Stratified Multivariate Mann-Whitney Estimators for the Comparison of Two Treatments with Randomization Based Covariance Adjustment

Atsushi Kawaguchi, Gary G. Koch, and Xiaofei Wang

Biostatistics Center, Kurume University, Fukuoka, Japan
Department of Biostatistics, University of North Carolina, Chapel Hill, NC
Duke University Medical Center, Durham, NC

Many Randomized Clinical Trials Have Stratified Designs

1. Multi-center clinical trials with centers as strata for which there is separate and independent randomization
2. Clinical trials with strata according to one or more baseline characteristics of patients
3. Stratification is applicable to both multi-visit clinical trials with parallel groups and to crossover clinical trials with sequences of periods for one or more treatments

Many clinical trials have strictly ordinal response variables

1. Ordered categories
   a. Ratings of pain severity as none, mild, moderate, and severe
   b. Ratings of control for respiratory symptoms as excellent, good fair, poor, and terrible

2. Essentially continuous determinations for which an interval scale is not applicable (e.g., a narrow range for better outcomes at one end versus a wide range for poorer outcomes at the opposite end)
   a. Visual analogue scales for assessments of pain from osteoarthritis
   b. Number of painful joints for rheumatoid arthritis
   c. Ratings of well being subsequent to ischemic stroke

Primary analyses for confirmatory clinical trials need protocol specified methods that have minimal assumptions

1. Mann-Whitney (or Wilcoxon) statistics for comparisons between two groups through ranks of response variables have minimal assumptions in contrast to proportional odds models
2. Randomization-based covariance adjustment through constraints for no differences between groups for means of covariables have minimal assumptions in contrast to regression models with covariables as explanatory factors
3. Stratification adjustment through weighted averages of differences between groups for means of response variables has minimal assumptions (and sample size requirements) in contrast to a regression model including the stratification as an explanatory factor
4. Management of missing data through ranking principles can have weaker assumptions than other methods for missing data

Mann-Whitney estimators

1. Addresses the comparison between two groups through the proportion of pairwise comparisons of each member of group 1 with each member of group 2 that have better response for the member from group 1 (with ties managed as 0.5 in favor of group 1)

2. Its difference from 0.5 equals the difference between the mean ranks for the two groups divided by the pooled sample size for the two groups

3. Twice its value minus one equals the Somer's version of the Kendall tau for the association between groups and ordinal response (with ties for members in the same group ignored)

Stratified extension of the Mann-Whitney (Wilcoxon) test statistic for the comparison between two groups

1. The van Elteren test statistic addresses $d = \sum_{h=1}^{q} w_h \left(\bar{R}_{h1} - \bar{R}_{h2}\right) / n_h$ where $h = 1, 2, \ldots, q$ indexes strata, $\bar{R}_{h1}$ and $\bar{R}_{h2}$ are mean ranks for groups 1 and 2 in strata h, $\mathrm{w}_h = n_{h1} n_{h2} / (n_h + 1)$ with n_{h1} and n_{h2} as sample sizes for groups 1 and 2, and $n_h = (n_{h1} + n_{h2})$
2. The randomization based variance for d is $v_{(d,0)} = \sum_{h=1}^{q} (n_{h1} n_{h2} v_{h,0} / n_h)$ where $v_{h,0} = \sum_{i=1}^{2} \sum_{l=1}^{n_{hi}} \left(\frac{R_{hil}}{(n_h + 1)} - 0.5\right)^2 / (n_h - 1)$ with R_{hil} as the rank for the l-th patient in the i-th group of the h-th stratum (with larger ranks for better response)
3. The van Elteren test statistic $Q_{VE} = (d^2 / v_{d,0})$ approximately has the chi-squared distribution with d.f.=1when overall sample sizes for the two groups are sufficiently large (e.g., $n_{+i} = \sum_{h=1}^{q} n_{hi} \geq 30$ with at least minimal sample sizes $n_{h1} \geq 2$ within each stratum)

Advantages of Q_{VE} as primary method for the stratification adjusted comparison between two randomized groups for a strictly ordinal response variable

1. Has no assumptions (beyond valid stratified randomization) and is applicable to any sample size configuration for groups in strata

2. Can have exact assessment through its randomization distribution (via enumeration or simulation)

3. Locally most powerful, particularly for an essentially continuous distribution (with no ties for ranks within strata)

Limitations of Q_{VE}

1. Does not have an inherent counterpart for confidence interval estimation for a strictly ordinal response variable, mainly because the randomization based variance $v_{d,0}$ for d is only applicable under the global null hypothesis H_0 of no differences between the groups in all of the strata

