sanon : An R Package for Stratified Analysis with Nonparametric Covariable Adjustment

Atsushi Kawaguchi[1] and Gary G. Koch[2]

[1] Department of Biomedical Statistics and Bioinformatics
Kyoto University Graduate School of Medicine
54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
E-mail: akawa@kuhp.kyoto-u.ac.jp

[2] Biometric Consulting Laboratory, Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, 27599-7420 U.S.A.
E-mail: bcl@bios.unc.edu

Abstract

Kawaguchi et al. (2011) provided methodology and applications for a stratified Mann–Whitney estimator that addresses the same comparison between two randomized groups for a strictly ordinal response variable as the van Elteren test statistic for randomized clinical trials with strata. The sanon package provides the implementation of the method within the R programming environment (R Core Team, 2012). The usage of sanon is illustrated with five examples. The first example is a randomized clinical trial with eight strata and a univariate ordinal response variable. The second example is a randomized clinical trial with four strata, two covariables, and four ordinal response variables. The third example is a cross over design randomized clinical trial with two strata, one covariable, and two ordinal response variables. The fourth example is a randomized clinical trial with seven strata (which are managed as a categorical covariable), three ordinal covariables with missing values, and three ordinal response variables with missing values. The fifth example is a randomized clinical trial with six strata, a categorical covariable with three levels, and three ordinal response variables with missing values.

Keywords: missing completely at random, multivariate outcomes, randomization-based non-parametric covariance adjustment, randomized clinical trial, repeated measurement, stratification, strictly ordinal response variable, R

1   Introduction

The primary analyses for confirmatory randomized clinical trials (and particularly those with regulatory objectives) should have protocol specified methods that have minimal assumptions. A nonparametric approach such as the Wilcoxon (or Mann–Whitney) test statistic for the comparison between two treatments through the ranking of a response variable for all patients (in the pooled treatment groups) has essentially no assumptions (under the strong null hypothesis of no treatment differences in the sense that each patient has the same response to both treatments) beyond valid randomization as its basis. More generally, Mann-Whitney test statistics and the corresponding estimators address the global null hypothesis of equality of response distributions for two treatment groups through its implied null hypothesis for Pr(Group 1 response > Group 2 response)$= \xi = 0.5$; and their sensitivity and power to detect differences between two distributions depend jointly on the extent to which $|\xi - 0.5| > 0$ and the applicable sample size. Nevertheless, such methods have the recognized limitation that their power for comparing different distributions for which $\xi = 0.5$ equals the specified type 1 error regardless of the applicable sample sizes (although differences between such distributions are typically of relatively minimal interest for clinical trials that evaluate whether one treatment group has relatively more patients with better responses that the other treatment group). The `coin` package (Hothorn et al., 2013, 2008, 2006) in the `R` programming environment (R Core Team, 2012) can handle ordered and multivariate responses as well as stratification and provides both randomization/permutation tests and asymptotic tests based on the conditional inference. `NParCov3` (Zink and Koch, 2012) is a `SAS/IML` macro (`SAS` Institute Inc., 2011) written to conduct the nonparametric randomization-based covariance analyses of Koch et al. (1998).

For confirmatory randomized clinical trials with stratified designs for comparing two treatments, Kawaguchi et al. (2011) proposed stratified multivariate Mann-Whitney estimators as a useful structure for the analysis of strictly ordinal response variables. Their scope can address strata with at least minimal sample sizes (e.g., ≥ 16), and randomization-based covariance adjustment is possible. The method is based on the Mann-Whitney estimator for the probability that a randomly selected patient from one treatment group has better status for a response variable within a stratum than a randomly selected patient from the other treatment group (with ties being randomly broken with probability 0.5). Such Mann-Whitney estimators can be combined across the strata to provide stratified estimator counterparts that address the same comparisons between the two treatment groups as the van Elteren test statistic. The multivariate vector of such stratified Mann-Whitney estimators for multivariate response variables

can be considered for one or more response variables such as in repeated measurements and these can have missing completely at random (MCAR) data (or missing data for which direct imputation methods are applicable). Randomization-based covariable adjustment is possible for stratified multivariate Mann-Whitney estimators by expanding the vector of such estimators to include stratified differences between means of covariables. The latter estimators for the covariables then have constraints to 0's invoked. For this purpose, weighted least squares methods are applied with weights based on the estimated covariance matrix for the expanded vector from the methods for ratios of multivariate U-statistics; see Stokes et al. (2012, chap. 14) and Koch et al. (1998). The resulting estimators are stratified multivariate Mann-Whitney estimators with randomization based covariance adjustment, and their interpretation is the same as the original Mann-Whitney estimator mentioned above. With sufficiently large sample sizes, such estimators have an approximately multivariate normal distribution with the covariance matrix being essentially known through its corresponding consistent estimator. Accordingly, confidence intervals can have construction for linear functions of such fully adjusted Mann-Whitney estimators (with respect to both stratification and covariables), and the scope of such linear functions can include the separate response variables, averages across response variables, and contrasts among response variables.

The software of the method was developed in the `R` programming environment (R Core Team, 2012). The `sanon` package contains functions to implement the method as well as two example data sets which were used in Kawaguchi et al. (2011) and three sets from other papers. The main function in the package has a similar structure as functions in `R` for standard statistical methods such as `lm` for linear models. This paper illustrates the usage of the package with five example data sets. Section 2 describes the methods, Section 3 explains the code, and Section 5 illustrates those methods for five examples introduced in Section 4.

2 Methods

As previously outlined in Section 1, a stratified Mann-Whitney estimator and corresponding confidence interval can have construction so as to address the same comparison between two randomized groups as the van Elteren test statistic. The specifications for the formal structure for this method are in Section 2.1 for a multivariate set of r response variables. Section 2.2 discusses randomization based covariance adjustment for stratified multivariate Mann-Whitney estimators and it explains the corresponding constraints for no expected differences between randomized groups for stratified means of covariables.

2.1 Stratified multivariate Mann-Whitney estimator

Let $h = 1, 2, \ldots, q$ index a set of strata within which patients have randomization to two groups indexed by $i = 1, 2$. Let $k = 1, 2, \ldots, r$ index the response variables with observation for the n_{hi} patients in the i-th group of the h-th stratum; and some of these response variables can be baselines at times prior to any treatment for a group; or in crossover studies, they can be at washout times prior to the treatments which are subsequent to the first treatment. Let $j = 1, 2, \ldots, N$ index the patients for the pooling of all patients in the clinical trial regardless of their groups or strata. Let S_j denote the stratum for the j-th patient, and let t_j correspond to the group for the j-th patient with $t_j = 1$ if $i = 1$ for patient j and $t_j = -1$ if $i = 2$ for patient j. Let $\boldsymbol{Y}_j = (Y_{j1}, \ldots, Y_{jr})^\top$ denote the response vector for the j-th patient with Y_{jk} denoting the k-th strictly ordinal response variable for the j-th patient. Since some of the Y_{jk} may be missing (by a missing completely at random (MCAR) process), let $Z_{jk} = 1$ if Y_{jk} is not missing and let $Z_{jk} = 0$ if Y_{jk} is missing ; and let $\boldsymbol{Z}_j = (Z_{j1}, \ldots, Z_{jr})^\top$. In the subsequent discussion, any missing Y_{jk} operationally has replacement by 0 since the value used for such replacement has no role in the subsequently described processes for estimation.

The stratified Mann-Whitney estimator for the k-th response variable is $\hat{\xi}_k = (\hat{\theta}_{1k}/\hat{\theta}_{2k})$ for which $\hat{\theta}_{1k}$ and $\hat{\theta}_{2k}$ are defined in (1) and (2),

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

$$
\begin{aligned}
U_{1jj'k} &= I\left\{(S_j - S_{j'}) = 0\right\} \times [I\left\{(t_j - t_{j'})(Y_{jk} - Y_{j'k})Z_{jk}Z_{j'k} > 0\right\} \\
&\quad + 0.5 \times I\left\{(t_j - t_{j'})^2 Z_{jk}Z_{j'k} > 0\right\} \times I\left\{(Y_{jk} - Y_{j'k}) = 0\right\}]/(N_{jk} + N_{j'k} + 1)
\end{aligned}
\tag{2}
$$

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

Also, $I(A)$ has the value 1 if the condition A is satisfied or the value 0 if otherwise, and N_{jk} denotes the sample size for the k-th response variable for patients with the same stratum and group as the j-th patient. For (1) and (2), the numerator of the $U_{1jj'k}$ equals 1 for a pair of patients who are in the same stratum with different groups and for whom the k-th response variable is larger for the patient in group 1 than for the patient in group 2, and it equals 0 for all other pairs of patients; and the numerator of the $U_{2jj'k}$ equals 1 for a pair of patients who are in the same stratum with different groups and for whom the k-th response variable is observed (i.e., not missing), and it equals 0 for all other pairs of patients. Accordingly, the

numerator of $\hat{\theta}_{2k}$ equals $\left(\sum_{h=1}^{q} w_{hk}\right)$ for which $w_{hk} = (n_{h1k}n_{h2k})/(n_{h*k} + 1)$ with n_{hik} being the number of patients for whom the k-th response variable is observed among the n_{hi} patients for the i-th group in the h-th stratum and $n_{h*k} = (n_{h1k} + n_{h2k})$; and the numerator of $\hat{\theta}_{1k}$ equals $\left(\sum_{h=1}^{q} w_{hk}\hat{\xi}_{hk}\right)$ for which the $\hat{\xi}_{hk}$ for the respective strata are the within strata Mann-Whitney estimators for the proportions of pairs of patients with different groups and for whom the k-th response variable is larger for the group 1 patient than the group 2 patient. Thus, $\hat{\xi}_k = \left(\sum_{h=1}^{q} w_{hk}\hat{\xi}_{hk} \Big/ \sum_{h=1}^{q} w_{hk}\right)$ is a weighted average of the within strata, Mann-Whitney estimators $\hat{\xi}_{hk}$ with the w_{hk} as weights.

