

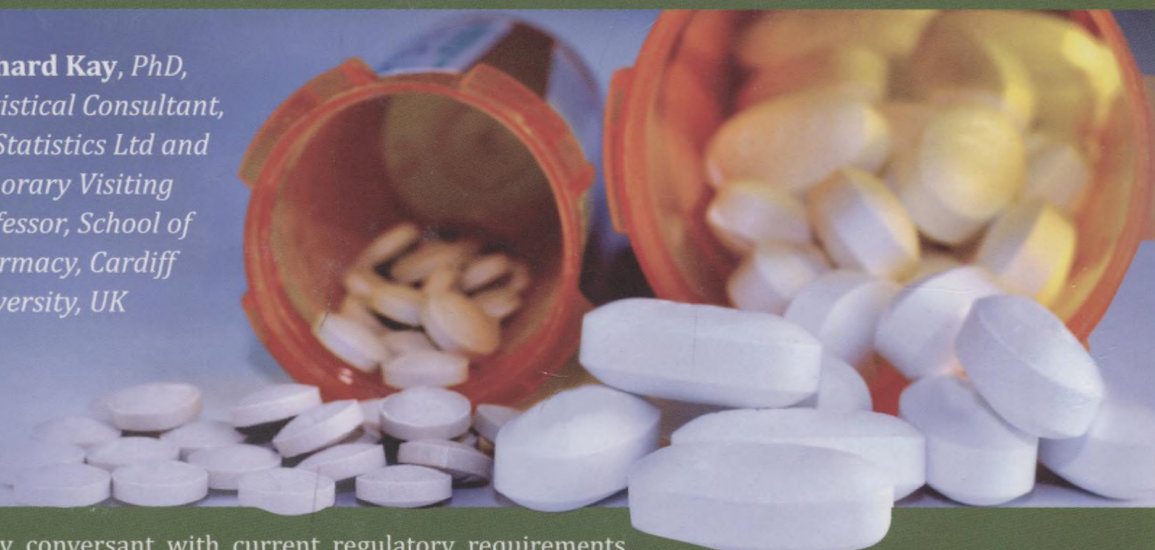
**Second Edition**

# Statistical Thinking for Non-Statisticians in Drug Regulation

*Statistical Thinking for Non-Statisticians in Drug Regulation, Second Edition*, is a need-to-know guide to understanding statistical methodology, statistical data and results within drug development and clinical trials.

It provides non-statisticians working in the pharmaceutical and medical device industries with an accessible introduction to the knowledge they need when working with statistical information and communicating with statisticians. It covers the statistical aspects of design, conduct, analysis and presentation of data from clinical trials in drug regulation and improves the ability to read, understand and critically appraise statistical methodology in papers and reports. As such, it is directly concerned with the day-to-day practice and the regulatory requirements of drug development and clinical trials.

**Richard Kay, PhD,**  
*Statistical Consultant,  
RK Statistics Ltd and  
Honorary Visiting  
Professor, School of  
Pharmacy, Cardiff  
University, UK*




Fully conversant with current regulatory requirements, this second edition includes five new chapters covering Bayesian statistics, adaptive designs, observational studies, methods for safety analysis and monitoring and statistics for diagnosis.

Authored by a respected lecturer and consultant to the pharmaceutical industry, *Statistical Thinking for Non-Statisticians in Drug Regulation* is an ideal guide for physicians, clinical research scientists, managers and associates, data managers, medical writers, regulatory personnel and for all non-statisticians working and learning within the pharmaceutical industry.

[www.wiley.com/wiley-blackwell](http://www.wiley.com/wiley-blackwell)