2. Does not have a convenient multivariate extension in situations with possibly missing data for one or more of the responsible variables

Stratified Mann-Whitney estimator

1. Addresses same stratified comparison between two groups as the van Elteren test statistic $\hat{\xi} = (\sum_{h=1}^{q} w_h \hat{\xi}_h / \sum_{h=1}^{q} w_h)$ with $w_h = (n_{h1} n_{h2}) / (n_h + 1)$ and $(\hat{\xi}_h - 0.5) = (\bar{R}_{h1} - \bar{R}_{h2}) / n_h$ so that $(\hat{\xi}_h - 0.5) = d / \sum_{h=1}^{q} w_h$.

2. For an essentially continuous response variable (with no ties among the ranks within the respective strata, $v_{h,0} = n_h / 12(n_h + 1)$ and so the $\hat{\xi}_h$ have variances $(n_h + 1) / 12(n_{h1} n_{h2}) = 1/12w_h$. Accordingly, $\hat{\xi}$ is a weighted average of the $\hat{\xi}_h$ with the reciprocals of their variances under H_0 as the weights

Stratified multivariate Mann-Whitney estimators

1. Multivariate extension is $\hat{\boldsymbol{\xi}} = (\hat{\xi}_1, ..., \hat{\xi}_r)'$ for r response variables, each of which can have one or more (completely at random) missing values $\hat{\xi}_k = (\sum_{h=1}^{q} w_{hk} \hat{\xi}_{hk} / \sum_{h=1}^{q} w_{hk})$ with $w_{hk} = (n_{h1k} n_{h2k}) / (n_{hk} + 1)$ and $(\hat{\xi}_{hk} - 0.5) = (\bar{R}_{h1k} - \bar{R}_{h2k}) / n_{hk}$ for $k = 1, 2, ..., r$

2. Consistent estimator for covariance matrix $\mathbf{V}_{\hat{\xi}}$ can have construction via methods for ratios of multivariate $\mathrm{U}-$statistics

Methods for stratified multivariate Mann-Whitney estimators

1. Sufficient overall sample size for each group (e.g., $n_{+ik} = \sum_{h=1}^{q} n_{hik} \geq 50$ and all n_{hik}) enables $\hat{\boldsymbol{\xi}}$ to have an approximately multivariate normal distribution with covariance matrix $\mathbf{V}_{\hat{\xi}}$
2. A two$-$sided $(1-2d)$ confidence interval for $\mathbf{c}'\hat{\boldsymbol{\xi}}$ for the comparison between the two groups is $\left(\mathbf{c}'\hat{\boldsymbol{\xi}} \pm z_a\sqrt{\mathbf{c}'\mathbf{V}_{\hat{\xi}}\mathbf{c}}\right)$ where z_a is 100(1-α) percentile of the standard normal distribution with mean 0 and variance 1
   a. $\mathbf{c} = \boldsymbol{\delta}_\mathbf{k}$ where $\boldsymbol{\delta}_\mathbf{k}$ has k-th element as 1 and all others as 0 so as to address k-th response variable
   b. $\mathbf{c} = \mathbf{1}'r / r$ so as to address average across response variables
3. A test statistic for $\mathbf{C}\hat{\boldsymbol{\xi}} = 0$ is $Q_C = \hat{\boldsymbol{\xi}}\mathbf{C}'(\mathbf{C}\mathbf{V}_{\hat{\xi}}\mathbf{C}')^{-1}\mathbf{C}\hat{\boldsymbol{\xi}}$ it has an approximately chi-squared distribution with d.f. = Rank($\mathbf{C}$). With $\mathbf{C} = [\mathbf{I}_{(r-1)}, \mathbf{-1}_{(r-1)}]$ homogeneity is addressed (i.e., no group x response variable interaction).

