`NParCov3`: A **SAS/IML** Macro for Nonparametric Randomization-Based Analysis of Covariance

Richard C. Zink
SAS Institute, Inc.

Gary G. Koch
University of North Carolina at Chapel Hill

Abstract

Analysis of covariance serves two important purposes in a randomized clinical trial. First, there is a reduction of variance for the treatment effect which provides more powerful statistical tests and more precise confidence intervals. Second, it provides estimates of the treatment effect which are adjusted for random imbalances of covariates between the treatment groups. The nonparametric analysis of covariance method of Koch, Tangen, Jung, and Amara (1998) defines a very general methodology using weighted least-squares to generate covariate-adjusted treatment effects with minimal assumptions. This methodology is general in its applicability to a variety of outcomes, whether continuous, binary, ordinal, incidence density or time-to-event. Further, its use has been illustrated in many clinical trial settings, such as multi-center, dose-response and non-inferiority trials.

`NParCov3` is a SAS/IML macro written to conduct the nonparametric randomization-based covariance analyses of Koch *et al.* (1998). The software can analyze a variety of outcomes and can account for stratification. Data from multiple clinical trials will be used for illustration.

Keywords: random imbalance of covariates, stratified analyses, variance reduction from covariance adjustment, weighted least squares.

1. Introduction

Analysis of covariance serves two important purposes in a randomized clinical trial. First, there is a reduction of variance for the treatment estimate which provides a more powerful statistical test and a more precise confidence interval. Second, it provides an estimate of the treatment effect which is adjusted for random imbalances of covariates between the treatment groups. Additionally, analysis of covariance models provide estimates of covariate effects on the outcome (adjusted for treatment) and can test for the homogeneity of the treatment effect within covariate-defined subgroups. The nonparametric randomization-based analysis of

covariance method of Koch *et al.* (1998) defines a very general methodology using weighted least-squares to generate covariate-adjusted treatment effects with minimal assumptions. This methodology is general in its applicability to a variety of outcomes, whether continuous, binary, ordinal, incidence density or time-to-event (Koch *et al.* 1998; Tangen and Koch 1999a,b, 2000; LaVange and Koch 2008; Moodie, Saville, Koch, and Tangen 2011; Saville, Herring, and Koch 2010; Saville, LaVange, and Koch 2011). Further, its use has been illustrated in many clinical trial settings, such as multi-center, dose-response and non-inferiority trials (Koch and Tangen 1999; LaVange, Durham, and Koch 2005; Tangen and Koch 2001).

This paper serves as an introduction and user's guide to NParCov3, a SAS/IML macro (SAS Institute Inc. 2011) that conducts the nonparametric randomization-based covariance analyses of Koch *et al.* (1998). We discuss the strengths and limitations of this methodology and provide complementary analysis strategies for these and more familiar parametric methods. Example data sets are introduced and the functionality of NParCov3 is illustrated through several examples. Finally, Appendix A provides a detailed description of the NParCov3 macro while Appendix B summarizes the theoretical and formulaic details of these methods.

2. Nonparametric randomization-based analysis of covariance

The nonparametric randomization-based analysis of covariance methodology of Koch *et al.* (1998) uses weighted least squares on the treatment differences of outcome and covariate means. The resulting model estimates represent treatment effects for the outcomes which have been adjusted for the accompanying covariates. Covariance-adjustment in the treatment-effect estimates is a result of the assumed null difference in covariate means which is a consequence of the underlying assumption of randomization to treatment. Full details of this methodology are left to Appendix B.

This methodology has several advantages:

1. Applicability to a variety of outcomes, whether continuous, binary, ordinal, event count or time-to-event.

2. Minimal assumptions.

    (a) Randomization to treatment is the principal assumption necessary for hypothesis testing.

    (b) Confidence intervals of treatment estimates require the additional assumption that subjects in the trial are a simple random sample of a larger population.

    (c) Survival analyses have the assumption of non-informative censoring.

3. Straightforward to accommodate stratification and generate essentially exact p values and confidence intervals.

4. Greater power of the adjusted treatment effect relative to the unadjusted.

However, there are some disadvantages:

1. No estimates for the effects of covariates or strata are available.

2. No direct means to test homogeneity of treatment effects within subgroups defined by the covariates.

3. Results are not directly generalizable to subjects with covariate distributions differing from those in the trial.

4. Is not appropriate when the same multivariate distribution of the covariates cannot be assumed to apply to both treatment groups.

This methodology can be used in a complementary fashion with model-based methods. First, since the specification of statistical methods in a protocol or analysis plan is typically required prior to the unblinding of study data, it is recommended that randomization-based nonparametric analysis of covariance be specified as the primary analysis because of minimal assumptions. Model-based methods, such as a logistic regression model in the case of binary outcomes, can be used in a supportive fashion to assess the effects of covariates, potential interactions of treatment with covariates and the generalizability of findings to populations not represented by the subjects in the current study.

3. Illustrative data sets

3.1. Respiratory disorder

The goal of this randomized clinical trial was to compare an active treatment versus placebo for 111 subjects with a respiratory disorder at two centers (Koch, Carr, Amara, Stokes, and Uryniak 1990; Stokes, Davis, and Koch 2000). Subjects were classified with an overall score 0, 1, 2, 3, 4 (terrible, poor, fair, good, excellent) at each of four visits. Covariates include a baseline score, gender and age (in years) at study entry. Treatment is coded 1 for active treatment, 0 for placebo, while gender is coded 1 for male, 0 for female. In addition to the ordinal scale, the dichotomous outcome of (good, excellent) versus (terrible, poor, fair) was of interest.