**WILEY** Blackwell

 Also available  
as an e-book

ISBN 978-1-118-47094-7



Preface to the second edition, xv

Preface to the first edition, xvii

Abbreviations, xxi

**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, 3

1.4.1 Unrestricted randomisation, 4

1.4.2 Block randomisation, 4

1.4.3 Unequal randomisation, 5

1.4.4 Stratified randomisation, 6

1.4.5 Central randomisation, 7

1.4.6 Dynamic allocation and minimisation, 8

1.4.7 Cluster randomisation, 9

1.5 Bias and precision, 9

1.6 Between- and within-patient designs, 11

1.7 Crossover trials, 12

1.8 Signal, noise and evidence, 13

1.8.1 Signal, 13

1.8.2 Noise, 13

1.8.3 Signal-to-noise ratio, 14

1.9 Confirmatory and exploratory trials, 15

1.10 Superiority, equivalence and non-inferiority trials, 16

1.11 Data and endpoint types, 17

1.12 Choice of endpoint, 18

1.12.1 Primary variables, 18

1.12.2 Secondary variables, 19

1.12.3 Surrogate variables, 20

1.12.4 Global assessment variables, 21

1.12.5 Composite variables, 21

1.12.6 Categorisation, 21

**2 Sampling and inferential statistics, 23**

2.1 Sample and population, 23

2.2 Sample statistics and population parameters, 24

- 2.2.1 Sample and population distribution, 24
- 2.2.2 Median and mean, 25
- 2.2.3 Standard deviation, 25
- 2.2.4 Notation, 26
- 2.2.5 Box plots, 27
- 2.3 The normal distribution, 28
- 2.4 Sampling and the standard error of the mean, 31
- 2.5 Standard errors more generally, 34
  - 2.5.1 The standard error for the difference between two means, 34
  - 2.5.2 Standard errors for proportions, 37
  - 2.5.3 The general setting, 37
- 3 Confidence intervals and  $p$ -values, 38**
  - 3.1 Confidence intervals for a single mean, 38
    - 3.1.1 The 95 per cent Confidence interval, 38
    - 3.1.2 Changing the confidence coefficient, 40
    - 3.1.3 Changing the multiplying constant, 40
    - 3.1.4 The role of the standard error, 41
  - 3.2 Confidence interval for other parameters, 42
    - 3.2.1 Difference between two means, 42
    - 3.2.2 Confidence interval for proportions, 43
    - 3.2.3 General case, 44
    - 3.2.4 Bootstrap Confidence interval, 45
  - 3.3 Hypothesis testing, 45
    - 3.3.1 Interpreting the  $p$ -value, 46
    - 3.3.2 Calculating the  $p$ -value, 47
    - 3.3.3 A common process, 50
    - 3.3.4 The language of statistical significance, 53
    - 3.3.5 One-sided and two-sided tests, 54
- 4 Tests for simple treatment comparisons, 56**
  - 4.1 The unpaired t-test, 56
  - 4.2 The paired t-test, 57
  - 4.3 Interpreting the t-tests, 60
  - 4.4 The chi-square test for binary data, 61
    - 4.4.1 Pearson chi-square, 61
    - 4.4.2 The link to a ratio of the signal to the standard error, 64
  - 4.5 Measures of treatment benefit, 64
    - 4.5.1 Odds ratio, 65
    - 4.5.2 Relative risk, 65
    - 4.5.3 Relative risk reduction, 66
    - 4.5.4 Number needed to treat, 66
    - 4.5.5 Confidence intervals, 67
    - 4.5.6 Interpretation, 68

- 4.6 Fisher's exact test, 69
- 4.7 Tests for categorical and ordinal data, 71
  - 4.7.1 Categorical data, 71
  - 4.7.2 Ordered categorical (ordinal) data, 73
  - 4.7.3 Measures of treatment benefit, 74
- 4.8 Extensions for multiple treatment groups, 75
  - 4.8.1 Between-patient designs and continuous data, 75
  - 4.8.2 Within-patient designs and continuous data, 76
  - 4.8.3 Binary, categorical and ordinal data, 76
  - 4.8.4 Dose-ranging studies, 77
  - 4.8.5 Further discussion, 77
- 5 Adjusting the analysis, 78**
  - 5.1 Objectives for adjusted analysis, 78
  - 5.2 Comparing treatments for continuous data, 78
  - 5.3 Least squares means, 82
  - 5.4 Evaluating the homogeneity of the treatment effect, 83
    - 5.4.1 Treatment-by-factor interactions, 83
    - 5.4.2 Quantitative and qualitative interactions, 85
  - 5.5 Methods for binary, categorical and ordinal data, 86
  - 5.6 Multi-centre trials, 87
    - 5.6.1 Adjusting for centre, 87
    - 5.6.2 Significant treatment-by-centre interactions, 87
    - 5.6.3 Combining centres, 88
- 6 Regression and analysis of covariance, 89**
  - 6.1 Adjusting for baseline factors, 89
  - 6.2 Simple linear regression, 89
  - 6.3 Multiple regression, 91
  - 6.4 Logistic regression, 94
  - 6.5 Analysis of covariance for continuous data, 94
    - 6.5.1 Main effect of treatment, 94
    - 6.5.2 Treatment-by-covariate interactions, 96
    - 6.5.3 A single model, 98
    - 6.5.4 Connection with adjusted analyses, 98
    - 6.5.5 Advantages of ANCOVA, 99
    - 6.5.6 Least squares means, 100
  - 6.6 Binary, categorical and ordinal data, 101
  - 6.7 Regulatory aspects of the use of covariates, 103
  - 6.8 Baseline testing, 105
- 7 Intention-to-treat and analysis sets, 107**
  - 7.1 The principle of intention-to-treat, 107
  - 7.2 The practice of intention-to-treat, 110
    - 7.2.1 Full analysis set, 110