Randomization based covariance adjustment

1. Let $m = 1, 2, \ldots M$ index M numeric covariables with observation prior to randomization of patients to groups (with scope including dichotomies)

2. Let $g_m = \sum_{h=1}^{q} \widetilde{w}_h (\bar{x}_{h1m} - \bar{x}_{h2m}) / (\sum_{h=1}^{q} \widetilde{w}_h)$ with $\widetilde{w}_h = n_{h1} n_{h2} / (n_{h1} + n_{h2})$ and $\bar{x}_{him}$ as the mean of the m-th covariable for the i-th group in the h-th stratum

3. Let $\mathbf{g} = (g_1, g_2, \ldots, g_M)'$ and let $\mathbf{f} = (\hat{\boldsymbol{\xi}}', g')'$. A consistent estimator for the covariance matrix $\mathbf{V_f}$ for $\mathbf{f}$ can have construction by methods for ratios of multivariate U−statistics

4. Since $\mathbf{g}$ would be expected to be null on the basis of randomization, one can fit $P = [\mathbf{I}_r, \mathbf{0}_{rM}]'$ to $\mathbf{f}$ by weighted least squares

5. The covariance adjusted counterpart $\mathbf{b}$ for $\hat{\boldsymbol{\xi}}$ is $\mathbf{b} = (\mathbf{P}'\mathbf{V_f^{-1}P})^{-1}\mathbf{P}'\mathbf{V_f^{-1}f} = (\hat{\boldsymbol{\xi}} - \mathbf{V}_{\hat{\boldsymbol{\xi}}\mathbf{g}}\mathbf{V_g^{-1}g})$ with covariance matrix $\mathbf{V_b} = (\mathbf{P}'\mathbf{V_f^{-1}P}) = (\mathbf{V}_{\hat{\boldsymbol{\xi}}} - \mathbf{V}_{\hat{\boldsymbol{\xi}}\mathbf{g}}\mathbf{V_g^{-1}}\mathbf{V}'_{\hat{\boldsymbol{\xi}}\mathbf{g}})$ where $\mathbf{V}_{\hat{\boldsymbol{\xi}}\mathbf{g}}$ corresponds to the covariances of $\hat{\boldsymbol{\xi}}$ with $\mathbf{g}$ and $\mathbf{V_g}$ is the covariance matrix of $\mathbf{g}$.

Example 1: Randomized clinical trial for chronic pain with univariate ordinal response

1. Cross-classification of two centers (as I, II) and four diagnoses (as A, B, C, D) produces 8 strata

2. Range of sample sizes for the strata is 10 to 34

3. Pain status after 4 weeks of treatment (as excellent, good, moderate, fair, poor) is the ordinal response variable for the comparison of test and control treatments

Table 1: Distributions of Pain Status After Treatment for 4 Weeks According to Center, Diagnosis and Treatment

Stratum	Center	Diagnosis	Treatment	Pain Status After Treatment for 4 Weeks				
				Excellent	Good	Moderate	Fair	Poor
1	I	A	Test	1	3	2	5	1
1	I	A	Control	2	4	3	4	3
2	I	B	Test	3	10	1	4	2
2	I	B	Control	2	4	1	5	2
3	I	C	Test	6	1	1	1	0
3	I	C	Control	0	5	1	1	3
4	I	D	Test	3	5	1	6	1
4	I	D	Control	3	3	2	4	5
5	II	A	Test	0	4	3	1	8
5	II	A	Control	0	3	3	0	5
6	II	B	Test	2	3	3	0	2
6	II	B	Control	1	8	0	0	5
7	II	C	Test	2	2	1	0	1
7	II	C	Control	1	1	0	1	1
8	II	D	Test	0	1	2	2	3
8	II	D	Control	1	1	1	0	7

Table 2. Mann-Whitney Estimators with Standard Errors and Sample Sizes for Strata According to Center and Diagnosis

Stratum	Center	Diagnosis	Sample Size	Mann-Whitney estimator	Standard Error
1	I	A	28	0.492	0.106
2	I	B	34	0.595	0.096
3	I	C	19	0.839	0.092
4	I	D	33	0.601	0.096
5	II	A	27	0.469	0.105
6	II	B	24	0.529	0.115
7	II	C	10	0.604	0.181
8	II	D	18	0.600	0.126

