7

Cox and Gompertz regression models
An assessment with empirical estimates

Abstract

To construct a life table (with or without covariates), age-specific transition rates are necessary. Transition rates can be derived directly from the data set or can be estimated by fitting regression models to the data. We used the Cox model, which is semi-parametric, and the Gompertz model, which is parametric. The age-specific transition rates obtained from the Cox and Gompertz models in the presence or absence of covariates and the consequences for the life table estimates have not been compared to the empirical rates. In this chapter, an assessment of these two widely used models is made by means of a comparison with age-specific *occurrence-exposure* rates. We used 48 years of follow-up of cardiovascular morbidity and mortality in the original Framingham Heart Study (FHS) cohort to illustrate and compare model estimates. We estimated the age-specific transition rates for both the null model (without covariates) and the model with covariates. Finally, the different estimates of the rates were transferred into life table outcomes. As expected, both the Cox and the observed rates were found to be the same in the null model. On adding covariates to the Cox model, the estimated age-specific rates did not overlap with the observed rates because of the proportionality assumption. The Gompertz estimate behaved in the same way in both the null model and in the model with covariates. The Cox null model could be applied instead of observed rates if the number of occurrences at each age group was sufficiently large, in which case the Cox null model and observed rates were seen to yield the same transition rates. The Gompertz model (with or without covariates) fit well with estimates of cardiovascular disease and mortality transition rates in the FHS population. Since the variability of the estimated transition rates in the Gompertz model was less than in the Cox model or empirical rates, life table estimates using Gompertz in the presence or absence of covariates is recommended.

age-specific transition rates. To estimate the age-specific transition rates for use as input in a life table, researchers prefer to use smoothed transition rates (e.g. a log-rate model) to avoid the stochastic variability in the transition rates across ages (Nusselder, 1998). In this thesis, we applied the Gompertz model (Chapter 9) as a regression model and as a smoothing technique (Chapter 6).

In this chapter, first the null-model is compared with the empirical rates. The transition rates obtained from the Cox null model will be the same as the empirical rates. The transition rates estimated applying the Cox model with covariates should be proportional to different values of the covariates. The Gompertz model smoothes the observed rates. Second, the age-specific transition rates are then used to construct life tables. The multistate life table is used to derive several life table indicators: life expectancy, life expectancy free of CVD and lifetime probability of CVD. The life table outcomes could demonstrate the differences in the Cox, Gompertz and observed rates. We used the first 48 years of follow-up of the original Framingham Heart Study (FHS) cohort for estimation of the transition rates and model comparisons.

We have described the Cox model and the Gompertz model with their limitations in Section 7.2. Sub-section 7.2.3 describes the data source used to illustrate the transition models. Transition rates are compared in Section 7.3. Life table estimates are compared in Section 7.4. Section 7.5 concludes this chapter.

7.2 Models and data

The objective of the transition rate models in lifetime analysis, survival analysis, failure time analysis, or in event history analysis is to study (describe and explain) the time, *T,* until an individual has changed state. In our study, the time scale was age, say *X,* instead of time *T*. Let *X* represent age. We regard *X* as a random variable with a cumulative distribution function $F(x)=\Pr(X<x)$, probability density function $f(x)=dF(x)/dx$, the survival function $S(x)= \Pr(X\geq x)=1-F(x)$ and the hazard rate at age *x*, conditional on survival to that age:

$$\mu(x) = \lim_{\Delta x \to 0} \frac{\Pr[(x \leq X < x + \Delta x) \mid X \geq x]}{\Delta x} = \frac{f(x)}{S(x)}$$

Here $\mu(x)$ represents the transition intensities estimated from the model. The empirical occurrence-exposure rates ($M(x)$) were calculated from the data using equation 2.7 (Chapter 2). We estimated the age-specific transition rates using Cox, Gompertz and observed occurrence-exposure rates. The Cox and Gompertz models, the estimations and their limitations are described in the below.