Let $U_{1jk} = \sum_{j'\neq j}^{N} U_{1jj'k}/(N-1)$ and $U_{2jk} = \sum_{j'\neq j}^{N} U_{2jj'k}/(N-1)$. Let $\boldsymbol{F}_j = (\boldsymbol{U}'_{1j}, \boldsymbol{U}'_{2j})^{\top}$ denote a compound vector for the j-th patient where $\boldsymbol{U}_{1j} = (U_{1j1}, \ldots, U_{1jr})^{\top}$ and $\boldsymbol{U}_{2j} = (U_{2j1}, \ldots, U_{2jr})^{\top}$. Let $\overline{\boldsymbol{F}} = \sum_{j=1}^{N} \boldsymbol{F}_j/N$ denote the sample mean vector for the $\boldsymbol{F}_j$. As noted in Davis and Quade (1968), Puri and Sen (1971), Quade (1974), Carr et al. (1989), Jung and Koch (1998), and Jung and Koch (1999), a consistent estimator for the covariance matrix for $\overline{\boldsymbol{F}}$ is given in (3).

$$\boldsymbol{V}_{\overline{\boldsymbol{F}}} = \frac{4}{N(N-1)} \sum_{j=1}^{N} (\boldsymbol{F}_j - \overline{\boldsymbol{F}})(\boldsymbol{F}_j - \overline{\boldsymbol{F}})^{\top}. \tag{3}$$

The stratified Mann-Whitney estimator for $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_r)^{\top}$ is $\hat{\boldsymbol{\xi}} = \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-1}\hat{\boldsymbol{\theta}}_1 = (\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_r)^{\top}$ where $\boldsymbol{D}_{\boldsymbol{a}}$ denotes a diagonal matrix with the elements of vector $\boldsymbol{a}$ on the main diagonal. Here $\hat{\boldsymbol{\theta}}_1 = (\hat{\theta}_{11}, \hat{\theta}_{12}, \ldots, \hat{\theta}_{1r})^{\top}$ corresponds to the first r elements of $\overline{\boldsymbol{F}}$ (as they represent numerators of the $\hat{\xi}_k$ for the r response variables); and $\hat{\boldsymbol{\theta}}_2 = (\hat{\theta}_{21}, \hat{\theta}_{22}, \ldots, \hat{\theta}_{2r})^{\top}$ corresponds to the remaining r elements of $\overline{\boldsymbol{F}}$; i.e., $\overline{\boldsymbol{F}} = (\hat{\boldsymbol{\theta}}_1^{\top}, \hat{\boldsymbol{\theta}}_2^{\top})^{\top}$. Based on the Taylor series linearization, a consistent estimator $\boldsymbol{V}_{\hat{\boldsymbol{\xi}}}$ for the covariance matrix for $\hat{\boldsymbol{\xi}}$ is given in (4).

$$\boldsymbol{V}_{\hat{\boldsymbol{\xi}}} = [\boldsymbol{I}_r, -\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-2}]\boldsymbol{V}_{\overline{\boldsymbol{F}}}[\boldsymbol{I}_r, -\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-2}]^{\top}. \tag{4}$$

As noted in Kawaguchi et al. (2011), with sufficient sample size for $\hat{\boldsymbol{\xi}}$ to have an approximately multivariate normal distribution (e.g., $n_{+ik} \geq 50$ and all $n_{hik} \geq 4$), then a two-sided $100(1 - 2\alpha)\%$ confidence interval for the linear statistic $\boldsymbol{c}^{\top}\hat{\boldsymbol{\xi}}$ for the comparison between the two groups is $\left(\boldsymbol{c}^{\top}\hat{\boldsymbol{\xi}} \pm z_{\alpha}\sqrt{\boldsymbol{c}^{\top}\boldsymbol{V}_{\hat{\boldsymbol{\xi}}}\boldsymbol{c}}\right)$ where z_{α} is the $100(1-\alpha)$ percentile of the standard normal distribution with mean 0 and variance 1. In this regard, the interpretation of the results from assessments of such contrasts may need some caution since Mann-Whitney estimators do not have explicitly transitive relationships with one another, although approximately transitive relationships can sometimes be applicable to their corresponding log odds $\hat{\psi}_k = \log_e[\hat{\xi}_k/(1 - \hat{\xi}_k)]$ through Bradley-Terry specifications for paired comparisons; see Jung and Koch (1999) and

Kawaguchi and Koch (2010).

2.2 Randomization Based Covariance Adjustment

Let $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jM})^\top$ denote the vector of M numeric covariables for the j-th patient with x_{jm} denoting the numeric value of the m-th covariable for the j-th patient. All of the M covariables have observation prior to the randomization of the patients to the two groups. Also, any categorical covariable has expression as a set of indicator variables that correspond to all except one of its categories with the excluded category serving as a reference category; and any ordinal covariable is managed in the same way as the response variables via (1)–(3). All of the covariables are assumed to have no missing data as is often realistic (although it is possible to manage a covariable with substantial missing data as a categorical covariable and to manage minimal missing data for other covariables in the same way as for response variables).

The stratification adjusted mean difference estimator for the m-th covariable is

$$g_m = (\hat{\varphi}_{1m}/\hat{\varphi}_2) = \left\{ \sum_{h=1}^q w_h(\overline{x}_{h1m} - \overline{x}_{h2m}) \middle/ \sum_{h=1}^q w_h \right\} \tag{5}$$

where $w_h = n_{h1}n_{h2}/(n_{h1} + n_{h2})$, and $\hat{\varphi}_{1m}$ and $\hat{\varphi}_2$ are defined as in (6).

$$\hat{\varphi}_{1m} = \frac{1}{N(N-1)} \sum_{j=1}^N \sum_{j' \neq j}^N \widetilde{U}_{1jj'm}, \qquad \hat{\varphi}_2 = \frac{1}{N(N-1)} \sum_{j=1}^N \sum_{j' \neq j}^N \widetilde{U}_{2jj'} \tag{6}$$

In (6), the $\widetilde{U}_{1jj'm}$ and $\widetilde{U}_{2jj'}$ have the structure shown in (7)

$$\widetilde{U}_{1jj'm} = \left[I\left\{ (S_j - S_{j'}) = 0 \right\} \times 0.5(t_j - t_{j'}) \times (x_{jm} - x_{j'm}) \right] /(N_j + N_{j'})$$

$$\tag{7}$$

$$\widetilde{U}_{2jj'} = \left[I\left\{ (S_j - S_{j'}) = 0 \right\} \times I\left\{ (t_j - t_{j'}) \neq 0 \right\} \right] /(N_j + N_{j'})$$

where $N_j = n_{hi}$ if patient j is from the h-th stratum and is in the i-th group. Also, the numerator of $\widetilde{U}_{1jj'm}$ (for a pair of patients who are from the same stratum with different groups) equals the difference between the patient in group 1 and the patient in group 2 for the m-th covariable, and it equals 0 for all other pairs of patients; and the numerator of $\widetilde{U}_{2jj'}$ has a comparable definition with respect to the covariables $\boldsymbol{x}_j$ as $U_{2jj'k}$ has for the k-th response variable.

Let $\boldsymbol{f} = (\hat{\boldsymbol{\xi}}^\top, \boldsymbol{g}^\top)^\top$ where $\hat{\boldsymbol{\xi}} = \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-1}\hat{\boldsymbol{\theta}}_1$ and $\boldsymbol{g} = \hat{\boldsymbol{\varphi}}_1/\hat{\varphi}_2$; here $\hat{\boldsymbol{\varphi}}_1 = (\hat{\varphi}_{11}, \ldots, \hat{\varphi}_{1M})^\top$. The first r elements of $\boldsymbol{f}$ address comparisons between the two groups for the r response variables with

stratification adjusted Mann-Whitney estimators, and the last M elements address comparisons between the two groups for the M covariables with stratification adjusted differences between their corresponding means. Since $\boldsymbol{g}$ would be expected to be null on the basis of randomization of patients to the two treatment groups, randomization-based covariance adjustment of $\hat{\boldsymbol{\xi}}$ is possible by fitting the model $\boldsymbol{P} = [\boldsymbol{I}_r, \boldsymbol{0}_{rM}]^\top$ to $\boldsymbol{f}$ by weighted least squares; see Koch et al. (1998) and LaVange et al. (2005). The weights are based on a consistent estimator for the covariance matrix of $\boldsymbol{f}$ which is derived as follows.

Let $\widetilde{U}_{1jm} = \sum_{j' \neq j}^N \widetilde{U}_{1jj'm}/(N-1)$ and $\widetilde{U}_{2j} = \sum_{j' \neq j}^N \widetilde{U}_{2jj'}/(N-1)$. Let $\boldsymbol{G}_j = (\boldsymbol{U}_{1j}^\top, \widetilde{\boldsymbol{U}}_{1j}^\top, \boldsymbol{U}_{2j}^\top, \widetilde{U}_{2j})^\top$ where $\widetilde{\boldsymbol{U}}_{1j}^\top = (\widetilde{U}_{1j1}, \ldots, \widetilde{U}_{1jM})^\top$. Let $\overline{\boldsymbol{G}} = \sum_{j=1}^N \boldsymbol{G}_j/N$ denote the sample mean vector for the $\boldsymbol{G}_j$. A consistent estimator for the covariance matrix of $\boldsymbol{G}$ is given in (8)

$$\boldsymbol{V}_{\overline{\boldsymbol{G}}} = \frac{4}{N(N-1)} \sum_{j=1}^N (\boldsymbol{G}_j - \overline{\boldsymbol{G}})(\boldsymbol{G}_j - \overline{\boldsymbol{G}})^\top \tag{8}$$

A consistent estimator for the covariance matrix of $\boldsymbol{f}$ is $\boldsymbol{V}_{\boldsymbol{f}} = \boldsymbol{H} \boldsymbol{V}_{\overline{\boldsymbol{G}}} \boldsymbol{H}^\top$ where $\boldsymbol{H}$ as shown in (9) is from Taylor series linearization.

$$\boldsymbol{H} = \boldsymbol{D}_{\boldsymbol{f}} \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0}_{rM} & -\boldsymbol{I}_r & \boldsymbol{0}_{r1} \\ \boldsymbol{0}_{Mr} & \boldsymbol{I}_M & \boldsymbol{0}_{Mr} & -\boldsymbol{1}_M \end{bmatrix} \boldsymbol{D}_{\widetilde{\boldsymbol{G}}}^{-1} = \begin{bmatrix} \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-1} & \boldsymbol{0}_{rM} & -\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2}^{-2} \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_1} & \boldsymbol{0}_{r1} \\ \boldsymbol{0}_{Mr} & \boldsymbol{I}_M \hat{\varphi}_2^{-1} & \boldsymbol{0}_{Mr} & -\hat{\varphi}_1 \hat{\varphi}_2^{-2} \end{bmatrix} \tag{9}$$

where $\widetilde{\boldsymbol{G}} = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\varphi}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top, \hat{\varphi}_2)^\top$.

