
Table of Contents

1	Fundamental issues	1
1.1	What is epidemiology?	1
1.2	Case studies: the work of Doll and Hill	2
1.3	Populations and samples	6
1.3.1	Populations	6
1.3.2	Samples	7
1.4	Measuring disease	7
1.4.1	Incidence and prevalence	9
1.5	Measuring the risk factor	10
1.6	Causality	11
1.6.1	Association	11
1.6.2	Problems with establishing causality	13
1.6.3	Principles of causality	14
1.7	Studies using routine data	14
1.7.1	Ecological data	15
1.7.2	National sources of data on disease	16
1.7.3	National sources of data on risk factors	17
1.7.4	International data	17
1.8	Study design	17
1.8.1	Intervention studies	18
1.8.2	Observational studies	19
1.9	Data analysis	20
	Exercises	21
2	Basic analytical procedures	23
2.1	Introduction	23
2.1.1	Inferential procedures	23
2.2	Case study	24
2.2.1	The Scottish Heart Health Study	24
2.3	Types of variables	25
2.3.1	Qualitative variables	26
2.3.2	Quantitative variables	26
2.3.3	The hierarchy of type	26
2.4	Tables and charts	27
2.4.1	Tables in reports	29
2.4.2	Diagrams in reports	33
2.5	Inferential techniques for categorical variables	33
2.5.1	Contingency tables	33

2.5.2	Binary variables: proportions and percentages	36	4	Confounding and interaction	125
2.5.3	Comparing two proportions or percentages	40	4.1	Introduction	125
2.6	Descriptive techniques for quantitative variables	41	4.2	The concept of confounding	126
2.6.1	The five-number summary	43	4.3	Identification of confounders	129
2.6.2	Quantiles	46	4.3.1	A strategy for selection	130
2.6.3	The two-number summary	48	4.4	Assessing confounding	131
2.6.4	Other summary statistics of spread	50	4.4.1	Using estimation	131
2.6.5	Assessing symmetry	50	4.4.2	Using hypothesis tests	132
2.6.6	Investigating shape	53	4.4.3	Dealing with several confounding variables	133
2.7	Inferences about means	57	4.5	Standardisation	134
2.7.1	Checking normality	58	4.5.1	Direct standardisation of event rates	135
2.7.2	Inferences for a single mean	60	4.5.2	Indirect standardisation of event rates	138
2.7.3	Comparing two means	61	4.5.3	Standardisation of risks	141
2.7.4	Paired data	64	4.6	Mantel–Haenszel methods	143
2.8	Inferential techniques for non-normal data	66	4.6.1	The Mantel–Haenszel relative risk	146
2.8.1	Transformations	66	4.6.2	The Cochran–Mantel–Haenszel test	147
2.8.2	Nonparametric tests	69	4.6.3	Further comments	148
2.8.3	Confidence intervals for medians	72	4.7	The concept of interaction	149
2.9	Measuring agreement	72	4.8	Testing for interaction	151
2.9.1	Quantitative variables	72	4.8.1	Using the relative risk	151
2.9.2	Categorical variables	74	4.8.2	Using the odds ratio	156
2.9.3	Ordered categorical variables	77	4.8.3	Using the risk difference	158
2.9.4	Internal consistency	78	4.8.4	Which type of interaction to use?	159
2.10	Assessing diagnostic tests	79	4.8.5	Which interactions to test?	159
2.10.1	Accounting for sensitivity and specificity	81	4.9	Dealing with interaction	160
Exercises		85	4.10	EPITAB commands in Stata	161
3	Assessing risk factors	89	Exercises		161
3.1	Risk and relative risk	89	5	Cohort studies	165
3.2	Odds and odds ratio	92	5.1	Design considerations	165
3.3	Relative risk or odds ratio?	94	5.1.1	Advantages	165
3.4	Prevalence studies	97	5.1.2	Disadvantages	165
3.5	Testing association	98	5.1.3	Alternative designs with economic advantages	167
3.5.1	Equivalent tests	99	5.1.4	Studies with a single baseline sample	168
3.5.2	One-sided tests	100	5.2	Analytical considerations	169
3.5.3	Continuity corrections	101	5.2.1	Concurrent follow-up	169
3.5.4	Fisher's exact test	102	5.2.2	Moving baseline dates	170
3.5.5	Limitations of tests	104	5.2.3	Varying follow-up durations	170
