Detailed contents

Symbols *xix*

1 Research design **1**

Introduction *2*
Introduction to research *3*
Research questions *4*
Interventional studies *6*
Randomized controlled trials *8*
Randomization in RCTs *10*
Patient consent in research studies *12*
Blinding in RCTs *14*
RCTs: parallel groups and crossover designs *16*
Zelen randomized consent design *18*
Superiority and equivalence trials *20*
Intention to treat analysis *22*
Case–control studies *24*
Cohort studies *28*
Cross-sectional studies *32*
Case study and series *34*
Deducing causal effects *36*
Designing an audit *38*
Data collection in audit *40*
Research versus audit *41*
Data collection: sources of data *42*
Data collection: outcomes *44*
Outcomes: continuous and categorical *46*
Collecting additional data *50*
The study protocol *52*
Sampling strategies *54*
Choosing a sample size *56*
Sample size for estimation studies: means *58*
Sample size for estimation studies: proportions *60*
Sample size for comparative studies *62*

Sample size for comparative studies: means *64*
Sample size for comparative studies: proportions *66*
Sample size calculations: further issues *68*
Sample size: other issues *70*
Using a statistical program to do the calculations *72*

2 Collecting data **75**
Data collection forms *76*
Form filling and coding *78*
Examples of questions with possible coding *80*
Data quality *81*
Questions and questionnaires *82*
Designing good questions *84*
Sensitive topics *86*
Designing questionnaires *88*
Example of a validated questionnaire *90*
Designing a new measurement tool: psychometrics *92*
Measuring reliability *94*
Questionnaire measurement scales *96*
Visual analogue scales *98*

3 Handling data: what steps are important **101**
Data entry *102*
Forms that can be automatically scanned for data entry *106*
Variable names and labels *108*
Joining datasets *110*
Joining datasets: examples *112*
Storing and transporting data *114*
Data checking and errors *116*
Data checking: examples *118*
Formal data monitoring *122*
Statistical issues in data monitoring *124*

4 Presenting research findings **127**
Communicating statistics *128*
Producing journal articles *130*
Research articles: abstracts *132*
Research articles: introduction and methods sections *134*

Research articles: results section *136*
Research articles: discussion section *138*
Presenting statistics: managing computer output *140*
Presenting statistics: numerical results *142*
Presenting statistics: P values and confidence intervals *144*
Presenting statistics: tables and graphs *146*
Statistics and the publication process *148*
Research articles: guidelines *150*
Statistical problems in medical papers *153*

5 Choosing and using statistical software for analysing data 155
Statistical software packages *156*
Choosing a package *158*
Using a package *160*
Examples of using statistical packages *162*
Using spreadsheets for analysis *166*
Transferring data between packages *168*
Common packages *170*

6 Summarizing data 173
Why summarize data? *174*
Types of data *176*
Quantitative data *178*
Categorical data *180*
Summarizing quantitative data *182*
Calculation of mean, SD *184*
Calculation of median, interquartile range *186*
Geometric mean, harmonic mean, mode *188*
Choosing a summary measure for quantitative data *190*
Summarizing categorical data *192*
Graphs: histogram, stem and leaf plot *194*
Graphs: box and whisker plot, dot plot *196*
Graphs: shapes of distributions *198*
Graphs: bar chart, pie chart *200*
Summary *202*

7 Probability and distributions 203
Independence: data and variables *204*

Probability: definitions *206*
Probability: properties *208*
Probability distributions *210*
Binomial distribution: formula *212*
Binomial distribution: derivation *214*
Poisson distribution *216*
Continuous probability distributions *220*
Normal distribution *222*
Normal distribution: calculating probabilities *224*
Normal distribution: percentage points *226*
Central limit theorem *228*
Other distributions: t, chi-squared, F, etc. *232*
Bayes' theorem *234*