3.2. Osteoporosis

Data from a randomized clinical trial in 214 postmenopausal women with osteoporosis at two study centers were used to evaluate the effectiveness of a new drug versus placebo (Stokes *et al.* 2000). Both treatment groups had access to calcium supplements, received nutritional counseling and were encouraged to exercise regularly. The number of fractures experienced each year were recorded through the trial's duration (3 years). Seven women discontinued during the third year and are considered to have a 6-month exposure for this year; otherwise, yearly exposure is considered to be 12 months. The outcome of interest was rate of fracture per year. Adjusting for age as a covariate would be appropriate.

3.3. Aneurysmal subarachnoid hemorrhage

Nicardipine is indicated for the treatment of high blood pressure and angina and there are oral and intravenous formulations. Data from 902 subjects from a two-arm intravenous study of high dose nicardipine are presented (Haley, Kassell, and Torner 1993). The goal of this trial

was to compare the rates of cerebral vasospasm between treatments that occurred following an aneurysmal subarachnoid hemorrhage. Subjects were treated with drug or placebo for 10 days on average after the hemorrhage occurred. Here, we will compare the number of treatment-emergent serious adverse events between nicardipine and placebo, accounting for the varying durations of treatment exposure.

3.4. Prostate cancer

Data in 38 subjects with stage III prostate cancer are presented from a larger randomized clinical trial to compare 1.0 mg of diethylstilbestrol (DES) to placebo (Collett 2003). Time origin is the date of randomization and the outcome of interest is death from prostate cancer. Subjects who died of other causes or were lost to follow-up were censored. Covariates include age at trial entry (years), serum hemoglobin level (gm/100 mL), primary tumor size (cm^2) and Gleason index, which is an indicator of tumor stage and grade with a larger score implying more advanced disease.

4. Examples

4.1. Linear model for ordered outcome scores

For the next several examples, we focus on the respiratory disorder data. Suppose we wish to compare the average response for the ordinal respiratory outcome at Visit 1 between the active and placebo treatments with adjustment for gender, age and baseline score. For our purposes, we perform a stratified analysis using centers as strata weighting each center using Mantel-Haenszel weights. Further, we combine treatment estimates across strata prior to covariance adjustment.

The `NParCov3` call for this example is:

```
%NParCov3(OUTCOMES = visit1, COVARS = gender age baseline, C=1,
  HYPOTH = ALT, STRATA = center, TRTGRPS = treatment, TRANSFORM = NONE,
  COMBINE = FIRST, DSNIN = RESP, DSNOUT = OUTDAT);
```

The data set `_OUTDAT_COVTEST` contains the criterion for covariate imbalance $Q = 6.95$ which, with sufficient sample size, has a χ^2 distribution with 3 degrees of freedom. The p value for the criterion is $p = 0.0737$ which is compatible with random imbalance for the distributions of covariates between the two treatments. In other words, randomization provided reasonably equal distributions of covariates across the strata. Data set `_OUTDAT_DEPTEST` provides results for the treatment difference:

```
outcomes   beta     sebeta    Q_j     pvalue
 visit1    0.4266   0.1666   6.5565   0.0105
```

with a 95% confidence interval for the treatment estimate provided in `_OUTDAT_CI` as (0.1001, 0.7531). Thus, for Visit 1, we see that there is a significant treatment effect when stratifying by center and adjusting for gender, age and baseline score. The treatment estimate is positive indicating that the active drug had a greater positive effect, about four-tenths-of-a-point on average, than placebo.

In contrast, the model without adjustment for covariates provides the following:

```
outcomes   beta    sebeta    Q_j     pvalue
 VISIT1   0.3935   0.1995   3.8891   0.0486
```

It is clear that adjustment for covariates provides a more powerful comparison between the two treatments.

Perhaps there is interest in testing whether the treatment effect differs across visits while adjusting for covariates. The following macro call:

```
%NParCov3(OUTCOMES = visit1 visit2 visit3 visit4, C=1, HYPOTH = ALT,
   COVARS = gender age baseline, STRATA = center, TRANSFORM = NONE,
   TRTGRPS = treatment, COMBINE = FIRST, DSNIN = RESP, DSNOUT = OUTDAT);
```

provides the output:

```
outcomes   beta    sebeta     Q_j     pvalue
 VISIT1   0.4266   0.1666    6.5565   0.0105
 VISIT2   1.0144   0.2005   25.6005   <.0001
 VISIT3   0.8786   0.2242   15.3619   <.0001
 VISIT4   0.6631   0.2305    8.2774   0.0040
```

These treatment estimates (from `_OUTDAT_DEPTEST`) and the covariance matrix from the data set `_OUTDAT_COVBETA` can be used to test the hypothesis that the treatment effect differs across study visits. The test $Q_{Visit} = \hat{\beta}^\top C^\top (C V_{\hat{\beta}} C^\top)^{-1} C \hat{\beta}$ where

$$
C = \begin{bmatrix}
1 & 0 & 0 & -1 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & -1
\end{bmatrix}
$$

is distributed $\chi^2_{(3)}$. The following SAS/IML code performs this analysis:

```
proc iml;
  use _OUTDAT_DEPTEST;
    read all var{beta} into B;
  close;
  use _OUTDAT_COVBETA;
    read all var{VISIT1 VISIT2 VISIT3 VISIT4} into V;
  close;
  C = {1 0 0 -1, 0 1 0 -1, 0 0 1 -1};
  Q = B` * C` * inv(C * V * C`) * C * B;
  p = 1 - probchi(Q, nrow(C));
  print Q p;
quit;
```

With $Q_{Visit} = 12.57$ and p value $= 0.006$, the analysis suggests that the treatment effect differs across visits.

4.2. Binary outcome

We analyze the binary response of (good, excellent) versus (terrible, poor, fair) at Visit 1. As mentioned in Koch *et al.* (1998), if the primary interest is in the test of the treatment effect, the most appropriate variance matrix is the one that applies to the randomization distribution under the null hypothesis of no treatment difference. For illustration, we obtain weighted estimates across strata prior to covariance adjustment.