3.6	Risk factors measured at several levels	105	5.2.4	Withdrawals	172
3.6.1	Continuous risk factors	107	5.3	Cohort life tables	173
3.6.2	A test for linear trend	108	5.3.1	Allowing for sampling variation	175
3.6.3	A test for nonlinearity	111	5.3.2	Allowing for censoring	176
3.7	Attributable risk	111	5.3.3	Comparison of two life tables	177
3.8	Rate and relative rate	116	5.3.4	Limitations	180
3.8.1	The general epidemiological rate	119	5.4	Kaplan–Meier estimation	181
3.9	Measures of difference	119	5.4.1	An empirical comparison	182
3.10	EPITAB commands in Stata	120	5.5	Comparison of two sets of survival probabilities	184
Exercises		121	5.5.1	Mantel–Haenszel methods	184

5.5.2	The log-rank test	186
5.5.3	Weighted log-rank tests	188
5.5.4	Allowing for confounding variables	190
5.5.5	Comparing three or more groups	190
5.6	Competing risk	190
5.7	The person-years method	193
5.7.1	Age-specific rates	194
5.7.2	Summarisation of rates	196
5.7.3	Comparison of two SERs	197
5.7.4	Mantel-Haenszel methods	199
5.7.5	Further comments	202
5.8	Period-cohort analysis	203
5.8.1	Period-specific rates	204
	Exercises	206
6	Case-control studies	211
6.1	Basic design concepts	211
6.1.1	Advantages	211
6.1.2	Disadvantages	212
6.2	Basic methods of analysis	214
6.2.1	Dichotomous exposure	214
6.2.2	Polytomous exposure	217
6.2.3	Confounding and interaction	218
6.2.4	Attributable risk	218
6.3	Selection of cases	220
6.3.1	Definition	220
6.3.2	Inclusion and exclusion criteria	220
6.3.3	Incident or prevalent?	221
6.3.4	Source	221
6.3.5	Consideration of bias	221
6.4	Selection of controls	222
6.4.1	General principles	222
6.4.2	Hospital controls	224
6.4.3	Community controls	226
6.4.4	Other sources	227
6.4.5	How many?	228
6.5	Matching	229
6.5.1	Advantages	229
6.5.2	Disadvantages	230
6.5.3	One-to-many matching	231
6.5.4	Matching in other study designs	231
6.6	The analysis of matched studies	231
6.6.1	1 : 1 Matching	232
6.6.2	1 : c Matching	234
6.6.3	1 : Variable matching	240
6.6.4	Many : many matching	242
6.6.5	A modelling approach	245

6.7	Nested case-control studies	245
6.7.1	Matched studies	247
6.7.2	Counter-matched studies	248
6.8	Case-cohort studies	248
6.9	Case-crossover studies	250
	Exercises	251
7	Intervention studies	257
7.1	Introduction	257
7.1.1	Advantages	259
7.1.2	Disadvantages	259
7.2	Ethical considerations	259
7.2.1	The protocol	260
7.3	Avoidance of bias	261
7.3.1	Use of a control group	261
7.3.2	Blindness	262
7.3.3	Randomisation	263
7.3.4	Consent before randomisation	264
7.3.5	Analysis by intention-to-treat	265
7.4	Parallel group studies	265
7.4.1	Number needed to treat	268
7.4.2	Cluster randomised trials	270
7.4.3	Stepped wedge trials	270
7.4.4	Non-inferiority trials	271
7.5	Cross-over studies	273
7.5.1	Graphical analysis	275
7.5.2	Comparing means	277
7.5.3	Analysing preferences	282
7.5.4	Analysing binary data	283
7.6	Sequential studies	284
7.6.1	The Haybittle-Peto stopping rule	285
7.6.2	Adaptive designs	286
7.7	Allocation to treatment group	286
7.7.1	Global randomisation	286
7.7.2	Stratified randomization	288
7.7.3	Implementation	291
7.8	Trials as cohorts	291
	Exercises	291
8	Sample size determination	295
8.1	Introduction	295
8.2	Power	296
8.2.1	Choice of alternative hypothesis	300
8.3	Testing a mean value	303
8.3.1	Common choices for power and significance level	305
8.3.2	Using a table of sample sizes	305

8.3.3	The minimum detectable difference	306	9.8	Model checking	383
8.3.4	The assumption of known standard deviation	307	9.9	Confounding	387
8.4	Testing a difference between means	307	9.9.1	Adjustment using residuals	391
8.4.1	Using a table of sample sizes	308	9.10	Splines	392