- 7.2.2 Per-protocol set, 112
- 7.2.3 Sensitivity, 112
- 7.3 Missing data, 113
  - 7.3.1 Introduction, 113
  - 7.3.2 Complete cases analysis, 114
  - 7.3.3 Last observation carried forward, 114
  - 7.3.4 Success/failure classification, 114
  - 7.3.5 Worst-case/best-case classification, 115
  - 7.3.6 Sensitivity, 115
  - 7.3.7 Avoidance of missing data, 116
  - 7.3.8 Multiple imputation, 117
- 7.4 Intention-to-treat and time-to-event data, 118
- 7.5 General questions and considerations, 120
- 8 Power and sample size, 123**
  - 8.1 Type I and type II errors, 123
  - 8.2 Power, 124
  - 8.3 Calculating sample size, 127
  - 8.4 Impact of changing the parameters, 130
    - 8.4.1 Standard deviation, 130
    - 8.4.2 Event rate in the control group, 130
    - 8.4.3 Clinically relevant difference, 131
  - 8.5 Regulatory aspects, 132
    - 8.5.1 Power >80 per cent, 132
    - 8.5.2 Powering on the per-protocol set, 132
    - 8.5.3 Sample size adjustment, 133
  - 8.6 Reporting the sample size calculation, 134
- 9 Statistical significance and clinical importance, 136**
  - 9.1 Link between  $p$ -values and Confidence intervals, 136
  - 9.2 Confidence intervals for clinical importance, 137
  - 9.3 Misinterpretation of the  $p$ -value, 139
    - 9.3.1 Conclusions of similarity, 139
    - 9.3.2 The problem with 0.05, 140
  - 9.4 Single pivotal trial and 0.05, 140
- 10 Multiple testing, 142**
  - 10.1 Inflation of the type I error, 142
    - 10.1.1 False positives, 142
    - 10.1.2 A simulated trial, 142
  - 10.2 How does multiplicity arise?, 143
  - 10.3 Regulatory view, 144
  - 10.4 Multiple primary endpoints, 145
    - 10.4.1 Avoiding adjustment, 145
    - 10.4.2 Significance needed on all endpoints, 145

- 10.4.3 Composite endpoints, 146
- 10.4.4 Variables ranked according to clinical importance:
  - Hierarchical testing, 146
- 10.5 Methods for adjustment, 149
  - 10.5.1 Bonferroni correction, 149
  - 10.5.2 Hochberg correction, 150
  - 10.5.3 Interim analyses, 151
- 10.6 Multiple comparisons, 152
- 10.7 Repeated evaluation over time, 153
- 10.8 Subgroup testing, 154
- 10.9 Other areas for multiplicity, 156
  - 10.9.1 Using different statistical tests, 156
  - 10.9.2 Different analysis sets, 156
  - 10.9.3 Pre-planning, 157
- 11 Non-parametric and related methods, 158**
  - 11.1 Assumptions underlying the t-tests and their extensions, 158
  - 11.2 Homogeneity of variance, 158
  - 11.3 The assumption of normality, 159
  - 11.4 Non-normality and transformations, 161
  - 11.5 Non-parametric tests, 164
    - 11.5.1 The Mann–Whitney U-test, 164
    - 11.5.2 The Wilcoxon signed rank test, 166
    - 11.5.3 General comments, 167
  - 11.6 Advantages and disadvantages of non-parametric methods, 168
  - 11.7 Outliers, 169
- 12 Equivalence and non-inferiority, 170**
  - 12.1 Demonstrating similarity, 170
  - 12.2 Confidence intervals for equivalence, 172
  - 12.3 Confidence intervals for non-inferiority, 173
  - 12.4 A *p*-value approach, 174
  - 12.5 Assay sensitivity, 176
  - 12.6 Analysis sets, 178
  - 12.7 The choice of  $\Delta$ , 179
    - 12.7.1 Bioequivalence, 179
    - 12.7.2 Therapeutic equivalence, 180
    - 12.7.3 Non-inferiority, 180
    - 12.7.4 The 10 per cent rule for cure rates, 182
    - 12.7.5 The synthesis method, 183
  - 12.8 Biocreep and constancy, 184
  - 12.9 Sample size calculations, 184
  - 12.10 Switching between non-inferiority and superiority, 186