7.2.1 The Cox regression model

The impact of various explanatory variables that might affect the transitions or event occurrences is examined using the Cox (1972; 1975) hazards model. Under this model, it is assumed that the hazard function is proportional for different values of an explanatory variable. More specifically, the rate of event occurrences for an individual with explanatory variable value given by the row vector $z^{(ij)}$ is assumed to be

$$\mu_{ij}(x) = \mu_{0ij}(x)\exp(z^{(ij)}\beta^{(ij)}) \tag{7.1}$$

where i is the state of origin, j is state of destination, $\mu_{0ij}(x)$ is the "baseline" rate of transition from i to j to which all force of events functions are proportional (i.e. shape is unspecified), $z^{(ij)}$ is a row vector of p measured covariates and $\beta^{(ij)}$ is a column vector of p parameters. The baseline hazard function $\mu_{0ij}(x)$ has been defined as the hazard for an individual with values of all explanatory variable equals to zero. The ability of this method to detect the differences of event rates associated with the different values of covariates is influenced by the extent to which the forces of event occurrences are proportional. In that case, the row vector $Z^{(ij)}$ is independent of time x.

The partial likelihood method gives an estimate of the parameters of $\beta^{(ij)}$ but no direct estimate of the underlying baseline hazard rate, i.e. without knowing the baseline hazard, we estimate $\beta^{(ij)}$ parameters. The partial likelihood method for the Cox model was developed under the assumption of continuous data, but real data sets often contain tied event times. Among the exposed people, tied events are those events that occur at the same time or age. Ties may occur because continuous event times are grouped into intervals or because the event time scale is discrete. Recently, Therneau and Grambsch (2000) summarized four variants of the computing algorithm that are commonly used to address the tie issues. (i) *Breslow approximation* - which is the simplest in terms of programming, and consequently was the only method available in the earliest Cox model routines, is still the default method in almost all packages. However, the solution is the least accurate. It counts failed individuals more than once in the denominator, which produces both a conservative bias and estimate regression coefficients too close to 0 in absolute value (Cox and Oakes, 1984); however, the method is fast. TDA's and SPSS's approach is based on a proposal made by Breslow. (ii) *Efron approximation*- is quite accurate unless the proportion of ties or the number of tied events relative to the size of the risk set is extremely large, and is as fast as the Breslow method. This approximation is the default in S-Plus. (iii) *Exact partial likelihood*- this approach engages an exhaustive account of the possible risk sets at each tied event time, and

can require an excessive amount of computation time if any of the individual death times has a large number of events (>10 say) (Thernau and Grambsch, 2000). S-Plus uses this as the *exact* option and SAS as the *discrete* option. (iv) *Average likelihood*- this is often very close to the Efron approximation. It also engages an exhaustive account, but a substitution allows this to be replaced by the numerical evaluation of an integral. This estimation is nearly untraceable computationally when the number of tied events at any time is even moderately large. In SAS, it is called the *exact* option.

Advantages and disadvantages

The application of the Cox model is very simple, software is available and no assumptions are made about the nature or shape of the hazard function. However, there are several disadvantages to this model. The relative lack of precision for the partial likelihood estimates of parameters compared with maximum likelihood estimates is expected to decrease with an increase in the sample size for most empirical situations (Efron, 1977; Oakes, 1977; Wong 1986). Another disadvantage of this model is the presence of tied events, although several approaches have been developed to approximate the tied events. The partial likelihood method does not permit us to analyze the form of time dependence directly. This method is relatively disadvantageous when the form of time dependence itself is of substantive interest (Yamaguchi, 1991). Finally, the partial likelihood method is based on weaker theoretical foundations than is the maximum likelihood method (Kalbfleisch and Prentice, 1980; Yamaguchi, 1991). In our application of the Cox model, the estimation of age-specific transition rates will be influenced by sample size (especially at older ages) and tied events as well.

7.2.2 The Gompertz regression model

The "Gompertz law" (Gompertz, 1825; 1827) of mortality states that the force of mortality (e.g. instantaneous death rate) increases exponentially with age because the 'resistance to death' declines with age. The change in resistance to death is a latent causal process that we cannot measure directly, only the effect (e.g death) can be measured (Willekens, 2001). The basic assumption of the Gompertz law is that the resistance to death declines exponentially with age. The basic notion is that at each time period or point, a person loses a constant fraction of his or her remaining 'vital force' or vitality. Gompertz summarized the effect of unobserved processes as a parametric form of age dependence. The Gompertz model dominated mortality forecasting for more than 100 years (Olshansky and Carnes, 1997), and it is also used in duration analysis in general (Blossfeld and Rohwer, 2002). In this study, we assumed that any disease or mortality transition (following the basic model of cardiovascular disease, Figure 3.1(b), Chapter 3) followed Gompertz's law. In

