Contents

Preface **xv**

1 Survival analysis **1**
1.1 Special features of survival data 1
1.1.1 Censoring 2
1.1.2 Independent censoring 3
1.1.3 Study time and patient time 3
1.2 Some examples 5
1.3 Survivor, hazard and cumulative hazard functions 10
1.3.1 The survivor function 10
1.3.2 The hazard function 12
1.3.3 The cumulative hazard function 13
1.4 Computer software for survival analysis 14
1.5 Further reading 15

2 Some non-parametric procedures **17**
2.1 Estimating the survivor function 17
2.1.1 Life-table estimate of the survivor function 19
2.1.2 Kaplan-Meier estimate of the survivor function 21
2.1.3 Nelson-Aalen estimate of the survivor function 24
2.2 Standard error of the estimated survivor function 25
2.2.1 Standard error of the Kaplan-Meier estimate 26
2.2.2 Standard error of other estimates 27
2.2.3 Confidence intervals for values of the survivor function 28
2.3 Estimating the hazard function 31
2.3.1 Life-table estimate of the hazard function 31
2.3.2 Kaplan-Meier type estimate 32
2.3.3 Estimating the cumulative hazard function 35
2.4 Estimating the median and percentiles of survival times 36
2.5 Confidence intervals for the median and percentiles 38
2.6 Comparison of two groups of survival data 40
2.6.1 Hypothesis testing 41
2.6.2 The log-rank test 44
2.6.3 The Wilcoxon test 48
2.6.4 Comparison of the log-rank and Wilcoxon tests 49

2.7 Comparison of three or more groups of survival data 50
2.8 Stratified tests 52
2.9 Log-rank test for trend 54
2.10 Further reading 56

3 The Cox regression model 57
3.1 Modelling the hazard function 57
3.1.1 A model for the comparison of two groups 58
3.1.2 The general proportional hazards model 59
3.2 The linear component of the model 60
3.2.1 Including a variate 61
3.2.2 Including a factor 61
3.2.3 Including an interaction 62
3.2.4 Including a mixed term 63
3.3 Fitting the Cox regression model 65
3.3.1 Likelihood function for the model 67
3.3.2 Treatment of ties 69
3.3.3 The Newton-Raphson procedure 71
3.4 Confidence intervals and hypothesis tests 72
3.4.1 Confidence intervals for hazard ratios 73
3.4.2 Two examples 73
3.5 Comparing alternative models 76
3.5.1 The statistic $-2 \log \hat{L}$ 77
3.5.2 Comparing nested models 78
3.6 Strategy for model selection 83
3.6.1 Variable selection procedures 84
3.7 Variable selection using the lasso 90
3.7.1 The lasso in Cox regression modelling 91
3.7.2 Data preparation 92
3.8 Non-linear terms 95
3.8.1 Testing for non-linearity 96
3.8.2 Modelling non-linearity 97
3.8.3 Fractional polynomials 98
3.9 Interpretation of parameter estimates 99
3.9.1 Models with a variate 99
3.9.2 Models with a factor 100
3.9.3 Models with combinations of terms 104
3.10 Estimating the hazard and survivor functions 107
3.10.1 The special case of no covariates 110
3.10.2 Some approximations to estimates of baseline functions 110
3.11 Risk adjusted survivor function 116
3.11.1 Risk adjusted survivor function for groups of individuals 117
3.12 Explained variation in the Cox regression model 120
3.12.1 Measures of explained variation 122
3.12.2 Measures of predictive ability 123

3.12.3 Model validation 124
3.13 Proportional hazards and the log-rank test 125
3.14 Further reading 128

4 Model checking in the Cox regression model 131
4.1 Residuals for the Cox regression model 131
4.1.1 Cox-Snell residuals 132
4.1.2 Modified Cox-Snell residuals 133
4.1.3 Martingale residuals 135
4.1.4 Deviance residuals 136
4.1.5 Schoenfeld residuals 137
4.1.6 Score residuals 138
4.2 Assessment of model fit 142
4.2.1 Plots based on the Cox-Snell residuals 142
4.2.2 Plots based on the martingale and deviance residuals 145
4.2.3 Checking the functional form of covariates 147
4.3 Identification of influential observations 152
4.3.1 Influence of observations on a parameter estimate 153
4.3.2 Influence of observations on the set of parameter estimates 155
4.3.3 Treatment of influential observations 158
4.4 Testing the assumption of proportional hazards 160
4.4.1 The log-cumulative hazard plot 161
4.4.2 Use of Schoenfeld residuals 163
4.4.3 Tests for non-proportional hazards 164
4.4.4 Adding a time-dependent variable 166
4.5 Recommendations 168
4.6 Further reading 169