- 13** The analysis of survival data, 189
  - 13.1 Time-to-event data and censoring, 189
  - 13.2 Kaplan-Meier curves, 190
    - 13.2.1 Plotting Kaplan-Meier curves, 190
    - 13.2.2 Event rates and relative risk, 192
    - 13.2.3 Median event times, 192
  - 13.3 Treatment comparisons, 193
  - 13.4 The hazard ratio, 196
    - 13.4.1 The hazard rate, 196
    - 13.4.2 Constant hazard ratio, 197
    - 13.4.3 Non-constant hazard ratio, 197
    - 13.4.4 Link to survival curves, 198
    - 13.4.5 Calculating Kaplan-Meier curves, 199
  - 13.5 Adjusted analyses, 199
    - 13.5.1 Stratified methods, 200
    - 13.5.2 Proportional hazards regression, 200
    - 13.5.3 Accelerated failure time model, 201
  - 13.6 Independent censoring, 202
  - 13.7 Sample size calculations, 203
- 14** Interim analysis and data monitoring committees, 205
  - 14.1 Stopping rules for interim analysis, 205
  - 14.2 Stopping for efficacy and futility, 206
    - 14.2.1 Efficacy, 206
    - 14.2.2 Futility and conditional power, 207
    - 14.2.3 Some practical issues, 208
    - 14.2.4 Analyses following completion of recruitment, 209
  - 14.3 Monitoring safety, 210
  - 14.4 Data monitoring committees, 211
    - 14.4.1 Introduction and responsibilities, 211
    - 14.4.2 Structure and process, 212
    - 14.4.3 Meetings and recommendations, 214
- 15** Bayesian statistics, 215
  - 15.1 Introduction, 215
  - 15.2 Prior and posterior distributions, 215
    - 15.2.1 Prior beliefs, 215
    - 15.2.2 Prior to posterior, 217
    - 15.2.3 Bayes theorem, 217
  - 15.3 Bayesian inference, 219
    - 15.3.1 Frequentist methods, 219
    - 15.3.2 Posterior probabilities, 219
    - 15.3.3 Credible intervals, 220
  - 15.4 Case study, 221