addition, all age-specific transition rates were assumed to depend on the incorporated covariates in the model. The transition rate at age x is given by the expression

$$\mu(x) = b\exp(cx) \tag{7.2}$$

Where, b and c are parameters to be estimated from the data. When c=0, the model trims down to the simple exponential model. The Gompertz model has two parameters and both of them can be used to include covariates. The model formulation for the transition rate from the origin state i to destination state j is

$$\mu_{ij}(x) = b^{(ij)}\exp(c^{(ij)}x) \tag{7.3}$$

where, $b^{(ij)} = \exp\{\beta^{(ij)}\}$. Following Blossfeld and Rohwer (2002), the associated coefficient vectors $\beta^{(ij)}$, linked exponentially and $c^{(ij)}$, linked linearly, are model parameters to be estimated. The $\beta^{(ij)}$ vector is linked exponentially to make sure that the estimated transition rate will not become negative. $c^{(ij)}$ is the shape parameter.

Following Rohwer and Potter (1999, TDA user's Manual), we estimated the parameters from the Gompertz regression model using the maximum likelihood method of estimation. Details of the derivation of the maximum likelihood estimates of the Gompertz model with and without covariates are given in Rohwer and Potter (1999).

Gompertz in the context

In this study, we considered the age at transition to an event or censoring as a time variable. The transition rates at each age interval depended on the age of an individual at that time and the associated covariates incorporated in the Gompertz model. Gompertz is one of the best known models, applied in a range of disciplines from botany to sociology. For mortality forecasting, Gompertz is still one of the best models in demographic literature. Since we estimated age-specific death rates either from non-cardiovascular disease or cardiovascular disease states, we assumed that the Gompertz mortality law is plausible in this case. As age increases (especially at older ages) the risk of cardiovascular disease increases irrespective of risk factor status at that age (Stamler et al., 1999). Therefore, the assumption that the shape of the transition rate from non-cardiovascular disease to cardiovascular disease follows Gompertz's law is also plausible. Some of the advantages and limitations of the Gompertz model are discussed here.

Advantages and disadvantages

The transition rates follow a Gompertz law, i.e. they increase exponentially with age, but the level and shape of the age-dependency is dependent on several covariates. Since our interest is to incorporate covariates, there will be an impact of these covariates. Therefore, we may assume that as more knowledge becomes available in the Gompertz model, the age dependency of the rate not only depends on the Gompertz law but also on individual characteristics (e.g. covariates). Manton et al. (1997) also argued that as more knowledge becomes available on disease and risk factors, the contribution of the Gompertz component is reduced. The simple Gompertz model without any covariates assumes that there is no important heterogeneity across individuals and each transition rate, as we are able to incorporate covariates with each parameter of this model that reduce the individual heterogeneity. The Gompertz regression model provides smooth transition rates and reduces the variability of measurement. The Gompertz function is monotonically increasing or decreasing. The Gompertz model can be used as a multidimensional context (Manton, et al., 1994).

7.2.3 Data source

We used the first 48 years of follow-up of the original Framingham Heart Study cohort consisting of 5209 respondents (45% male) from a sample of adults aged 28 through 62 years, residing in Framingham, Massachusetts between 1948 and 1951. The participants were tracked by standardised biennial cardiovascular examination, daily surveillance of hospital admissions, death information and information from physicians and other sources outside the clinic, ensuring highly accurate follow-up of death and clinically presenting cardiovascular disease. For this chapter, we used the data regarding age at onset of cardiovascular disease or death over forty-eight years of follow-up (exam rounds 1 to 24) for the 3481 participants who were free of cardiovascular disease at age 50.

For the purpose of illustration, we considered educational status as a covariate. We considered educational levels as a constant categorical covariate: *low education* (maximally high school without having graduated) and *high education* (at least graduation from high school). The educational status is described in Chapter 6.

