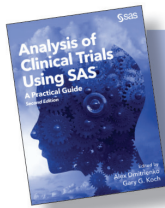


Analysis of Clinical Trials Using SAS®

A Practical Guide

Second Edition

Edited by
Alex Dmitrienko
Gary G. Koch

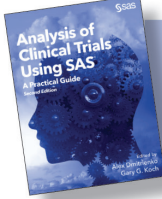


From *Analysis of Clinical Trials Using SAS*.
Full book available for purchase [here](#).

Contents

Preface	v
About This Book	xi
About These Authors	xii
1 Model-based and Randomization-based Methods	1
By Alex Dmitrienko and Gary G. Koch	
1.1 Introduction	1
1.2 Analysis of continuous endpoints	4
1.3 Analysis of categorical endpoints	20
1.4 Analysis of time-to-event endpoints	41
1.5 Qualitative interaction tests	56
References	61
2 Advanced Randomization-based Methods	67
By Richard C. Zink, Gary G. Koch, Yunro Chung and Laura Elizabeth Wiener	
2.1 Introduction	67
2.2 Case studies	70
2.3 %NParCov4 macro	73
2.4 Analysis of ordinal endpoints using a linear model	74
2.5 Analysis of binary endpoints	78
2.6 Analysis of ordinal endpoints using a proportional odds model	79
2.7 Analysis of continuous endpoints using the log-ratio of two means	80
2.8 Analysis of count endpoints using log-incidence density ratios	81
2.9 Analysis of time-to-event endpoints	82
2.10 Summary	86
3 Dose-Escalation Methods	101
By Guochen Song, Zoe Zhang, Nolan Wages, Anastasia Ivanova, Olga Marchenko and Alex Dmitrienko	
3.1 Introduction	101
3.2 Rule-based methods	103
3.3 Continual reassessment method	107
3.4 Partial order continual reassessment method	116
3.5 Summary	123
References	123

4	Dose-finding Methods	127
	By Srinand Nandakumar, Alex Dmitrienko and Ilya Lipkovich	
4.1	Introduction	127
4.2	Case studies	128
4.3	Dose-response assessment and dose-finding methods	132
4.4	Dose finding in Case study 1	145
4.5	Dose finding in Case study 2	160
	References	176
5	Multiplicity Adjustment Methods	179
	By Thomas Brechenmacher and Alex Dmitrienko	
5.1	Introduction	179
5.2	Single-step procedures	184
5.3	Procedures with a data-driven hypothesis ordering	189
5.4	Procedures with a prespecified hypothesis ordering	202
5.5	Parametric procedures	212
5.6	Gatekeeping procedures	221
	References	241
	Appendix	244
6	Interim Data Monitoring	251
	By Alex Dmitrienko and Yang Yuan	
6.1	Introduction	251
6.2	Repeated significance tests	253
6.3	Stochastic curtailment tests	292
	References	315
7	Analysis of Incomplete Data	319
	By Geert Molenberghs and Michael G. Kenward	
7.1	Introduction	319
7.2	Case Study	322
7.3	Data Setting and Methodology	324
7.4	Simple Methods and MCAR	334
7.5	Ignorable Likelihood (Direct Likelihood)	338
7.6	Direct Bayesian Analysis (Ignorable Bayesian Analysis)	341
7.7	Weighted Generalized Estimating Equations	344
7.8	Multiple Imputation	349
7.9	An Overview of Sensitivity Analysis	362
7.10	Sensitivity Analysis Using Local Influence	363
7.11	Sensitivity Analysis Based on Multiple Imputation and Pattern-Mixture Models	371
7.12	Concluding Remarks	378
	References	378
	Index	385



From *Analysis of Clinical Trials Using SAS*.
Full book available for purchase [here](#).

Chapter 2

Advanced Randomization-based Methods

Richard C. Zink (JMP Life Sciences, SAS Institute)

Gary G. Koch (University of North Carolina at Chapel Hill)

Yunro Chung (Fred Hutchinson Cancer Research Center)

Laura Elizabeth Wiener (University of North Carolina at Chapel Hill)

2.1	Introduction	67
2.2	Case studies	70
2.3	%NParCov4 macro	73
2.4	Analysis of ordinal endpoints using a linear model	74
2.5	Analysis of binary endpoints	78
2.6	Analysis of ordinal endpoints using a proportional odds model	79
2.7	Analysis of continuous endpoints using the log-ratio of two means	80
2.8	Analysis of count endpoints using log-incidence density ratios	81
2.9	Analysis of time-to-event endpoints	82
2.10	Summary	86

This chapter presents advanced randomization-based methods used in the analysis of clinical endpoints. The nonparametric, randomization-based analysis of covariance methodology, a versatile methodology that uses weighted least squares to generate covariate-adjusted treatment effects under minimal assumptions, is presented. These methods are implemented using a powerful SAS macro (%NParCov4) that is applicable to a variety of clinical outcomes with the ability to compute essentially-exact p -values and confidence intervals. Numerous clinical trial examples are presented to illustrate the methodology and software.

2.1 Introduction

As explained in Chapter 1, adjustments for important covariates in the analysis of clinical endpoints can be carried out using randomization-based and model-based methods. The two classes of methods play complementary roles in the sense that each method is useful within a particular setting, and the advantages of each method offset the limitations of the other method. Chapter 1 deals with both the model-based and basic randomization-based methods applied in clinical trials. This chapter focuses on a class of advanced randomization-based methods with broad applications to trials with different types of endpoints.

A detailed description of model-based approaches can be found in the beginning of Chapter 1. This includes, for example, logistic regression models used in the analysis of binary endpoints and the Cox proportional hazards model in settings with time-to-event endpoints. An important feature of model-based methods is that they rely on specific assumptions and thus can be quite sensitive to violations from these assumptions. For example, considering proportional hazards models, departures from the proportional hazards assumption are typically more evident for baseline characteristics and measurements than for treatments. Also, proportional hazards is applicable to treatments under the null hypothesis of equal treatment effects. The issue of bias in a treatment comparison through departures from the proportional hazards assumption for baseline characteristics and/or measurements can be addressed by a randomization-based method that invokes randomization-based covariance adjustment for the baseline characteristics and/or measurements for an estimated hazard ratio from a proportional hazards model that only includes treatments; see Saville and Koch (2013). Similar considerations apply to the proportional odds model with respect to how to address departures from the proportional odds assumption for baseline characteristics and/or measurements. See Tangen and Koch (1999a). Additionally, the scope of randomization-based methods for ordinal response includes Mann-Whitney rank measures of association. See Kawaguchi and Koch (2011, 2015).

Nonparametric randomization-based analysis of covariance

It was pointed out in Chapter 1 that 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, analysis of covariance 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 versatile methodology that uses 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. (See Koch et al., 1998; Tangen and Koch, 1999a, 1999b, 2000; LaVange and Koch, 2008; Moodie et al., 2010; Saville, Herring, and Koch, 2010; Saville, LaVange, and Koch, 2010; Saville and Koch 2013; and Zhao et al. 2016.) 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; and Tangen and Koch, 2001). Similar covariate adjustment for randomized studies has been applied to multivariate Mann-Whitney estimators of ordinal outcomes (Kawaguchi, Koch, and Wang, 2011; and Kawaguchi and Koch, 2015).

The nonparametric, randomization-based analysis of covariance methodology uses weighted least squares on the treatment differences of outcome and covariate means. The resulting model estimates represent treatment effects for the outcomes that 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, in turn, is a consequence of the underlying invocation of randomization to treatment. Full details of this methodology are left to Appendix A.

This methodology has several advantages:

1. Applicability to a variety of outcomes, whether continuous, binary, ordinal, incidence density, or time-to-event (See Appendices A.1 through A.5).
2. Minimal assumptions:
 - (a) Randomization to treatment is the principal structure 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 to generate essentially-exact p -values and confidence intervals (described in Appendices A.6 and A.7, respectively).
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 way to test homogeneity of treatment effects within subgroups that are defined by the covariates.
3. Results are not directly generalizable to subjects with covariate distributions that differ 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.

It is important to note that nonparametric, randomization-based analysis of covariance is not intended to replace likelihood-based methodologies such as generalized linear models (see Chapter 1). On the contrary, this methodology is expected to be used in a complementary fashion with model-based methods, as long as patients are randomized to treatment (see Tangen and Koch, 1999a, 2000; and Saville and Koch, 2013, as examples). In a regulatory setting, randomization-based, nonparametric analysis of covariance can establish whether a treatment effect exists under minimal statistical assumptions. Once a treatment effect has been identified, other scientific questions can be explored under some additional plausible assumptions using more traditional models. The examples in this chapter focus on the analysis and interpretation of data using nonparametric, randomization-based analysis of covariance. Additional analyses to address other scientific questions are not shown. (Interested readers might choose to use the model-based methods introduced in Chapter 1 to compare with the findings presented here.)

For example, as outlined in Tangen and Koch (1999a), 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 analysis of covariance be specified as the primary analysis because of minimal assumptions and the availability of essentially-exact p -values. 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. This last point is particularly important. For nonlinear models, such as logistic regression, the estimated treatment effect is specific to a population of patients with a specific set of values for the covariates. Randomization-based, nonparametric analysis of covariance estimates and tests the treatment effect with adjustment for covariates for the randomized population of patients.

This chapter provides a detailed description of the nonparametric, randomization-based analysis of covariance methodology. A description of key features of the advanced randomization-based framework is provided in Section 2.1. Case studies used throughout this chapter are presented in Section 2.2. Section 2.3 introduces the `%NParCov4` macro that performs these analyses with the ability to compute essentially exact p -values and confidence intervals. We discuss the strengths and limitations of the randomization-based methodology and provide complementary analysis strategies using the case studies in Sections 2.4 through 2.9. As stated above, technical details are left to the Appendix. Appendix A summarizes the theoretical details of these methods while Appendix B documents the `%NParCov4` macro.

The SAS code and data sets included in this chapter are available on the book's website at <http://support.sas.com/publishing/authors/dmitrienko.html>.

2.2 Case studies

The following four case studies, based on real clinical trials, will be used in this chapter to illustrate advanced randomization-based methods for the analysis of clinical endpoints.

2.2.1 Case study 1 (Respiratory disorder trial)

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 et al., 1990; and Stokes, Davis, and Koch, 2012). The primary endpoint was ordinal, and subjects were classified with an overall score of 0, 1, 2, 3, or 4 (terrible, poor, fair, good, or excellent) at each of four visits (V1 through V4). Covariates included a baseline score, gender, and age (in years) at study entry. Treatment was coded 1 for active treatment, 0 for placebo, while gender was coded 1 for male, and 0 for female. In addition to the ordinal scale, dichotomous outcomes, such as (good, excellent) versus (terrible, poor, fair) with values of 1 or 0, respectively, were of interest. The first few rows of the respiratory data are presented in Table 2.1.

TABLE 2.1 Respiratory disorder data

Center	Treatment	Gender	Age	Baseline	V1	V2	V3	V4
0	0	1	46	2	2	2	2	2
0	0	1	28	2	0	0	0	0
0	0	1	44	3	4	3	4	2
0	0	0	13	4	4	4	4	4
0	0	1	43	2	3	2	4	4
0	0	1	37	3	2	3	3	2
0	0	1	23	3	3	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

There are several ways to analyze the overall score from this clinical trial. For example, if the distribution of overall scores at a visit appears to have an approximately normal shape, a linear model might be appropriate to assess the effect of treatment with adjustment for covariates (Section 2.4). Alternatively, the clinical team might prefer to compare the proportion of patients with good or excellent response between the two treatments. Here, a logistic regression model can be used to model the effect of covariates on treatment response (Section 2.5). Finally, if the proportional odds assumption is appropriate, a proportional odds

model can be fit to the overall score with adjustment for covariates (Section 2.6). The proportional odds assumption implies that the model coefficients for all binary splits of the outcome (0 vs >0, ≤ 1 vs >1, ≤ 2 vs >2, and ≤ 3 vs 4) are identical.

2.2.2 Case study 2 (Osteoporosis trial)

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 (t) versus placebo (p) (Stokes et al., 2012). 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 (three 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 the rate of fracture per year. Adjusting for age as a covariate and for centers as strata would be appropriate.

Table 2.2 contains the first few rows of the osteoporosis data set. Year 1, Year 2, and Year 3 contain the number of fractures experienced in Years 1, 2, and 3, respectively. Total is the total number of fractures experienced over the duration of the trial. Exposure equals 3 years save for the 7 patients who discontinued early with 2.5 years of exposure. Yearly risk is defined as the total number of fractures (Total) divided by Exposure. Center A and B were recoded as 1 and 2, while Treatment p and t were recoded as 1 and 2 for the analysis. Yearly risk can be analyzed using a Poisson regression model to assess the effect of treatment with adjustment for covariates (Section 2.7).

TABLE 2.2 Osteoporosis data

Age	Center	Treatment	Year 1	Year 2	Year 3	Total	Exposure
56	A	p	0	0	0	0	3
71	A	p	1	0	0	1	3
60	A	p	0	0	1	1	3
71	A	p	0	1	0	1	3
78	A	p	0	0	0	0	3
67	A	p	0	0	0	0	3
49	A	p	0	0	0	0	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

2.2.3 Case study 3 (Aneurysmal subarachnoid hemorrhage trial)

Nicardipine is indicated for the treatment of high blood pressure and angina in both oral and intravenous formulations. Data from 902 subjects from a randomized, two-arm intravenous study of high dose nicardipine was presented in Haley, Kassell, and Torner (1993). The primary endpoint was improvement in patient recovery according to the Glasgow Outcome Scale, with the incidence of cerebral vasospasm and the incidence of death or disability due to vasospasm serving as important secondary endpoints (Jennett and Bond, 1975). Subjects were treated with drug or placebo up to 14 days after the hemorrhage occurred. Here, we compare the number of treatment-emergent, serious adverse events between nicardipine (Treatment = 1) and placebo (Treatment = 0), accounting for the varying durations of treatment exposure. Age in years and gender are available as covariates.

Table 2.3 contains the first few rows of the aneurysmal, subarachnoid hemorrhage data set. Count is the total number of treatment-emergent, serious adverse events

that occur over the duration of trial exposure (Days). Male is coded 1 if the patient is a male, and 0 if the patient is a female. The number of adverse events can be analyzed using a Poisson regression model using exposure as an offset to assess the effect of treatment with adjustment for covariates (Section 2.8).

TABLE 2.3 Aneurysmal, subarachnoid hemorrhage Data

Treatment	Count	Days	Age	Male
0	0	13	63	0
1	0	11	66	1
0	0	10	31	0
1	6	1	48	0
0	1	13	67	0
0	0	11	32	1
1	5	3	63	1
⋮	⋮	⋮	⋮	⋮

2.2.4 Case study 4 (Colorectal cancer trial)

Data from a randomized study of 110 subjects with colorectal cancer are presented to compare an experimental treatment to a control (Table 2.4). Treatment (Trt) is coded 1 or 0 for active treatment or control, respectively. Data for two endpoints are available: the time until death and the time until disease progression. Disease progression was defined using standardized RECIST criteria (Eisenhauer et al., 2009). Subjects who did not experience death or disease progression were censored on their last day of study participation. Overall survival (OS) and progression-free survival (PFS) are each represented by two variables. The variable OS_Time is the number of days from randomization until death. OS_Event equals 1 if the patient died. Otherwise, a value of 0 indicates that the patient was censored on their last day of study participation. For PFS, the variable PFS_Time is the number of days from randomization until disease progression. PFS_Event equals 1 if the patient experienced disease progression. Otherwise, a value of 0 indicates that the patient was censored on their last day of study participation.