Van Elteren Statistic $Q_{UE} = 3.89$, $p = 0.0486$

Stratified Mann-Whitney $\hat{\xi} = 0.5804$, $s.e. = 0.0417$

0.95 Confidence Interval for $\hat{\xi}$ is $(0.4988,\ 0.6621)$

Assessments of homogeneity are possible

1. Criterion is $Q_H = \sum_{\tilde{h}} \left\{ \left(\tilde{\xi}_{\tilde{h}} - \tilde{\xi} \right) \Big/ s.e.\left(\tilde{\xi}_{\tilde{h}} \right) \right\}^2$ with $\tilde{\xi}$ as weighted mean of $\tilde{\xi}_{\tilde{h}}$ with $\left(1 \Big/ s.e.\left(\tilde{\xi}_{\tilde{h}} \right) \right)^2$ as weights; Q_H approximately as chi-squared distribution with $d.f. = \left[\dim\left(\tilde{h} - 1 \right) \right]$.
2. For the two centers, $\tilde{\xi}_{\mathrm{I}*} = 0.536 \pm 0.064$ and $\tilde{\xi}_{\mathrm{II}*} = 0.612 \pm 0.054$; and so $Q_H = 0.88$ with $d.f. = 1$ and $p = 0.348$.
3. For the four diagnoses, $\tilde{\xi}_{*\mathrm{A}} = 0.601 \pm 0.077, \tilde{\xi}_{*\mathrm{B}} = 0.762 \pm 0.097,$ $\tilde{\xi}_{*\mathrm{C}} = 0.567 \pm 0.075,$ and $\tilde{\xi}_{*\mathrm{D}} = 0.481 \pm 0.075$; and so $Q_H = 5.39$ with $d.f. = 3$ and $p = 0.146$.

A proportional odds model with main effects for treatments, centers, and diagnoses has

1. $p = 0.072$ for the score test for the proportional odds assumption $d.f. = 15$ from SAS PROC LOGISTIC

2. The estimated odds ratio is 1.82 with (1.09, 3.05) as the 0.95 confidence interval and two-sided $p = 0.022$

3. An expanded model has $p = 0.032$ from treatment*diagnosis interaction, but $p = 0.018$ from the score test for the proportional odds assumption with $d.f. = 24$ (although small counts distort this result)

Example 2: Randomized clinical trial with ordinal responses at four visits for a respiratory disorder

1. Cross-classification of two centers and gender (as female or male) produces 4 strata for 111 patients

2. Patient global ratings of symptom control (as excellent, good, fair, poor, terrible) at 4 post-baseline visits are the 4 response variables for the comparison of test and control treatments

3. The baseline rating of symptom control is an ordinal covariable and age is a numeric covariable

The Stratified Mann-Whitney estimators for baseline and the four post-baseline visits and the stratified difference between mean ages with corresponding covariance matrix are as follows:

$$\mathbf{f} = [\hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4, g]' = [0.480, 0.600, 0.714, 0.656, 0.617, 1.050]$$

$$\mathbf{V_f} = \begin{bmatrix} 32.04 & 15.23 & 8.82 & 8.46 & 8.79 & 175.93 \\ 15.23 & 28.39 & 14.14 & 13.82 & 13.07 & -3.32 \\ 8.82 & 14.14 & 23.31 & 16.37 & 16.28 & -162.93 \\ 8.46 & 13.82 & 16.37 & 28.24 & 20.78 & -152.88 \\ 8.79 & 13.07 & 16.28 & 20.78 & 27.70 & -0.54 \\ 175.93 & -3.32 & -162.93 & -152.88 & -0.54 & 68220.05 \end{bmatrix} \times 10^{-4}$$

The criterion for random imbalance for

$\widehat{\xi}_0$ and g is $Q_0 = 0.33$ with $d.f. = 2$, $p = 0.8498$

The covariance and stratification adjusted Mann-Whitney estimators as departures from 0.5 and their covariance matrix are as follows:

$$\mathbf{b} = [0.1116, 0.2230, 0.1625, 0.1219]'$$

$$\mathbf{V_b} = \begin{bmatrix} 21.14 & 9.65 & 9.63 & 8.93 \\ 9.65 & 20.30 & 13.55 & 13.80 \\ 9.63 & 13.55 & 25.31 & 18.26 \\ 8.93 & 13.80 & 18.26 & 25.35 \end{bmatrix} \times 10^{-4}$$

With $\left\{ b_k / s.e.(b_k) \right\}^2$ as approximately having the chi-squared distribution with $d.f. = 1$, $p = 0.0152$, < 0.0001, 0.0012, and 0.0155 for Visits 1, 2, 3, 4. For assessment of homogeneity, $Q_c = 8.93$ with $d.f. = 3$ and $p = 0.0302$.