7.3 Transition rates comparison

Age-specific transition rates were estimated using the Cox and the Gompertz regression model. The estimated rates were compared to the age-specific observed occurrence-exposure rates. The observed age-specific occurrence-exposure rates were estimated using the method described in Section 3.4.4 of Chapter 3. To illustrate, age-specific transition rates were estimated for the basic model type 3(a)

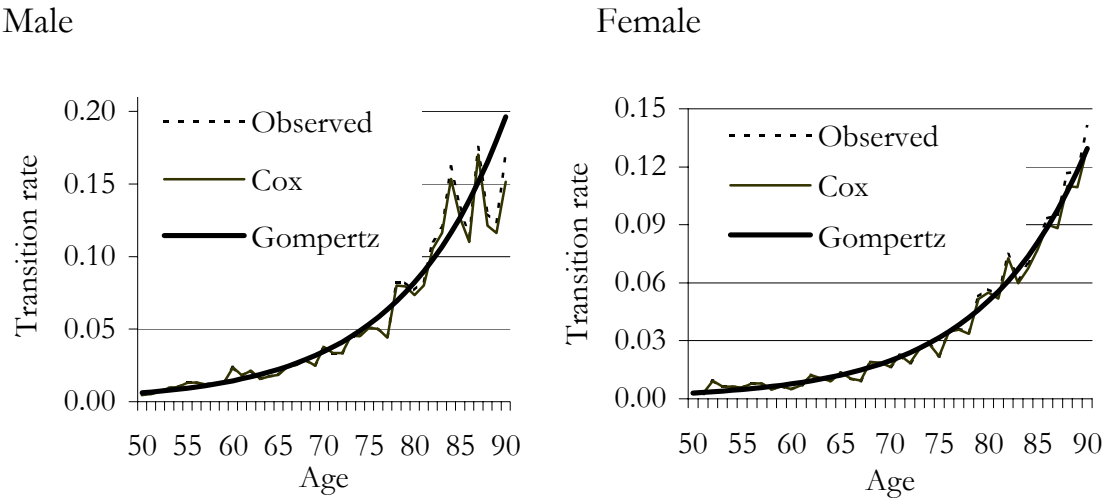
and 3(b) in Chapter 3, where, 3(a) was a two state (alive, dead) model and 3(b) a 3-state {NO-CVD, CVD, Dead} model. First, we compared different transition rates without any covariate i.e. the null model. Second, rates were compared with the presence of education, i.e. the model with covariate. Rates were compared from age 50 to 90.

7.3.1 Model without covariates (null model)

Transition rates *alive to death*

We used Cox and Gompertz models to estimate the age-specific death rates. The model estimates were compared to the observed rates in Figure 7.1. In the null model, we only estimated the baseline hazard. As expected, the estimated rates from the Cox model without any covariate was the same as the empirical rates, which showed that if the sample population was sufficiently large (before age 80), the Cox model and the observed rates were equal. Because of small sample size and estimation procedure, the Cox estimated transition rates at older ages differed slightly from the empirical rates. The estimated transition rates by the Gompertz model were smooth and increased exponentially. The Gompertz estimates fit quite well with both the empirical rates and the Cox model for males and females (Figure 7.1).

Figure 7.1 Comparison of age-specific death rates: Cox, Gompertz and observed occurrence-exposure rates, by sex