The `NParCov3` call for this example is:

```
%NParCov3(OUTCOMES = v1goodex, COVARS = gender age baseline, C = 1,
   HYPOTH = NULL, STRATA = center, TRTGRPS = treatment, TRANSFORM = NONE,
   COMBINE = FIRST, DSNIN = RESP, DSNOUT = OUTDAT);
```

The criterion for covariate imbalance $Q = 6.46$ has a χ^2 distribution with 3 degrees of freedom. The p value for the criterion is $p = 0.0911$. It appears randomization provided reasonably equal distributions of covariates across the strata. The model provides the following for the treatment estimate for the difference in proportions:

```
outcomes   beta    sebeta    Q_j     pvalue
 visit1    0.1839   0.0781   5.5455   0.0185
```

Therefore, active treatment provides a higher proportion of good or excellent response when adjusting for gender, age and baseline score.

However, if there was interest in confidence intervals for the odds ratio, we could specify `HYPOTH = ALT` and `TRANSFORM = LOGISTIC` in the macro call above. The odds ratio and confidence interval obtained from `_OUTDAT_RATIOCI` is 2.7193 (1.3097, 5.6458).

4.3. Proportional odds model for ordinal outcome

Rather than dichotomize an ordinal response, we can analyze it using a proportional odds model. There are few responses for poor and terrible so these categories will be combined for analysis. We can define 3 binary indicators to indicate excellent response or not, at least good response or not, or at least fair response or not. We use the following macro call:

```
%NParCov3(OUTCOMES = v1ex v1goodex v1fairgoodex,
   COVARS = gender age baseline, C = 1, HYPOTH = ALT, STRATA = center,
   TRTGRPS = treatment, TRANSFORM = PODDS, COMBINE = FIRST, DSNIN = RESP,
   DSNOUT = OUTDAT);
```

The test of proportional odds from `_OUTDAT_HOMOGEN` is $Q = 3.71$ which is distributed χ^2 with 2 degrees of freedom and provides a p value of 0.1563. Thus, we do not reject the assumption of proportional odds. Further, the covariate criterion, which in this case is a joint test of proportional odds and covariate imbalance is borderline ($p = 0.0586$).

The model provides the following treatment estimate:

```
outcomes    beta    sebeta   Q_j      pvalue
TREATMENT   0.6703   0.3193   4.4072   0.0358
```

Exponentiating the parameter estimate above provides odds ratio and 95% confidence interval 1.9548 (1.0455, 3.6548). Therefore, the odds of a positive response in the test treatment is greater than for placebo. Finally, had the test of homogeneity been statistically significant the proportional odds assumption would not have applied. In this case, the full model could be obtained by using `TRANSFORM = LOGISTIC` in the macro call above to estimate a separate odds ratio for each binary outcome.

4.4. Log-ratio of two means

For this example, we make use of the osteoporosis data by comparing the yearly risk of fractures between the two treatments using a risk ratio. Since prior testing has shown no difference in the risk ratio across the 3 years of the trial (not shown), we calculate the yearly risk for each subject as the total number of fractures divided by 3 years (or 2.5 years for the 7 subjects who discontinued early). Treatment (active = 2, placebo = 1) and center are recoded to be numeric. For illustration, we take weighted averages across the strata prior to taking the $\log_e$ transformation and applying covariance adjustment. The macro call is:

```
%NParCov3(OUTCOMES = YEARLYRISK, COVARS = age, C = 1, HYPOTH = ALT,
   STRATA = center, TRTGRPS = treatment, TRANSFORM = LOGRATIO,
   COMBINE = PRETRANSFORM, DSNIN = FRACTURE, DSNOUT = OUTDAT);
```

The covariate criterion for this example is $Q = 0.05$ with 1 degrees of freedom and p value 0.8155. It appears randomization provided a reasonably equal distribution of age between treatment groups.

The model provides treatment estimate:

```
outcomes        beta      sebeta        Q_j      pvalue
YEARLYRISK    -0.5798      0.3070     3.5666      0.0590
```

Exponentiating the treatment estimate provides the ratio 0.5600 which has 95% confidence interval (0.3068, 1.0222). Despite fractures being $(1/0.56-1) \times 100 = 79\%$ more likely per year for subjects on placebo compared to active drug, this difference is of borderline significance when adjusting for age. Had there been greater variability in the duration of exposure for each subject, we could have analyzed with the methods in the following section.

4.5. Log-incidence density ratios

Several of the 40 clinical sites for the nicardipine trial had low enrollment with less than two observations per treatment in some cases. In order to analyze under the alternative hypothesis, we'll ignore center as a stratification variable. Our endpoint will be the number of treatment-emergent serious adverse events that a subject experienced normalized by the number of days he or she was in the trial. Treatment estimates will be adjusted for age and gender.

```
%NParCov3(OUTCOMES = count, COVARS = age male, EXPOSURES = days,
   HYPOTH = ALT, TRTGRPS = trt, TRANSFORM = INCDENS, DSNIN = nic,
   DSNOUT = OUTDAT);
```

The covariate criterion for this example is $Q = 0.81$ with 2 degrees of freedom and p value 0.666. It appears randomization provided a reasonably equal distribution of age between treatment groups. The treatment estimate:

```
  outcomes    beta    sebeta  Q_j    pvalue
INC_COUNT  -0.0718 0.1448 0.2456   0.6202
```

Exponentiating the treatment estimate provides the incidence density ratio 0.9307 which has 95% confidence interval (0.7007, 1.2363). Despite treatment-emergent serious adverse events being $(1/0.9307 - 1) \times 100 = 7.4\%$ more likely per day for subjects on placebo compared to nicardipine, this difference is not statistically significant when adjusting for the covariates.