Results from other methods for Example 2

1. The p-values from van Elteren test statistics with centers*gender as strata are 0.0662, <0.0001, 0.0054, 0.0325, for visits 1, 2, 3, 4.
2. The 0.95 confidence intervals for odds ratios and p-values from proportional odds models with treatments, centers, gender, age, and baseline are (1.26, 5.69) with $p = 0.0106$, (2.64, 12.59) with $p < 0.0001$, (1.75, 7.70) with $p = 0.0006$, and (1.28, 5.57) with $p = 0.0089$ for visits 1, 2, 3, 4.
3. The p-values from score tests (with d.f. = 15) for the proportional odds assumption for the model in (2) are 0.3726, 0.4570, 0.3686, <0.0001 for visits 1, 2, 3, 4.
4. The results for treatment*visit from a GEE analysis with cumulative logits and "working independence" is

   $Q_c = 10.47$ with $p = 0.0150$

Notation

- Let h=1, 2, … q index strata
- Let i=1, 2 index groups
- Let j=1, 2,…, N index all patients
- Let n_{hi} = sample size for i-th group in h-th stratum
- Let k = 1, 2,…, r index response variables
- Let s_j = stratum for j-th patient
- Let t_j = 1, -1 according to patient j in group 1, 2
- Let $y_j = (y_{j1}, y_{j2}, \ldots, y_{jr})'$ denote response vector y for patient j
- Let $Z_j = (Z_{j1}, Z_{j2}, \ldots Z_{jr})$ with $Z_{jk} = 0, 1$ as y_{jk} is missing, observed

U-Statistic estimation

$$\hat{\theta}_{1k} = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{j' \neq j}^{N} U_{1jj'k} \text{ and } \hat{\theta}_{2k} = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{j' \neq j}^{N} U_{2jj'k}$$

$$U_{1jj'k} = \frac{I\left\{S_j - S_{j'} = 0\right\}\left[I\left\{\left(t_j - t_{j'}\right)\left(Y_{jk} - Y_{j'k}\right) Z_{jk} Z_{j'k} > 0\right\} + 0.5 I\left\{\left|t_j - t_{j'}\right| Z_{jk} Z_{j'k} > 0\right\} I\left\{\left(Y_{jk} - Y_{j'k}\right) = 0\right\}\right]}{\left(n_{jk} + n_{j'k} + 1\right)}$$

$$U_{2jj'k} = \frac{\left[I\left\{S_j - S_{j'} = 0\right\} \times I\left\{\left|t_j - t_{j'}\right| Z_{jk} Z_{j'k} > 0\right\}\right]}{\left(n_{jk} + n_{j'k} + 1\right)}$$

U-Statistic estimation (cont.)

$I(\Psi) = 1, 0$ if Ψ is satisfied or not

n_{jk} = sample size for k-th response variable for patients in same stratum and group as j-th patient

$= n_{hik}$ if patient j is from h-th stratum and i-th group

(as the number of non-missing y_{jk} for h-th stratum and i-th group)

Covariance matrix estimation for U-Statistics

$$\mathbf{V}_{\overline{\mathbf{F}}} = \frac{4}{N(N-1)} \sum_{j=1}^{N} \left(\mathbf{F}_j - \overline{\mathbf{F}}\right)\left(\mathbf{F}_j - \overline{\mathbf{F}}\right)'$$

$$\mathbf{F}_{\mathbf{j}} = \left(\mathbf{U}_{1j}^{'}, \mathbf{U}_{2j}^{'}\right)', \mathbf{U}_{1\mathbf{j}} = \left(U_{1j1}, \ldots U_{1jr}\right)', \mathbf{U}_{2\mathbf{j}} = \left(U_{2j1}, \ldots U_{2jr}\right)'$$

$$U_{1jk} = \frac{1}{(N-1)} \sum_{j' \neq j}^{N} U_{1jj'k}, U_{2jk} = \frac{1}{(N-1)} \sum_{j' \neq j}^{N} U_{2jj'k}$$

$$\overline{\mathbf{F}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{F}_j = \left(\hat{\boldsymbol{\theta}}_1^{'}, \hat{\boldsymbol{\theta}}_2^{'}\right)' = \left(\hat{\boldsymbol{\theta}}_{11}^{'}, \ldots \hat{\boldsymbol{\theta}}_{1r}^{'}, \hat{\boldsymbol{\theta}}_{21}^{'}, \ldots \hat{\boldsymbol{\theta}}_{2r}^{'}\right)'$$