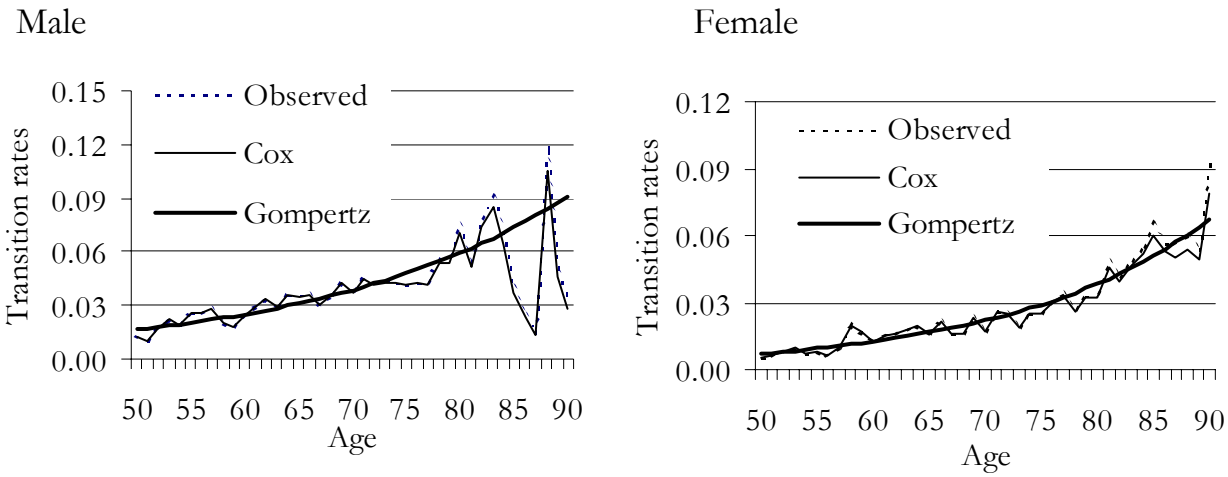


Transition rates *NO-CVD to CVD*

The age-specific transition rates NO-CVD to CVD of the model estimates and observed rates by sex are presented in Figure 7.2. Like mortality rates, cardiovascular disease incidence rates also increase as age increases. Both the

observed and the Cox rates show a similar pattern. The disease rates estimated applying the Gompertz model are shown to fit well compared to the Cox and the empirical rates.

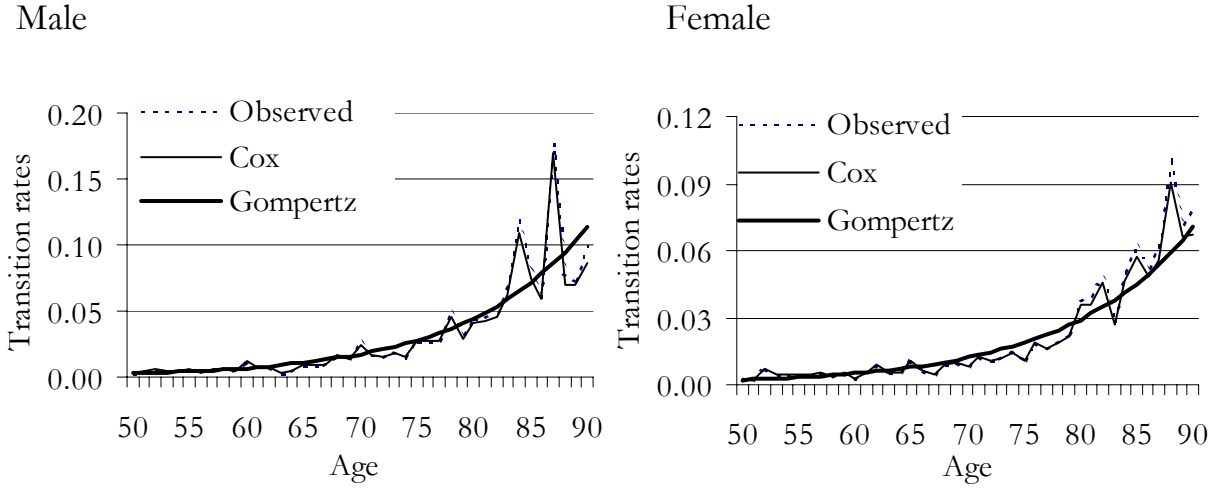
Figure 7.2 Comparison of age-specific rates of transition from NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex



Transition rates *NO-CVD to death*

The transition rates from NO-CVD to death are compared in Figure 7.3. The Cox null model fits with the empirical rates. At older ages, small differences were seen due to the limited number of transitions and the approximation of the Cox model (see Section 7.2.2, the limitation of Cox model). Like other transitions, Gompertz fits quite well for the NO-CVD to death transition.

Figure 7.3 Comparison of age-specific rates of transition NO-CVD to death: Cox, Gompertz and observed occurrence-exposure rates, by sex