4.6. Time-to-event outcome using logrank and Wilcoxon scores

We turn our attention to the prostate cancer data. Since this particular data set has 6 deaths (status = 1) out of 38 subjects, we perform our analysis assuming that status = 0 indicates an event, censored otherwise. First, we conduct the analysis using logrank scores adjusting for age, serum hemoglobin level, primary tumor size and Gleason index:

```
%NParCov3(OUTCOMES = event, COVARS = age serumhaem tumoursize gleason,
  EXPOSURES = survtime, HYPOTH = NULL, TRTGRPS = treatment,
  TRANSFORM = LOGRANK, DSNIN = PROSTATE, DSNOUT = OUTDAT);
```

The criterion for covariate imbalance $Q = 2.41$ has a χ^2 distribution with 4 degrees of freedom. The p value for the criterion is $p = 0.6610$. It appears randomization provided reasonably equal distributions of covariates between the two treatments. The model provides the following for the treatment estimate:

```
    outcomes       beta    sebeta  Q_j    pvalue
LOGRANK_EVENT   -0.1137    0.2612 0.1894   0.6634
```

The negative treatment estimate implies that DES has improved survival from prostate cancer than placebo. However, this result is not significant so that we conclude no difference between the treatments when adjusting for the covariates.

Performing the analysis with Wilcoxon scores with adjustment for covariates provides similar results:

```
    outcomes        beta    sebeta   Q_j    pvalue
WILCOXON_EVENT   -0.0806   0.1729   0.2174   0.6410
```

A joint test of the logrank and Wilcoxon scores can be performed with multiple calls to NParCov3. The _OUTDAT_SURV data sets from two runs can provide logrank (LOGRANK_EVENT) and Wilcoxon (WILCOXON_EVENT) scores. These data sets can be merged together and the following macro call

```
%NParCov3(OUTCOMES = LOGRANK_EVENT WILCOXON_EVENT,
  COVARS = age serumhaem tumoursize gleason, EXPOSURES = , HYPOTH = NULL,
  TRANSFORM = NONE, TRTGRPS = treatment, DSNIN = PROSTATE, DSNOUT = OUTDAT);
```

will generate the appropriate analysis. Notice that `EXPOSURES` is left unspecified since these values have already been accounted for in the `LOGRANK_EVENT` and `WILCOXON_EVENT` outcomes. Further, no transformation is needed so that `TRANSFORM = NONE`. Defining $C = I_2$, an identity matrix of order two, and using similar methods and code as those at the end of Section 4.1 provides the two degree-of-freedom bivariate test $Q_{Bivariate} = 0.22$ with p value $= 0.8969$.

5. Summary

`NParCov3` contains functionality for a majority of the nonparametric randomization-based analysis of covariance methods described to date. Future enhancements to the software may include the analysis of covariate-adjusted $\log_e$-hazard ratios (Moodie *et al.* 2011), multivariate proportional odds models and the computation of essentially exact p values and confidence intervals in situations where the sample size may not support large sample approximations.

Acknowledgments

References

Collett D (2003). *Modeling Survival Data in Medical Research.* 2nd edition. Chapman and Hall, Boca Raton.

Haley EC, Kassell NF, Torner JC (1993). "A Randomized Controlled Trial of High-Dose Intravenous Nicardipine in Aneurysmal Subarachnoid Hemorrhage." *Journal of Neurosurgery*, **78**, 537–47.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ (1990). "Categorical Data Analysis." In DA Berry (ed.), *Statistical Methodology in the Pharmaceutical Sciences*, pp. 389–473. Marcel Dekker, New York.

Koch GG, Tangen CM (1999). "Nonparametric Analysis of Covariance and Its Role in Noninferiority Clinical Trials." *Drug Information Journal*, **33**, 1145–59.

Koch GG, Tangen CM, Jung JW, Amara IA (1998). "Issues for Covariance Analysis of Dichotomous and Ordered Categorical Data from Randomzied Clinical Trials and Non-Parametric Strategies for Addressing Them." *Statistics in Medicine*, **17**, 1863–92.

LaVange LM, Durham TA, Koch GG (2005). "Randomization-Based Nonparametric Methods for the Analysis of Multicentre Trials." *Statistical Methods in Medical Research*, **14**, 281–301.

LaVange LM, Koch GG (2008). "Randomization-Based Nonparametric Analysis of Covariance (ANCOVA)." In RB D'Agostino, L Sullivan, J Massaro (eds.), *Wiley Encyclopedia of Clinical Trials.* John Wiley & Sons, Hoboken.

Moodie PF, Saville BR, Koch GG, Tangen CM (2011). "Estimating Covariate-Adjusted Log Hazard Ratios for Multiple Time Intervals in Clinical Trials Using Nonparametric Randomization Based ANCOVA." *Statistics in Biopharmaceutical Research*, **3**, 232–41.

SAS Institute Inc (2011). *SAS/IML 9.3 User's Guide.* SAS Institute Inc., Cary. URL http://www.sas.com/.

Saville BR, Herring AH, Koch GG (2010). "A Robust Method for Comparing Two Treatments in a Confirmatory Clinical Trial via Multivariate Time-to-Event Methods that Jointly Incorporate Information from Longitudinal and Time-to-Event Data." *Statistics in Medicine*, **29**, 75–85.

Saville BR, LaVange LM, Koch GG (2011). "Estimating Covariate-Adjusted Incidence Density Ratios for Multiple Time Intervals in Clinical Trials Using Nonparametric Randomization Based ANCOVA." *Statistics in Biopharmaceutical Research*, **3**, 242–52.