The resulting adjusted counterparts $\boldsymbol{b}$ for $\hat{\boldsymbol{\xi}}$ are shown in (10).

$$\boldsymbol{b} = (\boldsymbol{P}^\top \boldsymbol{V}_{\boldsymbol{f}}^{-1} \boldsymbol{P})^{-1} \boldsymbol{P}^\top \boldsymbol{V}_{\boldsymbol{f}}^{-1} \boldsymbol{f} = (\hat{\boldsymbol{\xi}} - \boldsymbol{V}_{\hat{\boldsymbol{\xi}}g} \boldsymbol{V}_{\boldsymbol{g}}^{-1} \boldsymbol{g}) \tag{10}$$

where $\boldsymbol{V}_{\hat{\boldsymbol{\xi}}g}$ corresponds to the covariances of $\hat{\boldsymbol{\xi}}$ with $\boldsymbol{g}$ and $\boldsymbol{V}_{\boldsymbol{g}}$ corresponds to the covariance matrix of $\boldsymbol{g}$. A consistent estimator for the covariance matrix of $\boldsymbol{b}$ is $\boldsymbol{V}_{\boldsymbol{b}}$ in (11)

$$\boldsymbol{V}_{\boldsymbol{b}} = (\boldsymbol{P}^\top \boldsymbol{V}_{\boldsymbol{f}}^{-1} \boldsymbol{P})^{-1} = (\boldsymbol{V}_{\hat{\boldsymbol{\xi}}} - \boldsymbol{V}_{\hat{\boldsymbol{\xi}}g} \boldsymbol{V}_{\boldsymbol{g}}^{-1} \boldsymbol{V}_{\hat{\boldsymbol{\xi}}g}^\top) \tag{11}$$

Additional models that address the variation of the elements of $\boldsymbol{b}$ across the r response variables can be fit by weighted least squares methods. One such model can invoke randomization based covariance adjustment for any strictly ordinal covariables among the response variables by having a structure like $\boldsymbol{P}$ with rows of 0's corresponding to the strictly ordinal covariables and rows of an identity matrix corresponding to the strictly ordinal response vari-

ables. Through the resulting estimator $\boldsymbol{b}_{adj}$ with full adjustment for all covariables and the strata, confidence intervals can have determination for linear statistics $\boldsymbol{c}^\top \boldsymbol{b}_{adj}$. Similarly, test statistics are applicable to assessments of homogeneity across the response variables for the adjusted estimators $\boldsymbol{b}_{adj}$ for the differences between the two groups.

Since the preceding discussion indicates different roles for strata and covariables, some clarifying comments for the distinction between them is useful. In this regard, strata typically correspond to the cross-classification of categorical factors for stratified randomization. However, sometimes only one or two such factors that more clearly need balanced allocation of treatments and better enable reduction of within stratum variance are pre-specified as the structure for strata in situations where the use of all such factors would produce too many strata with possibly uninformatively small sample sizes for rankings of responses within them (e.g., $(n_{h1} + n_{h2}) \leq 15$). In these cases, the other factors for stratification can be pre-specified as covariables. In contrast, covariables are typically pre-specified baseline variables which potentially have strong associations with response variables (even though they did not serve as a factor for stratification). For such baseline variables, randomization based covariance adjustment via (10) eliminates the spurious influence of random imbalances of their means for the treatment groups, and it provides variance reduction for treatment comparisons via (11). Also, as many as 10 baseline variables (with appropriate pre-specified clinical justification) can have randomization-based covariance adjustment via (10) for clinical trials with at least a moderately large, overall sample size (e.g., $N \geq 300$). In some situations, a categorical baseline variable that is not a factor for stratified randomization has pre-specified adjustment through stratification in order to remove the influence of random imbalances between treatment groups for its distributions and to have variance reduction for treatment comparisons more directly than provided by covariance adjustment. In summary, baseline variables can be either factors for stratification or covariables, and so pre-specification of which applies is very important to avoid potential bias from their roles having an arbitrary nature. Aspects of the scope of such analyses and the corresponding models have more specific discussion through the examples in Section 5.

2.3 Management of Missing Data

In addition to the method described in Section 2.1 to maintain missing values, three other options are newly introduced in this paper. The first option invokes the last observation carried forward (LOCF) convention for the kernels of U-statistics in (2). For this method,

the denominators in (2) are revised to $(N_j + N_{j'} + 1)$, and missing values for the numerators are replaced by the last observed value of that variable; that is, each $U_{1jj'k}$ for the nearest observed preceding k has its numerator carried forward to the missing k and $U_{2jj'k} = (N_j + N_{j'})\widetilde{U}_{2jj'}/(N_j + N_{j'} + 1)$ with $\widetilde{U}_{2jj'}$ as in (7). If there is a missing value in the first measurement, then this measurement is managed as tied with the (observed or missing) other measurements in the corresponding stratum; and so $U_{1jj'1} = 0.5U_{2jj'1} = 0.5(N_j + N_{j'})\widetilde{U}_{2jj'}/(N_j + N_{j'} + 1)$. Also, this management is more generally applied to all measurements prior to the first observed measurement by carrying forward $U_{1jj'1}$ and $U_{2jj'1}$. The second option is the LOCF based on the observed value of Y. As noted in the first LOCF method, the missing first value or values prior to first observed value is managed as tied with the (observed or missing) other measurements in the corresponding stratum. The third option for missing data manages missing values as tied with all other values in the same stratum. For this method, $U_{2jj'k} = (N_j + N_{j'})\widetilde{U}_{2jj'}/(N_j + N_{j'} + 1)$ with $\widetilde{U}_{2jj'}$ as in (7) and $U_{1jj'k}$ is modified to (12).

$$
\begin{aligned}
U_{1jj'k} \;=\; & I\left\{(S_j - S_{j'}) = 0\right\} \times [I\left\{(t_j - t_{j'})(Y_{jk} - Y_{j'k})Z_{jk}Z_{j'k} > 0\right\} \\
& +0.5 \times I\left\{(t_j - t_{j'})^2 Z_{jk}Z_{j'k} > 0\right\} \times I\left\{(Y_{jk} - Y_{j'k}) = 0\right\} \\
& +0.5 \times I\left\{(t_j - t_{j'})^2(1 - Z_{jk}Z_{j'k}) > 0\right\}]/(N_j + N_{j'} + 1)
\end{aligned}
\tag{12}
$$

Additionally, the method described in Section 2.1 and Section 2.2 is applicable with the method of multiple imputation for all missing data, but methods for multiple imputation are beyond the scope of this paper, particularly for strictly ordinal response variables. Nevertheless, adjusted Mann-Whitney estimates (10) and corresponding covariance matrices (11) from the respective invocations of multiple imputation can be combined to produce average estimators and their estimated covariance matrix by the methods of Rubin (2004) (as available, for example, in the `MIANALYZE Procedure` of the `SAS System`).

3   Code

The `R` package `sanon` can be installed from the Comprehensive R Archive Network (CRAN) by using the code `install.packages("sanon")`. Then, you need to type the `library(sanon)` command to load the package. The main function `sanon` can be accessed by your program by including the following line with the specification for the input and is formula based. The usage is:

```
sanon(formula, data, P = NULL, res.na.action = "default")
```

This function has four types of inputs (arguments), which are described in Table 1 and the details are below.

Table 1: Arguments for `sanon` function

Argument	Description
`formula`	a formula object, with the response on the left of a ~ operator, and the terms on the right, in which functions `strt`, `grp`, `covar`, and `catecovar` are required to specify the role of variables on the right hand side
`data`	a `data.frame` in which to interpret the variables named in the formula
`P`	a matrix object for weighted least squares estimation
`res.na.action`	character for setting `NA` actions. `"default"`, `"LOCF1"`, `"LOCF2"`, `"replace"`, and `"remove"` are available. The default is `"default"`. See Section 2.3 for the details.

The `formula` consists of variable names in your data set specified in the argument `data`, which has a similar structure as functions in R for standard statistical methods such as `lm` for linear models, but is more complicated. Let `y`, `g`, `s`, `cc`, and `c` be respectively the response, group, strata, categorical covariable, and continuous covariable names in the `data.frame`, and `"r"` is the reference group to be provided, the formula can be specified as follows:

```
formula = y ~ grp(g, ref = "r") + strt(s) + catecovar(cc, ref = "r")
                + covar(c)
```

These terms can be omitted except for the response and group if you don't need them. In this formula, the strata and group variables, and the continuous and categorical covariables are recognized by functions `strt`, `grp`, `covar` and `catecovar`, respectively. In other words, all terms on the right hand side should be specified for the role by using these functions. For example, if one used the input `sanon(y    grp(g) + cc)` with `cc` having no role specified, then `cc` would be ignored in the analysis. Covariables are distinguished on the basis of continuous or categorical type. The categorical covariable is internally converted into a set of dummy variables. The reference group for group variable and categorical covariable can be specified in the argument `ref` within each functions `grp` and `catecovar`. More than two strata variables can be in the formula object, and they are internally transformed into the cross-classification. The matrix $\boldsymbol{P}$ in (10) or (11) for the weighted least squares estimation is specified in the argument `P` by the R matrix object name.

The response variable can contain missing values, which should be coded by `NA`, and it can be multivariate (repeatedly measured). The function `sanon` has five options for deal-

ing with missing values for the response variable, with the default maintaining missing values and using the methods described in Section 2.1. It can be specified by no argument or `res.na.action="default"`, and it has the missing completely at random assumption (MCAR) for missing data. The second option is specified in the function sanon by `res.na.action="LOCF1"`; and it uses the LOCF convention for the kernels of U-statistics in (2), as described in Section 2.3. The function `sanon` recognizes the order in the left side of ~ as a measurement order, and an available baseline can be the first value in the order. The third option is specified in the function sanon by `res.na.action="LOCF2"`; and it applies the LOCF convention for Y as described in Section 2.3. The fourth is specified in the function `sanon` by `res.na.action="replace"`. The fifth option for managing missing data is the complete cases analysis, in which patients with missing values are removed, and it is specified in the function sanon by `res.na.action="remove"`.

The support functions for `sanon` are also available. First of all is the `print` function for the display for the output of the function `sanon`.

`print(object)`

The `object` is an object of class `sanon`, usually, a result of a call to `sanon`, which is usually obtained by the sentence `object = sanon(...)`. The `print` function is equivalent to typing `object` in the R console. The output consists of the input, the sample size, the strata, the response levels, the design matrix, and the stratificated adjusted Mann-Whitney estimates. The reference group and adjusting factor names are also represented to help the interpretation.

A generic function `coef` is available. It extracts the resulting adjusted estimates b defined in (10) from the `object` which is the output of `sanon`.

`coef(object)`

This is a part of the output of the function `print`, but one can obtain it as the vector object to be available for additional algebra calculation in R.

The corresponding estimated covariance matrix V_b can be extracted by using

`vcov(object)`

Because V_b is not represented as the output of the function `print`, this `vcov` function can be helpful to check it. This can be obtained as the matrix object to be available for additional algebra calculation in R.