Mann-Whitney estimators and covariance matrix estimation

$$\hat{\boldsymbol{\xi}} = \left(\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_r\right)' = \mathbf{D}_{\hat{\boldsymbol{\theta}}_2}^{-1} \hat{\boldsymbol{\theta}}_1 \text{ with } \hat{\xi}_k = \left(\hat{\theta}_{1k} / \hat{\theta}_{2k}\right)$$

$$\mathbf{V}_{\hat{\boldsymbol{\xi}}} = \mathbf{D}_{\hat{\boldsymbol{\xi}}} \left[D_{\hat{\theta}_1}^{-1}, -D_{\hat{\theta}_2}^{-1} \right] \mathbf{V}_{\overline{\mathbf{F}}} \left[D_{\hat{\theta}_1}^{-1}, -D_{\hat{\theta}_2}^{-1} \right]' \mathbf{D}_{\hat{\boldsymbol{\xi}}}$$

$$\hat{\xi}_k = \left(\sum_{h=1}^{q} w_{hk} \hat{\xi}_{hk} \Big/ \sum_{h=1}^{q} w_{hk} \right) \text{with } w_{hk} = n_{h1k} n_{h2k} / (n_{h1k} + n_{h2k} + 1)$$

$$\left(\hat{\xi}_{hk} - 0.5\right) = \left(\overline{R}_{h1k} - \overline{R}_{h2k}\right) / \left(n_{h1k} + n_{h2k}\right)$$

$\overline{R}_{hik}$ = mean rank for i - th group in h - th stratum for k - th response variable

Management of covariables

Let $\mathbf{x}_j = \left(x_{j1}, \ldots, x_{jM}\right)'$ denote baseline covariables with none missing

$$\tilde{U}_{1jj'm} = \frac{\left[I\left\{\left(S_j - S_{j'}\right) = 0\right\} \times 0.5\left(t_j - t_{j'}\right) \times \left(X_{jm} - X_{j'm}\right)\right]}{\left(n_j + n_{j'}\right)}$$

$$\tilde{U}_{2jj'} = \frac{\left[I\left\{\left(S_j - S_{j'}\right) = 0\right\} \times I\left\{\left(t_j - t_{j'}\right) \neq 0\right\}\right]}{\left(n_j + n_{j'}\right)}$$

$n_j = n_{hi}$ if patient i is in the i-th group and the h-th stratum

Management of covariables (cont)

$$\phi_{1m} = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{j' \neq j}^{N} \tilde{U}_{1jj'm}$$

$$\phi_{2m} = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{j' \neq j}^{N} \tilde{U}_{2jj'm}$$

Covariance matrix estimation for response variables and covariates jointly

$$\mathbf{V}_{\overline{\mathbf{G}}} = \frac{4}{N(N-1)} \sum_{j=1}^{N} \left(\mathbf{G}_j - \overline{\mathbf{G}}\right)\left(\mathbf{G}_{j'} - \overline{\mathbf{G}}\right)'$$

$$\mathbf{G}_j = \left(\mathbf{U}_{1j}^{'}, \tilde{\mathbf{U}}_{1j}^{'}, \mathbf{U}_{2j}^{'}, \tilde{U}_{2j}^{'}\right)', \tilde{\mathbf{U}}_{1j}^{'} = \left(\tilde{U}_{1j1}^{'}, \ldots, \tilde{U}_{1jM}^{'}\right)'$$

$$\tilde{U}_{1jm} = \frac{1}{N-1} \sum_{j' \neq j}^{N} \tilde{U}_{1jj'm}, \quad \tilde{U}_{2j} = \frac{1}{N-1} \sum_{j' \neq j}^{N} \tilde{U}_{2jj'}$$

$$\overline{\mathbf{G}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{G}_j = \left(\hat{\boldsymbol{\theta}}_1^{'}, \hat{\boldsymbol{\varphi}}_1^{'}, \hat{\boldsymbol{\theta}}_2^{'}, \hat{\boldsymbol{\varphi}}_2^{'}\right)' \text{ with } \hat{\boldsymbol{\varphi}}_1^{'} = \left(\hat{\boldsymbol{\varphi}}_{11}^{'}, \ldots, \hat{\boldsymbol{\varphi}}_{1M}^{'}\right)'$$

Mann-Whitney estimators for response variables and stratified differences between means for covariables jointly and their covariance matrix estimation

$$\mathbf{f} = \left(\hat{\boldsymbol{\xi}}^{'}, \mathbf{g}^{'}\right)^{'} \text{ where } \hat{\boldsymbol{\xi}} = D_{\hat{\boldsymbol{\theta}}_2}^{-1}\hat{\boldsymbol{\theta}}_1 \text{ and } \mathbf{g} = \hat{\boldsymbol{\varphi}}_1 / \hat{\varphi}_2$$

$$\mathbf{V}_{\mathbf{f}} = \mathbf{H}\,\mathbf{V}_{\bar{\mathbf{G}}}\,\mathbf{H}_{1}^{'}$$

$$\mathbf{H} = \mathbf{D}_{\mathbf{f}} \begin{bmatrix} \mathbf{I}_d & \mathbf{0}_{dM} & -\mathbf{I}_d & \mathbf{0}_{dM} \\ \mathbf{0}_{Md} & \mathbf{I}_M & \mathbf{0}_{Md} & -\mathbf{I}_M \end{bmatrix} \mathbf{D}_{\tilde{\mathbf{G}}}^{-1}$$