TABLE 2.4 Colorectal cancer data

Trt	OS_Time	OS_Event	PFS_Time	PFS_Event	KRAS	Days	Length
1	75	1	40	1	WT	805	121
1	406	0	68	1	WT	3186	160
1	225	1	118	1	M	2449	101
1	89	1	36	1	M	457	97
1	309	1	72	1	WT	1396	158
0	123	1	50	1	WT	2341	120
1	465	0	48	1	WT	1513	64
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Covariates include the number of days from initial cancer diagnosis to start of study drug administration (Days), and the sum of the longest lesion diameters in millimeters at baseline (Length). A flag as to whether the KRAS gene was wild-type (WT) or mutated (M) can be used as a stratification variable. For the analysis, a numeric variable mutant will be defined with values of 0 or 1 to indicate wild-type or mutated, respectively. Time-to-event endpoints are often analyzed using non-parametric Kaplan-Meier analysis (Section 2.9). Alternatively, Cox proportional-hazards regression is used when it is important to adjust for one or more covariates.

2.3 %NParCov4 macro

Analysis of clinical endpoints based on the nonparametric, randomization-based methodology is easily performed using the %NParCov4 macro, a SAS/IML macro that supports functionality for many of the analysis methods described to date, with the ability to compute essentially-exact p -values and confidence intervals. This macro will be used in Sections 2.4 through 2.9 to perform adjusted analyses of continuous, categorical, count-type, and time-to-event endpoints.

A macro call including all options is presented in Program 2.1.

PROGRAM 2.1 General %NParCov4 Macro Call

```
%NPARCOV4(OUTCOMES = , COVARS = , EXPOSURES = , TRTGRPS = , HYPOTH = ,  
TRANSFORM = , STRATA = , COMBINE = , C = , ALPHA = , EXACT = , SEED = ,  
NREPS = , SYMSIZE = , DETAILS = , DSNIN = , DSNOUT = );
```

The user can specify one or more numeric outcomes (OUTCOMES), one or more numeric covariates, or none if no covariate adjustment is required (COVARS). For time-to-event outcomes or for analyses of incidence densities, the user specifies one or more numeric exposure variables (EXPOSURES). A two-level numeric variable is required to define treatment (TRTGRPS).

Hypothesis (HYPOTH) is specified as either NULL or ALT and impacts whether the covariance matrices of responses and covariates in each strata are computed across treatments (under the null) or within treatments (under the alternative). HYPOTH = NULL by default. In general, the variance under the null would be the principal structure for producing p -values for hypothesis testing under the null when the null corresponded to no difference between treatments (or more formally, the strong null that each patient would have the same outcomes regardless of the assigned treatment). The variance under the alternative will be used for computing confidence intervals. Here, the standard errors address replicated random samples from a population for which the randomized population is arguably a random sample. P -values under the alternative might be useful since they will agree with the confidence interval, at least when both are based on normal approximations. Throughout this chapter, we take the approach that p -values will be computed under the null, with corresponding confidence intervals computed under the alternative hypothesis. The user can specify α (ALPHA) to determine the confidence-level of the $100(1 - \alpha)\%$ confidence intervals. By default, ALPHA = 0.05.

A single numeric variable can be provided to indicate strata (STRATA), and there are options to specify when covariate adjustment occurs relative to accounting for strata (COMBINE). By default, COMBINE = NONE, which assumes no stratification. When strata are present, the recommendation for when to account for covariate adjustment is typically determined by sample size. See Appendix A.6 for guidance on appropriate choices for COMBINE. How the estimates from the strata are combined is determined by C, which is the exponent for a Mantel-Haenszel weight. When C = 1 (default), Mantel-Haenszel weights are used to combine strata. This considers the sample size of each stratum when combining estimates. When C = 0, strata are weighted equally, effectively ignoring the size of each stratum. Values of C between 0 and 1 put lesser or greater emphasis on strata sample sizes, respectively.

For computing resampling-based exact analyses (EXACT, NO by default), the user can specify a seed (SEED, 0 by default) to generate reproducible analyses, the number of simulations to conduct (NREPS), and the amount of memory to allocate to the symbol space (SYMSIZE, 20000 by default). Whether macro details are written to the SAS log or not is controlled by the DETAILS option (YES by default). For lengthy simulations, the user should consider setting DETAILS = NO; specifying OPTIONS

NONOTES; and using a SAS macro variable to report simulation progress. See Program 2.18 for an example.

The input data set name (DSNIN) specifies a single data set containing analysis data. The prefix for output data sets (DSNOUT) enables the user to customize the names of the output data sets generated by %NParCov4; see Appendix B.1.15 for a complete list of output data sets.

Appendix B.1 provides details for all %NParCov4 macro options, and Appendix B.2 describes the contents of the generated output data sets.

2.4 Analysis of ordinal endpoints using a linear model

This section and the next five sections provide illustrations of the nonparametric, randomization-based methodology by applying it to the four clinical trial examples defined earlier in the chapter.

This section applies the randomization-based methodology to the analysis of an ordinal endpoint in Case study 1. Suppose that the trial's sponsor wants 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. We can perform a stratified analysis using centers as strata, weighting each center using Mantel-Haenszel weights ($C = 1$). Further, we combine treatment estimates across strata prior to covariance adjustment using `COMBINE = FIRST`, assuming that the sample size per stratum is small. The %NParCov4 call for this example is provided in Program 2.2.

PROGRAM 2.2 Linear Model with Adjustment for Covariates for Visit 1

```
%NParCov4(OUTCOMES = V1, COVARS = gender age baseline, C=1,
HYPOTH = NULL, STRATA = center, TRTGRPS = treatment, TRANSFORM = NONE,
COMBINE = FIRST, EXACT = YES, NREPS = 5000, SEED = 36, DSNIN = RESP,
DSNOUT = OUTDAT);
```

In Program 2.2, `HYPOTH = NULL` so that we can first obtain p -values of interest for our analysis. The option `DSNOUT = OUTDAT` sets the value `OUTDAT` as the prefix for the output data sets. Throughout this chapter, we assume `DSNOUT = OUTDAT` and $C = 1$. We refrain from repeating these options in the programs below unless the values change. For this particular example, the following data sets are created:

1. `_OUTDAT_BETASAMP`, which contains the observed and $NREPS = 5000$ permutation-based estimates of the treatment effect.
2. `_OUTDAT_COVBETA`, which contains the covariance matrix of the covariate-adjusted treatment effects for the outcomes.
3. `_OUTDAT_COVTEST`, which contains the criterion for covariate imbalance.
4. `_OUTDAT_DEPTEST`, which contains the treatment estimate, its standard error, and a two-sided asymptotic p -value.
5. `_OUTDAT_EXACT`, which contains one- and two-sided p -values for an essentially-exact analysis of $NREPS = 5000$ permutation samples using `SEED = 36`.

The data set `_OUTDAT_COVTEST` contains the criterion for covariate imbalance $Q=6.46$, which, with sufficient sample size, is distributed $\chi^2_{(3)}$. The p -value for the criterion is 0.0911. While the exact p -value based on 5000 permutation samples is 0.0920, which is not contrary to random imbalance for the distributions of covariates

between the two treatments. In other words, randomization provided statistically similar distributions of covariates across the strata.

This criterion should not be considered a test for validity of nonparametric, randomization-based analysis of covariance, but merely as a measure of the degree of imbalance among covariates between the treatments. Due to the randomization requirement of the methodology, significant imbalance detected among the covariates should be interpreted as a chance finding. Adjusting for the covariates in the analysis removes this imbalance; the results of an unadjusted analysis would be confounded with the imbalance among the covariates. Small p -values for the covariate imbalance criterion are more appropriately referred to as ‘atypical relative to “chance” rather than “significant.”

Data set `_OUTDAT_DEPTEST` summarizes results for the treatment difference that is provided in Table 2.5. Outcome, Treatment Effect, Standard Error, Test Statistic, and P-value are represented by the `_OUTDAT_DEPTEST` variables `OUTCOMES`, `BETA`, `SEBETA`, `Q_J`, and `PVALUE`.

TABLE 2.5 Linear Model with Adjustment for Covariates for Visit 1

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Visit 1	0.4008	0.1714	5.4690	0.0194

An exact p -value based on 5000 permutation data sets is 0.0162. If there was interest in confidence intervals, we could specify `HYPOTH = ALT` in Program 2.2 above. An additional data set, `_OUTDAT_CI`, is created. It contains asymptotic 95% confidence intervals. Exact 95% bias-corrected and accelerated (BC_a) and percentile intervals are included in the `_OUTDAT_EXACT` data set. The asymptotic 95% confidence interval for the treatment estimate is (0.1001, 0.7531). The corresponding bootstrap BC_a interval is (0.0974, 0.7749), and the percentile interval is (0.0901, 0.7646) based on 5000 bootstrap data sets.

The treatment effect in Table 2.5 represents the increase in the overall score for visit 1 for the active treatment compared to placebo. Based on our above analysis 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, by about 0.40 on average, than placebo. The treatment estimate of approximately 0.40 could also be interpreted as corresponding to a $0.40/4 = 0.10$ difference between treatments for the average of the cumulative probabilities of a score of at least 1, at least 2, at least 3, and at least 4. In contrast, the model without adjustment for covariates (Program 2.3) provides the results in Table 2.6.

PROGRAM 2.3 Linear Model without Adjustment for Covariates for Visit 1

```
%NParCov4(OUTCOMES = V1, COVARS = , HYPOTH = NULL, STRATA = center,
TRTGRPS = treatment, TRANSFORM = NONE, COMBINE = FIRST, EXACT = YES,
NREPS = 5000, SEED = 36, DSNIN = RESP);
```

TABLE 2.6 Linear Model without Adjustment for Covariates for Visit 1

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Visit 1	0.3935	0.2032	3.7497	0.0528

An exact p -value based on 5000 permutation data sets is 0.0542. For confidence intervals, we could specify `HYPOTH = ALT` in Program 2.3 above. The asymptotic 95% confidence interval for the treatment estimate is (0.0024, 0.7846). Here, the BC_a and percentile intervals are (0.0197, 0.7906) and (0.0131, 0.7799), respectively, based on 5000 bootstrap data sets.

Note that for this particular example, the interpretation of the p -value and the confidence intervals don't coincide. This phenomenon does occur from time to time. This is because the p -value is based on re-randomizations of the population, while the confidence interval is based on repeated random sampling. This phenomenon can also occur for 2×2 tables where the p -value for Fisher's exact test exceeds 0.05, but a confidence interval for the risk difference excludes 0. In this setting, Fisher's exact test is a randomization method based on re-randomizations of the population. But the confidence interval for the risk difference is based on repeated random sampling from a binomial distribution. The good news is that the result for the p -value and the confidence interval agree most of the time in terms of indicating whether a treatment effect exists. Otherwise, when the p -values and confidence intervals provide conflicting information, the p -value from the randomization test is based on more realistic and minimal assumptions, and should take precedence.

It is clear from the results of Tables 2.5 and 2.6 that adjustment for covariates provides a more sensitive comparison between the two treatments. To further illustrate the importance of covariate adjustment and the power of the nonparametric, randomization-based analysis of covariance methodology, Table 2.7 presents asymptotic 95% confidence intervals and interval widths for unstratified analyses similar to those presented in Programs 2.2 and 2.3, alongside comparable parametric analyses generated using PROC MIXED (Program 2.4). Intervals for both methods are non-significant or significant for the unadjusted or adjusted analysis, respectively. Further, in both the adjusted and unadjusted case, nonparametric, randomization-based analysis of covariance provides a narrower confidence interval than the model-based analysis, with greater precision observed when adjusting for covariates.

TABLE 2.7 Estimated Treatment Effects and Asymptotic 95% Confidence Intervals by Method

Method	Estimate	Lower Limit	Upper Limit	Width
Regression: Unadjusted	0.3996	-0.0129	0.8122	0.8251
NParCov: Unadjusted	0.3935	0.0024	0.7846	0.7822
Regression: Adjusted	0.4456	0.0995	0.7918	0.6923
NParCov: Adjusted	0.4266	0.1001	0.7531	0.6530

PROGRAM 2.4 Parametric Analyses with and without Adjustment for Covariates for Visit 1

```
proc mixed data = resp;
  ods output estimates = mixedc;
  model v1 = gender age baseline treatment;
  estimate "Adjusted Treatment Effect" treatment 1;
run;

proc mixed data = resp;
  ods output estimates = mixed;
  model v1 = treatment;
  estimate "Unadjusted Treatment Effect" treatment 1;
run;
```

Suppose there is interest in examining the treatment effect across all 4 visits while adjusting for covariates and accounting for strata (Program 2.5).

PROGRAM 2.5 Linear Model with Adjustment for Covariates for All Visits

```
%NParCov4(OUTCOMES = V1 V2 V3 V4, HYPOTH = NULL,
COVARS = gender age baseline, STRATA = center, TRANSFORM = NONE,
TRTGRPS = treatment, COMBINE = FIRST, DSNIN = RESP);
```

The output is provided in Table 2.8.

TABLE 2.8 Linear Model with Adjustment for Covariates for All Visits

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Visit 1	0.4008	0.1714	5.4690	0.0194
Visit 2	0.9516	0.2213	18.4901	< 0.0001
Visit 3	0.8160	0.2386	11.6948	0.0006
Visit 4	0.6175	0.2377	6.7513	0.0094

The treatment estimates from `_OUTDAT_DEPTEST` and the covariance matrix from the data set `_OUTDAT_COVBETA` can be used to test more complex hypotheses. We could test the global hypothesis of a non-null treatment effect at the 4 visits,

$$H_0 : f_1 = 0, f_2 = 0, f_3 = 0, f_4 = 0,$$

where f_i is the treatment effect at the i th visit. The test statistic

$$Q_{Global} = \hat{\beta}^\top C^\top (CV_{\hat{\beta}} C^\top)^{-1} C \hat{\beta}$$

is distributed $\chi^2_{(4)}$, where $C = I_4$ (an identity matrix of order four), and $V_{\hat{\beta}}$ is the covariance matrix of the covariate-adjusted treatment effects of visits 1 to 4 for the overall response. The SAS/IML code in Program 2.6 performs the analysis.