8 Statistical tests **237**
Introduction *238*
Samples and populations *240*
Confidence interval for a mean *242*
95% confidence interval for a proportion *244*
Tests of statistical significance *246*
P values *248*
Statistical significance and clinical significance *250*
t test for two independent means *252*
t test for two independent means: example *254*
t test for paired (matched) data *256*
t test for paired data: example *258*
z test for two independent proportions *260*
Chi-squared test *262*
Chi-squared test: calculations *264*
Fisher's exact test *266*
Estimates for tests of proportions *268*
Confidence intervals for tests of proportions *270*
Chi-squared test for trend *274*
McNemar's test for paired proportions *276*
Estimates and 95% confidence intervals for paired proportions *278*
One-way analysis of variance *280*
One-way analysis of variance: example *282*

Analysis of variance table *284*
Multiple comparisons *286*
Correlation and regression *288*
Pearson's correlation *290*
Correlation matrix *294*
Simple linear regression *296*
Simple linear regression: example *300*
Wilcoxon two-sample signed rank test (Mann Whitney U test) *303*
Wilcoxon two-sample signed rank test: calculations *304*
Wilcoxon two-sample signed rank test: example *306*
Wilcoxon matched pairs test *308*
Wilcoxon matched pairs test: example *310*
Rank correlation *312*
Rank correlation: example *314*
Survival data *316*
Kaplan–Meier curves *320*
Logrank test *322*
Logrank test: example *324*
Logrank test: interpreting the results *328*
Transforming data *330*
Transforming data: comparing means *332*
Transforming data: regression and correlation *334*
Transforming data: options *336*

9 Diagnostic studies **339**
Sensitivity and specificity *340*
Calculations for sensitivity and specificity *342*
Effect of prevalence *344*
Likelihood ratio, pre-test odds, post-test odds *346*
Receiver operating characteristic (ROC) curves *348*
Links to other statistics *350*

10 Other statistical methods **353**
Kappa for inter-rater agreement *354*
Extensions to kappa *358*
Bland–Altman method to measure agreement *360*
Chi-squared goodness of fit test *364*

Number needed to treat *366*
Life tables *370*
Direct standardization *372*
Indirect standardization *374*

11 Analysing multiple observations per subject **377**
Serial (longitudinal) data *378*
Summarizing serial data *380*
Calculating area under the curve *382*
Other summary measures for serial data *383*
Summary measures approach: key points *384*
Other approaches to serial data *386*
Cluster samples: units of analysis *388*
Cluster samples: analysis *390*

12 Analysing multiple variables per subject **393**
Multiple variables per subject *394*
Multifactorial methods: overview *396*
Multifactorial methods: model selection *398*
Multifactorial methods: challenges *400*
Missing data *402*
Generalized linear models *404*
Multiple regression *406*
Multiple regression: examples *408*
Multiple regression and analysis of variance *412*
Main effects and interactions *414*
Linear and non-linear terms *416*
How well the model fits *418*
Sample size for multiple regression *419*
Logistic regression *420*
Logistic regression: examples *422*
Logistic regression and ROC curves *424*
Extensions to logistic regression *426*
Cox proportional hazards regression *428*
Cox regression: example *430*
Poisson regression *432*
Poisson regression: example *434*

Multilevel models *436*
Generalized estimating equations (GEEs) *438*
GEEs: example *440*
Principal components analysis *441*
Principal components analysis: example *442*
Cluster analysis *444*
Factor analysis *445*

13 Meta-analysis **447**
Meta-analysis: introduction *448*
Searching for studies *450*
Combining estimates in meta-analyses *452*
Heterogeneity *454*
Overcoming heterogeneity *456*
Fixed effects estimates *458*
Random effects estimates *460*
Presenting meta-analyses *462*
Publication bias *464*
Detecting publication bias *466*
Adjusting for publication bias *468*
Independent patient data meta-analysis *472*
Challenges in meta-analysis *474*

14 Bayesian statistics **477**
Bayesian statistics *478*
How Bayesian methods work *480*
Prior distributions *482*
Likelihood; posterior distributions *484*
Summarizing and presenting results *486*
Using Bayesian analyses in medicine *488*
Software for Bayesian statistics *492*
Reading Bayesian analyses in papers *494*
Bayesian methods: a summary *496*

15 Glossary of terms **499**

Index *507*