The results of inference after weighted least squares estimation can be obtained through the function `summary`.

```
summary(object)
```

The output is a $p \times 4$ matrix with columns for the estimated coefficient, its standard error, chi-squared statistic and corresponding (two-sided) p-value. The estimate in this function is $\hat{\xi} - 0.5$ and the corresponding p-value is for the hypothesis H_0: $\xi = 0.5$.

The `confint` function for computing the confidence interval for $\boldsymbol{b}$ is also available.

```
confint(object, parm = NULL, level = 0.95)
```

A specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names is needed. If missing, all parameters are considered. The confidence level can be specified.

The linear statistics $\boldsymbol{c}^\top \boldsymbol{b}$ for assessments of homogeneity or average across the response variables and the corresponding confidence intervals for the average are computed by the function `contrast` where $\boldsymbol{c}$ is the contrast matrix and $\boldsymbol{b}$ is a weighted least squares estimator in (10) It is used for this purpose after subtraction of 0.5 from $\boldsymbol{cb}$.

```
contrast(object, C = diag(length(object$b)), confint = FALSE,
level = 0.95)
```

This function provides the inference based on contrast after applying the function `sanon`. The contrast matrix `C` should be defined by the user. If `C` has only one row, and the input is specified as `confint = TRUE` (default `FALSE`), the confidence interval for the estimator is produced.

More details for the usage of this function `sanon` will be explained through examples in Section 5. As another usage, one can specify R objects to outcome, group, and strata variables and covariables; see `help(sanon)` for details.

4 Illustrative data sets

There are five data sets in the `sanon` package. The first two example data sets were used in Kawaguchi et al. (2011), and the other three example data sets were used in other papers. As mentioned in the previous section, the following program codes are available after installing the package by using the code `install.packages("sanon")`.

4.1 Chronic pain data

The first example data is `cpain` in the `sanon` package, which can be considered as follows.

```
R> library("sanon")
R> data("cpain")
R> summary(cpain)

     treat           response   center   diagnosis
 active :97    poor      :49    I :114   A:55
 placebo:96    fair      :34    II: 79   B:58
               moderate  :25             C:29
               good      :58             D:51
               excellent:27

R> head(cpain)

   treat response center diagnosis
1 active     poor     II         D
2 active     poor     II         D
3 active     poor     II         D
4 active     fair     II         D
5 active     fair     II         D
6 active moderate     II         D
```

These data are from a multi-center randomized clinical trial to compare test and control treatments for the management of chronic pain, and they have had previous consideration in Stokes et al. (2012, chap. 14). This clinical trial has 8 strata (in correspondence to 2 centers × 4 diagnoses) for which the range of sample sizes is 10 to 34 and a univariate ordinal response variable with 5 categories (as excellent, good, moderate, fair, poor) for pain status after treatment for 4 weeks.

4.2 Respiratory disorder data

The second example data is `resp` in the `sanon` package, which can be considered as follows.

```
R> data("resp")
R> summary(resp)

 center treatment sex       age            baseline
 1:56   A:54      F:23   Min.   :11.00   Min.   :0.000
```

```
  2:55   P:57       M:88   1st Qu.:23.00   1st Qu.:2.000
                          Median :31.00   Median :2.000
                          Mean   :33.28   Mean   :2.378
                          3rd Qu.:43.00   3rd Qu.:3.000
                          Max.   :68.00   Max.   :4.000
       visit1           visit2           visit3           visit4
 Min.   :0.000    Min.   :0.000    Min.   :0.000    Min.   :0.00
 1st Qu.:2.000    1st Qu.:2.000    1st Qu.:2.000    1st Qu.:2.00
 Median :3.000    Median :3.000    Median :3.000    Median :3.00
 Mean   :2.721    Mean   :2.622    Mean   :2.631    Mean   :2.55
 3rd Qu.:4.000    3rd Qu.:4.000    3rd Qu.:4.000    3rd Qu.:4.00
 Max.   :4.000    Max.   :4.000    Max.   :4.000    Max.   :4.00
```

R> head(resp)

```
  center treatment sex age baseline visit1 visit2 visit3 visit4
1      1         1   A   F  32        1      2      2      4      2
2      2         2   A   F  37        1      3      4      4      4
3      3         1   A   F  47        2      2      3      4      4
4      4         2   A   F  39        2      3      4      4      4
5      5         1   A   M  11        4      4      4      4      2
6      6         2   A   F  60        4      4      3      3      4
```

This example is from a randomized clinical trial to compare a test treatment to placebo for a respiratory disorder, and listings of the data appear in Stokes et al. (2012, chap. 15, pp. 515–516) and Koch et al. (1990). This clinical trial has 111 patients from two centers, and it has four post-baseline visits with corresponding ordinal response variables for patient global ratings of symptom control according to 5 categories (as 4 = excellent, 3 = good, 2 = fair, 1 = poor, 0 = terrible).

4.3   Relief of heartburn data

The third example data is `heartburn` in the `sanon` package, which can be considered as follows.

R> data("heartburn")

R> summary(heartburn)

```
      OBS           center sequence        ID              age
 Min.   : 1.00   1:30   AP:30   Min.   : 1.00   Min.   :22.00
 1st Qu.:15.75   2:30   PA:30   1st Qu.: 8.00   1st Qu.:29.00
 Median :30.50                  Median :15.50   Median :32.00
 Mean   :30.50                  Mean   :15.40   Mean   :35.88
 3rd Qu.:45.25                  3rd Qu.:22.25   3rd Qu.:39.00
 Max.   :60.00                  Max.   :30.00   Max.   :65.00
                                                NA's   :1
      sex          freq            MD1             MD2          res1
 female:46   Min.   : 2.000   Min.   : 2.00   Min.   : 4.00   NR:32
 male  :14   1st Qu.: 2.500   1st Qu.: 9.00   1st Qu.:13.00   R :28
             Median : 3.000   Median :25.00   Median :35.00
             Mean   : 4.808   Mean   :30.28   Mean   :34.77
             3rd Qu.: 4.000   3rd Qu.:60.00   3rd Qu.:60.00
             Max.   :45.000   Max.   :60.00   Max.   :60.00


 res2
 NR:36
 R :24


R> head(heartburn)

  OBS center sequence ID age    sex freq MD1 MD2 res1 res2
1   1      1       AP  2 55 female  6.5   7  35    R   NR
2   2      1       AP  3 35 female  3.0   5  60    R   NR
3   3      1       AP  6 36 female  4.0  10  28    R   NR
4   4      1       AP  7 44 female  3.5  14  36    R   NR
5   5      1       AP  9 38 female  7.0  60  35   NR   NR
6   6      1       AP 11 46 female  3.5   5  15    R    R
```

The data are from a two period cross-over design clinical trial for relief of heartburn, and listings of the data appear in Koch et al. (1983). This clinical trial has 30 patients at each of two centers, 15 randomly assigned to the A:P sequence group (active treatment for the first period and placebo for the second period) and 15 to the P:A sequence group (placebo for the first period and active for the second period). The ordinal response (MD1 and MD2) is a

composite measure for time to relief. This composite measure has values between 1 and 20 if relief occurred within 20 minutes from first dose of treatment; between 21 and 40 if relief occurred within 20 minutes of the second dose (and there was no relief from the first dose within 15 minutes and use of the second dose shortly thereafter); and the value 60 if there was no relief from either the first or second dose within the respective 20 minute limits. Koch et al. (1983) used the binary response which is whether or not the patient experienced relief within 15 minutes of the first dose of treatment (`res1` and `res2`).

4.4 Seborrheic dermatitis data

The fourth example data is `sebor` in the `sanon` package, which can be considered as follows.

```
R> data("sebor")
R> summary(sebor)

     center        treat         score1          score2
 4      :34   placebo:84   Min.   :0.000   Min.   :0.000
 3      :30   test   :83   1st Qu.:1.000   1st Qu.:0.000
 2      :27                Median :1.000   Median :2.000
 5      :27                Mean   :1.795   Mean   :1.964
 1      :19                3rd Qu.:3.000   3rd Qu.:3.000
 8      :16                Max.   :5.000   Max.   :5.000
 (Other):14                NA's   :1       NA's   :84
     score3        severity1       severity2       severity3
 Min.   :0.0   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:0.0   1st Qu.:2.000   1st Qu.:1.500   1st Qu.:1.250
 Median :1.5   Median :2.000   Median :2.000   Median :2.000
 Mean   :2.0   Mean   :2.054   Mean   :1.816   Mean   :1.778
 3rd Qu.:3.0   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
 Max.   :5.0   Max.   :3.000   Max.   :3.000   Max.   :3.000
 NA's   :149   NA's   :1       NA's   :80      NA's   :149

R> head(sebor)

  center  treat score1 score2 score3 severity1 severity2 severity3
1      1   test      1     NA     NA         2        NA        NA
```

16

```
2      1 placebo     1       1      NA       2       2      NA
3      1    test      0       1      NA       3       2      NA
4      1 placebo      0       1      NA       2       1      NA
5      1 placebo      1       2      NA       2       2      NA
6      1    test      1       1      NA       3       2      NA
```

The data are from a randomized clinical trial to compare a test treatment to placebo for seborrheic dermatitis, and listings of the data appear in Ramaswamy et al. (1997). This clinical trial has 167 patients from eight centers, and it has three anatomical sites (face, scalp, and chest) corresponding to ordinal response variables for patient global scores according to 6 categories (as 0 = cleared, 1 = excellent improvement, 2 = moderate improvement, 3 = slight improvement, 4 = no change, 5 = exacerbation).

4.5   Skin conditions data

The fifth example data is `skin` in the `sanon` package, which can be considered as follows.

```
R> data("skin")
R> summary(skin)

 center       treat         stage           res1            res2
 1:37   placebo:84   Min.   :3.000   Min.   :1.00   Min.   :1.000
 2:33   test   :88   1st Qu.:3.000   1st Qu.:2.00   1st Qu.:1.000
 3:30                Median :4.000   Median :3.00   Median :2.000
 4: 4                Mean   :3.622   Mean   :2.74   Mean   :2.442
 5:35                3rd Qu.:4.000   3rd Qu.:4.00   3rd Qu.:3.000
 6:33                Max.   :5.000   Max.   :5.00   Max.   :5.000
                                     NA's   :3      NA's   :16
      res3
 Min.   :1.00
 1st Qu.:1.00
 Median :2.00
 Mean   :2.31
 3rd Qu.:3.00
 Max.   :5.00
 NA's   :30
```

```
R> head(skin)

  center treat stage res1 res2 res3
1      1  test     3    3   NA    3
2      1  test     3    3    2    2
3      1  test     4    3    2    2
4      1  test     3    2    2    1
5      1  test     3    3    2    2
6      1  test     4    2    1    3
```

The data are from a randomized clinical trial to compare a test treatment to placebo for skin conditions, and listings of the data appear in Stanish et al. (1978) and Stanish et al. (1978). This clinical trial has 172 patients from six centers, and it has three post-baseline visits with corresponding ordinal response variables for patient extent of improvement according to 5 categories (as $1 =$ rapidly improving, $2 =$ slowly improving, $3 =$ stable, $4 =$ slowly worsening, $5 =$ rapidly worsening).