Mann-Whitney estimators for response variables and stratified differences between means for covariables jointly and their covariance matrix estimation (cont)

$$\hat{\varphi}_{1m} / \hat{\varphi}_2 = \sum_{h=1}^{q} \tilde{w}_h \left(\bar{x}_{h1m} - \bar{x}_{h2m} \right) \Big/ \sum_{h=1}^{q} \tilde{w}_h$$

with $\tilde{w}_h = n_{h1} n_{h2} / \left(n_{h1} + n_{h2} \right)$

and $\bar{x}_{him} =$ mean of m-th covariable for i-th group and h-th stratum

Randomization-based covariance adjustment

Fit $P = \left[\mathbf{I}_r, \mathbf{0}_{rM} \right]'$ to $\mathbf{f}$ by weighted least squares

$$\mathbf{b} = \left(\mathbf{P}' \mathbf{V}_{\mathbf{f}}^{-1} \mathbf{P} \right)^{-1} \mathbf{P}' \mathbf{V}_{\mathbf{f}}^{-1} \mathbf{f} = \left(\hat{\boldsymbol{\xi}} - V_{\hat{\boldsymbol{\xi}}\mathbf{g}} \mathbf{V}_{\mathbf{g}}^{-1} \mathbf{g} \right)$$

$$\mathbf{V}_{\mathbf{b}} = \left(\mathbf{P}' \mathbf{V}_{\mathbf{f}}^{-1} \mathbf{P} \right)^{-1} = \left(\mathbf{V}_{\hat{\boldsymbol{\xi}}} - V_{\hat{\xi} g} \mathbf{V}_{g}^{-1} V_{\hat{\xi} g}' \right)$$

$V_{\hat{\boldsymbol{\xi}}\mathbf{g}}$ is submatrix of $\mathbf{V}_{\mathbf{f}}$ for covariances of $\hat{\boldsymbol{\xi}}$ with $\mathbf{g}$ and

$\mathbf{V}_{\mathbf{g}}$ is submatrix of $\mathbf{V}_{\mathbf{f}}$ for estimated covariance matrix for $\mathbf{g}$

Example 3: Randomized two period crossover clinical trial for osteoarthritis of the hip or knee with ordinal responses

1. There are four strata according to a pain severity index from baseline factors for 227 patients
2. There are two sequence groups as C:T and T:C for test (T) and control (C) treatments
3. There are five visits for screening (visit 1), baseline prior to period 1 of treatment (visit 2), end of period 1 (visit 3), baseline prior to period 2 (visit 4), end of period 2 (visit 5).
4. The response variable at each visit is pain according to ordinal visual analogue scale (with range of 0 to 100 for no pain to very severe pain)
5. Age is a numeric covariable.

Visit 1, Visit 2, and Age have no missing data;
Visit 3 has minimal missing data;
Visits 4 and Visit 5 have substantial missing data

Table 3: Sample Sizes for Ordinal Response Variables and Age for Two Sequence Groups of Two Treatments within Four Strata According to Pain Severity Index.

Pain Severity Index	Treatment Sequence	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Age
0	T:C	11	11	10	8	8	11
0	C:T	15	15	15	12	12	15
1	T:C	26	26	26	22	22	26
1	C:T	24	24	22	20	19	24
2	T:C	39	39	37	32	32	39
2	C:T	35	35	34	27	26	35
3	T:C	36	36	33	33	31	36
3	C:T	41	41	41	33	32	41

