

*Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model* takes standard survival models into a wider realm, greatly increasing their usefulness. Some applications include prognostic modeling, time-dependent effects of covariates, and relative survival. Royston–Parmar flexible parametric models are a key tool in the approach. They extend familiar parametric models (Weibull, loglogistic, and lognormal), offering a much wider range of distributional forms.

The starting point of the text is a basic understanding of survival analysis and how it is done in Stata. For instance, a reader is assumed to know how to plot a Kaplan–Meier curve and fit a Cox model. The aim of the text is for researchers to build on the illustrations and examples to apply the methodology to their own investigations of survival data. To that end, the necessary tools (ado-files) are provided. The authors also provide Stata code for many of the analyses and graphs in the examples. Presentation of the results of flexible parametric modeling is often best achieved by well-chosen graphs, which is an important message of the book. Many examples with real data are given, and in almost all instances, the datasets are available from the book's website.

Patrick Royston is a senior medical statistician at the Medical Research Council, London, UK. He has published research papers on a variety of topics in leading statistics journals. His key interests include multivariable modeling and validation, survival analysis, design and analysis of clinical trials, and statistical computing and algorithms. He is an associate editor of the *Stata Journal*.

Paul Lambert is a reader in medical statistics at Leicester University, UK. His main interest is in the development and application of statistical methods in population-based cancer research and related fields. He has published widely in leading statistical and medical journals.

Telephone: 979-696-4600

800-782-8272

800-STATAPC

Fax: 979-696-4601

[service@stata-press.com](mailto:service@stata-press.com)

[www.stata-press.com](http://www.stata-press.com)





# Contents

	List of tables	xiii
	List of figures	xv
	Preface	xxv
1	Introduction	1
1.1	Goals . . . . .	1
1.2	A brief review of the Cox proportional hazards model . . . . .	2
1.3	Beyond the Cox model . . . . .	2
1.3.1	Estimating the baseline hazard . . . . .	2
1.3.2	The baseline hazard contains useful information . . . . .	5
1.3.3	Advantages of smooth survival functions . . . . .	8
1.3.4	Some requirements of a practical survival analysis . . . . .	9
1.3.5	When the proportional-hazards assumption is breached . . . . .	10
1.4	Why parametric models? . . . . .	13
1.4.1	Smooth baseline hazard and survival functions . . . . .	13
1.4.2	Time-dependent HRs . . . . .	13
1.4.3	Modeling on different scales . . . . .	13
1.4.4	Relative survival . . . . .	13
1.4.5	Prediction out of sample . . . . .	14
1.4.6	Multiple time scales . . . . .	14
1.5	Why not standard parametric models? . . . . .	14
1.6	A brief introduction to stpm2 . . . . .	16
1.6.1	Estimation (model fitting) . . . . .	16
1.6.2	Postestimation facilities (prediction) . . . . .	17
1.7	Basic relationships in survival analysis . . . . .	17



1.8	Comparing models . . . . .	18
1.9	The delta method . . . . .	19
1.10	Ado-file resources . . . . .	20
1.11	How our book is organized . . . . .	21
<b>2</b>	<b>Using stset and stsplot</b>	<b>23</b>
2.1	What is the stset command? . . . . .	23
2.2	Some key concepts . . . . .	23
2.3	Syntax of the stset command . . . . .	24
2.4	Variables created by the stset command . . . . .	25
2.5	Examples of using stset . . . . .	25
2.5.1	Standard survival data . . . . .	26
2.5.2	Using the scale() option . . . . .	27
2.5.3	Date of diagnosis and date of exit . . . . .	27
2.5.4	Date of diagnosis and date of exit with the scale() option . . . . .	28
2.5.5	Restricting the follow-up time . . . . .	29
2.5.6	Left-truncation . . . . .	31
2.5.7	Age as the time scale . . . . .	32
2.6	The stsplot command . . . . .	33
2.6.1	Time-dependent effects . . . . .	33
2.6.2	Time-varying covariates . . . . .	34
2.7	Conclusion . . . . .	35
<b>3</b>	<b>Graphical introduction to the principal datasets</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Rotterdam breast cancer data . . . . .	37
3.3	England and Wales breast cancer data . . . . .	39
3.4	Orchiectomy data . . . . .	42
3.5	Conclusion . . . . .	45
<b>4</b>	<b>Poisson models</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Modeling rates with the Poisson distribution . . . . .	48