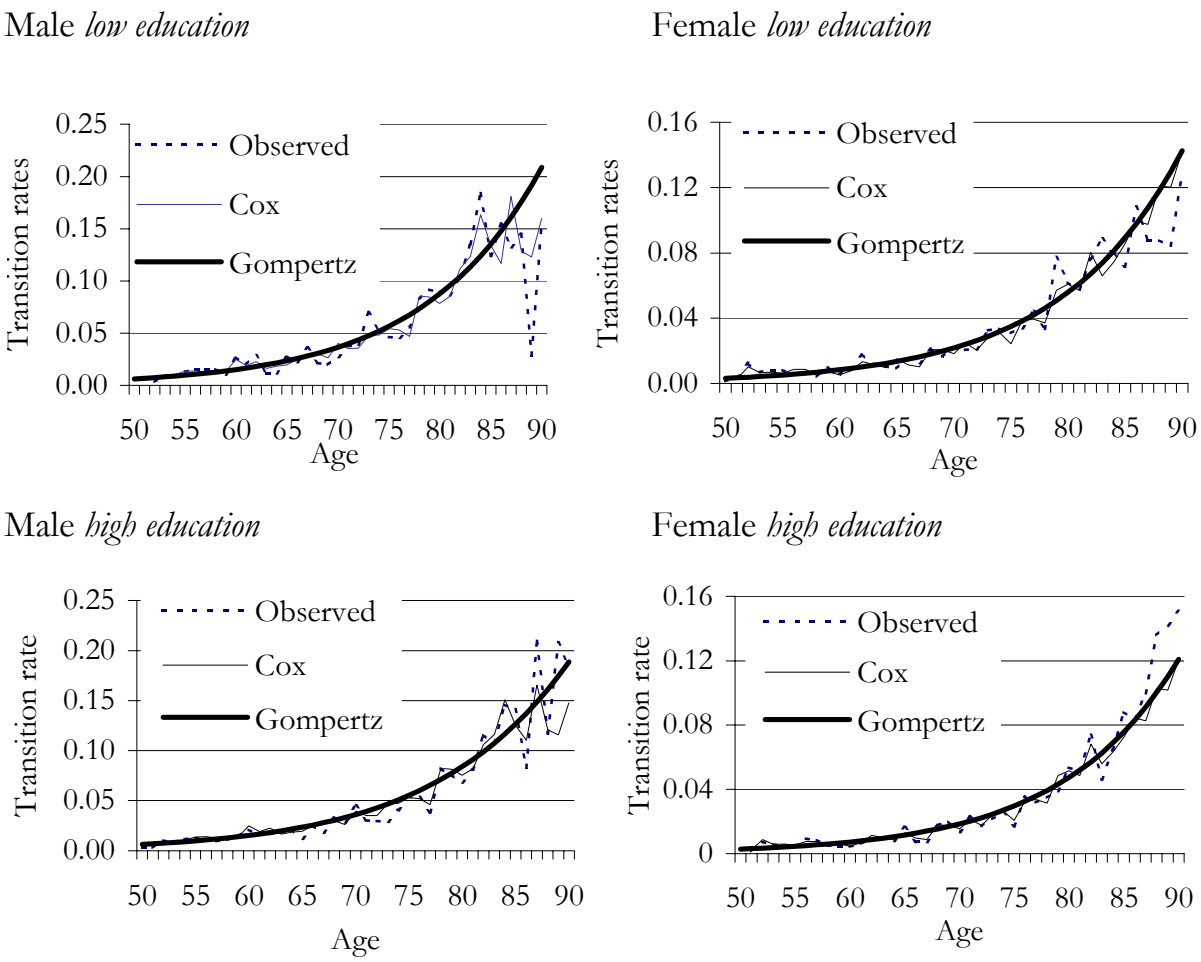
7.3.2 Model with covariates

The transition rates, described above were also compared by educational status. For the sake of simplicity, the educational levels are divided into the two categories low education and high education.

Transition rates *alive to death*

Incorporating educational status as a covariate, a comparison was then made between the model estimates and the empirical rates (see Figure 7.4). This figure clearly demonstrates that the age-specific transition rates estimated by the Cox model differ from the observed rates although the trend is same. The Cox estimates are less extreme than the observed rates because of the proportionality assumption. Given the exponential nature, Gompertz increased exponentially, despite the irregular changes of the observed rates and the Cox estimates. Mortality rates for the male low education group appear to level off after age 80, but the Gompertz estimates increase exponentially.