5 Examples

5.1 Chronic pain data

The first example is a randomized clinical trial with eight strata and a univariate ordinal response variable for its primary analysis. The function `sanon` is applied to this data set as follows.

```
R> out11 = sanon(response ~ grp(treat, ref="placebo") + strt(center)
+   + strt(diagnosis), data=cpain)
R> out11

Call:
sanon.formula(formula = response ~ grp(treat, ref = "placebo") +
    strt(center) + strt(diagnosis), data = cpain)


Sample size: 193


Strata ( center*diagnosis ): I*A, I*B, I*C, I*D, II*A, II*B, II*C, II*D
```

```
Response levels:

[response; 5 levels] (lower) poor, fair, moderate, good, excellent (higher)


Design Matrix:
         [,1]
response    1


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ active / placebo ] :
response
  0.5804
```

The argument `data=cpain` is for specification of data set. Each variable in the data set is given a role by the formula which has a similar nature as the standard `R` function such as `lm` for linear models. The outcome `response` is put in the left side of ˜. In the right side, the group variable `treat` and the strata variables `center` and `diagnosis` are specified by using the functions `grp` and `strt`, respectively, which are connected with "+". The reference group of `treat` can be specified in the function `grp` by `ref="placebo"`. Two strata variables are taken as the cross-classification of two centers (as I, II) and four diagnoses (as A, B, C, D) corresponding to 8 strata. The result is stored in the object `out11` and shown through the `print` function for the `sanon` function, which indicates the resulting stratified Mann-Whitney estimate of $\hat{\xi} = 0.5804$.

The `summary` function is also available. This function produces the inference of the stratified Mann-Whitney estimator based on the null hypothesis of $H_0 : \xi = 0.5$.

```
R> summary(out11)

Call:
sanon.formula(formula = response ~ grp(treat, ref = "placebo") +
    strt(center) + strt(diagnosis), data = cpain)



        Estimate Std.Err Chisq Pr(>Chisq)
response   0.0804  0.0417  3.72      0.054 .
```

```
---
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).
```

The Estimate column of the output is $\hat{\xi} - 0.5 = 0.0804$ and its robust standard error from the methods in Section 2 is s.e.$(\hat{\xi}) = 0.0417$. The test statistic $Q = \{(\hat{\xi} - 0.5)/s.e.(\hat{\xi})\}^2$ with p-value from an approximate chi-squared distribution with d.f.=1 is also produced.

The `confint` function for computing the confidence interval is also available.

```
R> confint(out11)

M-W Estimate and 95% Confidence Intervals
:

         Estimate  Lower  Upper
response   0.5804 0.4988 0.6621
```

The resulting two-sided 0.95 confidence interval for $\hat{\xi}$ is (0.4988, 0.6621).

For illustrative purposes, `diagnosis` is alternatively managed as a categorical covariable with `center` remaining as a factor for stratification. This structure is of exploratory interest since it has substantially larger sample sizes within its strata.

```
R> out12 = sanon(response ~ grp(treat, ref="placebo") + strt(center)
+   + catecovar(diagnosis, ref="D"), data=cpain)
R> out12

Call:
sanon.formula(formula = response ~ grp(treat, ref = "placebo") +
    strt(center) + catecovar(diagnosis, ref = "D"), data = cpain)


Sample size: 193


Strata ( center ): I, II


Response levels:
[response; 5 levels] (lower) poor, fair, moderate, good, excellent (higher)
```

```
Design Matrix:
                [,1]
response           1
diagnosis[C/D]     0
diagnosis[B/D]     0
diagnosis[A/D]     0


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ active / placebo ] :
(adjusted by diagnosis[C/D], diagnosis[B/D], diagnosis[A/D] )
response
   0.5729
```

The categorical covariable is specified by the `catecovar` function of the `sanon` package. Since `diagnosis` has four categories, the function `catecovar` produces three dummy variables with the reference of D; that is, the first dummy variable is 1 for `diagnosis` = C and 0 for others, the second dummy variable is 1 for `diagnosis` = B and 0 for others, and the third dummy variable is 1 for `diagnosis` = A and 0 for others. Note that, in default setting, $\boldsymbol{P} = [\boldsymbol{I}_r, \boldsymbol{0}_{rM}]^\top$ as in (10).

```
R> summary(out12)


Call:
sanon.formula(formula = response ~ grp(treat, ref = "placebo") +
    strt(center) + catecovar(diagnosis, ref = "D"), data = cpain)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by diagnosis[C/D], diagnosis[B/D], diagnosis[A/D] ):
         Estimate Std.Err Chisq Pr(>Chisq)
response    0.0729  0.0387  3.55       0.059 .
---
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).


R> confint(out12)
```

```
M-W Estimate and 95% Confidence Intervals
(adjusted by diagnosis[C/D], diagnosis[B/D], diagnosis[A/D] ):
        Estimate  Lower  Upper
response   0.5729 0.4971 0.6488
```

The result is very similar to the previous analysis with the adjusted Mann-Whitney estimator being 0.5729 (95%CI: 0.4971 – 0.6488; $p = 0.059$). As noted in Section 2.2, the management of a baseline factor as strata or as a covariable should be pre-specified.

5.2 Respiratory disorder data

The second example is a randomized clinical trial with four strata, two covariables, and four ordinal response variables. This example has gender (female or male) as an additional factor for stratification, and so there are 4 strata for center*gender; and it has age and the baseline rating of symptom control (with the same ordinal categories as the response variables) as two covariables. The application of function `sanon` to this data set is as follows.

```
R> out21 = sanon(cbind(baseline, visit1, visit2, visit3, visit4)
+  ~ grp(treatment, ref="P") + strt(center) + strt(sex) + covar(age)
+  , data=resp, P=diag(6))
R> out21

Call:
sanon.formula(formula = cbind(baseline, visit1, visit2, visit3,
    visit4) ~ grp(treatment, ref = "P") + strt(center) + strt(sex) +
    covar(age), data = resp, P = diag(6))


Sample size: 111


Strata ( center*sex ): 1*F, 1*M, 2*F, 2*M


Response levels:
[baseline; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit1; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit2; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit3; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
```

```
[visit4; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)


Design Matrix:
          [,1] [,2] [,3] [,4] [,5] [,6]
baseline    1    0    0    0    0    0
visit1      0    1    0    0    0    0
visit2      0    0    1    0    0    0
visit3      0    0    0    1    0    0
visit4      0    0    0    0    1    0
age         0    0    0    0    0    1


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ A / P ] :
baseline   visit1   visit2   visit3   visit4      age
  0.4799   0.6005   0.7139   0.6535   0.6155   1.0501
```

Multivariate responses in the left side of ˜ are specified by the R function `cbind`. The covariable is specified by the `covar` function of the `sanon` package. The matrix $\boldsymbol{P}$ in (10) is an identity matrix $\boldsymbol{I}_6$ specifed by `P=diag(6)`. Others are the same as the previous example.

The resulting vector of stratification adjusted estimators for the comparisons between the test treatment and placebo is in (13) containing $\hat{\xi}_0$, $\hat{\xi}_1$, $\hat{\xi}_2$, $\hat{\xi}_3$, $\hat{\xi}_4$ for the Mann-Whitney estimators that correspond to the baseline visit and visits 1, 2, 3, 4 and g for the difference between mean ages.

$$\boldsymbol{f} = [\hat{\xi}_0,\ \hat{\xi}_1,\ \hat{\xi}_2,\ \hat{\xi}_3,\ \hat{\xi}_4,\ g]^\top = [0.4799,\ 0.6005,\ 0.7139,\ 0.6535,\ 0.6155,\ 1.0501]^\top, \qquad (13)$$

The corresponding estimated covariance matrix from the methods in Section 2 can be obtained as follows.

```
R> round(out21$Vf, 5)

          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.00319   0.00152   0.00088   0.00088   0.00088   0.01717
[2,] 0.00152   0.00285   0.00141   0.00141   0.00132  -0.00044
[3,] 0.00088   0.00141   0.00234   0.00166   0.00164  -0.01633
```

```
[4,] 0.00088   0.00141   0.00166   0.00283   0.00208  -0.01578
[5,] 0.00088   0.00132   0.00164   0.00208   0.00278  -0.00044
[6,] 0.01717  -0.00044  -0.01633  -0.01578  -0.00044   6.82200
```

Both $(\hat{\xi}_0 - 0.5)$ and g have null expected values on the basis of randomization of patients to the two treatment groups. The function `contrast` is used for this purpose after subtraction of 0.5 from $\hat{\xi}_0$, $\hat{\xi}_1$, $\hat{\xi}_2$, $\hat{\xi}_3$ and $\hat{\xi}_4$.

```
R> contrast(out21, C=rbind(c(1,0,0,0,0,0), c(0,0,0,0,0,1)))

Contrast Matrix:
     baseline visit1 visit2 visit3 visit4 age
[1,]        1      0      0      0      0   0
[2,]        0      0      0      0      0   1


Contrast Inference:
     Chisq df Pr(>Chisq)
[1,] 0.325  2       0.85
```

$Q_C = b^\top C^\top (C^\top V_b C)^{-1} C b = 0.33$ with d.f. $= 2$ and $p = 0.85$ for the corresponding assessment of random imbalance was obtained.