Stokes ME, Davis CS, Koch GG (2000). *Categorical Data Analysis Using The SAS System.* 2nd edition. SAS Institute Inc., Cary. Data available at http://support.sas.com/rnd/app/da/cat/samples/chapter15.html.

Tangen CM, Koch GG (1999a). "Complementary Nonparmetric Analysis of Covariance for Logistic Regression in a Randomized Clinical Trial Setting." *Journal of Biopharmaceutical Statistics*, **9**, 45–66.

Tangen CM, Koch GG (1999b). "Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon Scores and Survival-Rate Estimation in a Randomized Clinical Trial." *Journal of Biopharmaceutical Statistics*, **9**, 307–38.

Tangen CM, Koch GG (2000). "Non-Parametric Covariance Methods for Incidence Density Analyses of Time-to-Event Data from a Randomized Clinical Trial and Their Complementary Roles to Proportional Hazards Regression." *Statistics in Medicine*, **19**, 1039–58.

Tangen CM, Koch GG (2001). "Non-Parametric Analysis of Covariance for Confirmatory Randomized Clinical Trials to Evaluate Dose Response Relationships." *Statistics in Medicine*, **20**, 2585–607.

A. The macro `NParCov3`

`NParCov3` can be accessed by your program by including the lines:

```
filename nparcov3 "directory of file NParCov3";
%include nparcov3(NParCov3) / nosource2;
```

Data must be set up so there is one line per experimental unit and there cannot be any missing data. Records with missing data should be deleted or made complete with imputed values prior to analysis. The program does not automatically write output to the SAS listing file. The results of individual data sets can be displayed using `PROC PRINT`. `PROC DATASETS` will list available data sets generated by `NParCov3` at the end of the log file.

`NParCov3` has the following options:

OUTCOMES: List of outcome variables for the analysis. All outcome variables are numeric and at least one outcome is required. Outcomes for `TRANSFORM = LOGISTIC`, `PODDS`, `LOGRANK` or `WILCOXON` can only have (0, 1) values.

COVARS: List of covariates for the analysis. All covariates are numeric and so categorical variables must have expression as dummy variables. `COVARS` may be left unspecified (blank) for stratified, unadjusted results.

EXPOSURES: Required numeric exposures for `TRANSFORM = LOGRANK`, `WILCOXON` or `INCDENS`. The same number of exposure variables as `OUTCOMES` is required and the order of the variables in `EXPOSURES` should correspond to the variables in `OUTCOMES`.

TRTGRPS: This is a required single two-level numeric variable which defines the two randomized treatments or groups of interest. Differences are based on higher number (`TRT2`) − lower number (`TRT1`). For example, if the variable is coded (1,2), differences computed are treatment 2 − treatment 1.

STRATA: A single numeric variable that defines strata, this variable can represent strata for cross-classified factors. Specify `NONE` (default) for one stratum (i.e., no stratification). Note that each stratum must have at least one observation per treatment when `HYPOTH = NULL` and at least two observations per treatment when `HYPOTH = ALT`.

C: Used in calculation of strata weights for all methods. $0 \leq C \leq 1$ (default = 1). A 0 assigns equal weights to strata while 1 is for Mantel-Haenszel weights.

COMBINE: Defines how the analysis will accommodate `STRATA`. One of

1. `NONE` (default) performs analysis as if data are from a single stratum.
2. `LAST` performs covariance adjustment within each stratum, then takes weighted averages of parameter estimates with weights based on stratum sample size (appropriate for larger strata).
3. `FIRST` obtains a weighted average of treatment group differences (post-transformation, if applicable) across strata, then performs covariance adjustment (appropriate for smaller strata).

    4. PRETRANSFORM (available for LOGISTIC, PODDS, LOGRATIO or INCDENS) obtains a weighted average of treatment group means (pre-transformation) across strata, performs a transformation, then applies covariance adjustment (appropriate for very small strata).

HYPOTH: Determines how variance matrices are computed.

    1. NULL applies the assumption that the means and covariance matrices of the treatment groups are equal and therefore computes a single covariance matrix for each stratum.

    2. ALT computes a separate covariance matrix for each treatment group within each stratum.

TRANSFORM: Applies transformations to outcome variables. One of

    1. NONE (default) applies no transformation and analyzes means for one or more outcome variables as-provided.

    2. LOGISTIC applies the logistic transformation to the means of one or more $(0, 1)$ outcomes.

    3. PODDS applies the logistic transformation to multiple means of $(0, 1)$ outcomes that together form a single ordinal outcome.

    4. RATIO applies the $\log_e$-transformation to one or more ratios of outcome means.

    5. INCDENS uses number of events experienced in OUTCOMES and EXPOSURES to calculate and analyze the $\log_e$-transformation of one or more ratios of incidence density.

    6. LOGRANK uses event flags in OUTCOMES (1 = event, 0 = censored) and EXPOSURES to calculate and analyze logrank scores for one or more time-to-event outcomes.

    7. WILCOXON uses event flags in OUTCOMES (1 = event, 0 = censored) and EXPOSURES to calculate and analyze Wilcoxon scores for one or more time-to-event outcomes.

ALPHA: Confidence limit for intervals (default $= 0.05$).

DSNIN: Input data set. Required.

DSNOUT: Prefix for output data sets. Required. For example, if DSNOUT = OUTDAT:

    1. _OUTDAT_COVTEST contains the criterion for covariate imbalance.

    2. _OUTDAT_DEPTEST contains tests, treatment estimates and ratios (for TRANSFORM = LOGISTIC, PODDS, LOGRATIO and INCDENS) for outcomes.

    3. _OUTDAT_CI contains confidence intervals for outcomes when HYPOTH = ALT.

    4. _OUTDAT_RATIOCI contains confidence intervals for odds ratios (TRANSFORM = PODDS or LOGISTIC), ratios (TRANSFORM = LOGRATIO) or incidence densities (TRANSFORM = INCDENS) when HYPOTH = ALT.

