Contents

Preface xvii

1 Introduction and Examples 1

1.1 Health-Related Quality of Life (HRQoL) 1

1.2 Measuring Health-Related Quality of Life 3

1.2.1 Health Status Measures 3

1.2.2 Patient Preference Measures 4

1.2.3 Objective versus Subjective Questions 5

1.2.4 Generic versus Disease-Specific Instruments 5

1.2.5 Global Index versus Profile of Domain-Specific Measures 6

1.2.6 Response Format 6

1.2.7 Period of Recall 7

1.3 Study 1: Adjuvant Breast Cancer Trial 8

1.3.1 Patient Selection and Treatment 9

1.3.2 Quality of Life Measure and Scoring 9

1.3.3 Timing of HRQoL Assessments 11

1.3.4 Questionnaire Completion/Missing Data 12

1.4 Study 2: Migraine Prevention Trial 12

1.4.1 Patient Selection and Treatment 13

1.4.2 Quality of Life Measure and Scoring 14

1.4.3 Timing of HRQoL Assessments 15

1.4.4 Questionnaire Completion/Missing Data 15

1.5 Study 3: Advanced Lung Cancer Trial 16

1.5.1 Treatment 16

1.5.2 Quality of Life Measure and Scoring 17

1.5.3 Timing of Assessments 18

1.5.4 Questionnaire Completion/Missing Data 19

1.6 Study 4: Renal Cell Carcinoma Trial 19

1.6.1 Patient Selection and Treatment 20

1.6.2 Quality of Life Measures and Scoring 20

1.6.3 Timing of HRQoL Assessments 21

1.6.4 Questionnaire Completion/Missing Data 22

1.7 Study 5: Chemoradiation (CXRT) Trial 23

1.7.1 Patient Selection and Treatment 23

1.7.2 Patient Reported Outcomes 24

1.7.3 Timing and Frequency of HRQoL Assessments 25
1.8 Study 6: Osteoarthritis Trial 25
1.8.1 Patient Selection and Treatment 25
1.8.2 Timing and Frequency of Assessments 25
1.9 Summary 27

2 Study Design and Protocol Development **29**
2.1 Introduction 29
2.1.1 Purpose of the Protocol 29
2.2 Background and Rationale 31
2.3 Research Objectives and Goals 31
2.3.1 Role of HRQoL in the Trial 33
2.3.2 Pragmatic versus Explanatory Inference 33
2.4 Selection of Subjects 34
2.5 Longitudinal Designs 35
2.5.1 Event- or Condition-Driven Designs 35
2.5.2 Time-Driven Designs 36
2.5.3 Timing of the Initial HRQoL Assessment 36
2.5.4 Timing of the Follow-Up HRQoL Assessments 36
2.5.5 Frequency of Evaluations 37
2.5.6 Duration of HRQoL Assessments 38
2.5.7 Assessment after Discontinuation of Therapy 38
2.6 Selection of Measurement Instrument(s) 39
2.6.1 Trial Objectives 40
2.6.2 Validity and Reliability 41
2.6.3 Appropriateness 42
2.7 Conduct of HRQoL Assessments 43
2.7.1 Order and Place of Administration 43
2.7.2 Mode of Administration and Assistance by Third Parties 44
2.7.3 Data Collection and Key Personnel 44
2.7.4 Avoiding Missing Data 45
2.8 Scoring Instruments 48
2.8.1 Reverse Coding 48
2.8.2 Scoring Multi-Item Scales 48
2.8.3 Item nonresponse 49
2.9 Summary 51

3 Models for Longitudinal Studies I **53**
3.1 Introduction 53
3.1.1 Repeated Measures Models 53
3.1.2 Growth Curve Models 54
3.1.3 Selection between Models 54
3.2 Building Models for Longitudinal Studies 56
3.2.1 Advantages of the General Linear Model (GLM) . . . 56

3.2.2 Building a General Linear Model 57
3.2.3 Statistics Guiding Model Reduction 57
3.3 Building Repeated Measures Models: The Mean Structure . . 58
3.3.1 Treatment and Time 58
3.3.2 Common Baseline 61
3.3.3 Change from Baseline 61
3.3.4 Covariates 61
3.3.5 Modeling the "Mean" Structure in SAS 62
3.3.6 Modeling the "Mean" Structure in SPSS 64
3.3.7 Modeling the "Mean" Structure in R 66
3.4 Building Repeated Measures Models: The Covariance Structure 67
3.4.1 Unstructured Covariance 68
3.4.2 Structured Covariance 69
3.4.3 Building the Covariance Structure in SAS 70
3.4.4 Building the Covariance Structure in SPSS 72
3.4.5 Building the Covariance Structure in R 72
3.5 Estimation and Hypothesis Testing 74
3.5.1 Estimation and Hypothesis Testing in SAS 75
3.5.2 Estimation and Hypothesis Testing in SPSS 77
3.5.3 Estimation and Hypothesis Testing in R 78
3.6 Summary 81

