SESUG Paper 139-2023

SAS Macro for Randomization-Based Methods for Covariance and Stratified Adjustment of Win Ratios and Win Odds for Ordinal Outcomes

Elaine K. Kowalewski, Ann Marie K. Weideman, Gary G. Koch, Department of Biostatistics, University of North Carolina at Chapel Hill

ABSTRACT

In a randomized clinical trial, participant responses to treatments may be measured on an ordinal, rather than interval, scale. The trial may also have a stratified design, such as randomization within clinics, and information may be collected across multiple visits to evaluate the treatment efficacy. Due to randomization, random imbalances for baseline measurements and covariates between treatment groups are expected to be minimal. Nevertheless, baseline covariates may be strongly associated with the outcome, and so adjustment for baseline covariates can improve the power for assessing the treatment effect. The win ratio (ignores ties) and the win odds (accounts for ties) can be useful when analyzing these types of clinical trial data. This work provides a SAS macro which implements randomization-based methodology for covariance and stratified adjustment of the win ratio and the win odds for ordinal outcomes from a multi-visit clinical trial with stratified randomization. Large and small sample within strata methodology is made available. The software is illustrated for two multi-visit clinical trials with ordinal outcomes.

INTRODUCTION

A multi-visit clinical trial may collect participant responses to treatment on an ordinal, rather than interval, measurement scale. The trial may also have a stratified design, such as randomization within clinics, and information may be collected across multiple visits to evaluate the treatment efficacy. Due to randomization, random imbalances for baseline measurements and covariates between treatment groups are expected to be minimal. Nevertheless, baseline covariates may be strongly associated with the outcome, and so adjustment for baseline covariates can improve the power for assessing the treatment effect. The win ratio (ignores ties) and the win odds (accounts for ties) can be useful when analyzing these types of clinical trial data. A SAS macro for these methods is illustrated for two multi-visit clinical trials with ordinal outcomes, for which one has stratified randomization and the other has missing data.

METHODS

Methods for an ordinal endpoint include the Mann-Whitney probability (win proportion, WP), the win ratio (WR), and the win odds (WO), defined as follows:

$$WP = P(T > C) + 0.5P(T = C) \quad (1)$$

$$WR = P(T > C)/P(C > T) \quad (2)$$

$$WO = WP/(1 - WP) \quad (3)$$

In the above, $P(T > C)$ is the probability of a better outcome for a patient on test treatment (T) compared to a patient on control treatment (C); and for (1), ties are managed as half wins. Of note, $(WR - 1)/(WR + 1)$ equals the Goodman-Kruskal gamma, a measure of rank correlation, and $(WO - 1)/(WO + 1) = (2 \times WP + 1) = [P(T > C) - P(C > T)]$ equals the Somers' D, a measure of association between two ordinal variables, in this case outcome and assigned treatment. The win ratio was popularized by Pocock et al. (2012) for analyzing composite outcomes in clinical trials based on clinical priorities.

STRATIFICATION ADJUSTMENT

Let w_h denote the van Elteren (1960) weight for the h-th stratum in (4), where the sum of the weights across strata is 1 for n_{hT} participants in the h-th stratum of the treatment arm and n_{hC} participants in the h-th stratum of the control arm.

$$w_h = \frac{\left\{\frac{n_{hT}n_{hC}}{n_{hT}+n_{hC}+1}\right\}}{\left\{\sum_{h'=1}^{q}\frac{n_{h'T}n_{h'C}}{n_{h'T}+n_{h'C}+1}\right\}} \tag{4}$$

Then, for random outcomes y_{Thij} and $y_{Chi'j}$ for patients i and i' in the test treatment and control arms attending their j-th visit and where $I(\cdot)$ is the indicator function for whether $(\cdot)$ applies or not, the win ratio (5) and win odds (6) can be estimated at each visit.