Correspondingly adjusted estimation for $(\hat{\xi}_1 - 0.5)$, $(\hat{\xi}_2 - 0.5)$, $(\hat{\xi}_3 - 0.5)$, $(\hat{\xi}_4 - 0.5)$ is possible by fitting the model $P = [0_4, I_4, 0_4]^\top$ to f (after subtraction of 0.5 from each of the $\hat{\xi}_k$) by weighted least squares; here 0_4 is the 4×1 vector of 0's and I_4 is the 4×4 identity matrix. The resulting adjusted estimators $b^\top$ from the methods in Section 2.2 can be computed as follows.

```
R> P = rbind(rep(0, 4), diag(4), rep(0, 4))
R> out22 = sanon(cbind(baseline, visit1, visit2, visit3, visit4)
+    ~ grp(treatment, ref="P") + strt(center) + strt(sex) + covar(age)
+    , data=resp, P=P)
R> out22

Call:
sanon.formula(formula = cbind(baseline, visit1, visit2, visit3,
    visit4) ~ grp(treatment, ref = "P") + strt(center) + strt(sex) +
    covar(age), data = resp, P = P)
```

```
Sample size: 111


Strata ( center*sex ): 1*F, 1*M, 2*F, 2*M


Response levels:
[baseline; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit1; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit2; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit3; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)
[visit4; 5 levels] (lower) 0, 1, 2, 3, 4 (higher)


Design Matrix:
          [,1] [,2] [,3] [,4]
baseline    0    0    0    0
visit1      1    0    0    0
visit2      0    1    0    0
visit3      0    0    1    0
visit4      0    0    0    1
age         0    0    0    0


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ A / P ] :
(adjusted by baseline, age )
visit1 visit2 visit3 visit4
0.6116 0.7230 0.6625 0.6219
```

The corresponding estimated covariance matrix V_b can be extracted by using the `vcov` function.

```
R> vcov(out22)


            visit1        visit2        visit3        visit4
visit1 0.0021137493 0.0009651028 0.0009633427 0.0008931803
visit2 0.0009651028 0.0020297172 0.0013550994 0.0013796796
```

```
visit3 0.0009633427 0.0013550994 0.0025311719 0.0018257579
visit4 0.0008931803 0.0013796796 0.0018257579 0.0025351549
```

The statistics $\{b_k/s.e.(b_k)\}^2$ have approximately chi-squared distributions with d.f. $= 1$ under the null hypotheses H_{0k}: $\xi_k = 0.5$.

```
R> summary(out22)

Call:
sanon.formula(formula = cbind(baseline, visit1, visit2, visit3,
    visit4) ~ grp(treatment, ref = "P") + strt(center) + strt(sex) +
    covar(age), data = resp, P = P)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by baseline, age ):
       Estimate Std.Err Chisq Pr(>Chisq)
visit1   0.1116  0.0460  5.89     0.0152 *
visit2   0.2230  0.0451 24.51    7.4e-07 ***
visit3   0.1625  0.0503 10.43     0.0012 **
visit4   0.1219  0.0504  5.86     0.0155 *
---
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).
```

The respective p-values are 0.0152, < 0.0001, 0.0012, 0.0155 for visits 1, 2, 3, 4. Homogeneity of the ξ_k across the four visits can be assessed with $Q_{C_1} = \boldsymbol{b}^\top \boldsymbol{C}_1{}^\top (\boldsymbol{C}_1 \boldsymbol{V}_{\boldsymbol{b}} \boldsymbol{C}_1{}^\top)^{-1} \boldsymbol{C}_1 \boldsymbol{b}$ where $\boldsymbol{C}_1 = [\boldsymbol{I}_3, -\boldsymbol{1}_3]$ and $\boldsymbol{1}_3$ is the (3×1) vector of 1's.

```
R> contrast(out22, C=cbind(diag(3), rep(-1, 3)))

Contrast Matrix:
     visit1 visit2 visit3 visit4
[1,]      1      0      0     -1
[2,]      0      1      0     -1
[3,]      0      0      1     -1
```

```
Contrast Inference:
     Chisq df Pr(>Chisq)
[1,]  8.93  3        0.03 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since $Q_{C_1} = 8.93$ has $p = 0.03$ with respect to its approximate chi-squared distribution with d.f. = 3, there is some suggestion of departures of the ξ_k from homogeneity; and from inspection of their estimates, it appears that the difference between test treatment and placebo tends to be larger at Visit 2 and Visit 3 than at Visit 1 and Visit 4.

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 4 visits is possible with $Q_{C_2} = \boldsymbol{b}^\top \boldsymbol{C_2}^\top (\boldsymbol{C_2} \boldsymbol{V_b} \boldsymbol{C_2}^\top)^{-1} \boldsymbol{C_2} \boldsymbol{b}$ with $\boldsymbol{C_2} = \boldsymbol{1}_4^\top / 4$;

```
R> contrast(out22, C=matrix(rep(1, 4)/4, ncol=4), confint = TRUE)

Contrast Matrix:
     visit1 visit2 visit3 visit4
[1,]   0.25   0.25   0.25   0.25


Contrast Inference:
     Chisq df Pr(>Chisq)
[1,]    16  1    6.4e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Contrast M-W Estimate (95% Confidence Interval)
0.6548 ( 0.5789 - 0.7306 )
```

Thus, the result for testing the null hypothesis $H_0 : \boldsymbol{C_2}\boldsymbol{\xi} = 0.5$ is $Q_{C_2} = 16$ for which two-sided p<0.0001 with respect to the approximate chi-squared distribution with d.f. = 1. Note that the `contrast` computes the Mann-Whitney estimate and its confidence interval (CI) adjusted for strata and covariables with the specification `confint = TRUE` in the case where $\boldsymbol{C}$ has one row (i.e., d.f.=1). The estimate averaged over visits was $\boldsymbol{C_2}\boldsymbol{b} + 0.5 = 0.6548$ (95%CI 0.5789 – 0.7306).

5.3 Relief of heartburn data

The third example is a cross over design randomized clinical trial with two strata, one continuous covariable, and two ordinal response variables. The two strata are for centers, and age is a continuous covariable. For illustrative convenience, a missing value in age is imputed with the mean for the corresponding stratum.

```
R> heartburn2 = heartburn
R> heartburn2$age[is.na(heartburn$age)] = mean(heartburn$age, na.rm=TRUE)
```

The application of function sanon to this data set is as follows.

```
R> out31 = sanon(cbind(MD1, MD2) ~ grp(sequence, ref="AP") + strt(center)
+  + covar(age), data=heartburn2)
R> out31

Call:
sanon.formula(formula = cbind(MD1, MD2) ~ grp(sequence, ref = "AP") +
    strt(center) + covar(age), data = heartburn2)


Sample size: 60


Strata ( center ): 1, 2


Response levels:
[MD1; 20 levels] (lower) 2, 3, 4, ..., 30, 32, 60 (higher)
[MD2; 15 levels] (lower) 4, 7, 8, ..., 35, 36, 60 (higher)


Design Matrix:
    [,1] [,2]
MD1    1    0
MD2    0    1
age    0    0


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ PA / AP ] :
```

```
(adjusted by age )
   MD1    MD2
0.5597 0.2199
```

The statistics $\{b_k/s.e.(b_k)\}^2$, where the b_k are the adjusted estimators for the $(\xi_k - 0.5)$ have approximately chi-squared distributions with d.f. = 1 under the null hypotheses H_{0k}: ξ_k = 0.5. Although both b_1 for the comparison between P and A for the first period and b_2 for the comparison between P and A for the second period indicate more favorable responses for active treatment A than for placebo P, only the second period has $p < 0.05$ for the comparison between A and P.

```
R> summary(out31)

Call:
sanon.formula(formula = cbind(MD1, MD2) ~ grp(sequence, ref = "AP") +
    strt(center) + covar(age), data = heartburn2)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by age ):
    Estimate Std.Err Chisq Pr(>Chisq)
MD1   0.0597  0.0752  0.63       0.43
MD2  -0.2801  0.0608 21.24    4.1e-06 ***
---
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).
```

Since it is recognized that the comparison between the AP and PA groups for period 1 pertains to A versus P whereas that for period 2 pertains to P versus A, treatment×period interaction is assessed for $(b_1 + b_2)$ with $Q_{C_1} = \boldsymbol{b}^\top \boldsymbol{C}_1{}^\top (\boldsymbol{C}_1 \boldsymbol{V_b} \boldsymbol{C}_1{}^\top)^{-1} \boldsymbol{C}_1 \boldsymbol{b}$ where $\boldsymbol{C}_1 = [1\ 1]$.

```
R> contrast(out31, C=cbind(1, 1))

Contrast Matrix:
     MD1 MD2
[1,]   1   1
```

```
Contrast Inference:

     Chisq df Pr(>Chisq)
[1,]  7.17  1     0.0074 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since $Q_{C_1} = 7.17$ has $p = 0.0074$ with respect to its approximate chi-squared distribution with d.f. $= 1$, there is some suggestion that the difference between test treatment and placebo is not homogeneous for the two periods by tending to be larger at period 2 than at period 1. Since the two periods of this crossover study were separated by a sufficiently long washout period, the treatment $\times$ period interaction corresponding to this departure from homogeneity is unlikely to be due to unequal pharmacologic carryover effects of A and P during the first period.

The Mann-Whitney estimates and their confidence intervals adjusted for strata and covariables for each period are computed as follows.

```
R> confint(out31)

M-W Estimate and 95% Confidence Intervals
(adjusted by age ):
     Estimate  Lower  Upper
MD1    0.5597 0.4122 0.7072
MD2    0.2199 0.1007 0.3390
```

5.4   Seborrheic dermatitis data

The fourth example is a randomized clinical trial with seven strata (after the pooling of centers 6 and 7 in the original data as in Ramaswamy et al. (1997)), three ordinal covariables with missing values, and three ordinal response variables with missing values. For this example, the MCAR assumption is realistic because missing sites didn't receive treatment.

```
R> sebor2 = sebor
R> sebor2$center = ifelse(sebor2$center == 6, 7, sebor2$center)
```