    5. _OUTDAT_HOMOGEN contains the test for homogeneity of odds ratios for TRANSFORM = PODDS.

    6. _OUTDAT_SURV contains survival scores for TRANSFORM = LOGRANK or WILCOXON. The scores from this data set from multiple calls of NParCov3 can be used with TRANSFORM = NONE to analyze logrank and Wilcoxon scores simultaneously or to analyze survival outcomes with non-survival outcomes.

7. `_OUTDAT_COVBETA` contains the covariance matrix for $\hat{\beta}$. The estimates from data set `_OUTDAT_DEPTEST` and the covariance matrix can be used to test hypotheses of linear combinations of the $\hat{\beta}$.

B. Theoretical details

B.1. Nonparametric randomization-based analysis of covariance

Suppose there are 2 randomized treatment groups and there is interest in a nonparametric randomization-based comparison of r outcomes adjusting for t covariates within a single stratum. Let n_i be the sample size, $\bar{y}_i$ be the r-vector of outcome means and $\bar{x}_i$ be the t-vector of covariate means of the ith treatment group, $i = 1, 2$. Let $f_i = \begin{bmatrix} g(\bar{y}_i) \\ \bar{x}_i \end{bmatrix}$, where $g(\cdot)$ is a twice-differentiable function. Further, define $\mathrm{cov}(f_i) = V_{f_i} = D_i V_i D_i$ where $V_i = \mathrm{cov}\left(\begin{bmatrix} \bar{y}_i \\ \bar{x}_i \end{bmatrix}\right)$ and D_i is a diagonal matrix with $g'(\bar{y}_i)$ corresponding to the first r cells of the diagonal with 1's for the remaining t cells of the diagonal.

Define $f = f_2 - f_1$ and $V_f = V_{f_1} + V_{f_2}$ and let I_r be an identity matrix of dimension r and $0_{(t \times r)}$ be a $t \times r$ matrix of 0's and fit the model $\mathsf{E}[f] = \begin{bmatrix} I_r \\ 0_{(t \times r)} \end{bmatrix} \hat{\beta} = X\hat{\beta}$ using weighted least squares with weights based on V_f^{-1}. The weighted least squares estimator is $\hat{\beta} = (X^\top V_f^{-1} X)^{-1} X^\top V_f^{-1} f$ and a consistent estimator for $\mathrm{cov}(\hat{\beta})$ is $V_{\hat{\beta}} = (X^\top V_f^{-1} X)^{-1}$. There are two ways to estimate V_i. Under the null hypothesis of no treatment difference,

$$V_i = \frac{1}{n_i(n_1 + n_2 - 1)} \left\{ \sum_{i=1}^{2} \sum_{j=1}^{n_i} \begin{bmatrix} (y_{ij} - \bar{y})(y_{ij} - \bar{y})^\top & (y_{ij} - \bar{y})(x_{ij} - \bar{x})^\top \\ (x_{ij} - \bar{x})(y_{ij} - \bar{y})^\top & (x_{ij} - \bar{x})(x_{ij} - \bar{x})^\top \end{bmatrix} \right\},$$

where $\bar{y}$ and $\bar{x}$ are the means for the outcomes and covariates across both treatment groups, respectively. Under the alternative,

$$V_i = \frac{1}{n_i(n_i - 1)} \left\{ \sum_{l=1}^{n_i} \begin{bmatrix} (y_{il} - \bar{y}_i)(y_{il} - \bar{y}_i)^\top & (y_{il} - \bar{y}_i)(x_{il} - \bar{x}_i)^\top \\ (x_{il} - \bar{x}_i)(y_{il} - \bar{y}_i)^\top & (x_{il} - \bar{x}_i)(x_{il} - \bar{x}_i)^\top \end{bmatrix} \right\}.$$

A test of the two treatments for outcome $y_j, j = 1, 2, ..., r$, with adjustment for covariates is $Q_j = \hat{\beta}_j^2 / V_{\hat{\beta}_j}$. With sufficient sample size Q_j is approximately distributed $\chi^2_{(1)}$. A criterion for the random imbalance of the covariates between the two treatment groups is

$$Q = (f - X\hat{\beta})^\top V_f^{-1}(f - X\hat{\beta})$$

which is approximately distributed $\chi^2_{(t)}$. Under the alternative hypothesis, asymptotic confidence intervals for $\hat{\beta}_j$ can be defined as $\hat{\beta}_j \pm Z_{(1-\alpha/2)} V_{\hat{\beta}_j}^{1/2}$, where $Z_{(1-\alpha/2)}$ is a quantile from a standard normal random variable.

In many situations, a linear transformation of $\bar{y}_i$ is appropriate, e.g., $g(\bar{y}_i) = \bar{y}_i$. Here, D_i will be an identity matrix of dimension $(r + t)$. In `NParCov3`, the above methods correspond to `TRANSFORM = NONE` and `COMBINE = NONE`.

B.2. Logistic transformation

Should $\bar{y}_i$ consist of $(0, 1)$ binary outcomes, a logistic transformation of the elements of $\bar{y}_i$ may be more appropriate. Let $g(\bar{y}_i) = \text{logit}(\bar{y}_i)$ where $\text{logit}(\cdot)$ transforms each element $\bar{y}_{ij}$ of $\bar{y}_i$ to $\log_e(\bar{y}_{ij}/(1 - \bar{y}_{ij}))$. The first r elements of f correspond to

$$\log_e \frac{\bar{y}_{2j}(1 - \bar{y}_{1j})}{\bar{y}_{1j}(1 - \bar{y}_{2j})},$$

which is the $\log_e$-odds ratio of the jth outcome. Under the null hypothesis, the first r elements of the diagonal of D_i are $\bar{y}_j^{-1}(1 - \bar{y}_j)^{-1}$, where $\bar{y}_j$ is the mean for outcome y_j across both treatment groups. Under the alternative, the first r elements of the diagonal of D_i are $\bar{y}_{ij}^{-1}(1 - \bar{y}_{ij})^{-1}$. The $\hat{\beta}_j$ and confidence intervals can be exponentiated to obtain estimates and confidence intervals for the odds ratios. These methods correspond to `TRANSFORM = LOGISTIC` and `COMBINE = NONE`.