$$\widehat{WR}_j = \frac{\sum_{h=1}^{q}\frac{w_h}{n_{hT}n_{hC}}\sum_{i=1}^{n_{hT}}\sum_{i'=1}^{n_{hC}} I\left(y_{Thij} > y_{Chi'j}\right)}{\sum_{h=1}^{q}\frac{w_h}{n_{hT}n_{hC}}\sum_{i=1}^{n_{hT}}\sum_{i'=1}^{n_{hC}} I\left(y_{Thij} < y_{Chi'j}\right)} \tag{5}$$

$$\widehat{WO}_j = \frac{\sum_{h=1}^{q}\frac{w_h}{n_{hT}n_{hC}}\sum_{i=1}^{n_{hT}}\sum_{i'=1}^{n_{hC}}\left[I\left(y_{Thij} > y_{Chi'j}\right) + 0.5I\left(y_{Thij} = y_{Chi'j}\right)\right]}{\sum_{h=1}^{q}\frac{w_h}{n_{hT}n_{hC}}\sum_{i=1}^{n_{hT}}\sum_{i'=1}^{n_{hC}}\left[I\left(y_{Thij} < y_{Chi'j}\right) + 0.5I\left(y_{Thij} = y_{Chi'j}\right)\right]} \tag{6}$$

In this regard, there is adjustment for strata through the weighted average of within stratum two-sample U statistics for numerators and denominators for both the win ratio and win odds.

RANDOMIZATION-BASED COVARIANCE ADJUSTMENT

For randomized multi-visit clinical trials, there can be covariance adjustment of the win ratio and win odds (or their stratified counterparts) by constraining baseline covariate differences to zero in the joint vector with logarithms of win ratios or win odds. 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 ratios or win odds have narrower confidence intervals than their unadjusted counterparts when covariates have at least moderately strong associations with response variables. Methodology for covariance adjustment for the stratified win ratio and win odds is provided in the Appendix.

INTRODUCTION TO EXAMPLES

RESPIRATORY DISORDER

The first illustrative example is for a dataset from a randomized clinical trial comparing a test treatment to control in the treatment of a chronic respiratory disorder (Koch et al., 1989; Stokes et al., 2012). In this trial, 111 patients (54 active, 57 placebo) at two centers were evaluated at baseline and four follow-up visits, and their respiratory status was assessed at every visit using an ordinal global rating (0 for terrible, 1 for poor, 2 for fair, 3 for good, and 4 for excellent). The two centers correspond to a stratification factor; and baseline covariables for patients enrolled in the respiratory study are age, sex, and baseline respiratory status.

Table 1 summarizes the number randomized at each center within each treatment arm and provides descriptive statistics for distributions of baseline covariates for the respiratory disorder dataset.

	Arm		
	Treatment (N=54)	**Control (N=57)**	**Total (N=111)**
Age			
N	54	57	111
Mean (SD)	32.9 (14.0)	33.6 (13.5)	33.3 (13.7)
Median	30.0	35.0	31.0
Range	11.0, 68.0	11.0, 66.0	11.0, 68.0
Sex, n (%)			
F	6 (11.1)	17 (29.8)	23 (20.7)
M	48 (88.9)	40 (70.2)	88 (79.3)
Center, n (%)			
1	27 (50.0)	29 (50.9)	56 (50.5)
2	27 (50.0)	28 (49.1)	55 (49.5)
Baseline, n (%)			
0=Terrible	3 (5.6)	0 (0.0)	3 (2.7)
1=Poor	9 (16.7)	11 (19.3)	20 (18.0)
2=Fair	18 (33.3)	20 (35.1)	38 (34.2)
3=Good	13 (24.1)	19 (33.3)	32 (28.8)
4=Excellent	11 (20.4)	7 (12.3)	18 (16.2)

Table 1. Descriptive Statistics for Baseline Covariates for the Respiratory Disorder Dataset