8.4.2	Power and minimum detectable difference	310	9.10.1	Choice of knots	395
8.4.3	Optimum distribution of the sample	310	9.10.2	Other types of splines	396
8.4.4	Paired data	311	9.11	Panel data	398
8.5	Testing a proportion	311	9.12	Non-normal alternatives	402
8.5.1	Using a table of sample sizes	312	Exercises		404
8.6	Testing a relative risk	313	10	Modelling binary outcome data	409
8.6.1	Using a table of sample sizes	315	10.1	Introduction	409
8.6.2	Power and minimum detectable relative risk	316	10.2	Problems with standard regression models	412
8.7	Case-control studies	317	10.2.1	The r-x relationship may well not be linear	412
8.7.1	Using a table of sample sizes	319	10.2.2	Predicted values of the risk may be outside the valid range	412
8.7.2	Power and minimum detectable relative risk	319	10.2.3	The error distribution is not normal	412
8.7.3	Comparison with cohort studies	321	10.3	Logistic regression	413
8.7.4	Matched studies	321	10.4	Interpretation of logistic regression coefficients	415
8.8	Complex sampling designs	324	10.4.1	Binary risk factors	415
8.9	Concluding remarks	325	10.4.2	Quantitative risk factors	417
Exercises		326	10.4.3	Categorical risk factors	419
9	Modelling quantitative outcome variables	331	10.4.4	Ordinal risk factors	424
9.1	Statistical models	331	10.4.5	Floating absolute risks	425
9.2	One categorical explanatory variable	332	10.5	Generic data	427
9.2.1	The hypotheses to be tested	332	10.6	Multiple logistic regression models	428
9.2.2	Construction of the ANOVA table	333	10.7	Tests of hypotheses	432
9.2.3	How the ANOVA table is used	336	10.7.1	Goodness of fit for grouped data	433
9.2.4	Estimation of group means	336	10.7.2	Goodness of fit for generic data	435
9.2.5	Comparison of group means	337	10.7.3	Effect of a risk factor	435
9.2.6	Fitted values	338	10.7.4	Information criteria	438
9.2.7	Using computer packages	341	10.7.5	Tests for linearity and nonlinearity	440
9.3	One quantitative explanatory variable	344	10.7.6	Tests based upon estimates and their standard errors	443
9.3.1	Simple linear regression	344	10.7.7	Problems with missing values	444
9.3.2	Correlation	352	10.8	Confounding	444
9.3.3	Nonlinear regression	355	10.9	Interaction	445
9.4	Two categorical explanatory variables	358	10.9.1	Between two categorical variables	445
9.4.1	Model specification	358	10.9.2	Between a quantitative and a categorical variable	449
9.4.2	Model fitting	359	10.9.3	Between two quantitative variables	452
9.4.3	Balanced data	359	10.10	Dealing with a quantitative explanatory variable	452
9.4.4	Unbalanced data	359	10.10.1	Linear form	453
9.4.5	Fitted values	362	10.10.2	Categorical form	453
9.4.6	Least squares means	363	10.10.3	Linear spline form	455
9.4.7	Interaction	364	10.10.4	Generalisations	459
9.5	Model building	365	10.11	Model checking	459
9.6	General linear models	371	10.11.1	Residuals	459
9.7	Several explanatory variables	377	10.11.2	Influential observations	462
9.7.1	Information criteria	381	10.12	Measurement error	462
9.7.2	Boosted regression	383			

10.12.1	Regression to the mean	463	11.5.1	Comparing two groups	521
10.12.2	Correcting for regression dilution	465	11.5.2	Comparing several groups	521
10.13	Case-control studies	467	11.5.3	Modelling with a quantitative variable	523
10.13.1	Unmatched studies	467	11.5.4	Modelling with several variables	524
10.13.2	Matched studies	468	11.5.5	Left-censoring	525
10.14	Outcomes with several levels	469	11.6	The Cox proportional hazards model	526
10.14.1	The proportional odds assumption	471	11.6.1	Time-dependent covariates	535
10.14.2	The proportional odds model	473	11.6.2	Recurrent events	536
