is a $(s + 2(r+1))$ vector of two sample U statistics, a consistent estimator for its covariance matrix is $\boldsymbol{V}_h$ in (A.7).

$$\begin{aligned} \boldsymbol{V}_h = &\left\{ \frac{1}{n_{Th}(n_{Th}-1)} \sum_{k=1}^{n_{Th}} (\boldsymbol{U}_{hk*} - \boldsymbol{U}_h)(\boldsymbol{U}_{hk*} - \boldsymbol{U}_h)' \right\} \\ &+ \left\{ \frac{1}{n_{Ch}(n_{Ch}-1)} \sum_{k'=1}^{n_{Ch}} (\boldsymbol{U}_{h*k'} - \boldsymbol{U}_h)(\boldsymbol{U}_{h*k'} - \boldsymbol{U}_h)' \right\}, \end{aligned} \tag{A.7}$$

where $\boldsymbol{U}_{hk*} = (\sum_{k'=1}^{n_{Ch}} \boldsymbol{U}_{hkk'}/n_{Ch})$ and $\boldsymbol{U}_{h*k'} = (\sum_{k=1}^{n_{Th}} \boldsymbol{U}_{hkk'}/n_{Th})$; see Brunner and Munzel (2000), Gasparyan et al. (2021a), and Puri and Sen (1971) as related references. For $\boldsymbol{V}_h$, the underlying assumption is that the participants in each treatment group within the h-th stratum hypothetically represent a corresponding population of similar participants in a sense comparable to

a simple random sample; and for this assumption, the corresponding populations for the two treatment groups have the same distributions for the baseline covariates as a consequence of the stratified randomization of the participants to the test treatment and control groups.

Let $\boldsymbol{U} = \sum_{h=1}^{q} w_h \boldsymbol{U}_h = (\boldsymbol{U}'_{\boldsymbol{x}*}, \boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*})'$ with w_h as the van Elteren weights specified in Section 1. Let (A.8), with $\boldsymbol{A} = [\boldsymbol{I}_{(r+1)},\ -\boldsymbol{I}_{(r+1)}]$, specify the relationship between $\boldsymbol{F}$ in Section 3 and $\boldsymbol{U}$.

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{x}*} \\ \boldsymbol{A} \log(\boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*})' \end{bmatrix}. \tag{A.8}$$

It then follows that a consistent estimator for the covariance matrix of $\boldsymbol{F}$ from methods for multivariate linear Taylor series approximations is $\boldsymbol{V}_{\boldsymbol{F}}$ in (A.9).

$$\boldsymbol{V}_{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{I}_s & \boldsymbol{0}_{s,2(r+1)} \\ \boldsymbol{0}_{(r+1),s} & \boldsymbol{A}\boldsymbol{D}^{-1} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{I}_s & \boldsymbol{0}_{s,2(r+1)} \\ \boldsymbol{0}_{(r+1),s} & \boldsymbol{A}\boldsymbol{D}^{-1} \end{bmatrix}'. \tag{A.9}$$

For $\boldsymbol{V}_{\boldsymbol{F}}$, $\boldsymbol{D}$ is a diagonal matrix with $[\boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*}]'$ as its diagonal elements, $\boldsymbol{0}_{(r+1),s}$ is an $(r+1) \times s$ matrix of zeros, and $\boldsymbol{0}_{s,2(r+1)}$ is a $s \times 2(r+1)$ matrix of zeros; see Koch et al. (1977), Koch and Wiener (2017), and Chapter 14 of Stokes et al. (2012) as related references. When the sample sizes n_{Th} and n_{Ch} within the strata and their totals $n_T = \sum_{h=1}^{q} n_{Th}$ and $n_C = \sum_{h=1}^{q} n_{Ch}$ for the combined strata are sufficiently large (e.g., all $n_{Th}, n_{Ch} \geq 15$ and $n_T, n_C \geq 40$), $\boldsymbol{U}$ has an approximately multivariate normal distribution via central limit theorems for U statistics (see Puri and Sen (1971)); and so $\boldsymbol{F}$ has an approximately multivariate normal distribution for studies with stratified randomization. Similarly, (A.8) and (A.9) can produce $\boldsymbol{F}$ and $\boldsymbol{V}_{\boldsymbol{F}}$ that correspond to the win odds by modifying the respective elements of $\boldsymbol{U}_{1hkk'}$ and $\boldsymbol{U}_{2hkk'}$ that pertain to (A.6) and (A.7) to be the $U^*_{1hjkk'}$ and $U^*_{2hjkk'}$ in (A.3) and (A.4) rather than the $U_{1hjkk'}$ and $U_{2hjkk'}$ in (A.1) and (A.2).

References

Brunner, E. and Munzel, U. (2000), "The nonparametric Behrens-Fisher problem: asymptotic theory and a small sample approximation," *Biometrical Journal*, 42, 17–25.

Brunner, E., Vandemeulebroecke, M., and Mütze, T. (2021), "Win odds: an adaptation of the win ratio to include ties," *Statistics in Medicine*, 40, 3367–3384.