SKIN CONDITIONS DISORDER

The second illustrative example is for a dataset from a randomized clinical trial comparing a test treatment to control for skin conditions (Stanish et al., 1978). In this trial, 172 patients (88 test, 84 placebo) at six clinics were evaluated at three follow-up visits, and their extent of improvement for their skin condition was recorded on a five-point scale (1 for rapidly improving, 2 for slowly improving, 3 for stable, 4 for slowly worsening, and 5 for rapidly worsening). Since clinic 9 only enrolled 4 patients, patients in clinics 8 and 9 are pooled. This pooling is further justified by clinics 8 and 9 having the smallest stratum sample sizes. The baseline covariable for patients enrolled in the skin conditions study is disease stage recorded at baseline (3 = Fair, 4 = Poor, 5 = Exacerbation). Unlike the respiratory dataset, this dataset is subject to missing data at the follow-up visits, and the extent of the missing data increases at each visit, with 3 (2%) missing observations at visit 1, 16 (9%) at visit 2, and 30 (17%) at visit 3.

Table 2 summarizes the number randomized at each clinic by treatment arm and provides descriptive statistics for the distribution of the baseline covariate for the skin conditions disorder dataset.

	Arm		
	Treatment (N=88)	**Placebo (N=84)**	**Total (N=172)**
Initial Disease Stage, n (%)			
3 = Fair	40 (45.5)	41 (48.8)	81 (47.1)
4 = Poor	40 (45.5)	35 (41.7)	75 (43.6)
5 = Exacerbation	8 (9.1)	8 (9.5)	16 (9.3)
Clinic[1], n (%)			
5	19 (21.6)	18 (21.4)	37 (21.5)
6	17 (19.3)	16 (19.0)	33 (19.2)
8-9	18 (20.5)	16 (19.0)	34 (19.8)
10	18 (20.5)	17 (20.2)	35 (20.3)
11	16 (18.2)	17 (20.2)	33 (19.2)

Table 2. Descriptive Statistics for Baseline Covariates for the Skin Condition Study

[1]Clinic 8-9 pools data from clinics 8 and 9 since clinic 9 enrolled only 4 patients.

MACRO INTRODUCTION

Two SAS macros are provided to implement the methodology described in Methods and in the Appendix. %Adj_WinRatio computes the randomization-based stratified adjusted win ratio, while %Adj_WinOdds is the counterpart for the win odds. The arguments for each macro are the same, so they will only be described for %Adj_WinRatio, but %Adj_WinOdds follows similarly. The call to %Adj_WinRatio is described as follows with details for each argument provided in Table 3:

```
%Adj_WinRatio(  DSNIN,
                DSNOUT,
                PID,
                OUTCOMES,
                ARM,
                BASELINE = NONE,
                COVARS = NONE,
                STRATA = NONE,
                METHOD = SMALL,
                DEBUG = 0);
```

Argument	Description
DSNIN	SAS dataset containing the analysis data. Must be in wide format such that a participant's repeated responses are in a single row and each response is a separate column.
DSNOUT	Name for output dataset.
PID	Variable corresponding to unique participant ID.
OUTCOMES	List of the variables (each separated by a space) corresponding to outcomes measured at each visit. The outcomes must have at least an ordinal measurement scale with larger values being better than smaller values. Thus, the outcome can be ordered categories or continuous measurements or dichotomies such as 0 or 1 or "no" or "yes."
ARM	Variable for treatment arm. Required to be a positive integer such that the test treatment arm is ALWAYS higher in value than the control arm.
BASELINE	Variable corresponding to outcome measurement at baseline. If not specified, no baseline adjustment is employed (which is default).
COVARS	List of the variables corresponding to the covariates (measured at baseline) to be used for adjustment. These covariates must be numeric, and can be measured on a binary, categorical, ordered categorical, or continuous scale. If not specified, no covariate adjustment is employed (which is default).
STRATA	Variable used for stratification. If not specified, no stratification is utilized (which is default).
METHOD	SMALL or LARGE used to denote the method employed. The small sample size method is recommended unless within-stratum sample size is reasonably large (e.g., >= 50), number of visits is small (e.g., <=6), and number of covariates is small (e.g., <=4). If not specified, the default is SMALL.
DEBUG	0 does not print analysis details to the log and 1 prints analysis details to the log. If not specified, the default is 0.