5 Parametric proportional hazards models 171
5.1 Models for the hazard function 171
5.1.1 The exponential distribution 172
5.1.2 The Weibull distribution 173
5.2 Assessing the suitability of a parametric model 177
5.3 Fitting a parametric model to a single sample 178
5.3.1 Likelihood function for randomly censored data 180
5.4 Fitting exponential and Weibull models 181
5.4.1 Fitting the exponential distribution 182
5.4.2 Fitting the Weibull distribution 186
5.4.3 Standard error of a percentile of the Weibull distribution 188
5.5 A model for the comparison of two groups 192
5.5.1 The log-cumulative hazard plot 192
5.5.2 Fitting the model 194
5.6 The Weibull proportional hazards model 199
5.6.1 Fitting the model 200

5.6.2 Standard error of a percentile in the Weibull model 201
5.6.3 Log-linear form of the model 203
5.6.4 Exploratory analyses 205
5.7 Comparing alternative Weibull models 208
5.8 Explained variation in the Weibull model 215
5.9 The Gompertz proportional hazards model 216
5.10 Model choice 218
5.11 Further reading 219

6 Accelerated failure time and other parametric models 221
6.1 Probability distributions for survival data 221
6.1.1 The log-logistic distribution 222
6.1.2 The lognormal distribution 222
6.1.3 The gamma distribution 224
6.1.4 The inverse Gaussian distribution 225
6.2 Exploratory analyses 225
6.3 Accelerated failure model for two groups 227
6.3.1 Comparison with the proportional hazards model 228
6.3.2 The percentile-percentile plot 231
6.4 The general accelerated failure time model 232
6.4.1 Log-linear form of the accelerated failure time model 234
6.5 Parametric accelerated failure time models 236
6.5.1 The Weibull accelerated failure time model 236
6.5.2 The log-logistic accelerated failure time model 239
6.5.3 The lognormal accelerated failure time model 240
6.5.4 Summary 241
6.6 Fitting and comparing accelerated failure time models 243
6.7 The proportional odds model 250
6.7.1 The log-logistic proportional odds model 253
6.8 Some other distributions for survival data 255
6.9 Flexible parametric models 256
6.9.1 The Royston and Parmar model 259
6.9.2 Number and position of the knots 262
6.9.3 Fitting the model 262
6.9.4 Proportional odds models 266
6.10 Modelling cure rates 268
6.11 Effect of covariate adjustment 270
6.12 Further reading 272

7 Model checking in parametric models 275
7.1 Residuals for parametric models 275
7.1.1 Standardised residuals 275
7.1.2 Cox-Snell residuals 276
7.1.3 Martingale residuals 277
7.1.4 Deviance residuals 277

7.1.5 Score residuals 277
7.2 Residuals for particular parametric models 278
7.2.1 Weibull distribution 279
7.2.2 Log-logistic distribution 279
7.2.3 Lognormal distribution 280
7.2.4 Analysis of residuals 280
7.3 Comparing observed and fitted survivor functions 284
7.4 Identification of influential observations 287
7.4.1 Influence of observations on a parameter estimate 287
7.4.2 Influence of observations on the set of parameter estimates 288
7.5 Testing proportional hazards in the Weibull model 291
7.6 Further reading 292

8 Time-dependent variables 295
8.1 Types of time-dependent variables 295
8.2 A model with time-dependent variables 296
8.2.1 Fitting the Cox model 297
8.2.2 Estimation of baseline hazard and survivor functions 300
8.3 Model comparison and validation 302
8.3.1 Comparison of treatments 303
8.3.2 Assessing model adequacy 303
8.4 Some applications of time-dependent variables 304
8.5 Three examples 306
8.6 Counting process format 316
8.7 Further reading 317

9 Interval-censored survival data 319
9.1 Modelling interval-censored survival data 319
9.2 Modelling the recurrence probability in the follow-up period 322
9.3 Modelling the recurrence probability at different times 325
9.4 Arbitrarily interval-censored survival data 332
9.4.1 Modelling arbitrarily interval-censored data 332
9.4.2 Proportional hazards model for the survivor function 334
9.4.3 Choice of the step times 337
9.5 Parametric models for interval-censored data 342
9.6 Discussion 343
9.7 Further reading 344

