

Contents

Sections marked with an asterisk refer to some more challenging sections of the book.

Preface	xiii
Abbreviations	xvii
1 Basic ideas in clinical trial design	1
1.1 Historical perspective	1
1.2 Control groups	2
1.3 Placebos and blinding	3
1.4 Randomisation	4
1.4.1 Unrestricted randomisation	5
1.4.2 Block randomisation	5
1.4.3 Unequal randomisation	6
1.4.4 Stratified randomisation	7
1.4.5 Central randomisation	8
1.4.6 Dynamic allocation and minimisation	9
1.4.7 Cluster randomisation	10
1.5 Bias and precision	11
1.6 Between- and within-patient designs	12
1.7 Cross-over trials	14
1.8 Signal and noise	15
1.8.1 Signal	15
1.8.2 Noise	15
1.8.3 Signal-to-noise ratio	15
1.9 Confirmatory and exploratory trials	16
1.10 Superiority, equivalence and non-inferiority trials	17
1.11 Data types	18
1.12 Choice of endpoint	20
1.12.1 Primary variables	20
1.12.2 Secondary variables	21
1.12.3 Surrogate variables	21
1.12.4 Global assessment variables	22
1.12.5 Composite variables	23
1.12.6 Categorisation	23

2	Sampling and inferential statistics	25
2.1	Sample and population	25
2.2	Sample statistics and population parameters	26
2.2.1	Sample and population distribution	26
2.2.2	Median and mean	27
2.2.3	Standard deviation	28
2.2.4	Notation	29
2.3	The normal distribution	29
2.4	Sampling and the standard error of the mean	32
2.5	Standard errors more generally	35
2.5.1	The standard error for the difference between two means	35
2.5.2	Standard errors for proportions	38
2.5.3	The general setting	38
3	Confidence intervals and p-values	39
3.1	Confidence intervals for a single mean	39
3.1.1	The 95 per cent confidence interval	39
3.1.2	Changing the confidence coefficient	41
3.1.3	Changing the multiplying constant	41
3.1.4	The role of the standard error	43
3.2	Confidence intervals for other parameters	44
3.2.1	Difference between two means	44
3.2.2	Confidence intervals for proportions	45
3.2.3	General case	46
3.3	Hypothesis testing	47
3.3.1	Interpreting the p -value	47
3.3.2	Calculating the p -value	49
3.3.3	A common process	52
3.3.4	The language of statistical significance	55
3.3.5	One-tailed and two-tailed tests	55
4	Tests for simple treatment comparisons	57
4.1	The unpaired t -test	57
4.2	The paired t -test	58
4.3	Interpreting the t -tests	61
4.4	The chi-square test for binary data	63
4.4.1	Pearson chi-square	63
4.4.2	The link to a signal-to-noise ratio	66
4.5	Measures of treatment benefit	67
4.5.1	Odds ratio (OR)	67
4.5.2	Relative risk (RR)	68
4.5.3	Relative risk reduction (RRR)	69
4.5.4	Number needed to treat (NNT)	69
4.5.5	Confidence intervals	70
4.5.6	Interpretation	71
4.6	Fisher's exact test	71
4.7	The chi-square test for categorical and ordinal data	73
4.7.1	Categorical data	73

4.7.2	Ordered categorical (ordinal) data	75
4.7.3	Measures of treatment benefit for categorical and ordinal data	76
4.8	Extensions for multiple treatment groups	77
4.8.1	Between-patient designs and continuous data	77
4.8.2	Within-patient designs and continuous data	78
4.8.3	Binary, categorical and ordinal data	79
4.8.4	Dose ranging studies	79
4.8.5	Further discussion	80
5	Multi-centre trials	81
5.1	Rationale for multi-centre trials	81
5.2	Comparing treatments for continuous data	82
5.3	Evaluating homogeneity of treatment effect	84
5.3.1	Treatment-by-centre interactions	84
5.3.2	Quantitative and qualitative interactions	87
5.4	Methods for binary, categorical and ordinal data	88
5.5	Combining centres	88
6	Adjusted analyses and analysis of covariance	91
6.1	Adjusting for baseline factors	91
6.2	Simple linear regression	92
*6.3	Multiple regression	94
6.4	Logistic regression	96
6.5	Analysis of covariance for continuous data	97
6.5.1	Main effect of treatment	97
6.5.2	Treatment-by-covariate interactions	99
*6.5.3	A single model	101
6.5.4	Connection with adjusted analyses	102
6.5.5	Advantages of analysis of covariance	102
6.6	Binary, categorical and ordinal data	104
6.7	Regulatory aspects of the use of covariates	106
*6.8	Connection between ANOVA and ANCOVA	109
6.9	Baseline testing	109
7	Intention-to-treat and analysis sets	111
7.1	The principle of intention-to-treat	111
7.2	The practice of intention-to-treat	115
7.2.1	Full analysis set	115
7.2.2	Per-protocol set	117
7.2.3	Sensitivity	117
7.3	Missing data	118
7.3.1	Introduction	118
7.3.2	Complete cases analysis	119
7.3.3	Last observation carried forward (LOCF)	119
7.3.4	Success/failure classification	120
7.3.5	Worst case/best case imputation	120
7.3.6	Sensitivity	121
7.3.7	Avoidance of missing data	121
7.4	Intention-to-treat and time-to-event data	122
7.5	General questions and considerations	124