This example has 7 centers and disease severities recorded at the baseline measurement. In the previous three examples, centers were managed as strata to account for the stratified randomization within centers. However, for this example they are managed as a categorical

covariable, mainly because the sample size of 18 patients for score 3 is insufficient to support stratified estimation (relative to the 167 patients in this clinical trial). Additionally, initial disease severities can be considered as covariables; and they are ordinal with some missing values. Thus, they are treated in the left side of ˜ in the function sanon as responses.

```
R> out41 =
+  sanon(cbind(score1, score2, score3, severity1, severity2, severity3)
+  ~ grp(treat, ref="placebo") + catecovar(center, ref="8"), data=sebor2)
R> out41

Call:
sanon.formula(formula = cbind(score1, score2, score3, severity1,
    severity2, severity3) ~ grp(treat, ref = "placebo") + catecovar(center,
    ref = "8"), data = sebor2)


Sample size: 167


Response levels:
[score1; 6 levels] (lower) 0, 1, 2, 3, 4, 5 (higher)
[score2; 6 levels] (lower) 0, 1, 2, 3, 4, 5 (higher)
[score3; 6 levels] (lower) 0, 1, 2, 3, 4, 5 (higher)
[severity1; 3 levels] (lower) 1, 2, 3 (higher)
[severity2; 3 levels] (lower) 1, 2, 3 (higher)
[severity3; 3 levels] (lower) 1, 2, 3 (higher)


Design Matrix:
              [,1] [,2] [,3] [,4] [,5] [,6]
score1          1    0    0    0    0    0
score2          0    1    0    0    0    0
score3          0    0    1    0    0    0
severity1       0    0    0    1    0    0
severity2       0    0    0    0    1    0
severity3       0    0    0    0    0    1
center[1/8]     0    0    0    0    0    0
center[2/8]     0    0    0    0    0    0
```

```
center[3/8]     0     0     0     0     0     0
center[4/8]     0     0     0     0     0     0
center[5/8]     0     0     0     0     0     0
center[7/8]     0     0     0     0     0     0


Mann-Whitney Estimate
 for comparison [ test / placebo ] :
(adjusted by center[1/8], center[2/8], center[3/8], center[4/8],
 center[5/8], center[7/8] )
   score1     score2     score3 severity1 severity2 severity3
   0.4395     0.3934     0.6477    0.5421    0.4752    0.4614

R> summary(out41)

Call:
sanon.formula(formula = cbind(score1, score2, score3, severity1,
    severity2, severity3) ~ grp(treat, ref = "placebo") + catecovar(center,
    ref = "8"), data = sebor2)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by center[1/8], center[2/8], center[3/8], center[4/8],
 center[5/8], center[7/8] ):
          Estimate Std.Err Chisq Pr(>Chisq)
score1     -0.0605  0.0403  2.25       0.13
score2     -0.1066  0.0587  3.29       0.07 .
score3      0.1477  0.1306  1.28       0.26
severity1   0.0421  0.0278  2.29       0.13
severity2  -0.0248  0.0499  0.25       0.62
severity3  -0.0386  0.1095  0.12       0.72
---
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).
```

Values of MW estimates less than 0.5 represent lower (better) scores for test. For each variable, there is no significant treatment effect.

To manage initial disease severities as covariables, the model $\boldsymbol{P} = [\boldsymbol{I}_3, \boldsymbol{0}_{3,9}]^\top$ is fitted to $\boldsymbol{f}$ (after subtraction of 0.5 from each of the $\hat{\xi}_k$) by weighted least squares; here $\boldsymbol{0}_{3,9}$ is a $3{\times}9$ matrix with all entries being 0.

```
R> P = rbind(diag(3), matrix(0, 9, 3))
R> out42 =
+  sanon(cbind(score1, score2, score3, severity1, severity2, severity3)
+  ~ grp(treat, ref="placebo") + catecovar(center, ref="8"), data=sebor2
+  , P=P)
R> out42$matP
```

```
              [,1] [,2] [,3]
score1           1    0    0
score2           0    1    0
score3           0    0    1
severity1        0    0    0
severity2        0    0    0
severity3        0    0    0
center[1/8]      0    0    0
center[2/8]      0    0    0
center[3/8]      0    0    0
center[4/8]      0    0    0
center[5/8]      0    0    0
center[7/8]      0    0    0
```

```
R> summary(out42)
```

```
Call:
sanon.formula(formula = cbind(score1, score2, score3, severity1,
    severity2, severity3) ~ grp(treat, ref = "placebo") + catecovar(center,
    ref = "8"), data = sebor2, P = P)
```

```
Randomization-Based Covariance Adjusted Analysis
 (adjusted by severity1, severity2, severity3, center[1/8], center[2/8],
 center[3/8], center[4/8], center[5/8], center[7/8] ):
        Estimate Std.Err Chisq Pr(>Chisq)
```

```
score1  -0.0696  0.0396  3.09        0.079 .

score2  -0.1137  0.0580  3.84        0.050 .

score3   0.1462  0.1262  1.34        0.246

---

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).
```

Homogeneity of the adjusted ξ_k across the three sites can be assessed with $Q_{C_1} = \boldsymbol{b}^\top \boldsymbol{C}_1^\top$ $(\boldsymbol{C}_1 \boldsymbol{V}_b \boldsymbol{C}_1^\top)^{-1} \boldsymbol{C}_1 \boldsymbol{b}$ where $\boldsymbol{C}_1 = [\boldsymbol{I}_2, -\boldsymbol{1}_2]$ and $\boldsymbol{1}_2$ is the (2×1) vector of 1's.

```
R> contrast(out42, C=cbind(diag(2), rep(-1, 2)))


Contrast Matrix:

     score1 score2 score3

[1,]      1      0     -1

[2,]      0      1     -1


Contrast Inference:

     Chisq df Pr(>Chisq)

[1,]  4.43  2       0.11
```

Since $Q_{C_1} = 4.43$ has $p = 0.11$ with respect to its approximate chi-squared distribution with d.f. $= 2$, there is no suggestion of departures of the $\hat{\xi}_k$ from homogeneity.

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 sites is possible with the model $\boldsymbol{P} = [1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^\top$, and it is fitted to $\boldsymbol{f}$ (after subtraction of 0.5 from each of the $\hat{\xi}_k$) by weighted least squares;

```
R> P = matrix(c(rep(1, 3), rep(0, 9)), ncol=1)

R> out43 =

+  sanon(cbind(score1, score2, score3, severity1, severity2, severity3)

+  ~ grp(treat, ref="placebo") + catecovar(center, ref="8"), data=sebor2

+  , P=P)

R> summary(out43)


Call:

sanon.formula(formula = cbind(score1, score2, score3, severity1,
```

```
      severity2, severity3) ~ grp(treat, ref = "placebo") + catecovar(center,
      ref = "8"), data = sebor2, P = P)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by severity1, severity2, severity3, center[1/8], center[2/8],
 center[3/8], center[4/8], center[5/8], center[7/8] ):
                         Estimate Std.Err Chisq Pr(>Chisq)
score1 + score2 + score3  -0.0662  0.0386  2.94       0.086 .
---
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).

R> confint(out43)

M-W Estimate and 95% Confidence Intervals
(adjusted by severity1, severity2, severity3, center[1/8], center[2/8],
 center[3/8], center[4/8], center[5/8], center[7/8] ):
                         Estimate  Lower  Upper
score1 + score2 + score3   0.4338 0.3581 0.5095
```

Its result for the overall comparison between treatments has two-sided $p = 0.086$. The estimate averaged over visits was $b + 0.5 = 0.4338$ (95%CI $0.3581 - 0.5095$). This comparison averages sites with more weight for those with larger sample size (and thereby less variance).

5.5   Skin conditions data

The fifth example is a randomized clinical trial with six strata, a categorical covariable with three levels, and three ordinal response variables with missing values. First of all, the centers 3 and 4 in the original data set are pooled, because there are only four observations in the center 4, and rankings are less informative when overall sample size $(n_{h_1} + n_{h_2}) = n_h$ within strata are < 10; see Kawaguchi et al. (2011).

```
R> skin2 = skin
R> skin2$center = ifelse(skin2$center == 4, 3, skin2$center)
```

This example has 5 strata for center and it has initial severity of the skin conditions as a categorical covariable. At three follow-up visits, patients were evaluated according to a five-

point ordinal response scale defining extent of improvement. There are some missing values in response variables. The application of function `sanon` to this data set is as follows.

```
R> out51 = sanon(cbind(res1, res2, res3) ~ grp(treat, ref="placebo")
+  + strt(center) + catecovar(stage, ref="3"), data=skin2)
R> out51

Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),
    data = skin2)


Sample size: 172


Strata ( center ): 1, 2, 3, 5, 6


Response levels:
[res1; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
[res2; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
[res3; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)


Design Matrix:
          [,1] [,2] [,3]
res1         1    0    0
res2         0    1    0
res3         0    0    1
stage[4/3]   0    0    0
stage[5/3]   0    0    0


Stratification Adjusted Mann-Whitney Estimate
 for comparison [ test / placebo ] :
(adjusted by stage[4/3], stage[5/3] )
  res1   res2   res3
0.1931 0.1537 0.1359
```

Note that since response variables are ordered so that lower values represent more improvement, values of MW estimates less than 0.5 represent better improvement for `test` treatment.

```
R> summary(out51)

Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),
    data = skin2)


Randomization-Based Covariance Adjusted Analysis
 (adjusted by stage[4/3], stage[5/3] ):
     Estimate Std.Err Chisq Pr(>Chisq)
res1  -0.3069  0.0331    86    <2e-16 ***
res2  -0.3463  0.0306   128    <2e-16 ***
res3  -0.3641  0.0319   130    <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).
```

Note that the estimates are for the $(\xi_k - 0.5)$. Homogeneity of the ξ_k across the three visits can have assessment with $Q_{C_1} = \boldsymbol{b}^\top \boldsymbol{C}_1^\top (\boldsymbol{C}_1 \boldsymbol{V}_b \boldsymbol{C}_1^\top)^{-1} \boldsymbol{C}_1 \boldsymbol{b}$ where $\boldsymbol{C}_1 = [\boldsymbol{I}_2, -\boldsymbol{1}_2]$ and $\boldsymbol{1}_2$ is the (2×1) vector of 1's.

```
R> contrast(out51, C=cbind(diag(2), rep(-1, 2)))

Contrast Matrix:
     res1 res2 res3
[1,]    1    0   -1
[2,]    0    1   -1


Contrast Inference:
     Chisq df Pr(>Chisq)
[1,]  3.58  2       0.17
```

Since $Q_{C_1} = 3.58$ has $p = 0.17$ with respect to its approximate chi-squared distribution with d.f. = 2, there is no suggestion of departures of the $\hat{\xi}_k$ from homogeneity.

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 visits is possible with $Q_{C_2} = \boldsymbol{b}^\top \boldsymbol{C_2}^\top (\boldsymbol{C_2} \boldsymbol{V_b} \boldsymbol{C_2}^\top)^{-1} \boldsymbol{C_2} \boldsymbol{b}$ with $\boldsymbol{C_2} = \boldsymbol{1}_3^\top / 3$;

```
R> contrast(out51, C=matrix(rep(1, 3)/3, ncol=3), confint = TRUE)

Contrast Matrix:
          res1      res2      res3
[1,] 0.3333333 0.3333333 0.3333333


Contrast Inference:
     Chisq df Pr(>Chisq)
[1,]   152  1    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Contrast M-W Estimate (95% Confidence Interval)
0.1609 ( 0.1069 - 0.2149 )
```

its result is $Q_{C_2} = 152$ for which two-sided p<0.0001 with respect to the approximate chi-squared distribution with d.f. = 1. The estimate averaged over visits was 0.1609 (95%CI 0.1069 – 0.2149). In the case of opposite order of response variables, it was 0.8391 (95%CI 0.7851 – 0.8931).