15.5 History and regulatory acceptance, 222

15.6 Discussion, 224

## **16 Adaptive designs, 225**

16.1 What are adaptive designs?, 225

16.1.1 Advantages and drawbacks, 225

16.1.2 Restricted adaptations, 226

16.1.3 Flexible adaptations, 227

16.2 Minimising bias, 228

16.2.1 Control of type I error, 228

16.2.2 Estimation, 229

16.2.3 Behavioural issues, 230

16.2.4 Exploratory trials, 232

16.3 Unblinded sample size re-estimation, 232

16.3.1 Product of  $p$ -values, 232

16.3.2 Weighting the two parts of the trial, 233

16.3.3 Rationale, 234

16.4 Seamless phase II/III studies, 234

16.4.1 Standard framework, 234

16.4.2 Aspects of the  $p$ -value calculation, 235

16.4.3 Logistical challenges, 236

16.5 Other types of adaptation, 236

16.5.1 Changing the primary endpoint, 236

16.5.2 Focusing on a sub-population, 237

16.5.3 Dropping the placebo arm in a non-inferiority trial, 237

16.6 Further regulatory considerations, 238

16.6.1 Impact on power, 238

16.6.2 Non-standard experimental settings, 239

## **17 Observational studies, 241**

17.1 Introduction, 241

17.1.1 Non-randomised comparisons, 241

17.1.2 Study types, 241

17.1.3 Sources of bias, 243

17.1.4 An empirical investigation, 244

17.1.5 Selection bias in concurrently controlled studies:

An empirical evaluation, 245

17.1.6 Selection bias in historically controlled studies:

An empirical evaluation, 246

17.1.7 Some conclusions, 246

17.2 Guidance on design, conduct and analysis, 247

17.2.1 Regulatory guidance, 247

17.2.2 Strengthening the Reporting of Observational Studies in Epidemiology, 248

- 17.3 Evaluating and adjusting for selection bias, 249
  - 17.3.1 Baseline balance, 249
  - 17.3.2 Adjusting for imbalances using stratification and analysis of covariance, 250
  - 17.3.3 Propensity scores, 250
  - 17.3.4 Different methods for adjustment:
    - An empirical evaluation, 253
  - 17.3.5 Some conclusions, 256
- 17.4 Case-control studies, 257
  - 17.4.1 Background, 257
  - 17.4.2 Odds ratio and Relative risk, 259
- 18 Meta-analysis, 261**
  - 18.1 Definition, 261
  - 18.2 Objectives, 263
  - 18.3 Statistical methodology, 264
    - 18.3.1 Methods for combination, 264
    - 18.3.2 Confidence intervals, 265
    - 18.3.3 Fixed and random effects, 265
    - 18.3.4 Graphical methods, 266
    - 18.3.5 Detecting heterogeneity, 266
    - 18.3.6 Robustness, 269
    - 18.3.7 Rare events, 269
    - 18.3.8 Individual patient data, 269
  - 18.4 Case study, 270
  - 18.5 Ensuring scientific validity, 271
    - 18.5.1 Planning, 271
    - 18.5.2 Assessing the risk of bias, 273
    - 18.5.3 Publication bias and funnel plots, 273
    - 18.5.4 Preferred Reporting Items for Systematic Reviews and Meta-Analyses, 275
  - 18.6 Further regulatory aspects, 275
- 19 Methods for the safety analysis and safety monitoring, 277**
  - 19.1 Introduction, 277
    - 19.1.1 Methods for safety data, 277
    - 19.1.2 The rule of three, 278
  - 19.2 Routine evaluation in clinical studies, 279
    - 19.2.1 Types of data, 280
    - 19.2.2 Adverse events, 281
    - 19.2.3 Laboratory data, 284
    - 19.2.4 ECG data, 287
    - 19.2.5 Vital signs, 288

- 19.2.6 Safety summary across trials, 288
- 19.2.7 Specific safety studies, 289
- 19.3 Data monitoring committees, 289
- 19.4 Assessing benefit–risk, 290
  - 19.4.1 Current approaches, 290
  - 19.4.2 Multi-criteria decision analysis, 291
  - 19.4.3 Quality-Adjusted Time without Symptoms or Toxicity, 297
- 19.5 Pharmacovigilance, 299
  - 19.5.1 Post-approval safety monitoring, 299
  - 19.5.2 Proportional reporting ratios, 300
  - 19.5.3 Bayesian shrinkage, 302
- 20** Diagnosis, 304
  - 20.1 Introduction, 304
  - 20.2 Measures of diagnostic performance, 304
    - 20.2.1 Sensitivity and specificity, 304
    - 20.2.2 Positive and negative predictive value, 305
    - 20.2.3 False positive and false negative rates, 306
    - 20.2.4 Prevalence, 306
    - 20.2.5 Likelihood ratio, 307
    - 20.2.6 Predictive accuracy, 307
    - 20.2.7 Choosing the correct cut-point, 307
  - 20.3 Receiver operating characteristic curves, 308
    - 20.3.1 Receiver operating characteristic, 308
    - 20.3.2 Comparing ROC curves, 309
  - 20.4 Diagnostic performance using regression models, 310
  - 20.5 Aspects of trial design for diagnostic agents, 312
  - 20.6 Assessing agreement, 313
    - 20.6.1 The kappa statistic, 313
    - 20.6.2 Other applications for kappa, 314
- 21** The role of statistics and statisticians, 316
  - 21.1 The importance of statistical thinking at the design stage, 316
  - 21.2 Regulatory guidelines, 317
  - 21.3 The statistics process, 321
    - 21.3.1 The statistical methods section of the protocol, 321
    - 21.3.2 The statistical analysis plan, 322
    - 21.3.3 The data validation plan, 322
    - 21.3.4 The blind review, 322
    - 21.3.5 Statistical analysis, 323

- 21.3.6 Reporting the analysis, 323
- 21.3.7 Pre-planning, 324
- 21.3.8 Sensitivity and robustness, 326

21.4 The regulatory submission, 327

21.5 Publications and presentations, 328

References, 331

Index, 339