PROGRAM 2.6 Global Test for Treatment Response Across Visits

```
proc iml;
  use _OUTDAT_DEPTEST;
    read all var{beta} into B;
  close;
  use _OUTDAT_COVBETA;
    read all var{V1 V2 V3 V4} into V;
  close;
  C = I(4);
  Q = t(B) * t(C) * inv(C * V * t(C)) * C * B;
  p = 1 - probchi(Q, nrow(C));
  print Q p;
quit;
```

With $Q_{Global} = 19.44$ and $p = 0.0006$, the analysis suggests that a non-null treatment effect exists for at least one visit. Note that the 4 degree-of-freedom test Q_{Global} would tend to have much less power than the 1 degree-of-freedom test that uses the average of the treatment differences across the visits; this would also be considered a global test of a treatment effect across the 4 visits.

An additional analysis can test the null hypothesis that the treatment effect does not differ across study visits,

$$H_0 : f_1 = f_2 = f_3 = f_4.$$

Here, the result of the test statistic and p -value for treatment \times visit should be based on the variance under the alternative rather than the variance under the null, since the treatment effects are already understood to be non-null. In this case, the test statistic

$$Q_{Visit} = \hat{\beta}^\top C^\top (CV_{\hat{\beta}} C^\top)^{-1} C \hat{\beta}$$

is distributed $\chi^2_{(3)}$, where

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

Changing HYPOTH = ALT in Program 2.5 and setting

```
C = {1 0 0 -1, 0 1 0 -1, 0 0 1 -1};
```

in Program 2.6 provides $Q_{Treatment \times Visit} = 12.57$ and $p = 0.0057$. This suggests that the treatment effect differs across visits. In Section 2.16, we demonstrate how to generate essentially exact results for analyses with more than one response.

2.5 Analysis of binary endpoints

Case study 1 will be used to illustrate the analysis of binary endpoints with appropriate covariate adjustment. Specifically, in this section, we will analyze the binary response of (good, excellent) versus (terrible, poor, fair) at Visit 1. If the primary interest is in the test of the treatment effect, the most appropriate covariance matrix is the one that applies to the randomization distribution under the null hypothesis of no treatment difference. An exact p -value can be obtained; under the null hypothesis, all resampled data sets have the same covariance matrix. For illustration, we obtain weighted estimates across strata prior to covariance adjustment. The %NParCov4 call for this example is provided in Program 2.7.

PROGRAM 2.7 Difference in Proportions with Adjustment for Covariates for Visit 1

```
data RESP;
  set RESP;
  v1goodex = (V1 in (3,4));
run;

%NParCov4(OUTCOMES = v1goodex, COVARS = gender age baseline,
HYPOTH = NULL, STRATA = center, TRTGRPS = treatment, TRANSFORM = NONE,
COMBINE = FIRST, EXACT = YES, NREPS = 5000, SEED = 78, DSNIN = RESP);
```

The criterion for covariate imbalance $Q = 6.46$ is distributed $\chi^2_{(3)}$. The asymptotic p -value for the criterion is 0.0911, and the exact p -value provided in _OUTDAT_COVTEST based on 5000 permutation samples is 0.0922. It appears that randomization provided reasonably equal distributions of covariates across the strata. With TRANSFORM = NONE, the model provides the treatment estimate for the difference in proportions in Table 2.9.

TABLE 2.9 Linear Model with Adjustment for Covariates for Visit 1

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Visit 1	0.1839	0.0781	5.5455	0.0185

The exact p -value for the treatment effect provided in _OUTDAT_EXACT, based on 5000 permutation samples, is 0.0164. Therefore, active treatment provides a higher proportion of good or excellent responses when adjusting for gender, age, and baseline score. If there was interest in confidence intervals for the odds ratio, we could specify HYPOTH = ALT and TRANSFORM = LOGISTIC in the macro call above. However, this analysis causes an error for %NParCov4! What went wrong? The outcome was rare enough that some of the resampled data sets had no successes for at least one treatment arm within strata, and this caused the estimation to fail. If the outcome is expected to be rare, caution should be taken in specifying the analysis. For example, the following analysis adjusting for center as a covariate generates exact results (Program 2.8).

PROGRAM 2.8 Logistic Model with the Addition of Center as a Covariate for Visit 1

```
%NParCov4(OUTCOMES = v1goodex, COVARS = center gender age baseline,
HYPOTH = ALT, STRATA = NONE, TRTGRPS = treatment, TRANSFORM = LOGISTIC,
EXACT = YES, NREPS = 5000, SEED = 78, DSNIN = RESP);
```

The data set `_OUTDAT_CI` is created for all models and contains asymptotic confidence intervals for the treatment parameters when `HYPOTH = ALT`. However, an additional data set `_OUTDAT_RATIOCI` is produced here. It contains the asymptotic confidence intervals for the odds ratios. The odds ratio and asymptotic 95% confidence interval obtained from the data set `_OUTDAT_RATIOCI` is 2.2707 (1.2086, 4.2665). The BC_a and percentile intervals for the odds ratio are (1.1510, 4.6950) and (1.1681, 4.7672), respectively, based on 5000 samples.

Ideally, the first analysis should be regenerated using the new options (center as a covariate) to be consistent with the analysis from Program 2.8. The appropriate code is presented in Program 2.9 (results not shown).

PROGRAM 2.9 Difference in Proportions with Center as a Covariate for Visit 1

```
%NParCov4(OUTCOMES = v1goodex, COVARS = center gender age baseline,
HYPOTH = ALT, STRATA = NONE, TRTGRPS = treatment, TRANSFORM = NONE,
EXACT = YES, NREPS = 5000, SEED = 78, DSNIN = RESP);
```

2.6 Analysis of ordinal endpoints using a proportional odds model

Case study 1 can also be used to illustrate an analysis strategy based on a proportional odds model. Rather than dichotomize the ordinal response in this clinical trial, we can analyze the outcome 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 (Program 2.10).

PROGRAM 2.10 Proportional Odds Model with Adjustment for Covariates for Visit 1

```
data RESP;
  set RESP;
  v1ex = (V1 in (4));
  v1goodex = (V1 in (3,4));
  v1fairgoodex = (V1 in (2,3,4));
run;

%NParCov4(OUTCOMES = v1ex v1goodex v1fairgoodex, COVARS = gender age baseline,
HYPOTH = NULL, STRATA = center, TRTGRPS = treatment, TRANSFORM = PODDS,
COMBINE = FIRST, DSNIN = RESP);
```

A model where `TRANSFORM = PODDS` generates an additional data set `_OUTDAT_HOMOGEN` that contains a test for the appropriateness of the proportional odds assumption. The test of proportional odds is $Q = 3.69$, which is distributed $\chi^2_{(2)}$ and provides a p -value of 0.1578. Thus, we do not reject the assumption of proportional odds. Further, the covariate imbalance criterion, which in this case is a joint test of proportional odds and covariate imbalance (for gender, age, and baseline), is borderline ($p=0.0709$). The model provides the following treatment estimate of the log odds ratio comparing active treatment to placebo in Table 2.10.

TABLE 2.10 Proportional Odds Model with Adjustment for Covariates for Visit 1

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Visit 1	0.6233	0.3046	4.1857	0.0408

When `HYPOTH = ALT` is changed in Program 2.10, the data set `_OUTDAT_RATIOCI` contains odds ratio and asymptotic 95% confidence interval of 1.9548 (1.0455, 3.6548). Therefore, the odds of a positive response in the test treatment are greater than for placebo. Finally, had the test of homogeneity been statistically significant, the proportional odds assumption would not have applied for this analysis. In this case, the full model could be obtained by using `TRANSFORM = LOGISTIC` in the macro call above to estimate separate odds ratios for each binary outcome. See Program 2.11 for a sample program (results not shown).

PROGRAM 2.11 Model with Adjustment for Covariates for Visit 1 - Proportional Odds Does Not Apply

```
%NParCov4(OUTCOMES = v1ex v1goodex v1fairgoodex, COVARS = gender age baseline,
HYPOTH = ALT, STRATA = center, TRTGRPS = treatment, TRANSFORM = LOGISTIC,
COMBINE = FIRST, DSNIN = RESP);
```

2.7 Analysis of continuous endpoints using the log-ratio of two means

Analysis of continuous endpoints in the context of the nonparametric, randomization-based methodology can be illustrated using the osteoporosis trial from Case study 2. The analysis focuses on comparing the yearly risk of fractures between the two treatments using a risk ratio. 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 transformation (based on the natural logarithm) and applying covariance adjustment (Program 2.12).

PROGRAM 2.12 Log-Ratio of Two Means with Adjustment for Covariates

```
%NParCov4(OUTCOMES = YEARLYRISK, COVARS = age, HYPOTH = NULL,
STRATA = center, TRTGRPS = treatment, TRANSFORM = LOGRATIO,
COMBINE = PRETRANSFORM, EXACT = YES, NREPS = 5000, SEED = 22,
DSNIN = FRACTURE);
```

The covariate imbalance criterion for this example is $Q = 0.0536$ that is distributed $\chi^2_{(1)}$ with $p = 0.8170$. The exact p -value is 0.8286. It appears that randomization provided a reasonably equal distribution of age between treatment groups. The model generates the treatment estimate of the log-ratio of the two means. It is presented in Table 2.11.

TABLE 2.11 Log-Ratio of Two Means with Adjustment for Covariates

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Yearly Risk	-0.5803	0.3006	3.7259	0.0536

The exact p -value is 0.0626 based upon 5000 samples. If there was interest in confidence intervals for the log-ratio of the means, we could specify `HYPOTH = ALT` in the macro call above. 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. The BC_a and percentile intervals for the risk ratio are (0.2823, 0.9999) and (0.2882, 1.0209), respectively, based on 5000 samples. Here, the BC_a intervals imply a significant difference between placebo and active drug that conflicts with the asymptotic and percentile intervals. Doubling the number of bootstrap samples provides a BC_a interval that includes 1. Note that had there been greater variability in the duration of exposure for each subject, we could have analyzed with the methods in the following section.

2.8 Analysis of count endpoints using log-incidence density ratios

Several of the 40 clinical sites in the aneurysmal subarachnoid hemorrhage trial from Case study 3 had low enrollment, with fewer than two patients per treatment in some cases. In order to analyze under the alternative hypothesis, we will ignore center as a stratification variable. For each treatment, the incidence density will be the number of treatment-emergent, serious adverse events that a subject experienced, normalized by the number of days that he or she was in the trial. The ratio of the incidence densities (nicardipine over placebo) will be treated as the primary endpoint in the trial. Treatment estimates are adjusted for age and gender in the macro call of Program 2.13.

PROGRAM 2.13 Log-Incidence Density Ratios with Adjustment for Covariates

```
%NParCov4(OUTCOMES = count, COVARS = age male, EXPOSURES = days,
HYPOTH = NULL, TRTGRPS = treatment, TRANSFORM = INCDENS,
EXACT = YES, NREPS = 5000, SEED = 77, DSNIN = nic);
```

The covariate imbalance criterion for this example is $Q = 0.81$, which is distributed $\chi^2_{(2)}$ with the p -value of 0.666. The exact p -value is 0.6810. It appears that randomization provided a reasonably equal distribution of age and gender between treatment groups. The treatment estimate of the log-incidence density ratio is presented in Table 2.12.

TABLE 2.12 Log-Incidence Density Ratios with Adjustment for Covariates

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Incidence Density	-0.0718	0.1428	0.2526	0.6153

The exact p -value is 0.6104 based on 5000 samples. When $HYPOTH = ALT$ is changed in Program 2.13, the data set `_OUTDAT_RATIOCI` contains incidence density ratio 0.9307 which has 95% confidence interval (0.7007, 1.2363). Treatment-emergent, serious adverse events are $(1/0.9307 - 1) \times 100\% = 7.4\%$ more likely per day for subjects on placebo compared to nicardipine. This difference is not statistically significant when adjusting for the covariates. The BC_a and percentile intervals for the incidence density ratio are (0.6973, 1.2228) and (0.7009, 1.2303), respectively, based on 5000 samples.

2.9 Analysis of time-to-event endpoints

This section focuses on the colorectal cancer trial from Case study 4. First, we conduct the analysis of the primary time-to-event endpoint in the trial (time to death) using log-rank scores and adjusting for the number of days from initial cancer diagnosis to start of study drug administration and the sum of the longest lesion diameters in millimeters at baseline. We perform a stratified analysis of overall survival based upon the KRAS gene, weighting each center using Mantel-Haenszel weights. Further, we combine treatment estimates across strata prior to covariance adjustment (Program 2.14). Note that for time-to-event outcomes, the time until the event occurs or when the patient is censored (os_time) is provided in EXPOSURES. The binary outcome of whether a patient experienced an event or not (here, os_event, where 1 implies an event occurred, and 0 implies the patient was censored) is provided in OUTCOMES.

PROGRAM 2.14 Overall Survival Using Log-rank Scores with Adjustment for Covariates

```
%NParCov4(OUTCOMES = os_event, COVARS = days length, EXPOSURES = os_time,
STRATA = mutant, HYPOTH = NULL, TRTGRPS = treatment, TRANSFORM =
LOGRANK, COMBINE = FIRST, EXACT = YES, NREPS = 5000, SEED = 45,
DSNIN = PROSTATE, DSNOUT = OS);
```

The criterion for covariate imbalance $Q = 1.19$ is distributed $\chi^2_{(2)}$. The asymptotic p -value for the criterion is 0.5514, and the exact p -value based on 5000 permutation samples is 0.5484. It appears that randomization provided reasonably equal distributions of covariates between the two treatments. The treatment estimate, which represents the difference in average log-rank scores for overall survival, is presented in Table 2.13.

TABLE 2.13 Overall Survival Using Log-rank Scores with Adjustment for Covariates

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
Overall Survival	0.0299	0.1749	0.0292	0.8642

The exact p -value based on 5000 permutation samples is 0.8712. The positive treatment estimate implies that the treatment has decreased overall survival from colorectal cancer, relative to control. However, this result is not significant so we conclude that there is no difference between the treatments when adjusting for the covariates. A similar analysis for the other time-to-event endpoint in the colorectal cancer trial (time to disease progression) is performed in Program 2.15.

PROGRAM 2.15 Progression-Free Survival Using Log-rank Scores with Adjustment for Covariates

```
%NParCov4(OUTCOMES = pfs_event, COVARS = days length, EXPOSURES = pfs_time,
STRATA = mutant, HYPOTH = NULL, TRTGRPS = treatment, TRANSFORM =
LOGRANK, COMBINE = FIRST, EXACT = YES, NREPS = 5000, SEED = 45,
DSNIN = PROSTATE, DSNOUT = PFS);
```

In this case, the criterion for covariate imbalance $Q = 1.19$, which is distributed $\chi^2_{(2)}$. The asymptotic p -value for the criterion is 0.5514, and the exact p -value based on 5000 permutation samples is 0.562. It appears that randomization provided reasonably equal distributions of covariates between the two treatments. The treatment estimate, which represents the difference in average log-rank scores for progression-free survival, is presented in Table 2.14.

TABLE 2.14 Progression-Free Survival Using Log-rank Scores with Adjustment for Covariates

Outcome	Treatment Effect	Standard Error	Test Statistic	P-value
PFS	0.0090	0.1750	0.0026	0.9592

The exact p -value based on 5000 permutation samples is 0.9600. The positive treatment estimate implies that the treatment has decreased progression-free survival from colorectal cancer, relative to control. However, this result is not significant, so we conclude that there is no difference between the treatments when adjusting for the covariates.