Table 3. Arguments for Macros %Adj_WinRatio and %Adj_WinOdds

APPLICATIONS TO ILLUSTRATIVE DATASETS

RESPIRATORY DISORDER DATASET

For the respiratory disorder dataset, suppose we wish to compute the win ratio at each of the four follow-up visits without any adjustment for stratification or baseline covariates, the call to `%Adj_WinRatio` would be as follows:

```
%Adj_WinRatio(  DSNIN = RESP,
                DSNOUT = OUT,
                PID = UniqID,
                OUTCOMES = Visit1 Visit2 Visit3 Visit4,
                ARM = Trt);
```

Table 4 provides the results produced by %Adj_WinRatio without any adjustment for stratification or baseline covariates for the respiratory dataset.

Visit	log(WR)	SE log(WR)	Chi-square	P-value	WR	WR 95% CI
Visit1	0.507	0.293	2.99	0.084	1.66	(0.93,2.95)
Visit2	1.218	0.308	15.66	<.001	3.38	(1.85,6.18)
Visit3	0.906	0.297	9.31	0.002	2.47	(1.38,4.43)
Visit4	0.629	0.286	4.85	0.028	1.88	(1.07,3.28)

Table 4. Unadjusted Win Ratio Results for the Respiratory Disorder Dataset

Suppose instead we wish to estimate the win ratio at each of the four follow-up visits with adjustment for center as a stratification factor and randomization-based adjustment for baseline as the baseline measurement and age and sex as baseline covariates. For this purpose, the call to %Adj_WinRatio would be as follows:

```
%Adj_WinRatio(  DSNIN = RESP,
                DSNOUT = OUT,
                PID = UniqID,
                OUTCOMES = Visit1 Visit2 Visit3 Visit4,
                ARM = Trt,
                BASELINE = Baseline,
                COVARS = Age SexNum,
                STRATA = Center);
```

Table 5 provides the fully adjusted (i.e., stratification adjusted with randomization-based adjustment for baseline as the baseline measurement and age and sex as baseline covariates) results produced by %Adj_WinRatio.

Visit	log(WR)	SE log(WR)	Chi-square	P-value	WR	WR 95% CI
Visit1	0.603	0.252	5.71	0.017	1.83	(1.11,3.00)
Visit2	1.315	0.282	21.74	<.001	3.72	(2.14,6.47)
Visit3	0.982	0.266	13.61	<.001	2.67	(1.58,4.50)
Visit4	0.754	0.275	7.52	0.006	2.13	(1.24,3.64)

Table 5. Fully Adjusted Win Ratio Results for the Respiratory Disorder Dataset

In comparing Table 4 and Table 5, we note that the standard error estimates for the log of the win ratios at each follow-up visit are smaller in Table 5 which yields smaller p-values and narrower confidence intervals for the fully adjusted win ratios compared to the unadjusted win ratios.

For the fully adjusted win odds at each of the four follow-up visits, the call to %Adj_WinOdds would be as follows:

```
%Adj_WinOdds(   DSNIN = RESP,
                DSNOUT = OUT,
                PID = UniqID,
                OUTCOMES = Visit1 Visit2 Visit3 Visit4,
                ARM = Trt,
                BASELINE = Baseline,
                COVARS = Age SexNum,
                STRATA = Center);
```

Table 6 provides the results produced by %Adj_WinOdds.

Visit	log(WO)	SE log(WO)	Chi-square	P-value	WO	WO 95% CI	WP
Visit1	0.437	0.185	5.57	0.018	1.55	(1.08,2.22)	0.607
Visit2	0.965	0.210	21.10	<.001	2.63	(1.74,3.96)	0.724
Visit3	0.726	0.200	13.13	<.001	2.07	(1.40,3.06)	0.674
Visit4	0.528	0.197	7.17	0.007	1.70	(1.15,2.50)	0.629

Table 6. Fully Adjusted Win Odds Results for the Respiratory Disorder Dataset

SKIN CONDITIONS DISORDER DATASET

For the skin conditions disorder dataset, suppose we wish to compute the win ratio at each of the three follow-up visits with adjustment for center as a stratification factor and randomization-based adjustment for stage as a baseline covariate. For this purpose, the call to %Adj_WinRatio would be as follows:

```
%Adj_WinRatio(  DSNIN = SKIN,
                DSNOUT = OUT,
                PID = ID,
                OUTCOMES = R1 R2 R3,
                ARM = Trt,
                COVARS = Stage,
                STRATA = Center);
```

Table 7 provides the results produced by %Adj_WinRatio.