Carr, G. J., Hafner, K. B., and Koch, G. G. (1989), "Analysis of rank measures of association for ordinal data from longitudinal studies," *Journal of the American Statistical Association*, 84, 797–804, doi: 10.1080/01621459.1989.10478840.

Dong, G., Hoaglin, D. C., Qiu, J., Matsouaka, R. A., Chang, Y.-W., Wang, J., and Vandemeulebroecke, M. (2020), "The win ratio: on interpretation and handling of ties," *Statistics in Biopharmaceutical Research*, 12, 99–106, doi: 10.1080/19466315.2019.1575279.

Dong, G., Qiu, J., Wang, D., and Vandemeulebroecke, M. (2018), "The stratified win ratio," *J Biopharm Stat*, 28, 778–796.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, London: Oliver and Boyd.

— (1992), "Statistical Methods for Research Workers," in *Breakthroughs in Statistics: Methodology and Distribution*, eds. Kotz, S. and Johnson, N. L., New York, NY: Springer New York, pp. 66–70.

Gasparyan, S. B., Folkvaljon, F., Bengtsson, O., Buenconsejo, J., and Koch, G. G. (2021a), "Adjusted win ratio with stratification: Calculation methods and interpretation," *Statistical Methods in Medical Research*, 30, 580–611.

Gasparyan, S. B., Kowalewski, E. K., Folkvaljon, F., Bengtsson, O., Buenconsejo, J., Adler, J., and Koch, G. G. (2021b), "Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials," *Journal of Biopharmaceutical Statistics*, 1–23.

Goodman, L. A. and Kruskal, W. H. (1963), "Measures of association for cross classifications III: approximate sampling theory," 58, 310.

— (1972), "Measures of association for cross classifications, IV: simplification of asymptotic variances," *Journal of the American Statistical Association*, 67, 415–421.

Kawaguchi, A. and Koch, G. G. (2015), "sanon: an R package for stratified analysis with nonparametric covariable adjustment," *Journal of Statistical Software*, 67, 1 – 37.

Kawaguchi, A., Koch, G. G., and Wang, X. (2011), "Stratified multivariate Mann–Whitney estimators for the comparison of two treatments with randomization based covariance adjustment," *Statistics in Biopharmaceutical Research*, 3, 217–231, doi: 10.1198/sbr.2010.10007.

Koch, G., Carr, G., Amara, I., Stokes, M., and Uryniak, T. (1989), "Categorical Data Analysis," in *Statistical Methodology in the Pharmaceutical Sciences*, eds. Berry, D. and Dekker, M., New York: CRC Press, pp. 389–473.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., and Lehnen, R. G. (1977), "A general methodology for the analysis of experiments with repeated measurement of categorical data," *Biometrics*, 33, 133.

Koch, G. G., Tangen, C. M., Jung, J. W., and Amara, I. A. (1998), "Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them," *Stat Med*, 17, 1863–92.

Koch, G. G. and Wiener, L. E. (2017), "Chi-Squared Tests: Basics," in *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons, Ltd, pp. 1–20.

Kowalewski, E. K., Weideman, A. M. K., and Koch, G. G. (2023), "SAS macro for randomization-based methods for covariance and stratified adjustment of win ratios and win odds for ordinal outcomes," Cary, N.C.: SAS Institute, Inc., Proceedings of the SouthEast SAS Users Group 2023 Conference.

Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, Oxford, England: Holden-Day.

Pocock, S. J., Ariti, C. A., Collier, T. J., and Wang, D. (2012), "The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities," *Eur Heart J*, 33, 176–82.

Puri, M. and Sen, P. (1971), *Nonparametric Methods in Multivariate Analysis*, New York: Wiley.

R Core Team (2021), *R: A Language and Environment for Statistical Computing*.

SAS Institute Inc. (2018), *SAS/STAT(R): 15.1 User's Guide*, Cary, NC.

Stokes, M., Davis, C., and Koch, G. (2012), *Categorical Data Analysis using SAS*, Cary, NC: SAS Institute, 3rd ed.

Sun, H., Kawaguchi, A., and Koch, G. (2017), "Analyzing multiple endpoints in a confirmatory randomized clinical trial—an approach that addresses stratification, missing values, baseline imbalance and multiplicity for strictly ordinal outcomes," *Pharmaceutical Statistics*, 16, 157–166.

van Elteren, P. H. (1960), "On the combination of independent two-sample tests of Wilcoxon," *Bulletin of the International Statistical Institute*, 37, 351–361.

Zink, R., Koch, G., Chung, Y., and Wiener, L. (2017), "Chapter 2: Advanced randomization-based methods," in *Analysis of Clinical Trials Using SAS*, eds. Dmitrienko, A. and Koch, G., Cary, NC: SAS Institute, Inc., 2nd ed.

Zink, R. C. and Koch, G. G. (2012), "NParCov3: A SAS/IML Macro for Nonparametric Randomization-Based Analysis of Covariance," *Journal of Statistical Software*, 50, 1 – 17.

Received: December 6, 2023

Accepted: May 12, 2024