10.14.3	Multinomial regression	475	11.7	The Weibull proportional hazards model	536
10.15	Longitudinal data	475	11.8	Model checking	541
10.16	Binomial regression	476	11.8.1	Log cumulative hazard plots	541
10.16.1	Adjusted risks	479	11.8.2	An objective test of proportional hazards for the Cox model	545
10.16.2	Risk differences	483	11.8.3	An objective test of proportional hazards for the Weibull model	545
10.16.3	Problems with binomial models	484	11.8.4	Residuals and influence	546
10.17	Propensity scoring	488	11.8.5	Nonproportional hazards	546
10.17.1	Pair-matched propensity scores	488	11.9	Competing risk	546
10.17.2	Stratified propensity scores	489	11.9.1	Joint modeling of longitudinal and survival data	548
10.17.3	Weighting by the inverse propensity score	490	11.10	Poisson regression	549
10.17.4	Adjusting for the propensity score	491	11.10.1	Simple regression	550
10.17.5	Deriving the propensity score	492	11.10.2	Multiple regression	553
10.17.6	Propensity score outliers	493	11.10.3	Comparison of standardised event ratios	555
10.17.7	Conduct of the matched design	493	11.10.4	Routine or registration data	556
10.17.8	Analysis of the matched design	494	11.10.5	Generic data	558
10.17.9	Case studies	495	11.10.6	Model checking	559
10.17.10	Interpretation of effects	498	11.11	Pooled logistic regression	559
10.17.11	Problems with estimating uncertainty	499	Exercises		561
10.17.12	Propensity scores in practice	499			
Exercises		501			
11	Modelling follow-up data	507	12	Meta-analysis	565
11.1	Introduction	507	12.1	Reviewing evidence	565
11.1.1	Models for survival data	507	12.1.1	The Cochrane Collaboration	567
11.2	Basic functions of survival time	507	12.2	Systematic review	567
11.2.1	The survival function	507	12.2.1	Designing a systematic review	567
11.2.2	The hazard function	507	12.2.2	Study quality	571
11.3	Estimating the hazard function	508	12.3	A general approach to pooling	572
11.3.1	Kaplan-Meier estimation	508	12.3.1	Inverse variance weighting	573
11.3.2	Person-time estimation	510	12.3.2	Fixed effect and random effects	573
11.3.3	Actuarial estimation	511	12.3.3	Quantifying heterogeneity	574
11.3.4	The cumulative hazard	512	12.3.4	Estimating the between-study variance	576
11.4	Probability models	512	12.3.5	Calculating inverse variance weights	577
11.4.1	The probability density and cumulative distribution functions	512	12.3.6	Calculating standard errors from confidence intervals	577
11.4.2	Choosing a model	514	12.3.7	Case studies	578
11.4.3	The exponential distribution	514	12.3.8	Pooling risk differences	582
11.4.4	The Weibull distribution	517	12.3.9	Pooling differences in mean values	583
11.4.5	Other probability models	520	12.3.10	Other quantities	583
11.5	Proportional hazards regression models	521	12.3.11	Pooling mixed quantities	583
			12.3.12	Dose-response meta-analysis	584

12.4	Investigating heterogeneity	584	13.6	Recalibration	643
12.4.1	Forest plots	585	13.6.1	Recalibration of the mean	643
12.4.2	Influence plots	586	13.6.2	Recalibration of scores in a fixed cohort	643
12.4.3	Sensitivity analyses	588	13.6.3	Recalibration of parameters from a Cox model	646
12.4.4	Meta-regression	588	13.6.4	Recalibration and discrimination	647
12.5	Pooling tabular data	591	13.7	The accuracy of predictions	648
12.5.1	Inverse variance weighting	591	13.7.1	The Brier score	648
12.5.2	Mantel-Haenszel methods	591	13.7.2	Comparison of Brier scores	650
12.5.3	The Peto method	592	13.8	Assessing an extraneous prognostic variable	651
12.5.4	Dealing with zeros	592	13.9	Reclassification	652
12.5.5	Advantages and disadvantages of using tabular data	593	13.9.1	The integrated discrimination improvement from a fixed cohort	653