Visit	log(WR)	SE log(WR)	Chi-square	P-value	WR	WR 95% CI
R1	1.937	0.301	41.35	<.001	6.94	(3.85,12.52)
R2	2.349	0.344	46.75	<.001	10.48	(5.34,20.55)
R3	2.383	0.37	41.45	<.001	10.84	(5.25,22.39)

Table 7. Fully Adjusted Win Ratio Results for the Skin Disorder Dataset

As mentioned previously, the skin conditions dataset includes missing data at follow-up visits, which is managed in the macro by introducing additional ties for missing data (i.e., missing values are managed as tied with observed values). Kawaguchi & Koch (2015) provide an R package, sanon, for stratified analysis with nonparametric covariable adjustment. For the skin dataset, sanon was invoked (with neither stratification nor covariable adjustment) and the same method for managing missing data via the use of 'replace' with the following call:

```
sanon(    cbind(R1, R2, R3) ~ grp(Trt, ref = "test"),
          data = skin,
          res.na.action = "replace")
```

Output 1 provides the results produced by sanon for this call.

```
Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "test"), data = skin, res.na.action = "replace")


     Estimate Std.Err Chisq Pr(>Chisq)
res1   0.2901  0.0328  78.0     <2e-16 ***
res2   0.2838  0.0288  97.2     <2e-16 ***
res3   0.2349  0.0278  71.4     <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Note that the estimates of responses are for the (MW estimate - 0.5).
```

Output 1. Output from Sanon R Package for the Skin Disorder Dataset

Table 8 provides the corresponding results (with neither stratification nor covariable adjustment) produced by %Adj_WinOdds for the corresponding call:

```
%Adj_WinOdds(   DSNIN = SKIN,
                DSNOUT = OUT,
                PID = ID,
                OUTCOMES = R1 R2 R3,
                ARM = Trt);
```

Visit	log(WO)	SE log(WO)	Chi-square	P-value	WO	WO 95% CI	WP
R1	1.326	0.199	44.55	<.001	3.76	(2.55,5.56)	0.790
R2	1.288	0.170	57.19	<.001	3.62	(2.60,5.06)	0.784
R3	1.020	0.143	50.77	<.001	2.77	(2.09,3.67)	0.735

Table 8. Unadjusted Win Odds Results for the Skin Disorder Dataset

Note that the results in the “WP” column from the results for the %Adj_WinOdds macro are exactly 0.5 larger than the results in the “Estimate” column from sanon, in correspondence to the note that sanon provides as “estimates of responses are for the (MW estimate – 0.5),” where “MW estimate” means Mann-Whitney probability or win proportion (WP). Moreover, the following SAS code takes the results produced by %Adj_WinOdds in OUT and produces the corresponding standard error, chi-square value, and p-value for (WP-0.5) to compare to the results from sanon:

```
data out_sanon;
    set out;
    SE_WP = SE_logWO * WP * (1-WP);
    Chi_Square_WP = ((WP-0.5) / SE_WP)**2;
    p_WP = 1-probchi(Chi_Square_WP, 1);
run;
```

The corresponding results for the win proportion are shown in Table 9. A reason why SE WP and Chi-square (WP) in Table 9 differ slightly from their counterparts in Figure 1 from sanon is that those in Table 9 are based on estimated variances corresponding to the use of two-sample U statistics as described in the Appendix whereas those from sanon are based on estimated variances corresponding to one-sample U statistics.