4.3	Splitting the time scale . . . . .	50
4.3.1	The piecewise exponential model . . . . .	53
4.3.2	Time as just another covariate . . . . .	57
4.4	Collapsing the data to speed up computation . . . . .	57
4.5	Splitting at unique failure times . . . . .	59
4.5.1	Technical note: Why the Cox and Poisson approaches are equivalent* . . . . .	61
4.6	Comparing a different number of intervals . . . . .	62
4.7	Fine splitting of the time scale . . . . .	66
4.8	Splines: Motivation and definition . . . . .	67
4.8.1	Calculating splines* . . . . .	69
4.8.2	Restricted cubic splines . . . . .	70
4.8.3	Splines: Application to the Rotterdam data . . . . .	71
4.8.4	Varying the number of knots . . . . .	74
4.8.5	Varying the location of the knots . . . . .	78
4.8.6	Estimating the survival function* . . . . .	79
4.9	FPs: Motivation and definition . . . . .	81
4.9.1	Application to Rotterdam data . . . . .	83
4.9.2	Higher order FP models . . . . .	87
4.9.3	FP function selection procedure . . . . .	89
4.10	Discussion . . . . .	90
<b>5</b>	<b>Royston–Parmar models</b>	<b>91</b>
5.1	Motivation and introduction . . . . .	92
5.1.1	The exponential distribution . . . . .	92
5.1.2	The Weibull distribution . . . . .	95
5.1.3	Generalizing the Weibull . . . . .	96
5.1.4	Estimating the hazard function . . . . .	100
5.2	Proportional hazards models . . . . .	101
5.2.1	Generalizing the Weibull . . . . .	101
5.2.2	Example . . . . .	103



5.2.3	Comparing parameters of PH(1) and Weibull models . . . . .	104
5.3	Selecting a spline function . . . . .	108
5.3.1	Knot positions . . . . .	108
	Example . . . . .	109
5.3.2	How many knots? . . . . .	110
5.4	PO models . . . . .	111
5.4.1	Introduction . . . . .	111
5.4.2	The loglogistic model . . . . .	112
5.4.3	Generalizing the loglogistic model . . . . .	113
5.4.4	Comparing parameters of PO(1) and loglogistic models . . .	113
	Example . . . . .	114
5.5	Probit models . . . . .	114
5.5.1	Motivation . . . . .	114
5.5.2	Generalizing the probit model . . . . .	115
5.5.3	Comparing parameters of probit(1) and lognormal models .	116
5.5.4	Comments on probit and POs models . . . . .	117
5.6	Royston–Parmar (RP) models . . . . .	118
5.6.1	Models with $\theta$ not equal to 0 or 1 . . . . .	119
5.6.2	Example . . . . .	119
5.6.3	Likelihood function and parameter estimation* . . . . .	120
5.6.4	Comparing regression coefficients . . . . .	121
5.6.5	Model selection . . . . .	121
5.6.6	Sensitivity to number of knots . . . . .	122
5.6.7	Sensitivity to location of knots . . . . .	123
5.7	Concluding remarks . . . . .	124
6	<b>Prognostic models</b>	<b>125</b>
6.1	Introduction . . . . .	125
6.2	Developing and reporting a prognostic model . . . . .	126
6.3	What does the baseline hazard function mean? . . . . .	127
6.3.1	Example . . . . .	128



6.4	Model selection . . . . .	129
6.4.1	Choice of scale and baseline complexity . . . . .	130
	Example . . . . .	130
6.4.2	Selection of variables and functional forms . . . . .	131
	Example . . . . .	132
6.5	Quantitative outputs from the model . . . . .	134
6.5.1	Survival probabilities for individuals . . . . .	134
6.5.2	Survival probabilities across the risk spectrum . . . . .	137
6.5.3	Survival probabilities at given covariate values . . . . .	138
6.5.4	Survival probabilities in groups . . . . .	140
6.5.5	Plotting adjusted survival curves . . . . .	142
6.5.6	Plotting differences between survival curves . . . . .	143
6.5.7	Centiles of the survival distribution . . . . .	145
6.6	Goodness of fit . . . . .	147
6.6.1	Example . . . . .	148
6.7	Discrimination and explained variation . . . . .	149
6.7.1	Example . . . . .	151
6.7.2	Harrell's C index of concordance . . . . .	152
6.8	Out-of-sample prediction: Concept and applications . . . . .	153
6.8.1	Extrapolation of survival functions: Basic technique . . . . .	153
6.8.2	Extrapolation of survival functions: Further investigations . . . . .	155
6.8.3	Validation of prognostic models: Basics . . . . .	157
6.8.4	Validation of prognostic models: Further comments . . . . .	160
6.9	Visualization of survival times . . . . .	161
6.9.1	Example . . . . .	161
6.10	Discussion . . . . .	164
7	<b>Time-dependent effects</b>	<b>167</b>
7.1	Introduction . . . . .	167
7.2	Definitions . . . . .	168
7.3	What do we mean by a TD effect? . . . . .	169