A joint test of the overall and progression-free survival can be performed with multiple calls to %NParCov4 to test the hypothesis

$$H_0 : f_{os} = 0, f_{pfs} = 0,$$

where f_{os} and f_{pfs} are the treatment effects for overall and progression-free survival, respectively. The _OUTDAT_SURV data sets from the two runs above provide the log-rank scores (LOGRANK_OS_EVENT or LOGRANK_PFS_EVENT). These data sets can be merged together, and a joint analysis of overall and progression-free survival can be performed using Program 2.16.

PROGRAM 2.16 Joint Analysis of Overall and Progression-Free Survival Using Log-rank Scores with Adjustment for Covariates

```
%NParCov4(OUTCOMES = logrank_os_event logrank_pfs_event, COVARS = days length,
EXPOSURES =, STRATA = mutant, HYPOTH = NULL, TRTGRPS = treatment,
TRANSFORM = NONE, COMBINE = FIRST, EXACT = YES, NREPS = 5000, SEED = 45,
DSNIN = PROSTATE);
```

Notice that EXPOSURES is left unspecified since these values have already been accounted for in the LOGRANK_OS_EVENT and LOGRANK_PFS_EVENT outcomes. Further, no transformation is needed, so TRANSFORM = NONE. Defining $C = I_2$, an identity matrix of order two, and using similar methods and code as those for Program 2.6, the two degree-of-freedom bivariate test is computed in Program 2.17.

PROGRAM 2.17 Bivariate Test of Overall and Progression-Free Survival

```
proc iml;
  use _OUTDAT_DEPTEST;
    read all var{beta} into B;
  close;
  use _OUTDAT_COVBETA;
    read all var{logrank_os_event logrank_pfs_event} into V;
  close;
  C = I(2);
  Q = t(B) * t(C) * inv(C * V * t(C)) * C * B;
  p = 1 - probchi(Q, nrow(C));
  print Q p;
quit;
```

Here, $Q_{Bivariate} = 0.0295$ with asymptotic $p = 0.9854$. It is natural to ask if an exact p -value is available for this example. As described in Appendix B, the exact methodologies of %NParCov4 are limited to a single outcome variable. However, it is possible to write additional code to repeatedly call %NParCov4. Though repeated calls to %NParCov4 might not be the most efficient procedure programmatically, it does enable the analyst to compute exact results for more general problems.

In Program 2.18, the $Q_{Bivariate}$ statistic is computed for each permutation sample, and the proportion of these $Q_{Bivariate}$ statistics that is greater than or equal to the $Q_{Bivariate}$ statistic computed from the observed sample constitutes the p -value.

PROGRAM 2.18 Exact Bivariate Test of Overall and Progression-Free Survival

```
%let nreps = 5000;
%let seed = 50;

options nonotes;
proc sort data = both;
  by treatment patient;
run;

proc multtest permutation nocenter noprint seed=&seed nsample=&nreps
  outsamp=samp_null1(drop=_obs_) data=both;
  class treatment;
  strata mutant;
  test mean(logrank_os_event logrank_pfs_event days length);
run;

data samp_null2;
  set samp_null1;
  *** convert treatment back to numeric ***;
  treatment = _class_+0;
  *** convert strata back to numeric ***;
  mutant = _stratum_+0;
  drop _class_;
run;

*** sample 0 is the observed sample ***;
data samp_null3;
  set both(in=a) samp_null2;
  if a then _sample_=0;
run;

%macro BIVARIATE;
  %do rep = 0 %to &nreps;
    data temp;
      set samp_null3;
      where _sample_ = &rep;
    run;

    %put PROGRESS = &rep;

    %NPARCOV4(outcomes = logrank_os_event logrank_pfs_event,
      covars = days length,
      trtgrps = treatment,
      strata = mutant,
      hypoth = NULL,
      transform = NONE,
      combine = FIRST,
      c = 1,
      dsnin = temp,
      dsnout = outdat,
      exact = NO,
      details = NO);

    %if &rep = 0 %then %do;
```

```

data allbetas;
  set _outdat_deptest(keep = outcomes beta);
  _sample_ = &rep;
run;

data allcovs;
  set _outdat_covbeta;
  _sample_ = &rep;
run;
%end;
%else %do;
  data allbetas;
    set allbetas _outdat_deptest(keep = outcomes beta in = inb);
    if inb then _sample_ = &rep;
  run;

  data allcovs;
    set allcovs _outdat_covbeta(in = inb);
    if inb then _sample_ = &rep;
  run;
%end;

proc datasets nolist;
  delete temp;
quit;
%end;
%mend BIVARIATE;

%BIVARIATE;

proc iml;
  use allbetas;
  read all var{beta} into B;
  close;
  use allcovs;
  read all var{logrank_os_event logrank_pfs_event} into V;
  close;
  C = I(2);
  Q0 = t(B[1:2]) * C * inv(C * V[1:2,] * t(C)) * C * B[1:2];
  twosided = 0;
  do i = 2 to &nreps+1;
    Btemp = B[1+(i-1)*ncol(V):i*ncol(V)];
    Vtemp = V[1+(i-1)*ncol(V):i*ncol(V),];
    Q = t(Btemp) * t(C) * inv(C * Vtemp * t(C)) * C * Btemp;
    twosided = twosided + (Q >= Q0);
  end;
  twosided = twosided / &nreps;
  print Q0 twosided;
quit;

options notes;

```

Program 2.18 generates 5000 permutation samples to compute an essentially exact p -value = 0.9836. DETAILS = NO and OPTIONS NONOTES are used to minimize the amount of information saved to the SAS log while these simulations are running. Note that in practice, an exact p -value is typically not useful when the asymptotic p -value exceeds 0.15. Computing an exact p -value here was for illustration more than anything else.

This entire analysis could have been conducted using Wilcoxon scores. An alternative analysis that might be of interest is to conduct a joint analysis of an endpoint using both Wilcoxon and log-rank scores. For example, in the above description, merely replace the progression-free survival call with an additional analysis for overall survival using Wilcoxon scores (`TRANSFORM = WILCOXON`) to test the null hypothesis

$$H_0 : f_{Wilcoxon} = 0, f_{logrank} = 0,$$

where $f_{Wilcoxon}$ and $f_{logrank}$ are the treatment effects for overall survival. They are computed using the Wilcoxon or log-rank scores, respectively. What is the rationale for such an analysis? In some instances, it may not be clear which test might be most appropriate for a given data set. For example, the log-rank test assumes proportional hazards between the two treatments that might not hold in practice. Wilcoxon scores place greater emphasis on earlier event times (Collett, 2015). If the proportional hazards assumption holds, it might be unclear which test might be more appropriate. In oncology, the log-rank test is often more powerful to assess long-term survival. However, in analgesic studies where interest is in earlier pain relief, the Wilcoxon test might be preferred. See Tangen and Koch (1999b) for an illustration of this particular bivariate test.

2.10 Summary

This chapter presented the nonparametric, randomization-based analysis of covariance methodology and its application to the analysis of clinical endpoints. The methodology was implemented using the `%NParCov4` macro that contains functionality for many of the analysis methods described to date, with the ability to compute essentially-exact p -values and confidence intervals. It is recommended that advanced randomization-based methods be specified as the primary analysis because of their minimal assumptions and the availability of essentially-exact p -values. 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 distributions of subjects in a clinical trial.

Acknowledgements

The authors extend their sincerest thanks to Elaine Kearney, Todd Schwartz, Hengrui Sun, and Laura Zhou for their careful review of this chapter.

References

- Collett D. (2015). *Modelling Survival Data in Medical Research, Third Edition*. Boca Raton, FL: CRC Press.
- Davison A.C., Hinkley D.V. (1997). *Bootstrap Methods and Their Applications*. Cambridge, UK: Cambridge University Press.
- Efron B., Tibshirani R.J. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman and Hall.

- Eisenhauer E.A., Therasse P., Bogaerts J., Schwartz L.H., Sargent D., Ford R., Dancey J., Arbuck S., Gwyther S., Mooney M., Rubinstein L., Shankar L., Dodd L., Kaplan R., Lacombe D., Verweij J. (2009). New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1). *European Journal of Cancer* 45, 228-247.
- Haley E.C., Kassell N.F., Torner J.C. (1993). A randomized controlled trial of high-dose intravenous nicardipine in aneurysmal subarachnoid hemorrhage. *Journal of Neurosurgery* 78, 537-547.
- Jennett B., Bond M. (1975). Assessment of outcome after severe brain damage: A practical scale. *Lancet* 1, 480-484.
- Kalbfleisch J.D., Prentice R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Kawaguchi A., Koch G.G., Wang X. (2011). Stratified multivariate Mann-Whitney estimators for the comparison of two treatments with randomization based covariance adjustment. *Statistics in Biopharmaceutical Research* 3, 217-231.
- Kawaguchi A., Koch G.G. (2015). sanon: An R package for stratified analysis with nonparametric covariable adjustment. *Journal of Statistical Software* 67:9, 1-37. Available at <http://www.jstatsoft.org/v67/i09/>.
- Koch G.G., Carr G.J., Amara I.A., Stokes M.E., Uryniak T.J. (1990). Categorical data analysis. In: Berry D.A., ed. *Statistical Methodology in the Pharmaceutical Sciences*. New York: Marcel Dekker.
- Koch G.G., Gillings D.B. (1983). Inference, design-based vs. model-based. In: Kotz S., Johnson N.L., Read C.B., eds. *Encyclopedia of Statistical Sciences*. New York: Wiley.
- Koch G.G., Tangen C.M. (1999). Nonparametric analysis of covariance and its role in non-inferiority clinical trials. *Drug Information Journal* 33, 1145-1159.
- Koch G.G., Tangen C.M., Jung J.W., Amara I.A. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine* 17, 1863-1892.
- LaVange L.M., Durham T.A., Koch G.G. (2005). Randomization-based nonparametric methods for the analysis of multicentre trials. *Statistical Methods in Medical Research* 14, 281-301.
- LaVange L.M., Koch G.G. (2008). Randomization-based nonparametric analysis of covariance (ANCOVA). In: D'Agostino R.B., Sullivan L., Massaro J., eds. *Wiley Encyclopedia of Clinical Trials, First Edition*. Hoboken, NJ: Wiley.
- Moodie P.F., Saville B.R., Koch G.G., Tangen C.M. (2011). Estimating covariate-adjusted log hazard ratios for multiple time intervals in clinical trials using nonparametric randomization based ANCOVA. *Statistics in Biopharmaceutical Research* 3, 232-241.
- Saville B.R., Herring A.H., Koch G.G. (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 B.R., Koch G.G. (2013). Estimating covariate-adjusted log hazard ratios in randomized clinical trials using Cox proportional hazards models and nonparametric randomization based analysis of covariance. *Journal of Biopharmaceutical Statistics* 23, 477-490.
- Saville B.R., LaVange L.M., Koch G.G. (2011). Estimating covariate-adjusted incidence density ratios for multiple time intervals in clinical trials using nonpara-

metric randomization based ANCOVA. *Statistics in Biopharmaceutical Research* 3, 242-252.

Stokes M.E., Davis C.S., Koch G.G. (2012). *Categorical Data Analysis Using The SAS System, Third Edition*. Cary, NC: SAS Institute Inc. Data available at <http://support.sas.com/rnd/app/stat/cat/edition3/samples/index.html>.

Tangen C.M., Koch G.G. (1999a). Complementary nonparametric analysis of covariance for logistic regression in a randomized clinical trial setting. *Journal of Biopharmaceutical Statistics* 9, 45-66.

Tangen C.M., Koch G.G. (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, 307-338.

Tangen C.M., Koch G.G. (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, 1039-1058.

Tangen C.M., Koch G.G. (2001). Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose response relationships. *Statistics in Medicine* 20, 2585-2607.

Zhao Y., Saville B.R., Zhou H., Koch G.G. (2016). Sensitivity analysis for missing outcomes in time-to-event data with covariate adjustment. *Journal of Biopharmaceutical Statistics* 26, 269-279.

Appendix

A Theoretical Details

A.1 Non-Parametric Randomization-Based Analysis of Covariance

Suppose there are two 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 from a population of $N_i \gg n_i$ patients, \bar{y}_i be the r -vector of outcome means, and \bar{x}_i be the t -vector of covariate means of the i th 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 $\text{cov}(f_i) = V_{f_i} = D_i V_i D_i$, where

$$V_i = \text{cov} \left(\begin{bmatrix} \bar{y}_i \\ \bar{x}_i \end{bmatrix} \right)$$

and D_i is a diagonal matrix with the first derivatives $g'(\bar{y}_i)$ corresponding to the first r cells of the diagonal with ones 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 zeros and fit the model

$$E[f] \hat{=} \begin{bmatrix} I_r \\ 0_{(t \times r)} \end{bmatrix} \hat{\beta} = X \hat{\beta}$$

using weighted least squares with weights based on V_f^{-1} , with $\hat{=}$ meaning "is estimated by". $\hat{\beta}$ is an $r \times 1$ vector of covariate-adjusted treatment effects. 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 $\text{cov}(\hat{\beta})$ is $(X^\top V_f^{-1} X)^{-1}$. There are two ways to estimate V_f . Under the null hypothesis of no treatment difference,

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

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

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

In general, the variance under the null would be the principal structure for producing p -values for hypothesis testing under the null when the null corresponded to no difference between treatments (or more formally, the strong null that each patient would have the same outcome, regardless of the assigned treatment). The variance under the alternative will be used for computing confidence intervals. Here, the standard errors address replicated random samples from a population for which the randomized population is arguably a random sample. P -values under the alternative might be useful since they will agree with the confidence interval, at least when both are based on normal approximations. Throughout this chapter, we take the approach that p -values will be computed under the null, with corresponding confidence intervals computed under the alternative hypothesis.

A test to compare the two treatments for outcome $y_j, j = 1, 2, \dots, r$, with adjustment for covariates is

$$Q_j = \frac{\hat{\beta}_j^2}{V_{\hat{\beta}_j}}.$$

With sufficient sample size, Q_j is approximately distributed $\chi_{(1)}^2$. 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 $\chi_{(t)}^2$. This criterion should not be considered a test for validity of nonparametric, randomization-based analysis of covariance, but merely as a measure of the degree of imbalance among covariates between the treatments. Due to the randomization requirement of the methodology, significant imbalance detected among the covariates should be interpreted as a chance finding. Adjusting for the covariates in the analysis removes this imbalance. The results of an unadjusted analysis would be confounded with the imbalance among the covariates. Small p -values for the covariate imbalance criterion are more appropriately referred to as "atypical relative to chance" rather than "significant."

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 $(r \times t)$. In %MParCov4, the above methods correspond to TRANSFORM = NONE and COMBINE = NONE.