8	Power and sample size	127
8.1	Type I and type II errors	127
8.2	Power	128
8.3	Calculating sample size	131
8.4	Impact of changing the parameters	134
8.4.1	Standard deviation	134
8.4.2	Event rate in the control group	135
8.4.3	Clinically relevant difference	135
8.5	Regulatory aspects	136
8.5.1	Power > 80 per cent	136
8.5.2	Powering on the per-protocol set	137
8.5.3	Sample size adjustment	137
8.6	Reporting the sample size calculation	138
9	Statistical significance and clinical importance	141
9.1	Link between p -values and confidence intervals	141
9.2	Confidence intervals for clinical importance	143
9.3	Misinterpretation of the p -value	144
9.3.1	Conclusions of similarity	144
9.3.2	The problem with 0.05	145
10	Multiple testing	147
10.1	Inflation of the type I error	147
10.2	How does multiplicity arise	148
10.3	Regulatory view	148
10.4	Multiple primary endpoints	149
10.4.1	Avoiding adjustment	149
10.4.2	Significance needed on all endpoints	149
10.4.3	Composite endpoints	150
10.4.4	Variables ranked according to clinical importance	150
10.5	Methods for adjustment	152
10.6	Multiple comparisons	153
10.7	Repeated evaluation over time	154
10.8	Subgroup testing	155
10.9	Other areas for multiplicity	157
10.9.1	Using different statistical tests	157
10.9.2	Different analysis sets	158
11	Non-parametric and related methods	159
11.1	Assumptions underlying the t-tests and their extensions	159
11.2	Homogeneity of variance	160
11.3	The assumption of normality	160
11.4	Transformations	163
11.5	Non-parametric tests	166
11.5.1	The Mann–Whitney U-test	166
11.5.2	The Wilcoxon signed rank test	168
11.5.3	General comments	169
11.6	Advantages and disadvantages of non-parametric methods	169
11.7	Outliers	170

12	Equivalence and non-inferiority	173
12.1	Demonstrating similarity	173
12.2	Confidence intervals for equivalence	175
12.3	Confidence intervals for non-inferiority	176
12.4	A p -value approach	178
12.5	Assay sensitivity	180
12.6	Analysis sets	182
12.7	The choice of Δ	182
12.7.1	Bioequivalence	183
12.7.2	Therapeutic equivalence	183
12.7.3	Non-inferiority	184
12.7.4	The 10 per cent rule for cure rates	185
12.7.5	Biocrep and constancy	186
12.8	Sample size calculations	187
12.9	Switching between non-inferiority and superiority	189
13	The analysis of survival data	193
13.1	Time-to-event data and censoring	193
13.2	Kaplan–Meier (KM) curves	195
13.2.1	Plotting KM curves	195
13.2.2	Event rates and relative risk	196
13.2.3	Median event times	196
13.3	Treatment comparisons	197
13.4	The hazard ratio	200
13.4.1	The hazard rate	200
13.4.2	Constant hazard ratio	201
13.4.3	Non-constant hazard ratio	201
13.4.4	Link to survival curves	202
*13.4.5	Calculating KM curves	203
*13.5	Adjusted analyses	204
13.5.1	Stratified methods	204
13.5.2	Proportional hazards regression	204
13.5.3	Accelerated failure time model	207
13.6	Independent censoring	208
13.7	Sample size calculations	209
14	Interim analysis and data monitoring committees	213
14.1	Stopping rules for interim analysis	213
14.2	Stopping for efficacy and futility	214
14.2.1	Efficacy	214
14.2.2	Futility and conditional power	215
14.2.3	Some practical issues	216
14.2.4	Analyses following completion of recruitment	217
14.3	Monitoring safety	218
14.4	Data Monitoring Committees	219
14.4.1	Introduction and responsibilities	219
14.4.2	Structure	220
14.4.3	Meetings and recommendations	222

14.5	Adaptive designs	223
14.5.1	Sample size re-evaluation	223
14.5.2	Flexible designs	224
15	Meta-analysis	229
15.1	Definition	229
15.2	Objectives	231
15.3	Statistical methodology	232
15.3.1	Methods for combination	232
15.3.2	Confidence Intervals	233
15.3.3	Fixed and random effects	234
15.3.4	Graphical methods	234
15.3.5	Detecting heterogeneity	236
15.3.6	Robustness	236
15.4	Ensuring scientific validity	237
15.4.1	Planning	237
15.4.2	Publication bias and funnel plots	238
15.5	Meta-analysis in a regulatory setting	240
15.5.1	Retrospective analyses	240
15.5.2	One pivotal study	241
16	The role of statistics and statisticians	245
16.1	The importance of statistical thinking at the design stage	245
16.2	Regulatory guidelines	247
16.3	The statistics process	249
16.3.1	The Statistical Methods section of the protocol	250
16.3.2	The statistical analysis plan	250
16.3.3	The data validation plan	251
16.3.4	The blind review	251
16.3.5	Statistical analysis	252
16.3.6	Reporting the analysis	252
16.3.7	Pre-planning	253
16.3.8	Sensitivity and robustness	255
16.4	The regulatory submission	256
16.5	Publications and presentations	257
	References	261
	Index	267