12.6	Individual participant data	593	13.9.2	The net reclassification improvement from a fixed cohort	656
12.7	Dealing with aspects of study quality	594	13.9.3	The integrated discrimination improvement from a variable cohort	659
12.8	Publication bias	595	13.9.4	The net reclassification improvement from a variable cohort	660
12.8.1	The funnel plot	596	13.9.5	Software	662
12.8.2	Consequences of publication bias	597	13.10	Validation	662
12.8.3	Correcting for publication bias	597	13.11	Presentation of risk scores	663
12.8.4	Other causes of asymmetry in funnel plots	599	13.11.1	Point scoring	664
12.9	Advantages and limitations of meta-analysis	600	13.12	Impact studies	674
Exercises		600	Exercises		675
13	Risk scores and clinical decision rules	605	14	Computer-intensive methods	679
13.1	Introduction	605	14.1	Rationale	679
13.1.1	Individual and population level interventions	605	14.2	The bootstrap	679
13.1.2	Scope of this chapter	607	14.2.1	Bootstrap distributions	681
13.2	Association and prognosis	608	14.3	Bootstrap confidence intervals	684
13.2.1	The concept of discrimination	610	14.3.1	Bootstrap normal intervals	685
13.2.2	Risk factor thresholds	611	14.3.2	Bootstrap percentile intervals	686
13.2.3	Risk thresholds	615	14.3.3	Bootstrap bias-corrected intervals	688
13.2.4	Odds ratios and discrimination	616	14.3.4	Bootstrap bias-corrected and accelerated intervals	690
13.3	Risk scores from statistical models	618	14.3.5	Overview of the worked example	691
13.3.1	Logistic regression	618	14.3.6	Choice of bootstrap interval	692
13.3.2	Multiple variable risk scores	620	14.4	Practical issues when bootstrapping	692
13.3.3	Cox regression	621	14.4.1	Software	692
13.3.4	Risk thresholds	623	14.4.2	How many replications should be used?	693
13.3.5	Multiple thresholds	624	14.4.3	Sensible strategies	696
13.4	Quantifying discrimination	625	14.5	Further examples of bootstrapping	696
13.4.1	The area under the curve	626	14.5.1	Complex bootstrap samples	701
13.4.2	Comparing AUCs	629	14.6	Bootstrap hypothesis testing	703
13.4.3	Survival data	631	14.7	Limitations of bootstrapping	705
13.4.4	The standardised mean effect size	632	14.8	Permutation tests	706
13.4.5	Other measures of discrimination	637	14.8.1	Monte Carlo permutation tests	707
13.5	Calibration	637	14.8.2	Limitations	709
13.5.1	Overall calibration	638	14.9	Missing values	709
13.5.2	Mean calibration	638			
13.5.3	Grouped calibration	639			
13.5.4	Calibration plots	641			

14.9.1	Dealing with missing values	711
14.9.2	Types of missingness	713
14.9.3	Complete case analyses	714
14.10	Naïve imputation methods	716
14.10.1	Mean imputation	716
14.10.2	Conditional mean and regression imputation	716
14.10.3	Hot deck imputation and predictive mean matching	718
14.10.4	Longitudinal data	719
14.11	Univariate multiple imputation	720
14.11.1	Multiple imputation by regression	720
14.11.2	The three-step process in MI	721
14.11.3	Imputer's and analyst's models	722
14.11.4	Rubin's equations	723
14.11.5	Imputation diagnostics	728
14.11.6	Skewed continuous data	729
14.11.7	Other types of variables	731
14.11.8	How many imputations?	731
14.12	Multivariate multiple imputation	733
14.12.1	Monotone imputation	733
14.12.2	Data augmentation	734
14.12.3	Categorical variables	742
14.12.4	What to do when DA fails	742
14.12.5	Chained equations	743
14.12.6	Longitudinal data	747
14.13	When is it worth imputing?	747
	Exercises	748
Appendix A Materials available on the website for this book		755
Appendix B Statistical tables		759
Appendix C Additional datasets for exercises		785
References		799
Index		821