A.2 Logistic Transformation

If \bar{y}_i consists of (0,1) binary outcomes, a logistic transformation of the elements of \bar{y}_i might 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 j th outcome. Under the null hypothesis, the first r elements of the diagonal of D_i are $\{\bar{y}_j(1 - \bar{y}_j)\}^{-1}$, 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 $\{\bar{y}_{ij}(1 - \bar{y}_{ij})\}^{-1}$. The $\hat{\beta}_j$ and associated 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

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

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 a vector of length $(r - 1)$ containing elements -1 . With sufficient sample size, Q_c is approximately distributed $\chi^2_{(r-1)}$.

If homogeneity applies, you can use the simplified model

$$E[f] = \begin{bmatrix} 1_r \\ 0_t \end{bmatrix} \hat{\beta}_R = X_R \hat{\beta}_R.$$

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**. **%NParCov4** computes the test of homogeneity and the reduced model when **PODDS** is selected. Should the p -value for Q_c be such that it is $\leq \alpha$, the proportional odds assumption is violated, which implies that the full model is more appropriate (obtained with **TRANSFORM = LOGISTIC**).

A.3 Log-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

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

which is the \log_e -ratio of the j th 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 associated 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**.

A.4 Log-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

$$M_0 = \begin{bmatrix} I_r & -I_r & 0_{(r \times t)} \\ 0_{(t \times r)} & 0_{(t \times r)} & I_t \end{bmatrix},$$

where $0_{(r \times t)}$ is an $r \times t$ matrix of zeros 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 $\text{cov}(g_i) = V_{g_i} = M_i V_i M_i^\top$, where

$$V_i = \text{cov} \left(\begin{bmatrix} \bar{y}_i \\ \bar{e}_i \\ \bar{x}_i \end{bmatrix} \right)$$

and

$$M_i = \begin{bmatrix} D_{\bar{y}_i} & -D_{\bar{e}_i} & 0_{(r \times t)} \\ 0_{(t \times r)} & 0_{(t \times r)} & I_t \end{bmatrix}.$$

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 j th 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 associated 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`.

A.5 Wilcoxon and Log-rank Scores for Time-to-Event Outcomes

Unlike Sections A.2, A.3, and A.4 where a transformation is applied to the means of one or more outcomes, `%NParCov4` can transform subject-level event flags and times into Wilcoxon or log-rank scores to perform analyses of time-to-event outcomes. Scores are defined from binary outcome flags (`OUTCOMES`) that define whether an event occurs ($= 1$) or not ($= 0$) and the exposure 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). The definitions are from the appendix of Tangen and Koch (1999b). The Wilcoxon scores are closely related to the Prentice generalized Wilcoxon scores presented in Kalbfleisch and Prentice (1980).

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

Log-rank 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 $\zeta = 1, 2, \dots, L$ 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 log-rank 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 $\zeta = 1, 2, \dots, L$ 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**. **%NParCov4** computes survival scores, summarizes as \bar{y}_i and analyzes as in Section A.1 using a linear transformation.

A.6 Stratification

Stratification can be handled in one of three ways:

1. Estimate $\hat{\beta}_h$ for each stratum h , $h = 1, 2, \dots, H$, and combine across strata to obtain

$$\bar{\beta} = \frac{\sum_h w_h \hat{\beta}_h}{\sum_h w_h},$$

where

$$w_h = \left(\frac{n_{h1}n_{h2}}{n_{h1} + n_{h2}} \right)^c$$

and $0 \leq c \leq 1$, 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 test 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

$$f_w = \frac{\sum_h w_h f_h}{\sum_h w_h}$$

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 A.2, A.3, and A.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**.

A.7 Resampling-Based Exact Methods

%NParCov4 contains options for computing essentially exact p -values under HYPOTH = NULL and confidence intervals under HYPOTH = ALT for a single outcome. Under EXACT = YES and HYPOTH = NULL, %NParCov4 uses PROC MULTTEST with the PERMUTATION option to compute NREPS data sets. In this case, the sampling of records to either treatment (within strata) is performed without replacement. Under EXACT = YES and HYPOTH = ALT, %NParCov4 uses PROC MULTTEST with the BOOTSTRAP option and a BY statement for treatment to compute NREPS data sets where the sampling of records with replacement by treatment (within strata). Further, under EXACT = YES and HYPOTH = ALT, n jackknife data sets are created where a single observation is deleted from each jackknife data set within strata, where n is the total number of observations in the analysis. The jackknife data sets are used in the computation of acceleration for the bias-corrected and accelerated (BC_a) intervals.

Exact p -values are obtained for EXACT = YES and HYPOTH = NULL. An exact criterion for the random imbalance of covariates between the two treatment groups computes Q_m , $m = 1, 2, \dots, B$, for each of B resampled data sets. The p -value is calculated as $\#(Q_m \geq Q_0)/B$, where Q_0 is the criterion for random imbalance for the observed data set. Under the null hypothesis, the treatment parameter $\hat{\beta}_m$, $m = 1, 2, \dots, B$, is computed for each resampled data set and is compared to the treatment parameter of the observed data set ($\hat{\beta}_0$) to calculate p -values:

1. two-sided: $\#(|\hat{\beta}_m| \geq |\hat{\beta}_0|)/B$,
2. one-sided lower: $\#(\hat{\beta}_m \leq \hat{\beta}_0)/B$,
3. one-sided upper: $\#(\hat{\beta}_m \geq \hat{\beta}_0)/B$.

Under EXACT = YES and HYPOTH = ALT, bias-corrected and accelerated intervals (BC_a) are computed using the B bootstrap samples and the n jackknife samples (Efron and Tibshirani, 1993). The $100(1 - \alpha)\%$ BC_a interval is the α_1 and α_2 percentiles of all B bootstrap estimates and is defined as $BC_a = (\hat{\beta}_{(\alpha_1)}, \hat{\beta}_{(\alpha_2)})$, where

$$\alpha_1 = \Phi \left(\hat{b} + \frac{\hat{b} + \Phi^{-1}(\alpha/2)}{1 - \hat{a}(\hat{b} + \Phi^{-1}(\alpha/2))} \right), \quad \alpha_2 = \Phi \left(\hat{b} + \frac{\hat{b} + \Phi^{-1}(1 - \alpha/2)}{1 - \hat{a}(\hat{b} + \Phi^{-1}(1 - \alpha/2))} \right),$$

$\hat{b} = \Phi^{-1} \left(\frac{\#(\hat{\beta}_m \leq \hat{\beta}_0)}{B} \right)$ is the bias-correction,

$$\hat{a} = \frac{\sum_{h=1}^H n_h^{-3} \sum_{i=1}^{n_h} (\hat{\beta}_{hi}^* - \bar{\beta}_h^*)^3}{6 \left(\sum_{h=1}^H n_h^{-2} \sum_{i=1}^{n_h} (\hat{\beta}_{hi}^* - \bar{\beta}_h^*)^2 \right)^{3/2}}$$

is the acceleration (Davis and Hinkley 1997); $\hat{\beta}_{hi}^*$ is the estimate from the i th jackknife sample in the h th strata; $\bar{\beta}_h^*$ is the mean of the n_h jackknife estimates from the h th strata, $\Phi(\cdot)$ is the CDF and $\Phi^{-1}(\cdot)$ is the inverse-CDF of the standard normal distribution, and α is the type I error.

$100(1 - \alpha)\%$ percentile confidence intervals are also calculated where $(\hat{\beta}_{(\alpha/2)}, \hat{\beta}_{(1-\alpha/2)})$ are the $\alpha/2$ and $1 - \alpha/2$ percentiles of the B estimates. When COMBINE = PRETRANSFORM or COMBINE = FIRST, acceleration is computed as if in a single stratum using $\hat{\beta}_w$. An exact two-sided p -value criterion for covariate imbalance (COVXACT) is calculated as $\#(|Q_m| \geq |Q_0|)/B$. While exact analyses are limited to a single outcome in %NParCov4, Program 2.18 in Section 2.9 illustrates how the macro can be repeatedly called to generate exact results for more general problems.

B SAS Macro %NParCov4

B.1 General Use and Options

%NParCov4 can be accessed by your program by including the lines:

```
filename nparcov4 "&path";
%include nparcov4(NParCov4) / nosource2;
```

or

```
%include "&path./nparcov4.sas".
```

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. The results of individual data sets can be displayed using PROC PRINT. PROC DATASETS will list available data sets generated by %NParCov4 at the end of the SAS log. %NParCov4 has the following options:

1. **OUTCOMES:** List of outcome variables for the analysis. All outcome variables are numeric and at least one outcome is required. Binary or time-to-event outcomes that use **TRANSFORM = LOGISTIC**, **PODDS**, **LOGRANK**, or **WILCOXON** can only have (0,1) values. Here, 1 should indicate that an event occurs, while 0 indicates that an event does not occur (or, in the case of a survival endpoint, that an individual is censored).
2. **COVARS:** List of covariates for the analysis. All covariates must be numeric; categorical variables must be re-expressed using one or more indicator variables. **COVARS** can be left unspecified (blank) for stratified, unadjusted results.
3. **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**.
4. **TRTGRPS:** This is a required, single two-level numeric variable that 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.
5. **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**.
6. **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.
7. **COMBINE:** Defines how the analysis will accommodate **STRATA**. Choose one.
 - (a) **NONE** (default) performs analysis as if data are from a single stratum.
 - (b) **LAST** performs covariance adjustment within each stratum, and then takes weighted averages of parameter estimates with weights based on stratum sample size (appropriate for larger strata).
 - (c) **FIRST** obtains a weighted average of treatment group differences (post-transformation, if applicable) across strata, then performs covariance adjustment (appropriate for smaller strata).

- (d) **PRETRANSFORM** (available for **TRANSFORM** = **LOGISTIC**, **PODDS**, **LOGRATIO**, or **INCDENS**) obtains a weighted average of treatment group means (pre-transformation) across strata, performs a transformation, and then applies covariance adjustment (appropriate for very small strata).
8. **HYPOTH**: Determines how variance matrices are computed.
 - (a) **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.
 - (b) **ALT** computes a separate covariance matrix for each treatment group within each stratum.
 9. **TRANSFORM**: Applies transformations to outcome variables. Choose one.
 - (a) **NONE** (default) applies no transformation and analyzes means for one or more outcome variables as provided.
 - (b) **LOGISTIC** applies the logistic transformation to the means of one or more (0,1) outcomes.
 - (c) **PODDS** applies the logistic transformation to multiple means of (0,1) outcomes that together form a single ordinal outcome.
 - (d) **LOGRATIO** applies the log-transformation based on the natural logarithm to one or more ratios of outcome means.
 - (e) **INCDENS** uses number of events experienced in **OUTCOMES** and **EXPOSURES** to calculate and analyze the log-transformation of one or more ratios of incidence density.
 - (f) **LOGRANK** uses event flags in **OUTCOMES** (1 = event, 0 = censored) and **EXPOSURES** to calculate and analyze log-rank scores for one or more time-to-event outcomes.
 - (g) **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.
 10. **ALPHA**: Confidence limit for intervals (default = 0.05).
 11. **EXACT**: Provides essentially exact analyses for a single outcome. Choose one.
 - (a) **NO** (default) does not perform resampling analyses;
 - (b) **YES** performs additional resampling-based analysis. If **HYPOTH** = **ALT**, bootstrap resampling is performed with observations sampled with replacement within treatment. If **HYPOTH** = **NULL**, permutation resampling is performed with observations sampled without replacement across treatments. This is equivalent to shuffling the treatment codes of the observations.
 12. **NREPS**: Number of random data sets to generate when **EXACT** = **YES** (default = 1000).
 13. **SEED**: Random seed when **EXACT** = **YES** (default = 0).
 14. **DSNIN**: Input data set. Required.
 15. **DSNOUT**: Prefix for output data sets. Required. For example, if **DSNOUT** is set to **OUTDAT**, the following data sets will be created:
 - (a) **_OUTDAT_COVTEST** contains the criterion for covariate imbalance. When **EXACT** = **YES** and **HYPOTH** = **NULL**, the exact criterion for covariate imbalance (**COVXACT**) is calculated.
 - (b) **_OUTDAT_DEPTEST** contains tests, treatment estimates, and ratios (for **TRANSFORM** = **LOGISTIC**, **PODDS**, **LOGRATIO**, or **INCDENS**) for outcomes.
 - (c) **_OUTDAT_CI** contains confidence intervals for outcomes when **HYPOTH** = **ALT**.
 - (d) **_OUTDAT_RATIOCI** contains confidence intervals for odds ratios (**TRANSFORM** =

PODDS or LOGISTIC), ratios (TRANSFORM = LOGRATIO), or incidence densities (TRANSFORM = INCDENS) when HYPOTH = ALT.

- (e) _OUTDAT_EXACT provided when EXACT = YES
 - i. When HYPOTH = NULL, the data set contains essentially exact, one- and two-sided p -values for the outcome.
 - ii. When HYPOTH = ALT, the data set contains percentile, bias-corrected, and accelerated confidence intervals. For TRANSFORM = LOGISTIC, PODDS, LOGRATIO, or INCDENS, the exact percentile, bias-corrected, and accelerated confidence intervals for the ratios are also provided.
 - (f) _OUTDAT_HOMOGEN contains the test for homogeneity of odds ratios for TRANSFORM = PODDS.
 - (g) _OUTDAT_BETASAMP contains the resampled values of the treatment parameter (BETASAMP). The ratios for TRANSFORM = LOGISTIC, PODDS, LOGRATIO, or INCDENS, are also provided (EXPBETASAMP). The observed treatment parameter is provided (labeled OBSERVED) and for either of the following:
 - i. HYPOTH = NULL, parameters from the permutation data sets are labeled PERMUTATION.
 - ii. HYPOTH = ALT, parameters from bootstrap data sets are labeled BOOTSTRAP, while parameters from jackknife data sets are labeled JACKKNIFE.
 - (h) _OUTDAT_SURV contains survival scores for TRANSFORM = LOGRANK or WILCOXON. The scores from this data set from multiple calls of %NParCov4 can be used with TRANSFORM = NONE to analyze log-rank and Wilcoxon scores simultaneously or to analyze survival outcomes with non-survival outcomes.
 - (i) _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 $\hat{\beta}$.
 - (j) _OUTDAT_STRATABETA contains the strata-specific estimates for $\hat{\beta}$ when COMBINE = LAST. These estimates can be used to compute acceleration for BC_a intervals (Appendix A.7) when repeatedly calling %NParCov4 for more general problems.
16. SYMSIZE: Define space requirements for SAS/IML. Should be increased for larger data sets or large NREPS (default = 20000).
 17. DETAILS: Print basic analysis information and the PROC DATASETS of analysis results to the SAS log (default = YES).

B.2 Content of Output Data Sets

Output data sets contain the following variables:

1. _OUTDAT_COVTEST
 - (a) TYPE = COVTEST.
 - (b) TRT1 is the value for control (lower value numerically).
 - (c) TRT2 is the value for treatment (higher value numerically).
 - (d) STRATA contains the strata variable.
 - (e) COVARIATES contains the list of covariates.
 - (f) OUTCOMES contains the list of outcomes.
 - (g) EXPOSURES contains the list of exposures for time-to-event endpoints.
 - (h) HYPOTHESIS is either NULL or ALT.