The stratified Mann-Whitney estimators for Visits 1, 2, 3, 4, 5 and the stratified difference between mean ages with corresponding covariance matrix are as follows:

$$\mathbf{f} = [\hat{\xi}_0, \hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4, g]' = [0.5184, 0.5943, 0.6467, 0.6058, 0.3302, -0.3847]'$$

$$\mathbf{V_f} = \begin{bmatrix} 16.08 & 5.48 & 2.06 & 3.54 & 1.67 & -82.24 \\ 5.48 & 14.65 & 4.21 & 6.87 & 2.81 & -1.44 \\ 2.06 & 4.21 & 15.50 & 9.89 & 2.96 & 30.22 \\ 3.54 & 6.87 & 9.89 & 19.55 & 6.30 & -20.89 \\ 1.67 & 2.81 & 2.96 & 6.30 & 18.01 & -27.84 \\ -82.24 & -1.44 & 30.22 & -20.89 & -27.84 & 20405.11 \end{bmatrix} \times 10^{-4}$$

The criterion for random imbalances for

$\widehat{\xi}_0$, $\widehat{\xi}_1$, and g is $Q_0 = 1.50$ with $d.f. = 3$, $p = 0.6821$

The covariance and stratification adjusted Mann-Whitney estimators as departures from 0.5 for Visits 3, 4, 5 and their covariance matrix are as follows:

$$\mathbf{b} = [0.1210, 0.0624, -0.1877]'$$

$$\mathbf{V_b} = \begin{bmatrix} 14.23 & 7.91 & 2.17 \\ 7.91 & 16.25 & 4.92 \\ 2.17 & 4.92 & 17.42 \end{bmatrix} \times 10^{-4}$$

With $\left\{ b_k / s.e.(b_k) \right\}^2$ as approximately having the chi-squared distribution with $d.f. = 1$, $p = 0.0014, 0.1232, <0.0001$, for Visits 3, 4, 5.

The covariance and stratification adjusted estimators as departures from 0.5 for Visits 3, 5 (with Visit 4 adjusted to null) and their covariance matrix are as follows:

$$\tilde{b} = \begin{bmatrix} 0.0906 \\ -0.2066 \end{bmatrix}, \mathbf{V}'_{\tilde{\mathbf{b}}} = \begin{bmatrix} 32.86 & -4.15 \\ -4.15 & 45.62 \end{bmatrix}$$

For Visit 3, $p = 0.0049$; for visit 5, $p =< 0.0001$;

For (Visit 3 + Visit 5), $p = 0.0225$.

For (Visit 3 - Visit 5)/2, $p =< 0.0001$;

and the corresponding fully adjusted

Mann-Whitney estimator is 0.6486 with

0.95 confidence interval $(0.557, 0.7403)$.

Other ways to manage missing data

1. Remove $z_{jk}z_{j'k}$ from the definition of the $U_{2jj'k}$ so that $U_{2jj'k} = U_{2jj'} = 1/(n_{h1} + n_{h2} + 1)$ for all k when $s_j = s_{j'}$ and $t_j \neq t_{j'}$ with $U_{2jj'} = 0$ if otherwise.

2. Impute value of $U_{1jj'k}$ when $z_{jk}z_{j'k} = 0$ for either y_{jk} or $y_{j'k}$ missing
   - a. $U_{1jj'k} = 0.5$ for tied status by adding $0.5(1 - z_{jk}z_{j'k})$
   - b. $U_{1jj'k} = U_{1jj'(k-1)}$ for the previous visit by adding $U_{1jj'(k-1)}(1 - z_{jk}z_{j'k})$ and proceeding recursively by visit.
   - c. $U_{1jj'k} = U_{1jj'1}$ for baseline by adding $U_{1jj'1}(1 - z_{jk}z_{j'k})$

Ways to improve applicability of approximately multivariate normal distributions and behavior of test statistics

1. Use of logit transformations $\hat{\lambda}_k = \log_e \left\{ \hat{\xi}_k \Big/ \left(1 - \hat{\xi}_k\right) \right\}$ with covariance matrix $\mathbf{V}_{\hat{\boldsymbol{\lambda}}} = D_{\hat{\boldsymbol{\eta}}}^{-1} \mathbf{V}_{\hat{\boldsymbol{\xi}}} D_{\hat{\boldsymbol{\eta}}}^{-1}$ for which $\hat{\boldsymbol{\eta}} = D_{\hat{\boldsymbol{\xi}}} \left(1 - \hat{\boldsymbol{\xi}}\right)$ and $\hat{\boldsymbol{\lambda}} = \left(\hat{\lambda}_1, \ldots, \hat{\lambda}_r\right)'$; such logits are like Fisher transformations of Somer's version of Kendall tau.

2. Multiplication of covariance matrices by $(N-1)/(N - \#\text{ estimators})$

3. Use F-distribution with $d.f. = (c, N - \#\text{ estimators})$ for constrasts with rank c or confidence intervals with $c = 1$ estimator