A single ordinal outcome with $(r+1)$ levels can be analyzed as r binary outcomes (cumulative across levels) using a proportional odds model. Logistic transformations are applied to the r outcomes as described above. A test of the homogeneity of the r odds ratios would be computed as

$$Q_c = \hat{\beta}^\top C^\top (C V_{\hat{\beta}} C^\top)^{-1} C \hat{\beta}$$

where $C = [\ I_{(r-1)} \quad -1_{(r-1)}\ ]$, $I_{(r-1)}$ is an identity matrix of dimension $(r-1)$ and $-1_{(r-1)}$ is an vector of length $(r-1)$ containing elements -1. With sufficient sample size, Q_c is approximately distributed $\chi^2_{(r-1)}$.

If homogeneity applies, one can use the simplified model $\mathsf{E}[f] = \begin{bmatrix} 1_r \\ 0_t \end{bmatrix} \hat{\beta}_R = X_R \hat{\beta}_R$. For ordinal outcomes, $\hat{\beta}_R$ would be the common odds ratio across cumulative outcomes in a proportional odds model. The criterion for random imbalance for this simplified model will be a joint test of random imbalance and proportional odds which is approximately $\chi^2_{(t+r-1)}$. These methods correspond to `TRANSFORM = PODDS` and `COMBINE = NONE`. `NParCov3` computes the test of homogeneity and the reduced model when `PODDS` is selected. Should the p value for $Q_c \leq \alpha$, the proportional odds assumption is violated implying the full model is more appropriate (obtained with `TRANSFORM = LOGISTIC`).

B.3. Log$_e$-ratio of two outcome vectors

Suppose there was interest in the analyses of the ratio of means while adjusting for the effect of covariates. Let $g(\bar{y}_i) = \log_e(\bar{y}_i)$ where $\log_e(\cdot)$ transforms each element $\bar{y}_{ij}$ of $\bar{y}_i$ to $\log_e(\bar{y}_{ij})$. The first r elements of f correspond to

$$\log_e(\bar{y}_{2j}) - \log_e(\bar{y}_{1j}) = \log_e \frac{\bar{y}_{2j}}{\bar{y}_{1j}},$$

which is the $\log_e$-ratio of the jth outcome. Under the null hypothesis, the first r elements of the diagonal of D_i are $\bar{y}_j^{-1}$, where $\bar{y}_j$ is the mean for outcome y_j across both treatment groups. Under the alternative, the first r elements of the diagonal of D_i are $\bar{y}_{ij}^{-1}$. The $\hat{\beta}_j$ and confidence intervals can be exponentiated to obtain estimates and confidence intervals

for the ratios of means. These methods correspond to `TRANSFORM = LOGRATIO` and `COMBINE = NONE`.

B.4. $\log_e$-incidence density ratios

Suppose each of the r outcomes has an associated exposure so that $\bar{e}_i$ is an r-vector of average exposures. Let $f_i = \begin{bmatrix} \log_e(\bar{y}_i) \\ \log_e(\bar{e}_i) \\ \bar{x}_i \end{bmatrix}$, where $\log_e(\cdot)$ transforms each element $\bar{y}_{ij}$ of $\bar{y}_i$ to $\log_e(\bar{y}_{ij})$ and each element $\bar{e}_{ij}$ of $\bar{e}_i$ to $\log_e(\bar{e}_{ij})$ and let $M_0 = \begin{bmatrix} I_r & -I_r & 0_{(r \times t)} \\ 0^\top_{(r \times t)} & 0^\top_{(r \times t)} & I_t \end{bmatrix}$, where $0_{(r \times t)}$ is an $r \times t$ matrix of 0's and I_r and I_t are identity matrices of dimension r and t, respectively. Define $g_i = M_0 f_i = \begin{bmatrix} \log_e(\bar{y}_i) - \log_e(\bar{e}_i) \\ \bar{x}_i \end{bmatrix}$ so that $\mathrm{cov}(g_i) = V_{g_i} = M_i V_i M_i^\top$, where $V_i = \mathrm{cov}\left(\begin{bmatrix} \bar{y}_i \\ \bar{e}_i \\ \bar{x}_i \end{bmatrix} \right)$ and $M_i = \begin{bmatrix} D_{\bar{y}_i} & -D_{\bar{e}_i} & 0_{(r \times t)} \\ 0^\top_{(r \times t)} & 0^\top_{(r \times t)} & I_t \end{bmatrix}$. The first r elements of g_i are $\log_e$-incidence densities so that if $g = g_2 - g_1$, the first r elements of g correspond to

$$\log_e\left(\frac{\bar{y}_{2j}}{\bar{e}_{2j}} \right) - \log_e\left(\frac{\bar{y}_{1j}}{\bar{e}_{1j}} \right) = \log_e\left(\frac{\left(\frac{\bar{y}_{2j}}{\bar{e}_{2j}} \right)}{\left(\frac{\bar{y}_{1j}}{\bar{e}_{1j}} \right)} \right),$$

which is the $\log_e$-incidence density ratio of the jth outcome. Under the null hypothesis, $D_{\bar{y}_i}$ is a diagonal matrix with diagonal elements $\bar{y}_j^{-1}$ and $D_{\bar{e}_i}$ is a diagonal matrix with diagonal elements $\bar{e}_j^{-1}$ where $\bar{y}_j$ and $\bar{e}_j$ are the means for outcome y_j and exposure e_j across both treatment groups, respectively. Under the alternative hypothesis, $D_{\bar{y}_i}$ is a diagonal matrix with diagonal elements $\bar{y}_{ij}^{-1}$ and $D_{\bar{e}_i}$ is a diagonal matrix with diagonal elements $\bar{e}_{ij}^{-1}$. The $\hat{\beta}_j$ and confidence intervals can be exponentiated to obtain estimates and confidence intervals for the incidence density ratios. These methods correspond to `TRANSFORM = INCDENS` and `COMBINE = NONE`.