- (i) H is the number of strata levels.
 - (j) DF is the degrees of freedom.
 - (k) Q is the test statistic.
 - (l) PVALUE is the p -value.
 - (m) COVXACT is the exact p -value when EXACT = YES and HYPOTH = NULL.
2. _OUTDAT_DEPTEST.
- (a) TYPE = DEPTEST.
 - (b) TRANSFORM is the transformation applied to the outcomes.
 - (c) TRT1 is the value for control (lower value numerically).
 - (d) TRT2 is the value for treatment (higher value numerically).
 - (e) STRATA contains the strata variable.
 - (f) COVARIATES contains the list of covariates.
 - (g) OUTCOMES contains the list of outcomes.
 - (h) EXPOSURES contains the list of exposures for time-to-event endpoints.
 - (i) HYPOTHESIS is either NULL or ALT.
 - (j) H is the number of strata levels.
 - (k) DF is the degrees of freedom.
 - (l) BETA is the treatment estimate.
 - (m) SEBETA is the standard error of the treatment estimate.
 - (n) RATIO is the exponentiated treatment estimate for TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS.
 - (o) Q_J is the test statistic.
 - (p) PVALUE is the p -value.
3. _OUTDAT_CI.
- (a) TYPE = CI.
 - (b) TRT1 is the value for control (lower value numerically).
 - (c) TRT2 is the value for treatment (higher value numerically).
 - (d) STRATA contains the strata variable.
 - (e) COVARIATES contains the list of covariates.
 - (f) OUTCOMES contains the list of outcomes.
 - (g) EXPOSURES contains the list of exposures for time-to-event endpoints.
 - (h) HYPOTHESIS is either NULL or ALT.
 - (i) H is the number of strata levels.
 - (j) BETA is the treatment estimate.
 - (k) SEBETA is the standard error of the treatment estimate
 - (l) LOWER is the lower limit of the $100(1 - \alpha)\%$ confidence interval for the treatment estimate.
 - (m) UPPER is the upper limit of the $100(1 - \alpha)\%$ confidence interval for the treatment estimate.
 - (n) ALPHA is the value of α .
4. _OUTDAT_RATIOCI
- (a) TYPE = RATIOCI.
 - (b) TRT1 is the value for control (lower value numerically).
 - (c) TRT2 is the value for treatment (higher value numerically).
 - (d) STRATA contains the strata variable.

- (e) **COVARIATES** contains the list of covariates.
 - (f) **OUTCOMES** contains the list of outcomes.
 - (g) **EXPOSURES** contains the list of exposures for time-to-event endpoints.
 - (h) **HYPOTHESIS** is either **NULL** or **ALT**.
 - (i) **H** is the number of strata levels.
 - (j) **RATIO** is the exponentiated treatment estimate.
 - (k) **RATIO_LOWER** is the lower limit of the $100(1 - \alpha)\%$ confidence interval for the exponentiated treatment estimate.
 - (l) **RATIO_UPPER** is the upper limit of the $100(1 - \alpha)\%$ confidence interval for the exponentiated treatment estimate.
 - (m) **ALPHA** is the value of α .
5. **_OUTDAT_EXACT**.
- (a) **TYPE = EXACT**.
 - (b) **TRT1** is the value for control (lower value numerically).
 - (c) **TRT2** is the value for treatment (higher value numerically).
 - (d) **STRATA** contains the strata variable.
 - (e) **COVARIATES** contains the list of covariates.
 - (f) **OUTCOMES** contains the list of outcomes.
 - (g) **HYPOTHESIS** is either **NULL** or **ALT**.
 - (h) **H** is the number of strata levels.
 - (i) **TWOSIDED** is the two-sided, exact p -value to test the treatment effect.
 - (j) **ONE_LOWER** is the lower one-sided, exact p -value to test the treatment effect.
 - (k) **ONE_UPPER** is the upper one-sided exact p -value to test the treatment effect.
 - (l) **ALPHA** is the value of α .
 - (m) **NREPS** is the number of resamples.
 - (n) **SEED** is the value of seed.
 - (o) **BIAS** is \hat{b} , the bias used in the calculation of the BC_a interval.
 - (p) **ACCEL** is \hat{a} , the acceleration used in the calculation of the BC_a interval.
 - (q) **ALPHA_LOW** is α_1 used in the calculation of the BC_a interval.
 - (r) **ALPHA_HI** is α_2 used in the calculation of the BC_a interval.
 - (s) **BCA_LOWER** is the lower limit of the BC_a interval for the treatment effect.
 - (t) **BCA_UPPER** is the upper limit of the BC_a interval for the treatment effect.
 - (u) **PCT_LOWER** is the lower limit of the percentile interval for the treatment effect.
 - (v) **PCT_UPPER** is the upper limit of the percentile interval for the treatment effect.
 - (w) **BCA_RATIO_LOWER** is the lower limit of the BC_a interval for the exponentiated treatment effect for **TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS**.
 - (x) **BCA_RATIO_UPPER** is the upper limit of the BC_a interval for the exponentiated treatment effect for **TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS**.
 - (y) **PCT_RATIO_LOWER** is the lower limit of the percentile interval for the exponentiated treatment effect for **TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS**.
 - (z) **PCT_RATIO_UPPER** is the upper limit of the percentile interval for the exponentiated treatment effect for **TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS**.

6. `_OUTDAT_HOMOGEN`.

- (a) `TYPE = HOMOGEN`.
- (b) `TRT1` is the value for control (lower value numerically).
- (c) `TRT2` is the value for treatment (higher value numerically).
- (d) `STRATA` contains the strata variable.
- (e) `COVARIATES` contains the list of covariates.
- (f) `OUTCOMES` contains the list of outcomes.
- (g) `DF` is the degrees of freedom.
- (h) `BETA` is the treatment estimate.
- (i) `Q_C` is the test statistic.
- (j) `PVALUE` is the p -value.

7. `_OUTDAT_BETASAMP`.

- (a) `FLAG = OBSERVED, PERMUTATION, BOOTSTRAP, JACKKNIFE` for type of resampled treatment estimate.
- (b) `BETASAMP` is the treatment estimate.
- (c) `EXPBETASAMP` is the exponentiated treatment estimate for `TRANSFORM = LOGISTIC, PODDS, LOGRATIO, INCDENS`.

8. `_OUTDAT_SURV`.

- (a) Same as the input data set but adds survival scores according to `TRANSFORM`. For example, a variable called `VAR` will have one or the other of `LOGRANK_VAR` or `WILCOXON_VAR` added to the data set.

9. `_OUTDAT_COVBETA`.

- (a) `TYPE = COVBETA`.
- (b) Other variables correspond to outcomes and covariates to identify rows and columns of the covariance matrix.

10. `_OUTDAT_STRATABETA`.

- (a) `STRATA_H` contains the value for a strata level.
- (b) `BETA_H` contains the strata-specific treatment estimate.
- (c) `N_H` contains the sample size for a strata level.

About This Book

What Does This Book Cover?

The main goal of this book is to introduce popular statistical methods used in clinical trials and to discuss their implementation using SAS software. To help bridge the gap between modern statistical methodology and clinical trial applications, the book includes numerous case studies based on real trials at all stages of drug development. It also provides a detailed discussion of practical considerations and relevant regulatory issues as well as advice from clinical trial experts.

The book focuses on fundamental problems arising in the context of clinical trials such as the analysis of common types of clinical endpoints and statistical approaches most commonly used in early- and late-stage clinical trials. The book provides detailed coverage of approaches utilized in Phase I/Phase II trials, e.g., dose-escalation and dose-finding methods. Important trial designs and analysis strategies employed in Phase II/Phase III include multiplicity adjustment methods, data monitoring methods and methods for handling incomplete data.

Is This Book for You?

Although the book was written primarily for biostatisticians, the book includes high-level introductory material that will be useful for a broad group of pre-clinical and clinical trial researchers, e.g., drug discovery scientists, medical scientists and regulatory scientists working in the pharmaceutical and biotechnology industries.

What Are the Prerequisites for This Book?

General experience with clinical trials and drug development, as well as experience with SAS/STAT procedures, will be desirable.

What's New in This Edition?

The second edition of this book has been thoroughly revised based on the feedback provided by numerous readers of the first edition. The topics covered in the book have been grouped into three parts. The first part provides detailed coverage of general statistical methods used across the three stages of drug development. The second and third parts focus on the topics specific to early-phase and late-phase clinical trials, respectively.

The chapters from the first edition have been expanded to cover new approaches to addressing the statistical problems introduced in the original book. Numerous revisions have been made to improve the explanations of key concepts, add more examples and case studies. A detailed discussion of new features of SAS procedures has been provided and, in some cases, new procedures are introduced that were not available when the first edition was released.

What Should You Know about the Examples?

The individual chapters within this book include tutorial material along with multiple examples to help the reader gain hands-on experience with SAS/STAT procedures used in the analysis of clinical trials.

Software Used to Develop the Book's Content

The statistical methods introduced in this book are illustrated using numerous SAS/STAT procedures, including PROC GLM, PROC FREQ, PROC LOGISTIC, PROC GENMOD, PROC LIFETEST and PROC PHREG (used in the analysis of different types of clinical endpoints), PROC MIXED, PROC NLMIXED and PROC GENMOD (used in dose-finding trials), PROC MULTTEST (used in clinical trials with multiple objectives), PROC SEQDESIGN and PROC SEQTEST (used in group-sequential trials), PROC MIXED, PROC GLIMMIX, PROC GEE, PROC MI and PROC MIANALYZE (used in clinical trials with missing data). These procedures are complemented by multiple SAS macros written by the chapter authors to support advanced statistical methods.

Example Code and Data

You can access the example code, SAS macros and data sets used in this book by linking to its author page at <http://support.sas.com/publishing/authors/dmitrienko.html>.

SAS University Edition



This book is compatible with SAS University Edition. If you are using SAS University Edition, then begin here: <https://support.sas.com/ue-data>.

Output and Graphics

The second edition takes full advantage of new graphics procedures and features of SAS software, including PROC SGPLOT, PROC SGPANEL and ODS graphics options.

We Want to Hear from You

SAS Press books are written *by* SAS Users *for* SAS Users. We welcome your participation in their development and your feedback on SAS Press books that you are using. Please visit sas.com/books to do the following:

- Sign up to review a book
- Recommend a topic
- Request authoring information
- Provide feedback on a book

Do you have questions about a SAS Press book that you are reading? Contact the author through saspress@sas.com or https://support.sas.com/author_feedback.

SAS has many resources to help you find answers and expand your knowledge. If you need additional help, see our list of resources: sas.com/books.

About These Authors



Alex Dmitrienko, PhD, is Founder and President of Mediana Inc. He is actively involved in biostatistical research with an emphasis on multiplicity issues in clinical trials, subgroup analysis, innovative trial designs, and clinical trial optimization. Dr. Dmitrienko coauthored the first edition of *Analysis of Clinical Trials Using SAS®: A Practical Guide*, and he coedited *Pharmaceutical Statistics Using SAS®: A Practical Guide*.



Gary G. Koch, PhD, is Professor of Biostatistics and Director of the Biometrics Consulting Laboratory at the University of North Carolina at Chapel Hill. He has been active in the field of categorical data analysis for fifty years. Professor Koch teaches classes and seminars in categorical data analysis, consults in areas of statistical practice, conducts research, and trains many biostatistics students. He is coauthor of *Categorical Data Analysis Using SAS®, Third Edition*.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:

<http://support.sas.com/dmitrienko>

<http://support.sas.com/koch>

Index

A

Abelson, R.P. 135
ADJUST statement 133
adjusted analyses 2
advanced randomization-based methods
 about 67–68
 analysis of binary endpoints 78–79
 analysis of continuous endpoints using log-ratio
 of two means 80–81
 analysis of count endpoints using log-incidence
 density ratios 81
 analysis of ordinal endpoints using linear models
 74–78
 analysis of ordinal endpoints using proportional
 odds model 79–80
 analysis of time-to-event endpoints 82–86
 case studies 70–72
 nonparametric-based analysis of covariance 68–
 70
 %NParCov4 macro 73–74
Agresti, A. 20, 28, 35, 128
AIC (Akaike information criterion) 141
Akaike information criterion (AIC) 141
Allison, P.D. 4, 51
Alosh, M. 221–222
ALPHA option, %NParCov4 macro 95
ALR (alternating logistic regressions) 332
alternating logistic regressions (ALR) 332
analysis
 of binary endpoints 78–79
 of continuous endpoints using log-ratio of two
 means 80–81
 of count endpoints using log-incidence density
 ratios 81
 of incomplete data 319–378
 of ordinal endpoints using linear models 74–78
 of ordinal endpoints using proportional odds
 model 79–80
 of time-to-event endpoints 82–86
analysis of covariance, non-parametric
 randomization-based 88–89
ANCOVA models, Multiple Comparison-Modeling
 (MCP-Mod) procedure based on 155–159
Andersen, P.K. 42
Anderson, G.L. 293, 297, 301
Anderson, J.S. 293, 307
ANOVA models, Multiple Comparison-Modeling
 (MCP-Mod) procedure based on 145–155

ANOVA procedure 6
association, measures of 21
asymptotic model-based tests 35–38
asymptotic randomization-based tests 25–28

B

Baker, S.G. 321, 363
Bancroft, T.A. 15
BASELINE statement 53
Bates, D.M. 326
Bauer, P.J. 252
%BayesFutilityBin macro 310
%BayesFutilityCont macro 305–307
Bayesian CRM 110
Bayesian framework 118
Bayesian information criterion (BIC) 141
Bayesian two-parameter logistic models,
 implementation of 115–116
Beal, S.L. 372
Beckman, R.J. 366
Benjamini, Y. 199
%BetaModel macro 144
Betensky, R.A. 293, 297, 301
Beunckens, C. 338
BIC (Bayesian information criterion) 141
binary endpoints 78–79, 309–310
Birch, M.W. 29
Blackburn, P.R. 306–307
Bonferroni method 184–187
Bonferroni-based gatekeeping procedure 228–233
Bortey, E.B. 15
Brannath, W. 252, 313
Branson, M. 116, 138, 140
Brechenmacher, T. 238
Breslow, N.E. 26–27, 29, 44, 54, 327
Breslow-Day test 28, 61
Bretz, F. 138, 140, 141, 148, 181, 209, 221–222, 237
Buckland, S.T. 142
BY statement 361
Byar, D.P. 36

C

C option, %NParCov4 macro 94
calendar time 259
candidate models, as a step in Multiple Comparison-
 Modeling (MCP-Mod) procedure 138–
 139
Cantor, A. 4, 51
Carey, V.C. 332