The function sanon has four other options for dealing with missing values for the response variable, although the default maintains missing values and uses the method described in Section 2 which can be specified by no argument or res.na.action="default". The second is the last observation carried forward (LOCF) based on the kernels of U-statistics in (2), which is specified in the function sanon by res.na.action="LOCF1".

```
R> out52 = sanon(cbind(res1, res2, res3) ~ grp(treat, ref="placebo")
+ + strt(center) + catecovar(stage, ref="3"), data=skin2
+ , res.na.action="LOCF1")
R> out52

Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),
```

```
     data = skin2, res.na.action = "LOCF1")
```

Sample size: 172

Strata (center): 1, 2, 3, 5, 6

Response levels:

[res1; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

[res2; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

[res3; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

Design Matrix:

```
          [,1] [,2] [,3]
res1         1   0    0
res2         0   1    0
res3         0   0    1
stage[4/3]   0   0    0
stage[5/3]   0   0    0
```

Stratification Adjusted Mann-Whitney Estimate

 for comparison [test / placebo] :

(adjusted by stage[4/3], stage[5/3])

```
  res1   res2   res3
0.2034 0.1666 0.1459
```

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 visits is

R> contrast(out52, C=matrix(rep(1, 3)/3, ncol=3), confint = TRUE)

Contrast Matrix:

```
          res1      res2      res3
[1,] 0.3333333 0.3333333 0.3333333
```

Contrast Inference:

```
    Chisq df Pr(>Chisq)
```

```
[1,]  130  1      <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Contrast M-W Estimate (95% Confidence Interval)
0.1720 ( 0.1155 - 0.2284 )
```

In this example, the p-value from LOCF1 was similar to that from default (the method described in Section 2).

The third option is the LOCF based on the observed value of Y, which is specified in the function sanon by res.na.action="LOCF2".

```
R> out53 = sanon(cbind(res1, res2, res3) ~ grp(treat, ref="placebo")
+  + strt(center) + catecovar(stage, ref="3"), data=skin2
+  , res.na.action="LOCF2")
R> out53

Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),
    data = skin2, res.na.action = "LOCF2")


Sample size: 172


Strata ( center ): 1, 2, 3, 5, 6


Response levels:
[res1; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
[res2; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
[res3; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)


Design Matrix:
          [,1] [,2] [,3]
res1         1    0    0
res2         0    1    0
```

```
res3            0    0    1
stage[4/3]      0    0    0
stage[5/3]      0    0    0
```

```
Stratification Adjusted Mann-Whitney Estimate
 for comparison [ test / placebo ] :
(adjusted by stage[4/3], stage[5/3] )
  res1   res2   res3
0.2034 0.1707 0.1485
```

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 visits is

```
R> contrast(out53, C=matrix(rep(1, 3)/3, ncol=3), confint = TRUE)
```

```
Contrast Matrix:
          res1      res2      res3
[1,] 0.3333333 0.3333333 0.3333333
```

```
Contrast Inference:
    Chisq df Pr(>Chisq)
[1,]   130  1    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Contrast M-W Estimate (95% Confidence Interval)
0.1742 ( 0.1181 - 0.2303 )
```

In this example, the p-value from `LOCF2` was similar to those from both `LOCF1` and `default` (the method described in Section 2).

The fourth method manages missing values as tied with all other values in the same stratum, which is specified in the function `sanon` by `res.na.action="replace"`.

```
R> out54 = sanon(cbind(res1, res2, res3) ~ grp(treat, ref="placebo")
+  + strt(center) + catecovar(stage, ref="3"), data=skin2
+  , res.na.action="replace")
R> out54
```

```
Call:

sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,

    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),

    data = skin2, res.na.action = "replace")
```

Sample size: 172

Strata (center): 1, 2, 3, 5, 6

Response levels:

[res1; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

[res2; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

[res3; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)

Design Matrix:

```
           [,1] [,2] [,3]
res1          1    0    0
res2          0    1    0
res3          0    0    1
stage[4/3]    0    0    0
stage[5/3]    0    0    0
```

Stratification Adjusted Mann-Whitney Estimate

 for comparison [test / placebo] :

(adjusted by stage[4/3], stage[5/3])

```
  res1    res2    res3
0.2034 0.2153 0.2533
```

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 visits is

R> contrast(out54, C=matrix(rep(1, 3)/3, ncol=3), confint = TRUE)

Contrast Matrix:

```
           res1      res2      res3
[1,] 0.3333333 0.3333333 0.3333333
```

```
Contrast Inference:

      Chisq df Pr(>Chisq)
[1,]   121  1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Contrast M-W Estimate (95% Confidence Interval)
0.2240 ( 0.1749 - 0.2731 )
```

The result shows that the p-value after managing missing values as tied with all other values in the same stratum was similar to those for LOCF1, LOCF2, and default (the method described in Section 2).

The fifth option for managing missing data is the complete cases analysis, in which patients with missing values are removed, and it is specified in the function sanon by res.na.action="remove".

```
R> out55 = sanon(cbind(res1, res2, res3) ~ grp(treat, ref="placebo")
+  + strt(center) + catecovar(stage, ref="3"), data=skin2
+  , res.na.action="remove")
R> out55

Call:
sanon.formula(formula = cbind(res1, res2, res3) ~ grp(treat,
    ref = "placebo") + strt(center) + catecovar(stage, ref = "3"),
    data = skin2, res.na.action = "remove")


Sample size: 135 ( 37 samples removed)


Strata ( center ): 1, 2, 3, 5, 6


Response levels:
[res1; 4 levels] (lower) 1, 2, 3, 4 (higher)
[res2; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
[res3; 5 levels] (lower) 1, 2, 3, 4, 5 (higher)
```

```
Design Matrix:

          [,1] [,2] [,3]

res1         1   0   0

res2         0   1   0

res3         0   0   1

stage[4/3]   0   0   0

stage[5/3]   0   0   0


Stratification Adjusted Mann-Whitney Estimate

 for comparison [ test / placebo ] :

(adjusted by stage[4/3], stage[5/3] )

  res1   res2   res3

0.1808 0.1409 0.1216
```

This output reports that 37 patients have missing values in at least one of the three responses and have their data removed from the analysis.

A comparison between treatments for the average of the $\hat{\xi}_k$ across the 3 visits is

```
R> contrast(out55, C=matrix(rep(1, 3)/3, ncol=3), confint = TRUE)

Contrast Matrix:

          res1        res2        res3

[1,] 0.3333333 0.3333333 0.3333333


Contrast Inference:

    Chisq df Pr(>Chisq)

[1,]  149  1    <2e-16 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Contrast M-W Estimate (95% Confidence Interval)

0.1478 ( 0.0913 - 0.2043 )
```

The result shows that the p-value after removing subjects with missing values was similar to that for the method described in Section 2.

Note that the complete case analysis makes the same MCAR assumption as the default method, but has the limitation of only using complete cases rather than all of the data with missing maintained as missing.

6 Summary

The R package `sanon` contains functions to implement the methods for stratified Mann-Whitney estimators. These methods address the same comparisons between two randomized groups for a strictly ordinal response variable as the van Elteren test statistic. Since the functions have similar structures to standard R functions, it would be accessible for R users. The role of variables in the analysis can be specified by using functions in the `sanon`. For example, the stratification variable is specified by the function `strt`. Among these, functions `treat` and `catecovar` have arguments to set the reference group. The function `sanon` can deal with missing values in five ways, according to the user's intent. This function `sanon` is also applicable to data without stratification. The output has an orderly nature for the interpretation of the results. Thus, the package `sanon` would be helpful for the analysis of randomized clinical trials with ordinal response variables. The details of the functions can be seen in the help file through the command `help(package="sanon")` in R.

Acknowledgements

References

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

Davis, C. E. and Quade, D. (1968). On comparing the correlations within two pairs of variables. *Biometrics* pages 987–995.

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2006). A lego system for conditional inference. *The American Statistician* **60,** 257–263.

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2008). Implementing a class of permutation tests: The coin package. *Journal of Statistical Software* **28,** 1–23.

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2013). `coin: Conditional Inference Procedures in a Permutation Test Framework`. R package version 1.0-23.

Jung, J. and Koch, G. (1998). A linear model method for rank measures of association from longitudinal studies with fixed conditions (visits) for data collection and more than two groups. *Journal of Biopharmaceutical Statistics* **8,** 299–316.

Jung, J. and Koch, G. (1999). Multivariate non-parametric methods for mann–whitney statistics to analyse cross-over studies with two treatment sequences. *Statistics in Medicine* **18,** 989–1017.

Kawaguchi, A., Koch, 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.

Kawaguchi, A. and Koch, G. G. (2010). Multivariate mann–whitney estimators for the comparison of two treatments in a three-period crossover study with randomly missing data. *Journal of biopharmaceutical statistics* **20,** 720–744.

Koch, G., Gitomer, S., Skalland, L., and Stokes, M. (1983). Some non-parametric and categorical data analyses for a change-over design study and discussion of apparent carry-over effects. *Statistics in Medicine* **2,** 397–412.

Koch, G., Tangen, C., Jung, J., and Amara, I. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine* **17,** 1863–1892.

Koch, G. G., Carr, G. J., Amara, I. A., Stokes, M. E., and Uryniak, T. J. (1990). Categorical data analysis. In Berry, D. A., editor, *Statistical Methodology in Pharmaceutical Sciences*, pages 291–475. New York: Marcel Dekker.

LaVange, L., Durham, T., and Koch, G. (2005). Randomization-based nonparametric methods for the analysis of multicentre trials. *Statistical Methods in Medical Research* **14,** 281–301.

Puri, M. and Sen, P. (1971). *Nonparametric Methods in Multivariate Analysis.* Krieger Pub. Co.

Quade, D. (1974). Nonparametric partial correlation. In Blalock, H. M. J., editor, *Measurement in the Social Sciences*, pages 369–398. Chicago: Aldine Publishing.

R Core Team (2012). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramaswamy, R., Koch, G., and Amara, I. (1997). Application of rank analysis of covariance methods to analysis of multiple anatomical regions with treatment for seborrheic dermatitis. *Journal of Biopharmaceutical Statistics* **7,** 403–416.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

SAS Institute Inc. (2011). *SAS/IML 9.3 User's Guide.* Cary, NC.

Stanish, W. M., Gillings, D. B., and Koch, G. G. (1978). An Application of Multivariate Ratio Methods for the Analysis of a Longitudinal Clinical Trial with Missing Data. *Biometrics* **34,** 305–317.

Stanish, W. M., Koch, G. G., and Landis, J. R. (1978). A Computer Program for Multivariate Ratio Analysis (MISCAT). *Computer Programs in Biomedicine* **8,** 197–207.

Stokes, M., Davis, C., and Koch, G. (2012). *Categorical Data Analysis Using The SAS System, 3rd Edition.* SAS Inst.

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.