4 Models for Longitudinal Studies II **83**
4.1 Introduction 83
4.2 Building Growth Curve Models: The "Mean" (Fixed Effects) Structure 84
4.2.1 Polynomial Models 84
4.2.2 Piecewise Linear Regression 85
4.2.3 Modeling the "Mean" Structure in SAS 87
4.2.4 Modeling the "Mean" Structure in SPSS 88
4.2.5 Modeling the "Mean" Structure in R 88
4.3 Building Growth Curve Models: The Covariance Structure . . 89
4.3.1 Variance of Random Effects ($\mathcal{G}$) 90
4.3.2 Variance of Residual Errors ($\mathcal{R}_i$) 91
4.3.3 Building the Covariance Structure in SAS 93
4.3.4 Building the Covariance Structure in SPSS 94
4.3.5 Building the Covariance Structure in R 95
4.4 Model Reduction 95
4.5 Hypothesis Testing and Estimation 96
4.5.1 Estimation and Testing in SAS 99
4.5.2 Estimation and Testing in SPSS 99
4.5.3 Estimation and Testing in R 100
4.6 An Alternative Covariance Structure 101
4.6.1 Model for the Means 102

4.6.2 Model for the Variance . . . 102
4.6.3 Estimation and Testing . . . 103
4.7 Summary . . . 104

5 Moderation and Mediation **105**
5.1 Introduction . . . 105
5.2 Moderation . . . 107
5.2.1 Moderation across Repeated Measures . . . 107
5.2.2 Change from Baseline . . . 112
5.2.3 Centering Covariates . . . 113
5.3 Mediation . . . 115
5.3.1 Mediation with Treatment as the Predictor . . . 116
5.3.2 Mediation with Time as the Predictor . . . 118
5.4 Other Exploratory Analyses . . . 120
5.4.1 Mediation in Mixed Effects Models . . . 120
5.4.2 Non-Linear Relationships . . . 121
5.5 Summary . . . 123

6 Characterization of Missing Data **125**
6.1 Introduction . . . 125
6.1.1 Terminology . . . 126
6.1.2 Why are Missing Data a Problem? . . . 126
6.1.3 How Much Data Can Be Missing? . . . 126
6.1.4 Prevention . . . 127
6.2 Patterns and Causes of Missing Data . . . 128
6.3 Mechanisms of Missing Data . . . 130
6.3.1 The Concept . . . 130
6.3.2 Notation . . . 131
6.4 Missing Completely at Random (MCAR) . . . 132
6.4.1 The Concept . . . 132
6.4.2 Covariate-Dependent Dropout . . . 133
6.4.3 Identifying Covariate-Dependent Missingness . . . 133
6.4.4 Analytic Methods . . . 134
6.5 MAR: Missing at Random . . . 135
6.5.1 The Concept . . . 135
6.5.2 Identifying Dependence on Observed Data (Y_i^{obs}) . . . 135
6.5.3 Analytic Methods . . . 139
6.6 MNAR: Missing Not at Random . . . 139
6.6.1 The Concept . . . 139
6.6.2 Identifying Dependence on Unobserved Data (Y_i^{mis}) . . . 140
6.6.3 Analytic Methods . . . 143
6.7 Example for Trial with Variation in Timing of Assessments . . . 143
6.8 Example with Different Patterns across Treatment Arms . . . 145
6.9 Summary . . . 146

7 Analysis of Studies with Missing Data **149**
7.1 Introduction . . . 149
7.2 Missing Completely at Random . . . 149
7.2.1 Complete Case Analysis (MANOVA) . . . 151
7.2.2 Repeated Univariate Analyses . . . 151
7.3 Ignorable Missing Data . . . 154
7.3.1 Maximum Likelihood Estimation (MLE) . . . 155
7.3.2 Empirical Bayes Estimates . . . 156
7.3.3 Multiple Imputation . . . 157
7.3.4 Lung Cancer Trial (Study 3) . . . 157
7.3.5 Baseline Assessment as a Covariate . . . 157
7.3.6 Adding Other Baseline Covariates . . . 158
7.3.7 Final Comments . . . 159
7.4 Non-Ignorable Missing Data . . . 160
7.4.1 Selection Models . . . 160
7.4.2 Mixture Models . . . 161
7.4.3 Final Comments . . . 161
7.5 Summary . . . 162