- Carpenter, J.R. 349, 353, 354, 363, 374–375
- case studies
- advanced randomization-based methods 70–72
 - categorical endpoints 21–22
 - continuous endpoints 4–6, 17
 - dose-finding methods 128–132, 144–176
 - incomplete data, analysis of 322–324
 - POCRM 118–123
 - repeated significance tests 257–258
 - stochastic curtailment tests 293–294
- categorical endpoints
- analysis of 20–40
 - asymptotic model-based tests 35–38
 - asymptotic randomization-based tests 25–28
 - case studies 21–22
 - exact model-based tests 38–40
 - exact randomization-based tests 28–32
 - minimum risk tests 32–34
- CATMOD procedure, model-based inferences and 35
- CC (complete case analysis) 320
- %Chain macro 210
- chain procedure 208–211, 244–249
- Chakravorti, S.R. 15
- Chang, C.K. 188, 201
- Cheung, Y.K. 111, 114, 116, 119
- Chinchilli, V.M. 15
- Chuang-Stein, C. 128, 313
- Ciminera, J.L. 15, 28, 57, 61
- Claeskens, G. 141
- CLASS statement 37–38, 52, 357
- Clayton, D.G. 327
- clinical information 182
- closed family of hypotheses 189
- closure principle 189–191
- CMH (Cochran-Mantel-Haenszel) procedure 25–28
- Cochran, W.G. 2, 25
- Cochran-Armitage permutation test 29–30, 31
- Cochran-Mantel-Haenszel (CMH) procedure 25–28, 29
- Collett, D. 4, 44, 51
- COMBINE option, %NParCov4 macro 94–95
- complete case analysis (CC) 320
- Conaway, M.R. 107, 117
- conditional independence model 325
- conditional power 294–312
- continual reassessment method (CRM)
- about 107–108
 - implementation of 111–116
 - modeling frameworks 108–111
- continuous endpoints
- analysis of 4–20, 80–81
 - analysis of using log-ratio of two means 80–81
 - case studies 17
 - fixed effects models 6–15
 - nonparametric tests 16–19
 - random effects models 15–16
- CONTRAST statement 135–136
- contrast-based tests 134–136
- Cook, R.D. 365–367
- copy reference 377
- count endpoints, analysis of using log-incidence density ratios 81
- covariance, nonparametric-based analysis of 68–70
- COVARS option, %NParCov4 macro 94
- Cox, D.R. 4, 51, 54
- Cox proportional hazards model 2, 3
- %CriticalValue macro 148–149, 158, 163, 166
- CRM
- See* continual reassessment method (CRM)
- cumulative cohort method 104
- %Custom macro 238
- D**
- D'Agostino, R.B. 181, 206, 237
- Daniel, R. 353
- Danielson, L. 253
- data monitoring committee charter 265
- data-driven hypothesis ordering
- about 189
 - closure principle 189–191
 - procedures with 189–202
- Davis, B.R. 297, 303
- Davis, C.S. 35
- Day, N.E. 29, 36
- DeCani, J.S. 261
- decision matrix algorithm
- about 191
 - Holm procedure 191–193
 - power comparisons 201–202
 - stepwise procedures 193–198
 - weighted procedures 199–201
- DeMets, D.L. 253, 258, 264, 281, 289
- Dempster, A.P. 337
- DerSimonian, R. 321, 363
- DETAILS option, %NParCov4 macro 96
- Diggle P.J. 321, 332, 363–365
- direct Bayesian analysis (ignorable Bayesian analysis) 341–344
- DISCRETE method 54
- distributional information 182, 183–184
- DLT (dose-limiting toxicity) 102–103
- Dmitrienko, A. 128, 137, 181–182, 184, 189, 191, 206, 221–222, 236–238, 289, 293–294, 298, 305, 309
- dose-escalation methods
- about 101
 - continual reassessment method (CRM) 107–116
 - partial order continual reassessment method 116–123
 - rule-based methods 103–107
 - trials 102–103
- dose-finding methods
- about 127–128

- case studies 128–132, 144–176
- dose-response assessment and 132–144
- dose-limiting toxicity (DLT) 102–103
- dose-placebo tests 132–134
- dose-response analysis 127, 132–144
- double robustness 349
- %dropout macro 346
- %dropwgt macro 346
- DSNIN option, %NParCov4 macro 95
- DSNOUT option, %NParCov4 macro 95–96
- Dunne, A. 372
- Dunnett, C.W. 215
- Dunnett test 133–134, 140

E

- Edwards, S. 29
- efficiency robust tests 32
- Efron, B. 54
- Ellenberg, S.S. 253
- %Emax macro 143, 146
- Emerson, S.S. 275, 289
- endpoints
 - See also* categorical endpoints
 - See also* continuous endpoints
 - binary 78–79, 309–310
 - count 81
 - normally distributed 305–306, 338–339
 - ordinal 74–80
 - primary 180
 - secondary 180
 - time-to-event 41–55, 43–51, 51–55, 82–86, 91–92
- error spending functions 261–262
- ESTIMATE statement 10, 336
- estimation, following sequential testing 289–291
- EXACT method 54
- exact model-based tests 38–40
- EXACT option, %NParCov4 macro 95
- exact randomization-based tests 28–32
- EXACT statement 23
- EXACTONLY option, LOGISTIC procedure 38
- exit probabilities 267
- %Exponential macro 143, 146, 161–162
- EXPOSURES option, %NParCov4 macro 94

F

- F* statistic 14
- fallback procedure 206–208
- familywise error rate (FWER) 181–182
- Faries, D. 111
- Fisher's exact test for binary endpoints 2
- Fitzmaurice, G.M. 320
- fixed data monitoring strategies 258–259
- fixed effects models 6–15
- fixed-sequence procedure 203–206
- Fleiss, J.L. 14–15, 25–26, 28
- Fleming, T.R. 44, 253, 259, 275, 289

- flexible data monitoring strategies 258, 259–261
- Follman, D.A. 281
- frameworks
 - Bayesian 118
 - modeling 108–111
- Freedman, L.S. 306–307
- FREQ procedure
 - about 4
 - binary data and 29
 - obtaining test statistics using 27
 - odds ratio and 22–23
 - Output Delivery System (ODS) and 24–25
 - relative risk and 22–23
 - risk difference and 22–23
 - van Elteren statistics and 19
- Freund, J.L. 4, 7
- futility stopping rules 294
- FWDLINK statement 36
- FWER (familywise error rate) 181–182

G

- Gail, M.H. 2, 28, 57–61
- Gail-Simon test 57–61
- Gallo, P.P. 14, 15, 253
- Gastwirth, J.L. 32
- gatekeeping procedures
 - about 221–222
 - implementation of mixture-based procedures 227–236
 - inferences based on mixture method 223–236
 - modified mixture method 236–240
 - ordered families of hypotheses in clinical trials 222–223
- Gatsonis, C. 116
- GEE
 - See* generalized estimating equations (GEE)
- GEE procedure
 - data analysis 320
 - incomplete data analysis 341, 348–349, 357, 360, 361
 - LOCF and 338
- Gehan, E.A. 44
- Geisser, S. 293
- generalized estimating equations (GEE)
 - about 320, 327, 331–332
 - methodological detail 332–333
 - weighted (WGEE) 344–349
- generalized linear mixed models (GLMM) 325–330
- %GeneralizedOptimalContrasts macro 147, 155–156, 162, 166
- GENMOD procedure
 - about 4
 - covariance matrix and 156
 - data analysis 320
 - dose-placebo tests 133
 - incomplete data analysis 341, 346–347, 357, 360, 361

- interim data monitoring 300
- LOCF and 338
- model-based inferences and 35, 37–38
- Genz, A. 141, 148
- GLIMMIX procedure
 - data analysis 320
 - generalized linear mixed models 329–330
 - incomplete data analysis 333, 341, 350, 356, 357, 360, 361
 - LOCF and 338
 - parametric procedures 215
 - step-down Dunnett procedure 218–220
- GLM procedure
 - about 4, 6, 8–9
 - Bonferroni method 184
 - dose-placebo tests 133
 - dose-response tests 148, 157
 - ESTIMATE statement 10
 - parametric procedures 215
 - Šidák method 184
 - in Type II, II, and III analysis 13–15
 - Type III analysis and 12
- GLMM (generalized linear mixed models) 325–330
- global null hypothesis 181
- Glynn, R.J. 321
- Goldberg, J.D. 4
- Goodman, S.N. 111
- Gould, A.L. 4
- Govindarajulu, Z. 289
- Greenhouse, J.B. 116
- Greenland, S. 26–27
- Grizzle, J.E. 15
- group sequential designs
 - about 251, 254
 - comparing 262–263
 - for detecting futility 255
 - for detecting superior efficacy 254–255
 - for simultaneous efficacy and futility testing 255–256
- Gsponder, T. 116
- Guo, W. 105

H

- Hackney, O.P. 6
- Haenszel, W. 25, 26
- Hájek, J. 44
- Halperin, M. 292, 295, 296
- Hardy, R.J. 297, 303
- Harrington, D.P. 44, 332
- Harrington-Fleming test 44–45, 49, 51
- Hartley, H.O. 15
- Hartzel, J. 20
- Herson, J. 307, 310
- Hives Severity (HS) score 129–132
- Hjorth, N.L. 141
- Hochberg, Y. 181, 199
- Hochberg procedure 193, 197–198

- Hocking, R.R. 6, 11
- Hogan, J.W. 371–372
- Holm, S. 199
- Holm procedure 191–193
- Hommel, G. 195
- Hommel procedure 193, 195–197, 245
- Hommel-based gatekeeping procedure 233–236, 246–249
- Hosmane, B. 17
- HS (Hives Severity) score 129–132
- Hsu, J.C. 181
- Hunsberger, S.A. 313
- Hwang, I.k. 261
- HYPOTH option, %NParCov4 macro 95
- hypothesis testing 352

I

- ignorability 331
- ignorable likelihood (direct likelihood) 338–341
- implication relationships 189–190
- imputation
 - mechanisms for 353–356
 - simple methods of 337–338
 - as a step 354
 - strategies for 166–176
- incomplete data, analysis of
 - case studies 322–324
 - data setting and methodology 324–333
 - direct Bayesian analysis (ignorable Bayesian analysis) 341–344
 - ignorable likelihood (direct likelihood) 338–341
 - multiple imputation 349–362
 - sensitivity analysis 362–378
 - sensitivity analysis based on multiple imputation and pattern-mixture models 371–378
 - sensitivity analysis using local influence 363–371
 - simple methods and MCAR 334–338
 - weighted generalized estimating equations 344–349
- Inference Task 351
- information time 259
- initial step 354
- interim data monitoring
 - about 251–253
 - repeated significance tests 253–292
 - stochastic curtailment tests 293–315
- INVAR procedure 34
- INVAR test 32, 33
- inverse probability weighting (IPW) 349
- IPW (inverse probability weighting) 349
- Ivanova, A. 104

J

- Jansen, I. 321, 363
- Jennison, C. 252, 254, 260, 261, 286–287, 290, 297
- Ji, Y. 104–107

Johns, D. 293, 307
 Johnson, D.E. 4, 6, 7
 Johnson, W. 293
 Jones, B. 14, 15

K

Kalbfleisch, J.D. 4, 48, 51
 Kaplan-Meier estimate 42–43
 Kass, R.E. 142
 Kawaguchi, A. 68
 Kenward, M.G. 321, 334, 349, 353–354, 363–365, 367, 371–372, 374–375
 Kim, K. 289
 Klein, J.P. 42
 Koch, G.G. 17, 29, 35, 68–69
 Koury, K.J. 4

L

Lachin, J.M. 2, 4, 20, 35, 44
 Lagakos, S.W. 2
 Laird, N.M. 321, 332, 371–372
 Lan, K.K.G. 252, 253, 258, 264, 281, 292, 295–296
 Laplace method 329
 large-strata asymptotics 28–29
 last observation carried forward (LOCF) 320, 337–338
 LaVange, L. 253
 Lee, J.J. 293
 Lee, S.M. 111, 119
 Lehmann, E.L. 17
 Lesaffre, E. 321, 366
 Li, K.H. 352
 Liang, K.-Y. 331, 333, 344
 LIFEREG procedure 51
 LIFETEST procedure
 about 4
 inferences performed by 50–51
 Kaplan-Meier estimate and 42–43
 qualitative information from 53
 STRATA statement 45–46
 TEST statement 46–48
 warning messages and 39
 likelihood ratio test, randomization-based tests and 43–51
 likelihood-based approaches 330–331
 linear mixed models 324–325
 linear models, analysis of ordinal endpoints using 74–78
 linear rank tests 44
 %LinRank macro 49, 51
 Lipkovich, I. 166, 350, 355
 Lipsitz, S.R. 332
 Littell, R.C. 4, 7, 15
 Little, R.J.A. 320, 324, 331, 337, 349, 371
 Liu, D.D. 293
 Liu, W. 189

LOCF (last observation carried forward) 320
 Logan, B.R. 137
 logical restrictions 182, 183
 log-incidence density ratios 81, 90–91
 %Logistic macro 143–144
 LOGISTIC procedure
 about 4
 exact inferences and 38–40
 model-based inferences and 35, 37–38
 logistic regression models 3
 logistic transformation 89–90
 logit-adjusted estimate 26
 LogLinear model 149–156
 log-rank scores, for time-to-event outcomes 91–92
 log-rank test
 randomization-based tests and 43–51
 for time-to-event end points 2
 log-ratio of two means, analysis of continuous endpoints using 80–81
 log-ratio of two outcome vectors 90
 Lu, K. 375

M

Mallinckrodt, C.H. 166, 331, 350, 355
 Mantel, N. 25, 26, 29
 Mantel-Fleiss criterion 29
 Mantel-Haenszel test 3, 17, 26, 28, 44
 MAR (missing at random) 320
 marginal quasi-likelihood estimates (MQL) 327
 MCAR (missing completely at random) 320
 MCAR, simple methods and 334–338
 MCMC procedure 350, 354, 355
 MCP-Mod
 See Multiple Comparison-Modeling (MCP-Mod) procedure
 MED (minimum effective dose) 128, 141–142, 152–156
 median unbiased estimate 290
 Mehrotra, D.V. 28, 32, 33
 Mehta, C.R. 289–290
 Menon, S. 128
 MI
 See multiple imputation (MI)
 MI procedure
 incomplete data analysis 356–359
 NIMPUTE option 168–169
 MIANALYZE procedure 169, 350, 356, 358, 359, 362
 Milliken, G.A. 4, 6, 7, 15
 minimum effective dose (MED) 128, 141–142, 152–156
 "minimum regret" procedure 33
 minimum risk tests 32–34
 %MinRisk macro 33
 missing at random (MAR) 320
 missing completely at random (MCAR) 320
 missing not at random (MNAR) 168