B.5. Wilcoxon or logrank scores for time-to-event outcomes

Unlike Sections B.2, B.3 and B.4 where a transformation is applied to the means of one or more outcomes, `NParCov3` can 'transform' subject-level event flags and times into Wilcoxon or logrank scores to perform analyses of time-to-event outcomes. Scores are defined from binary outcome flags (`OUTCOME`) that define whether an event occurs ($= 1$) or not ($= 0$) and the times (`EXPOSURES`) when the event occurs or the last available time at which an individual is censored (e.g., when the subject does not experience the event).

Suppose there are L unique failure times such that $y_{(1)} < y_{(2)} < ... < y_{(L)}$. Assume there are g_ζ failures at y_ζ and that there are $y_{\zeta_1}, y_{\zeta_2}, ...y_{\zeta_{\nu_\zeta}}$ censored values in $[y_{(\zeta)}, y_{(\zeta+1)})$, where ν_ζ is the number of censored subjects in the interval and $\zeta = 1, 2, ..., L$. Further, assume $y_{(0)} = 0, y_{(L+1)} = \infty$ and $g_0 = 0$.

Logrank scores are expressed as

$$c_\zeta = 1 - \sum_{\zeta'=1}^{\zeta} \frac{g_{\zeta'}}{N_{\zeta'}} \quad \text{and} \quad C_\zeta = - \sum_{\zeta'=1}^{\zeta} \frac{g_{\zeta'}}{N_{\zeta'}}$$

and $C_0 = 0$, where c_ζ and C_ζ are the scores for the uncensored and censored subjects, respectively, in the interval $[y_{(\zeta)}, y_{(\zeta+1)})$ for $\zeta = 1, 2, ..., L$ where $N_\zeta = n - \sum_{\zeta'=0}^{\zeta-1}(\nu_{\zeta'} + g_{\zeta'})$ is the number of subjects at risk at $t_{(\zeta)} - 0$ and $n = n_1 + n_2$ is the sample size. To conduct an unstratified analysis using logrank scores set `TRANSFORM = LOGRANK` and `COMBINE = NONE`.

Wilcoxon scores are expressed as

$$c_\zeta = 2 \prod_{\zeta'=1}^{\zeta} \left(\frac{N_{\zeta'} - g_{\zeta'}}{N_{\zeta'}} \right) - 1 \quad \text{and} \quad C_\zeta = \prod_{\zeta'=1}^{\zeta} \left(\frac{N_{\zeta'} - g_{\zeta'}}{N_{\zeta'}} \right) - 1$$

and $C_0 = 0$, where c_ζ and C_ζ are the scores for the uncensored and censored subjects, respectively, in the interval $[y_{(\zeta)}, y_{(\zeta+1)})$ for $\zeta = 1, 2, ..., L$ where N_ζ is the number of subjects at risk at $t_{(\zeta)} - 0$. To conduct an unstratified analysis using Wilcoxon scores set `TRANSFORM = WILCOXON` and `COMBINE = NONE`. `NParCov3` computes survival scores, summarizes as $\bar{y}_i$ and analyzes as in Section B.1 using a linear transformation.

B.6. Stratification

Stratification can be handled in one of three ways:

1. Estimate $\hat{\beta}_h$ for each stratum h, $h = 1, 2, ..., H$, and combine across strata to obtain $\bar{\beta} = \sum_h \frac{w_h \hat{\beta}_h}{\sum_h w_h}$, where $w_h = \left(\frac{n_{h1} n_{h2}}{n_{h1} + n_{h2}} \right)^c$ and $0 \leq$ `C` ≤ 1, with `C` $= 1$ for Mantel-Haenszel weights and `C` $= 0$ for equal strata weights. A criterion for random imbalance is performed within each stratum and then summed to give an overall criterion of imbalance across all strata. This method is more appropriate for large sample size per stratum ($n_{hi} \geq 30$ and preferably $n_{hi} \geq 60$). This corresponds to `COMBINE = LAST`.

2. Form $f_w = \sum_h \frac{w_h f_h}{\sum_h w_h}$ first and fit a model for covariance adjustment to determine $\hat{\beta}_w$. This method is more appropriate for small to moderate sample size per stratum. This corresponds to `COMBINE = FIRST`.

3. For transformations in Sections B.2, B.3 and B.4, define $\begin{bmatrix} \bar{y}_{w_i} \\ \bar{x}_{w_i} \end{bmatrix} = \left(\sum_h w_h \right)^{-1} \times$

   $\sum_h \left(w_h \begin{bmatrix} \bar{y}_{hi} \\ \bar{x}_{hi} \end{bmatrix} \right)$ first, then proceed with the transformation and finally apply covariance adjustment. This method is more appropriate for small sample size per stratum and corresponds to `COMBINE = PRETRANSFORM`.

Affiliation:

Richard C. Zink
JMP Life Sciences
SAS Institute, Inc.
Cary, North Carolina, United States of America
E-mail: richard.zink@jmp.com

Gary G. Koch
Biometric Consulting Laboratory
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, United States of America
E-mail: bcl@bios.unc.edu