Visit	WP	SE WP	Chi-square (WP)	P-value (WP)
R1	0.790	0.033	77.58	<.001
R2	0.784	0.029	96.68	<.001
R3	0.735	0.028	70.99	<.001

Table 9. Unadjusted Win Proportion Results for the Skin Disorder Dataset

Although the chi-square statistics in Table 9 that pertain to the win probability are much larger than those that pertain to the win odds (in Table 8), they essentially have the same interpretation in the sense of having two-sided p<0.001 for contradicting the null hypothesis of no difference between the two treatments. In this regard, simulations in Carr et al. (1989) and discussion in Kawaguchi et al. (2011) support that the statistical properties of methods pertaining to the natural logarithm of the win odds are better than those pertaining to the win probability for the intended coverage of confidence intervals and control of Type I error.

CONCLUSIONS

Many multi-visit randomized clinical trials for the comparison of two treatments have ordinal outcomes. The presented methods are potentially useful for such clinical trials through enabling the comparisons between the two treatments to have adjustment for stratification factors and baseline covariates; and for the illustrated examples, the fully adjusted CIs were narrower and had lower limits further above 1.0 in comparison to the other methods. Moreover, the results from such comparisons can be straightforward to interpret through the win ratio or the win odds; and although all estimates of the win ratios for the illustrated example exceed those for the corresponding win odds, interpretations for the extent to which lower limits of the CIs exceed 1.0 are similar for the win ratios and the win odds.

REFERENCES

Carr, G. J., Hafner, K. B., & Koch, G. G. (1989). Analysis of Rank Measures of Association for Ordinal Data from Longitudinal Studies. *Journal of the American Statistical Association*, *84*(407), 797-804. https://doi.org/10.1080/01621459.1989.10478840

Kawaguchi, A., & Koch, G. G. (2015). sanon: An R Package for Stratified Analysis with Nonparametric Covariable Adjustment. *Journal of Statistical Software*, *67*(9), 1 - 37. https://doi.org/10.18637/jss.v067.i09

Kawaguchi, A., Koch, G. G., & Wang, X. (2011). Stratified Multivariate Mann–Whitney Estimators for the Comparison of Two Treatments with Randomization Based Covariance Adjustment. *Statistics in Biopharmaceutical Research*, *3*(2), 217-231. https://doi.org/10.1198/sbr.2010.10007

Koch, G. G., Carr, G. J., Amara, I. A., Stokes, M. E., & Uryniak, T. J. (1989). Categorical Data Analysis. In D. A. Berry & M. Dekker (Eds.), *Statistical Methodology in the Pharmaceutical Sciences* (pp. 389-473). CRC Press.

Koch, G. G., Tangen, C. M., Jung, J. W., & 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*(15-16), 1863-1892. https://doi.org/10.1002/(sici)1097-0258(19980815/30)17:15/16<1863::aid-sim989>3.0.co;2-m

Pocock, S. J., Ariti, C. A., Collier, T. J., & 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*(2), 176-182. https://doi.org/10.1093/eurheartj/ehr352

Puri, M., & Sen, P. (1971). *Nonparametric methods in multivariate analysis*. Wiley.

Stanish, W. M., Gillings, D. B., & Koch, G. G. (1978). An Application of Multivariate Ratio Methods for the Analysis of a Longitudinal Clinical Trial with Missing Data. *Biometrics*, *34*(2), 305. https://doi.org/10.2307/2530021

Stokes, M. E., Davis, C. S., & Koch, G. G. (2012). *Categorical data analysis using SAS* (Vol. 3rd edition). SAS Institute.

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

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Elaine Kowalewski
elainekkowalewski@gmail.com

The macros described in this paper are available on GitHub: https://github.com/elaineek/adj-wrwo.