8 Simple Imputation **163**
8.1 Introduction to Imputation . . . 163
8.1.1 Simple versus Multiple Imputation . . . 164
8.1.2 Imputation in Multivariate Longitudinal Studies . . . 164
8.2 Missing Items in a Multi-Item Questionnaire . . . 165
8.3 Regression Based Methods . . . 167
8.3.1 Mean Value Substitution . . . 167
8.3.2 Explicit Regression Models . . . 168
8.4 Other Simple Imputation Methods . . . 172
8.4.1 Last Value Carried Forward (LVCF) . . . 172
8.4.2 δ-Adjustments . . . 174
8.4.3 Arbitrary High or Low Value . . . 174
8.4.4 Hot Deck and Other Sampling Procedures . . . 175
8.5 Imputing Missing Covariates . . . 176
8.6 Underestimation of Variance . . . 176
8.7 Final Comments . . . 178
8.7.1 Sensitivity Analysis . . . 178
8.8 Summary . . . 179

9 Multiple Imputation **181**
9.1 Introduction . . . 181
9.2 Overview of Multiple Imputation . . . 181
9.3 Explicit Univariate Regression . . . 183
9.3.1 Identification of the Imputation Model . . . 183
9.3.2 Computation of Imputed Values . . . 184
9.3.3 Practical Considerations . . . 185

9.3.4 Extensions to Longitudinal Studies 186
9.3.5 Extensions to Multiple HRQoL Measures 186
9.3.6 Assumptions 187
9.3.7 Lung Cancer Trial (Study 3) 187
9.3.8 Implementation 189
9.4 Closest Neighbor and Predictive Mean Matching 192
9.4.1 Closest Neighbor 192
9.4.2 Predictive Mean Matching 192
9.5 Approximate Bayesian Bootstrap (ABB) 194
9.5.1 Practical Considerations 194
9.5.2 The Assumptions 196
9.5.3 A Variation for Non-Ignorable Missing Data 196
9.6 Multivariate Procedures for Non-Monotone Missing Data . . 196
9.6.1 Implementation in SAS 197
9.6.2 Implementation in SPSS 197
9.6.3 Implementation in R 199
9.7 Analysis of the M Datasets 200
9.7.1 Univariate Estimates and Statistics 200
9.7.2 Multivariate Tests 202
9.7.3 Analysis of M Datasets in SAS 202
9.7.4 Analysis of M Datasets in SPSS 203
9.7.5 Analysis of M Datasets in R 204
9.8 Miscellaneous Issues 205
9.8.1 Sensitivity Analyses 205
9.8.2 Imputation after Death 207
9.8.3 Imputation versus Analytic Models 207
9.8.4 Implications for Design 207
9.9 Summary 208

10 Pattern Mixture and Other Mixture Models **209**
10.1 Introduction 209
10.1.1 General Approach 209
10.1.2 Illustration 210
10.2 Pattern Mixture Models 213
10.2.1 Specifying Patterns 213
10.3 Restrictions for Growth Curve Models 214
10.3.1 Linear Trajectories over Time 215
10.3.2 Estimation of the Parameters 217
10.3.3 Nonlinear Trajectories over Time 219
10.3.4 Implementation in SAS 220
10.3.5 Implementation in R 223
10.4 Restrictions for Repeated Measures Models 226
10.4.1 Bivariate Data (Two Repeated Measures) 226
10.4.2 Monotone Dropout 231
10.5 Standard Errors for Mixture Models 235

10.5.1 Delta Method 236
10.5.2 Bootstrap Methods 237
10.6 Summary 238

11 Random Effects Dependent Dropout **239**
11.1 Introduction 239
11.2 Conditional Linear Model 241
11.2.1 Assumptions 242
11.2.2 Testing MAR versus MNAR under the Assumptions of Conditional Linear Model 243
11.2.3 Lung Cancer Trial (Study 3) 243
11.2.4 Estimation of the Standard Errors 248
11.3 Varying Coefficient Models 249
11.3.1 Assumptions 250
11.3.2 Application 251
11.3.3 General Application 252
11.4 Joint Models with Shared Parameters 253
11.4.1 Joint versus Conditional Linear Model 255
11.4.2 Testing MAR versus MNAR under the Assumptions of the Joint Model 255
11.4.3 Alternative Parameterizations 255
11.4.4 Implementation 256
11.4.5 Implementation in SAS 257
11.4.6 Implementation in R 262
11.4.7 Multiple Causes of Dropout 263
11.4.8 Other Model Extensions 265
11.5 Summary 266