7.4	Proportional on which scale? . . . . .	176
7.5	Poisson models with TD effects . . . . .	179
7.5.1	Piecewise models . . . . .	180
7.5.2	Using restricted cubic splines . . . . .	184
7.6	RP models with TD effects . . . . .	190
7.6.1	Piecewise HRs . . . . .	190
7.6.2	Continuous TD effects . . . . .	193
7.6.3	More than one TD effect . . . . .	201
7.6.4	Stratification is the same as including TD effects . . . . .	203
7.7	TD effects for continuous variables . . . . .	205
7.8	Attained age as the time scale . . . . .	211
7.8.1	The orchiectomy data . . . . .	211
7.8.2	Proportional hazards model . . . . .	212
7.8.3	TD model . . . . .	214
7.9	Multiple time scales . . . . .	218
7.10	Prognostic models with TD effects . . . . .	219
7.10.1	Example . . . . .	220
7.11	Discussion . . . . .	224
8	<b>Relative survival</b> . . . . .	<b>227</b>
8.1	Introduction . . . . .	227
8.2	What is relative survival? . . . . .	227
8.3	Excess mortality and relative survival . . . . .	228
8.3.1	Excess mortality . . . . .	228
8.3.2	Relative survival is a ratio . . . . .	230
8.4	Motivating example . . . . .	231
8.5	Life-table estimation of relative survival . . . . .	233
8.5.1	Using strs . . . . .	234
8.6	Poisson models for relative survival . . . . .	235
8.6.1	Piecewise models . . . . .	235
8.6.2	Restricted cubic splines . . . . .	241



8.7	RP models for relative survival . . . . .	246
8.7.1	Likelihood for relative survival models . . . . .	247
8.7.2	Proportional cumulative excess hazards . . . . .	247
8.7.3	RP models on other scales . . . . .	248
8.7.4	Application to England and Wales breast cancer data . . . . .	248
8.7.5	Relative survival models on other scales . . . . .	250
8.7.6	Time-dependent effects . . . . .	253
8.8	Some comments on model selection . . . . .	259
8.9	Age as a continuous variable . . . . .	267
8.10	Concluding remarks . . . . .	272
9	<b>Further topics</b>	<b>273</b>
9.1	Introduction . . . . .	273
9.2	Number needed to treat . . . . .	273
9.2.1	Example . . . . .	274
9.3	Average and adjusted survival curves . . . . .	275
9.3.1	Renal data . . . . .	277
9.4	Modeling distributions with RP models . . . . .	283
9.4.1	Example 1: Rotterdam breast cancer data . . . . .	283
9.4.2	Example 2: CD4 lymphocyte data . . . . .	285
9.4.3	Example 3: Prostate cancer data . . . . .	294
9.5	Multiple events . . . . .	296
9.5.1	Introduction . . . . .	296
9.5.2	The AG model . . . . .	297
9.5.3	The WLW model . . . . .	298
9.5.4	The PWP model . . . . .	298
9.5.5	Multiple events in RP models . . . . .	298
9.5.6	Summary . . . . .	304
9.6	Bayesian RP models . . . . .	304
9.6.1	Introduction . . . . .	304
9.6.2	The “zeros trick” in WinBUGS . . . . .	305



9.6.3	Fitting a RP model . . . . .	305
9.6.4	Summary . . . . .	310
9.7	Competing risks . . . . .	310
9.7.1	Summary . . . . .	316
9.8	Period analysis . . . . .	317
9.8.1	Introduction . . . . .	317
9.8.2	What is period analysis? . . . . .	317
9.8.3	Application to England and Wales breast cancer data . . . . .	319
9.9	Crude probability of death from relative survival models . . . . .	322
9.9.1	Introduction . . . . .	322
9.9.2	Application to England and Wales breast cancer data . . . . .	323
9.9.3	Conclusion . . . . .	329
9.10	Final remarks . . . . .	329
	<b>References</b>	<b>331</b>
	<b>Author index</b>	<b>341</b>
	<b>Subject index</b>	<b>345</b>