APPENDIX

Estimators are produced for a vector of two-sample U statistics for each stratum in (7),

$$\boldsymbol{U}_h = \frac{1}{n_{hT} n_{hC}} \sum_{i=1}^{n_{hT}} \sum_{i'=1}^{n_{hC}} \left(\boldsymbol{U}'_{\boldsymbol{x}hii'}, \boldsymbol{U}'_{1hii'}, \boldsymbol{U}'_{2hii'}\right)' \tag{7}$$

where $\boldsymbol{U}_{\boldsymbol{x}hii'} = (\boldsymbol{x}_{hTi} - \boldsymbol{x}_{hCi'})$ is a vector of differences in baseline covariates; for the win ratio, $\boldsymbol{U}_{1hii'} = (U_{1hii'0}, U_{1hii'1}, \ \ldots, \ U_{1hii'r})'$ is a vector of indicators $U_{1hii'j} = I(y_{Thij} > y_{Chi'j})$, and $\boldsymbol{U}_{2hii'} = (U_{2hii'0}, U_{2hii'1}, \ \ldots, \ U_{2hii'r})'$ is a vector of indicators $U_{2hii'j} = I\left(y_{Thij} < y_{Chi'j}\right)$ for visits $j = 0, \ldots, r$; whereas, $0.5I(y_{Thij} = y_{Chi'j})$ is added to the $\boldsymbol{U}_{1hii'}$ and the $\boldsymbol{U}_{2hii'}$ for the win odds.

The estimated covariance matrix $\boldsymbol{V}_h$ for $\boldsymbol{U}_h$ is computed as in (8),

$$\begin{aligned} \boldsymbol{V}_h = &\left\{ \frac{1}{n_{hT}(n_{hT}-1)} \sum_{i=1}^{n_{hT}} (\boldsymbol{U}_{hi*} - \boldsymbol{U}_h)(\boldsymbol{U}_{hi*} - \boldsymbol{U}_h)' \right\} \\ &+ \left\{ \frac{1}{n_{hC}(n_{hC}-1)} \sum_{i'=1}^{n_{hC}} (\boldsymbol{U}_{h*i'} - \boldsymbol{U}_h)(\boldsymbol{U}_{h*i'} - \boldsymbol{U}_h)' \right\} \end{aligned} \tag{8}$$

where $\boldsymbol{U}_{hi*} = (\sum_{i'=1}^{n_{hC}} \boldsymbol{U}_{hii'}/n_{hC})$ and $\boldsymbol{U}_{h*i'} = (\sum_{i=1}^{n_{hT}} \boldsymbol{U}_{hii'}/n_{hT})$ for $\boldsymbol{U}_{hii'} = \left(\boldsymbol{U}'_{\boldsymbol{x}hii'}, \boldsymbol{U}'_{1hii'}, \boldsymbol{U}'_{2hii'}\right)'$ for patients i and i' in the test treatment and control arms. Adjustment for strata is produced by constructing $\sum_{h=1}^{q} w_h \boldsymbol{U}_h = (\boldsymbol{U}'_{\boldsymbol{x}*}, \boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*})'$ and the corresponding estimate $\boldsymbol{V} = \sum_{h=1}^{q} w_h^2 \boldsymbol{V}_h$ for its covariance matrix.

Randomization based covariance adjustment (Koch et al., 1998) is applicable to $\boldsymbol{F}$ in (9) by using weighted least squares methods to fit the linear model in (9),

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{x}*} \\ \boldsymbol{Alog}_{\boldsymbol{e}}([\boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*}]') \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{x}*} \\ \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{x}*} \\ f_0 \\ \boldsymbol{f}_* \end{bmatrix} \widehat{=} \begin{bmatrix} \boldsymbol{0}_s \\ 0 \\ \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0}_{(s+1),r} \\ \boldsymbol{I}_r \end{bmatrix} \boldsymbol{b} = \boldsymbol{L}\boldsymbol{b} \tag{9}$$

where $\boldsymbol{U}_{x*} = \sum_{h=1}^{q} w_h(\overline{\boldsymbol{x}}_{hT} - \overline{\boldsymbol{x}}_{hC})$, $\boldsymbol{A} = [\boldsymbol{I}_{(r+1)}, \quad -\boldsymbol{I}_{(r+1)}]$, and "≙" denotes "is estimated by." In (9), $\boldsymbol{Alog_e}([\boldsymbol{U}'_{1*}, \boldsymbol{U}'_{2*}]')$ is the vector of natural logarithms of the $\widehat{WR}_j$ in (5) as the win ratios (or the $\widehat{WO}_j$ in (6) as the win odds). On the basis of randomization, the expectation of the difference in sample means of baseline covariates in the test treatment and control arm within the h-th stratum is 0, i.e., $E[\boldsymbol{U}_{x*}] = \boldsymbol{0}_s$. Also, the asymptotic expected value for f_0 is 0. A consistent estimator $\boldsymbol{V_F}$ for the covariance matrix of $\boldsymbol{F}$ can be derived via methods for multivariate linear Taylor series approximations as in (10),