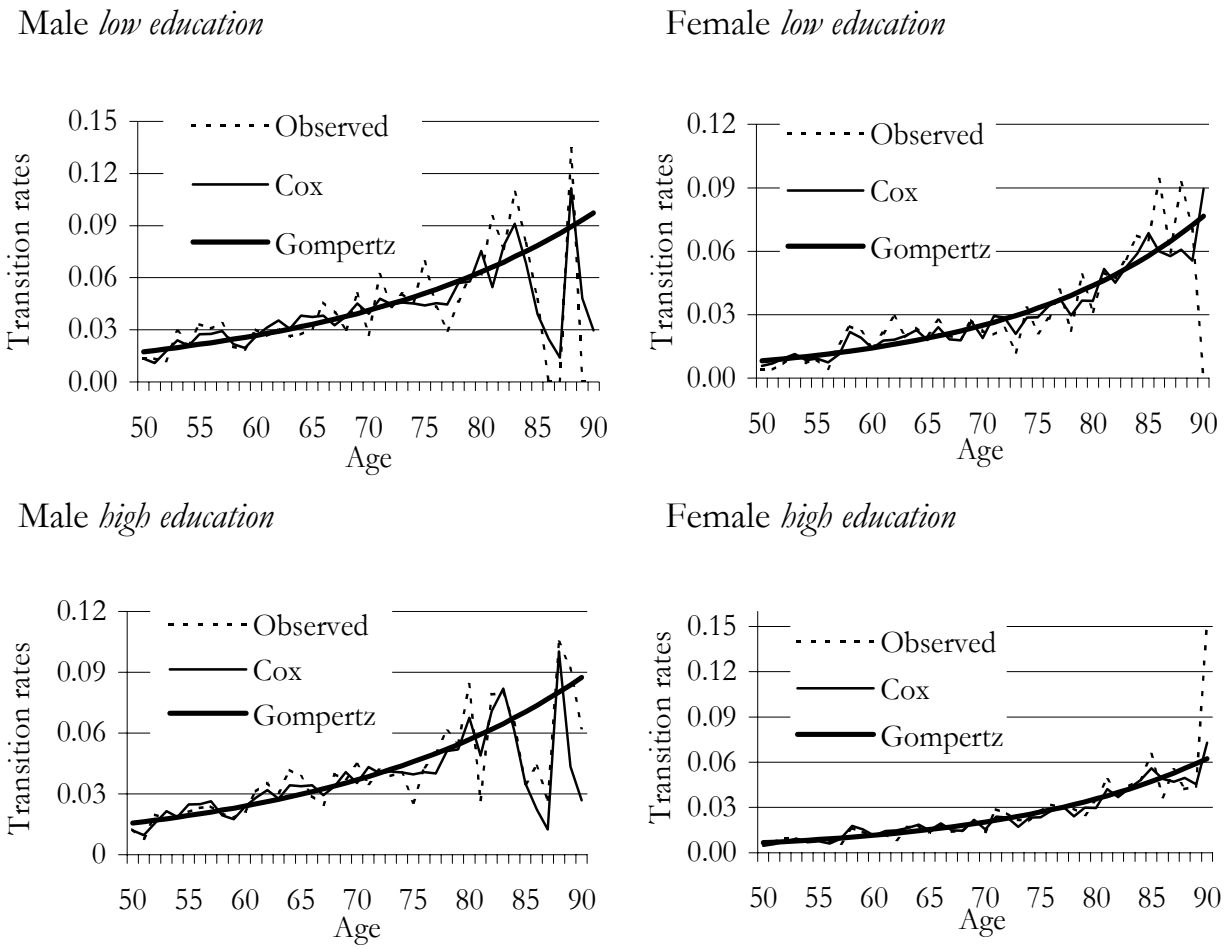
Figure 7.4 Comparison of age-specific death rates: Cox, Gompertz and observed occurrence-exposure rates, by sex and education



Transition rate *NO-CVD to CVD*

Similarly, the cardiovascular disease transition rates are compared by educational status in Figure 7.5. We observed the same behavior of the Cox model because of proportionality assumption. The Gompertz rates increased exponentially, as usual.