10 Frailty models 345
10.1 Introduction to frailty 345
10.1.1 Random effects 346
10.1.2 Individual frailty 346
10.1.3 Shared frailty 347
10.2 Modelling individual frailty 348

10.2.1 Frailty distributions 349
10.2.2 Observable survivor and hazard functions 351
10.3 The gamma frailty distribution 352
10.3.1 Impact of frailty on an observable hazard function 353
10.3.2 Impact of frailty on an observable hazard ratio 354
10.4 Fitting parametric frailty models 356
10.4.1 Gamma frailty 357
10.5 Fitting semi-parametric frailty models 363
10.5.1 Lognormal frailty effects 363
10.5.2 Gamma frailty effects 365
10.6 Comparing models with frailty 366
10.6.1 Testing for the presence of frailty 366
10.7 The shared frailty model 372
10.7.1 Fitting the shared frailty model 373
10.7.2 Comparing shared frailty models 374
10.8 Some other aspects of frailty modelling 377
10.8.1 Model checking 377
10.8.2 Correlated frailty models 378
10.8.3 Dependence measures 378
10.8.4 Numerical problems in model fitting 378
10.9 Further reading 379

11 Non-proportional hazards and institutional comparisons 381
11.1 Non-proportional hazards 381
11.2 Stratified proportional hazards models 383
11.2.1 Non-proportional hazards between treatments 385
11.3 Restricted mean survival 389
11.3.1 Use of pseudo-values 391
11.4 Institutional comparisons 393
11.4.1 Interval estimate for the *RAFR* 397
11.4.2 Use of the Poisson regression model 400
11.4.3 Random institution effects 402
11.5 Further reading 403

12 Competing risks 405
12.1 Introduction to competing risks 405
12.2 Summarising competing risks data 406
12.2.1 Kaplan-Meier estimate of survivor function 407
12.3 Hazard and cumulative incidence functions 409
12.3.1 Cause-specific hazard function 409
12.3.2 Cause-specific cumulative incidence function 410
12.3.3 Some other functions of interest 413
12.4 Modelling cause-specific hazards 414
12.4.1 Likelihood functions for competing risks models 415
12.4.2 Parametric models for cumulative incidence functions 418

12.5 Modelling cause-specific incidence 419
12.5.1 The Fine and Gray competing risks model 419
12.6 Model checking 422
12.7 Further reading 428

13 Multiple events and event history modelling 429
13.1 Introduction to counting processes 429
13.1.1 Modelling the intensity function 430
13.1.2 Survival data as a counting process 431
13.1.3 Survival data in the counting process format 433
13.1.4 Robust estimation of the variance-covariance matrix 434
13.2 Modelling recurrent event data 435
13.2.1 The Anderson and Gill model 436
13.2.2 The Prentice, Williams and Peterson model 437
13.3 Multiple events 443
13.3.1 The Wei, Lin and Weissfeld model 443
13.4 Event history analysis 448
13.4.1 Models for event history analysis 449
13.5 Further reading 455

14 Dependent censoring 457
14.1 Identifying dependent censoring 457
14.2 Sensitivity to dependent censoring 458
14.2.1 A sensitivity analysis 459
14.2.2 Impact of dependent censoring 461
14.3 Modelling with dependent censoring 463
14.3.1 Cox regression model with dependent censoring 464
14.4 Further reading 470

15 Sample size requirements for a survival study 471
15.1 Distinguishing between two treatment groups 471
15.2 Calculating the required number of deaths 472
15.2.1 Derivation of the required number of deaths 474
15.3 Calculating the required number of patients 479
15.3.1 Derivation of the required number of patients 480
15.3.2 An approximate procedure 483
15.4 Further reading 484

A Maximum likelihood estimation 487
A.1 Inference about a single unknown parameter 487
A.2 Inference about a vector of unknown parameters 489

B Additional data sets **491**
B.1 Chronic active hepatitis 491
B.2 Recurrence of bladder cancer 492
B.3 Survival of black ducks 492
B.4 Bone marrow transplantation 495
B.5 Chronic granulomatous disease 495

Bibliography **499**

Index of Examples **521**

Index **523**