$$\boldsymbol{V_F} = \begin{bmatrix} \boldsymbol{I}_s & \boldsymbol{0}_{s,2(r+1)} \\ \boldsymbol{0}_{(r+1),s} & \boldsymbol{AD}^{-1} \end{bmatrix} V \begin{bmatrix} \boldsymbol{I}_s & \boldsymbol{0}_{s,2(r+1)} \\ \boldsymbol{0}_{(r+1),s} & \boldsymbol{AD}^{-1} \end{bmatrix}' \tag{10}$$

for s baseline covariate, r post-baseline visits, and $\boldsymbol{D}$ a diagonal matrix with the respective elements of $[\boldsymbol{U}'_{1*}, \quad \boldsymbol{U}'_{2*}]'$ as its diagonal elements.

Accordingly (Koch et al., 1998), the $(r \times 1)$ vector $\boldsymbol{b}$ of the covariance adjusted stratified estimators for the $\log_e(WR_j)$ for the combined strata is given in (11), and a consistent estimator for its $(r \times r)$ covariance matrix $\boldsymbol{V_b}$ is given in (12),

$$\boldsymbol{b} = (\boldsymbol{L}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{F} = \boldsymbol{f}_* - \boldsymbol{V}'_{\boldsymbol{F},12}\boldsymbol{V}_{\boldsymbol{F},11}^{-1}\begin{bmatrix} \boldsymbol{U}_{x*} \\ f_0 \end{bmatrix} \tag{11}$$

$$\boldsymbol{V_b} = (\boldsymbol{L}'\boldsymbol{V}_{\boldsymbol{F}}^{-1}\boldsymbol{L})^{-1} = \boldsymbol{V}_{\boldsymbol{f}_*} - \boldsymbol{V}'_{\boldsymbol{F},12}\boldsymbol{V}_{\boldsymbol{F},11}^{-1}\boldsymbol{V}_{\boldsymbol{F},12} \tag{12}$$

where $\boldsymbol{V}_{\boldsymbol{F},11}$ is the $(s+1) \times (s+1)$ upper left hand part for $\boldsymbol{V_F}$, $\boldsymbol{V}_{\boldsymbol{F},12}$ is the $(s+1) \times r$ upper right hand part of $\boldsymbol{V_F}$, and $\boldsymbol{V}_{\boldsymbol{f}_*}$ is the $(r \times r)$ lower right hand part of $\boldsymbol{V_F}$. When the overall sample size $n = \sum_{h=1}^{q} n_h$ for the combined strata is sufficiently large (e.g., all $n_h \geq 20$ and $n \geq 100$), $\boldsymbol{b} = (b_1, \ldots, b_r)'$ approximately has a multivariate normal distribution via central limit theorems for U statistics (Puri & Sen, 1971) and a $100(1-\alpha)\%$ confidence interval (CI) can be constructed as in (13).

$$CI = \exp\left(b_j \pm Z_{\frac{\alpha}{2}}\sqrt{v_{b_j}}\right) \tag{13}$$

One can specify a contrast matrix, $\boldsymbol{C}$, with dimension $c \times r$ and full rank c to test the linear hypothesis $H_0: \boldsymbol{Cb} \triangleq \boldsymbol{0}$ with the chi-squared statistic in (14) with c degrees of freedom. This statistic can be calculated directly from $\boldsymbol{b}$ and $\boldsymbol{V_b}$ as provided by the macros, although it is not in scope for the macros.

$$Q_{\boldsymbol{Cb}} = \boldsymbol{b}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{V_b}\boldsymbol{C}')^{-1}\boldsymbol{Cb} \tag{14}$$