Table 1.2 Summary of survival and event history models

Class and type of model	Description	Advantages	Disadvantages
Non-parametric	*(Chapter 4)*		
– Life table estimates – Kaplan-Meier (Product-Limit) estimator	– Makes no assumption about: • shape of hazard function • how covariates affect shape of hazard function – Effects of covariates shown by stratifying data into groups	– Good method to understand basics and produce descriptive results – Life table: good for large data and crude measurement of event times – KM: good for smaller data and precisely measured event times	– Can only compare limited number of groups – Does not allow inclusion of multiple covariates and multivariate controls
Semi-parametric	*(Chapter 5)*		
– Cox model (most prominent) – Piecewise constant exponential model	– Makes no assumption about shape of hazard function – Makes strong assumption about how covariates affect shape of hazard function by assuming proportional hazard between groups over time – Partial-likelihood estimation	– Flexible model, often initial exploratory choice in analyses – Allows inclusion of multiple covariates, multivariate analysis – Results often similar to parametric models, but without (often) restrictive assumptions	– Less appropriate for testing hypotheses about time-dependence (i.e. how hazard varies over time) – Less precise than parametric models – Sometimes called 'overfitted'
Parametric	*(Chapter 6)*		
– Exponential, Weibull, logistic, Gamma, Gaussian, complementary log-log, log-logistic, log-normal, Gompertz, Makeham, extreme value, Rayleigh and others	– Researcher needs to decide in advance shape of the hazard function and how covariates impact the hazard function – Maximum likelihood estimation – Preferred when researcher wants to study the nature of time dependence and when time is meaningful in an independent variable – Continuous and discrete-time models	– More precise parameter estimates (if correct model assumptions) – Allows multivariate analysis – Allows analysis of discrete and continuous explanatory variables – Specifies the shape of the hazard function, allowing for predictive modelling	– If the hazard-function shape is incorrectly specified, parameter estimates can be seriously biased – Needs preliminary work to first define shape of hazard function and understand how covariates affect the hazard function – Very sensitive to included or omitted covariates
Discrete time and count	*(Chapter 9)*		
– Logit, probit, logistic, negative binomial regression (NBR models) – Poisson	– Analyzes the number of events since a defined starting time – Often for analysis of rare events (Poisson)	– Useful for analysis of rare events (Poisson, NBR) when number of zeros (i.e. number of trials without any events) is large	– Does not examine duration, but rather event-counts

Gompertz, Makeham, extreme value, Rayleigh and others	– Preferred when researcher wants to study the nature of time dependence and when time is meaningful in an independent variable – Continuous and discrete-time models	continuous explanatory variables – Specifies the shape of the hazard function, allowing for predictive modelling	define shape of hazard function and understand how covariates affect the hazard function – Very sensitive to included or omitted covariates
Discrete time and count	*(Chapter 9)*		
– Logit, probit, logistic, negative binomial regression (NBR models) – Poisson	– Analyzes the number of events since a defined starting time – Often for analysis of rare events (Poisson)	– Useful for analysis of rare events (Poisson, NBR) when number of zeros (i.e. number of trials without any events) is large	– Does not examine duration, but rather event-counts

Multilevel, frailty or recurrent event models	*(Chapter 8)*		
– Recurrent event or multiple episode models – Frailty models, conditional frailty models (sometimes also referred to as multilevel models, random effect models	– Some subjects more likely to experience repeated event due to unmeasured cause (unobserved heterogeneity) – Understanding how covariate effects change across episodes – Frailty: model as random effect – Conditional frailty: modifies frailty model to adjust for event dependence, stratifies cases by event number	– Goes beyond single-episode models that only compare effects between covariates to examine how covariate effects change across episodes – By estimating frailty as cause of unobserved heterogeneity as a random effect, coefficients for measured variables are less biased	– Frailty models may be badly biased if frailty is correlated with the covariates or the wrong distribution is assumed
Competing risk models	*(Chapter 10)*		
– Competing risk and multiple destination models: use one of the models described above (e.g. Cox) and make adjustments to risk group depending on whether risks are independent of one another	– Episode can end in two or more different outcomes – Central assumption is often conditional independence of the risks under analysis	– Considers more complex destination states – Treats different reasons as different events, allowing comparison of hazard functions across competing risks	– Problem if competing risks are not properly identified – Hard to cope with assumption of conditional independence of the risks under analysis
Multistate models	*(Chapter 10)*		
– Multistate models (also overlaps with competing risk, recurrent event and alternating state models)	– Model for a stochastic process, which at any time point occupies one set of discrete states – Specify state structure and form of hazard function for each transition	– Appropriate for event-related dependence	– Considers states, not events (problem for recurrent events) – All data considered longitudinal; less useful for repeated measurements
Sequence analysis	*(Chapter 11)*		
– Discrete Markov models and optimal-matching-based clustering	– Obtain a matrix of proximities between sequences via optimal matching (or other metric) and cluster sequences via multidimensional scaling methods	– Provides a holistic view of entire event history – Derives prominent characteristics of complete trajectories	– Remains highly descriptive if clusters not used as predictors in regression model