- missing observations, Multiple Comparison-Modeling (MCP-Mod) procedure in presence of 165–176
 - MISSMODEL statement 348
 - MIXED procedure
 - about 4, 6
 - analysis of ordinal endpoints using linear models 76
 - Bonferroni method 184
 - contrast-based tests 135
 - data analysis 320
 - dose-placebo tests 133
 - dose-response tests 163, 172
 - incomplete data analysis 341, 350, 357, 360
 - LOCF and 338
 - parametric procedures 215
 - random effects models and 15–16
 - Šidák method 184
 - in Type II, II, and III analysis 13–15
 - Type III analysis and 12
 - %MixGate macro 227–229, 236–238
 - mixture method, general principles of gatekeeping inferences based on 223–226
 - mixture-based gatekeeping procedures, implementation of 227–236
 - MNAR (missing not at random) 168
 - MNAR statement 376
 - model selection, as a step in Multiple Comparison-Modeling (MCP-Mod) procedure 141
 - MODEL statement 53, 325, 330
 - model-based tests 1–4, 51–55
 - modeling frameworks 108–111
 - modified mixture method 236–240, 247–249
 - modified toxicity probability interval method (mTPI) 104–107
 - Molenberghs, G. 320–321, 325, 332, 334, 344, 349, 363, 367–368, 371–372
 - MONOTONE statement 353, 354, 357
 - Morris, D. 17
 - MQL (marginal quasi-likelihood estimates) 327
 - mTPI (modified toxicity probability interval method) 104–107
 - MTPs (multiple testing procedures) 180–184
 - multi-contrast tests 136–137
 - Multiple Comparison-Modeling (MCP-Mod) procedure
 - about 128, 137–138
 - based on ANCOVA models 155–159
 - based on repeated-measures models 161–165
 - based on simple ANOVA models 145–155
 - candidate models 138–139
 - dose-finding on 138
 - dose-response tests 140–141
 - implementation of 143–144
 - model selection 141
 - optimal contrasts 139–140
 - in presence of missing observations 165–176
 - target dose selection 141–142
 - multiple imputation (MI)
 - about 349–350
 - efficiency 352–353
 - hypothesis testing 352
 - imputation mechanisms 353–356
 - pooling information 351–352
 - SAS code for 358–362
 - SAS for 356–358
 - sensitivity analysis based on pattern-mixture models and 371–378
 - theoretical justification 350–351
 - multiple testing procedures (MTPs) 180–184
 - multiplicity adjustment methods
 - about 179–181
 - control of familywise error rate 181–182
 - gatekeeping procedures 221–241
 - multiple testing procedures 182–184
 - parametric procedures 212–221
 - procedures with data-driven hypothesis ordering 189–202
 - procedures with prespecified hypothesis ordering 202–212
 - single-step procedures 184–188
 - multivariate modeling 354
 - %MultSeqSimul macro 237
 - MULTTEST procedure
 - Bonferroni method 184, 193
 - Cochran-Armitage permutation and 39
 - Fisher test and 29–30
 - Hochberg procedure 197–198
 - Holm procedure 193
 - Hommel procedure 196
 - permutations and 31–32
 - resampling-based exact methods 93
 - Šidák method 184
 - Simes method 187
 - weighted procedures 200
- N**
- Nachtsheim, C.J. 366
 - Nelder, J.A. 14
 - Neuenschwander, B. 116
 - NIMPUTE option, MI procedure 168–169
 - NLMIXED procedure
 - data analysis 320
 - generalized linear mixed models 325, 329–330
 - incomplete data analysis 341, 350, 356, 357, 360, 361
 - LOCF and 338
 - model selection 149, 158–159, 164, 173–174
 - model-based inferences and 35
 - target dose selection 165
 - non-Gaussian outcomes 339–340
 - non-parametric randomization-based analysis of covariance 88–89
 - nonparametric tests 16–19

nonparametric-based analysis of covariance 68–70
 non-prognostic factors 1–2
 non-random (MNAR) 320
 non-responder imputation 166–167
 non-testable hypotheses 224
 normally distributed endpoints/outcomes 305–306,
 338–339
 %NParCov4 macro
 about 73–74
 content of output data sets 96–99
 general use and options for 94–96
 NREPS option, %NParCov4 macro 95

O

Oakes, D.O. 4, 51
 O'Brien, P.C. 259, 275
 O'Brien-Fleming stopping boundary 258, 263–270,
 275–276, 280–284, 287–288
 O'Connell, M. 327
 odds ratio 21
 ODS (Output Delivery System), FREQ procedure
 and 24–25
 ODS statement 358, 361
 Öhashi, Y. 42
 O'Kelly, M. 166, 167, 349
 one-parameter logistic model 108–110
 optimal contrasts, as a step in Multiple Comparison-
 Modeling (MCP-Mod) procedure 139–
 140
 O'Quigley, J. 107, 117
 ordered families of hypotheses in clinical trials 222–
 223
 ordinal endpoints
 analysis of using linear models 74–78
 analysis of using proportional odds model 79–80
 OUTCOMES option, %NParCov4 macro 94
 Output Delivery System (ODS), FREQ procedure
 and 24–25

P

Pampallona, S. 261
 PANSS (Positive and Negative Syndrome Scale)
 128–129
 parametric procedures
 about 212–214
 single-step Dunnett procedure 214–217
 stepwise Dunnett procedures 217–220
 partial likelihood 51
 partial order 116–117
 partial order continual reassessment method
 (POCRM)
 about 107–108, 116–117
 Bayesian framework 118
 in case studies 118–123
 partial order 117
 partitioning principle 191
 patient populations, multiple 180

patient's latent toxicity 119
 Patra, K. 313
 pattern-mixture imputation, with placebo imputation
 167–176
 pattern-mixture models (PMM) 320, 371–378
 penalized quasi-likelihood estimates (PQL) 327
 Pepe, M.S. 293, 297, 301
 permutation *t*-test for continuous endpoints 2
 Peto, J. 44
 Peto, R. 44
 PHREG procedure
 about 4, 41–42
 proportional hazards models and 51–53
 ties and 54
 Piantadosi, S. 2, 111
 Pinheiro, J.C. 138, 140, 155, 326
 PMM (pattern-mixture models) 320, 371–378
 PMM (predictive mean matching) 353
 Pocock, S.J. 259
 Pocock stopping boundary 258, 271–276, 280–281,
 284–285
 POCRM
 See partial order continual reassessment method
 (POCRM)
 %POCRM macro 120–121
 Poisson regression methods 3
 pooling information 351–352
 Positive and Negative Syndrome Scale (PANSS)
 128–129
 posterior step 354
 power comparisons 201–202
 power model 108–110
 PQL (penalized quasi-likelihood estimates) 327
 precision 2
 predictive mean matching (PMM) 353
 predictive power 294–312
 predictive probability, futility rules based on 304–
 309
 Prentice, R.L. 4, 44, 48, 51, 332
 prespecified hypothesis ordering
 about 202–203
 chain procedure 208–211
 fallback procedure 206–208
 fixed-sequence procedure 203–206
 procedures with 202–212
 pre-testing, Type III analysis with 14–15
 primary endpoints, multiple 180
 PROBIT procedure, model-based inferences and 35
 prognostic factors 1
 propensity scores 353
 proportional odds model 3, 79–80
 Proschan, M.A. 252, 281, 313
 pseudo-quasi-likelihood 327
 Pulkstenis, E. 236, 238, 294
 pushback test 61
 %PvalProc macro 200, 207
p-values 225–226

Q

Qaqish, B. 332
 %Quadratic macro 144
 qualitative interaction tests
 about 56–57, 61
 Gail-Simon test 57–61
 quantitative interactions 15

R

R() notation
 Type I analysis and 7–9
 Type II analysis and 9–11
 Rabe-Hesketh, S. 326
 Radhakrishna, S. 32, 33
 Raftery, A.E. 142
 Raghunathan, T.E. 352
 Railkar, R. 32, 33
 random effects models 15–16
 RANDOM statement 320, 330, 341
 randomization-based methods 1–4
 See also advanced randomization-based methods
 randomization-based tests 43–51
 Rao J.N.K. 4, 15
 Ratitch, B. 166, 167, 349
 %RegularOptimalContrasts macro 146–147
 relative risk 21
 REML (restricted maximum likelihood) estimation
 327
 repeated confidence intervals 286–288
 repeated significance tests
 about 253
 case studies 257–258
 design and monitoring stages 258
 design stage 263–280
 estimation following sequential testing 289–291
 fixed and flexible data monitoring strategies
 258–263
 group sequential trial designs 254–256
 monitoring stage 280–285
 relationship with conditional power tests 297–
 298
 repeated confidence intervals 286–288
 REPEATED statement 325, 330, 333, 347
 repeated-measures models, Multiple Comparison-
 Modeling (MCP-Mod) procedure based on
 161–165
 resampling-based exact methods 93
 responder imputation 167
 restricted maximum likelihood (REML) estimation
 327
 risk difference 21
 Robins, J.M. 26–27, 321, 344, 363
 Rodriguez, R. 14
 Roger, J.H. 363, 374
 Rosenberger, W.F. 321, 363
 Rosner, G.L. 289–290
 Rotnitzky, A. 321, 344, 363

Royston, P. 353
 Ruberg, S.J. 128, 137
 Rubin, D.B. 320–321, 324, 331, 337, 349, 352, 355,
 361
 Rüger, B. 187
 rule-based methods
 about 103
 modified toxicity probability interval method
 (mTPI) 104–107
 up-and-down methods 103–104

S

Sarkar, S.K. 188, 201
 Saville, B.R. 68
 Schafer, J.L. 349, 354
 Scharfstein, D.O. 321, 363
 Scheffe, H. 6, 135
 Schenker, N. 349
 Schoenfeld, D.A. 2
 Searle, S.R. 4, 7, 15
 secondary endpoints, multiple 180
 SEED option, %NParCov4 macro 95
 selection modeling 320
 Senn, S. 15, 16, 57, 237
 sensitivity analysis
 about 362–363
 based on multiple imputation and pattern-mixture
 models 371–378
 using local influence 363–371
 separable procedures 225
 SEQDESIGN procedure, interim data monitoring
 252–253, 258, 261, 264, 271, 277, 299
 %SeqPower macro 302–303
 SEQTEST procedure, interim data monitoring 252–
 253, 258, 261, 271, 281–285, 287, 290–
 292, 297, 298, 300–301
 sequential rejection principle 191
 sequential testing, estimation following 289–291
 Sheiner, L.B. 372
 Shih, W.J. 261
 Shu, V. 17
 Šidák, Z. 44
 Šidák method 184–187
 Siegmund, D. 289–290
 %SigEmax macro 144, 161–162
 Simes method 187–188
 Simon, R. 28, 57–61, 292, 295–296
 simple order 116
 single-contrast tests 136
 single-step Dunnett procedure 214–217
 single-step procedures
 about 184
 Bonferroni method 184–187
 Šidák method 184–187
 Simes method 187–188
 Skrondal, A. 326
 sparse-data asymptotics 28–29

Spector, P.C. 4, 7
 Speed, F.M. 6, 11
 Spiegelhalter, D.J. 306–307
 SQL procedure, dose-response tests 147–148, 157
 SSIZE procedure 33, 34
 standard mixture method, Hommel-based
 gatekeeping procedure using 246–247
 %StartUp macro 144, 145, 162
 statistical information 182
 STDERR statement 358
 step-down algorithm 193–195
 step-down Dunnett procedure 217–220
 step-up Dunnett procedure 220
 stepwise Dunnett procedures 217–220
 stepwise procedures 193–198
 Stern, H.S. 321
 Stewart, W.H. 137
 stochastic curtailment tests
 about 292–293
 applications of 312–313
 case studies 293–294
 futility rules based on conditional and predictive
 power 294–312
 Stokes, M.E. 35
 stopping probabilities 267
 Storer, B.E. 115–116
 STRATA option, %NParCov4 macro 94
 STRATA statement 39, 45–46, 52
 stratification 92
 stratified analysis, of time-to-event data 50–51
 stratified log-rank test 3
 strong control 182
 Stroup, W.W. 15
 SYMSIZE option, %NParCov4 macro 96

T

TABLE statement 19
 Tamhane, A.C. 137, 138, 181, 222
 Tan, W.Y. 2
 Tangen, C.M. 68, 69
 Tanner, M.A. 349
 target dose selection, as a step in Multiple
 Comparison-Modeling (MCP-Mod)
 procedure 141–142
 Tarone, R.E. 44, 45, 49
 Tarone-Ware test 44–45, 49, 51
 TEST statement, LIFETEST procedure 46–48, 358
 testable hypotheses 224
 Thijs, H. 321, 372
 ties, analysis of time-to-event data with 53–55
 time-to-event data
 analysis of with ties 53–55
 stratified analysis of 50–51
 time-to-event endpoints/outcomes
 analysis of 41–55, 82–86
 log-rank scores for 91–92
 model-based tests 51–55

 randomization-based tests 43–51
 Wilcoxon scores for 91–92
 Ting, N. 128
 tipping point analysis 375
 Tobias, R.D. 14
 TRANSFORM option, %NParCov4 macro 95
 TRANSFORM statement 357
 treatment comparisons, multiple 179–180
 TRTGRPS option, %NParCov4 macro 94
 truncated Hommel procedure 245
 truncated procedures 225
 Tsiatis, A.A. 261, 275, 289–290
t-statistic method 104
 Tuerlinckx, F. 326
 Tukey, J.W. 135, 137
 Turnbull, B.W. 252, 254, 260–261, 286–287, 290,
 297
 Type I analysis 7–9, 13–15
 Type II analysis 9–11, 13–15
 Type II error rate control 303
 Type III analysis
 about 11–13
 compared with Type I and Type III 13–15
 with pre-testing 14–15

U

unit probability mass (UPM) 105–107
 up-and-down methods 103–104
 UPM (unit probability mass) 105–107

V

van Buuren, S. 349
 Van Den Berghe, G. 289
 van Elteren, P.H. 16–17
 Van Elteren test 3
 Van Steen, K. 321
 Vansteelandt, S. 349
 Vehovar, V. 321
 Verbeke, G. 321, 325, 332, 334, 366

W

Waclawiw, M.A. 281
 Wages, N.A. 107, 117
 Wald chi-square statistic 36
 Wang, M.D. 293, 298, 309, 313
 Wang, S.J. 105
 Wang, S.K. 275
 Wang, Y. 313
 Ware, J. 44–45, 49
 Wassmer, G. 252, 313
 Wedderburn, R.W.M. 331
 weighted generalized estimating equations (WGEE)
 344–349
 weighted procedures 199–201
 Westfall, P.H. 29, 181, 184, 192
 White, I.R. 353
 Whitehead, J. 253, 289

394 *Index*

Wieand, S. 2
Wilcoxon rank sum test 2, 16–17, 43–51
Wilcoxon scores, for time-to-event outcomes 91–92
Wittes, J. 252, 297
Wolfinger, R.D. 14, 15, 327
Wong, W.H. 349
working correlation matrix 332

Y

Yamaguchi, T. 42
Young, S.S. 184, 192

Z

Zeger, S.L. 331–333, 344
Zhang, M. 42
Zhao, L.P. 344
Zink, R. 128

Ready to take your SAS[®] and JMP[®] skills up a notch?



Be among the first to know about new books,
special events, and exclusive discounts.

support.sas.com/newbooks

Share your expertise. Write a book with SAS.

support.sas.com/publish

 sas.com/books
for additional books and resources.


THE POWER TO KNOW.®

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies. © 2017 SAS Institute Inc. All rights reserved. M1588358 US.0217