12 Selection Models **267**
12.1 Introduction 267
12.2 Outcome Selection Model for Monotone Dropout 268
12.2.1 Lung Cancer Trial (Study 3) 270
12.3 Summary 274

13 Multiple Endpoints **275**
13.1 Introduction 275
13.1.1 Aims and Goals/Role of HRQoL 276
13.1.2 Other Critical Information 276
13.2 General Strategies for Multiple Endpoints 277
13.2.1 Limiting the Number of Confirmatory Tests 278
13.2.2 Summary Measures and Statistics 279
13.2.3 Multiple Testing Procedures 279
13.3 Background Concepts and Definitions 280
13.3.1 Univariate versus Multivariate Test Statistics 280
13.3.2 Global Tests 281

13.3.3 Error Rates 281
13.4 Single Step Procedures 282
13.4.1 Global Tests Based on Multivariate Test Statistics . . 282
13.4.2 Equally Weighted Univariate Statistics 283
13.4.3 Importance Weighting/ Spending α 284
13.5 Sequentially Rejective Methods 285
13.5.1 Equal Weighting of Endpoints 285
13.5.2 Implementation in SAS 286
13.5.3 Implementation in SPSS 286
13.5.4 Implementation in R 287
13.5.5 Importance Weighting 287
13.6 Closed Testing and Gatekeeper Procedures 287
13.6.1 Closed Testing Based on a Bonferroni Correction . . . 288
13.6.2 Sequential Families 289
13.6.3 Shortcuts for Closed Testing Procedures 291
13.6.4 A Closed Testing Procedure Based on a Multivariate Test 293
13.6.5 Summary and Composite Measures 294
13.7 Summary 294

14 Composite Endpoints and Summary Measures 295
14.1 Introduction 295
14.1.1 Composite Endpoints versus Summary Measures . . . 295
14.1.2 Strengths and Weaknesses 297
14.2 Choosing a Composite Endpoint or Summary Measure 298
14.3 Summarizing across HRQoL Domains or Subscales 299
14.3.1 Weighting Proportional to the Number of Questions . 301
14.3.2 Factor Analytic Weights 301
14.3.3 Patient Weights 303
14.3.4 Statistically Derived Weights: Inverse Correlation . . . 303
14.4 Summary Measures across Time 305
14.4.1 Notation 306
14.4.2 Simple Linear Functions 306
14.4.3 Area Under the Curve (AUC) 309
14.5 Composite Endpoints across Time 311
14.5.1 Notation 312
14.5.2 Missing Data 313
14.5.3 Average Rate of Change (Slopes) 314
14.5.4 Area Under the Curve (AUC) 315
14.5.5 Average of Ranks 318
14.5.6 Analysis of Composite Endpoints 319
14.6 Summary 320

15 Quality Adjusted Life-Years (QALYs) and Q-TWiST 323
15.1 Introduction . . . 323
15.2 QALYs . . . 323
15.2.1 Estimation of $QALY_i$ . . . 324
15.2.2 Analysis of $QALY_i$ . . . 327
15.3 Q-TWiST . . . 328
15.3.1 Kaplan-Meier Estimates of Time in Health States . . . 329
15.3.2 Case 1: Known Estimates of U_{TOX} and U_{REL} . . . 331
15.3.3 Case 2: Two Unknown Weights That Are Equal across Treatments . . . 332
15.3.4 Proportional Hazards Estimates of Time in Health States . . . 334
15.4 Summary . . . 335

16 Analysis Plans and Reporting Results 337
16.1 Introduction . . . 337
16.2 General Analysis Plan . . . 338
16.2.1 Goals of the Trial and Role of HRQoL . . . 338
16.2.2 Primary versus Secondary Endpoints . . . 338
16.2.3 Composite Endpoints or Summary Measures . . . 339
16.2.4 Comparison Procedures . . . 339
16.2.5 Who Is Included? . . . 340
16.2.6 Models for Longitudinal Data . . . 341
16.2.7 Missing Data . . . 341
16.3 Sample Size and Power . . . 342
16.3.1 Simple Linear Combinations of β . . . 344
16.3.2 Longitudinal Studies with Repeated Measures . . . 345
16.3.3 Longitudinal Data and Growth Curve Model . . . 350
16.3.4 Other Considerations . . . 352
16.4 Reporting Results . . . 356
16.5 Summary . . . 356

Appendix C: Cubic Smoothing Splines 357

Appendix P: PAWS/SPSS Notes 359

Appendix R: R Notes 363

Appendix S: SAS Notes 369

References 377

Index 395