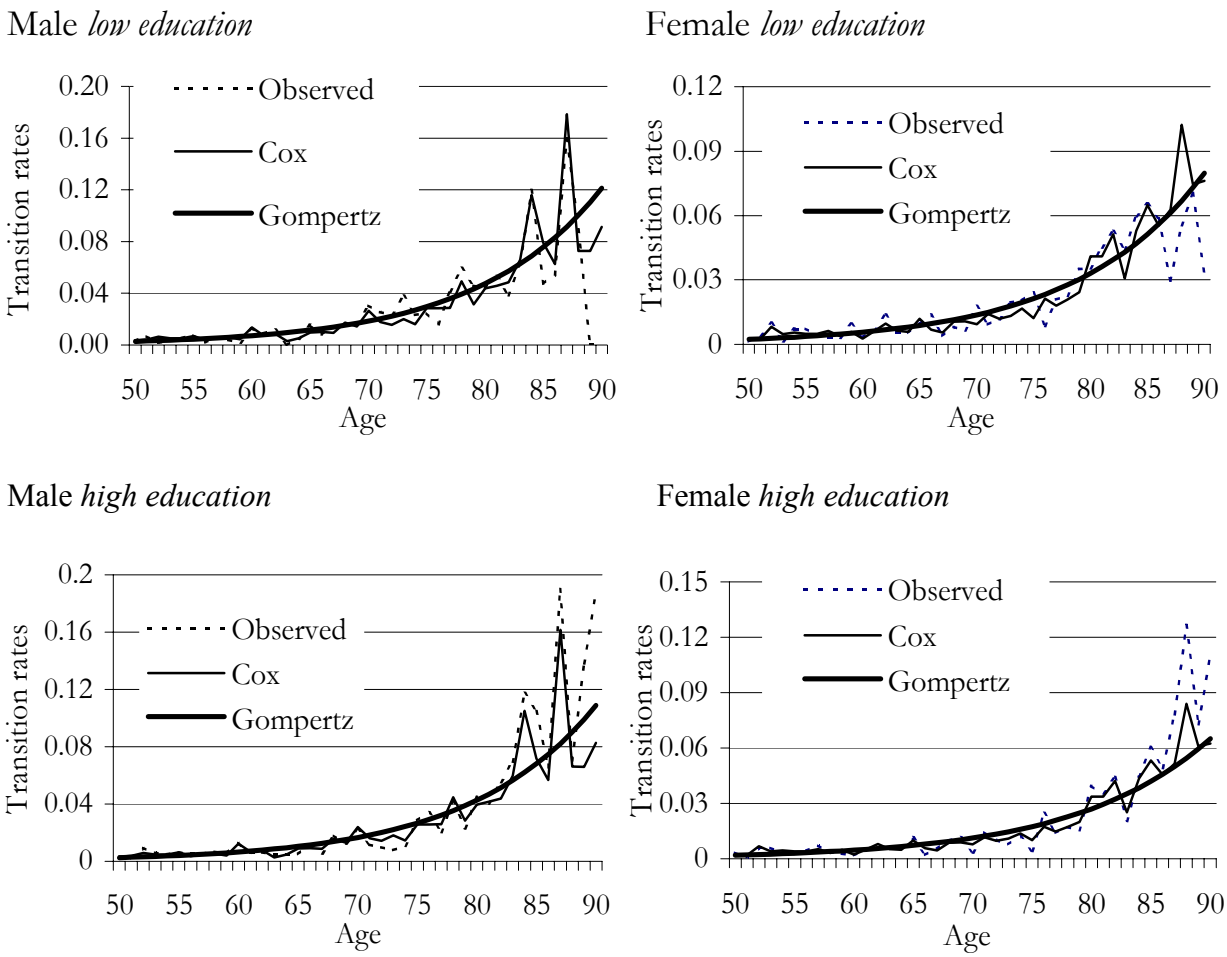
Figure 7.5 Comparison of age-specific rates of transition NO-CVD to CVD: Cox, Gompertz and observed occurrence-exposure rates, by sex and education



Transition *NO-CVD to death*

The transition rates from NO-CVD to death by educational status are shown in Figure 7.6. Both the observed and Cox estimates show irregular patterns. The Gompertz estimates increase exponentially by nature, whether the observed rates increase, decrease or remain at the same level (Figure 7.5, Female: high education). This irregular behavior of the observed rates was probably due to the small number of transitions or sample selection.

Figure 7.6 Comparison of age-specific transition rates of NO-CVD to death: Cox, Gompertz and observed occurrence-exposure rates, by sex and education



Similarly, we can also compare the age-specific transition rates for the transition from CVD to death. We did not estimate the post-disease transition rates for the Cox model, since our applied software (SPSS or TDA) was unable to handle delay entry or left censored cases, although estimating the baseline hazard for the Cox model was very simple. On the other hand, using a number of advanced statistical packages (STATA, S-plus), gave us the option to deal with delay entry or left censoring, but made it difficult to estimate the baseline hazard for the Cox model. To construct the basic CVD life table using the Cox model, CVD to death transition rates were replaced by the empirical rates.

7.4 Life table outcomes

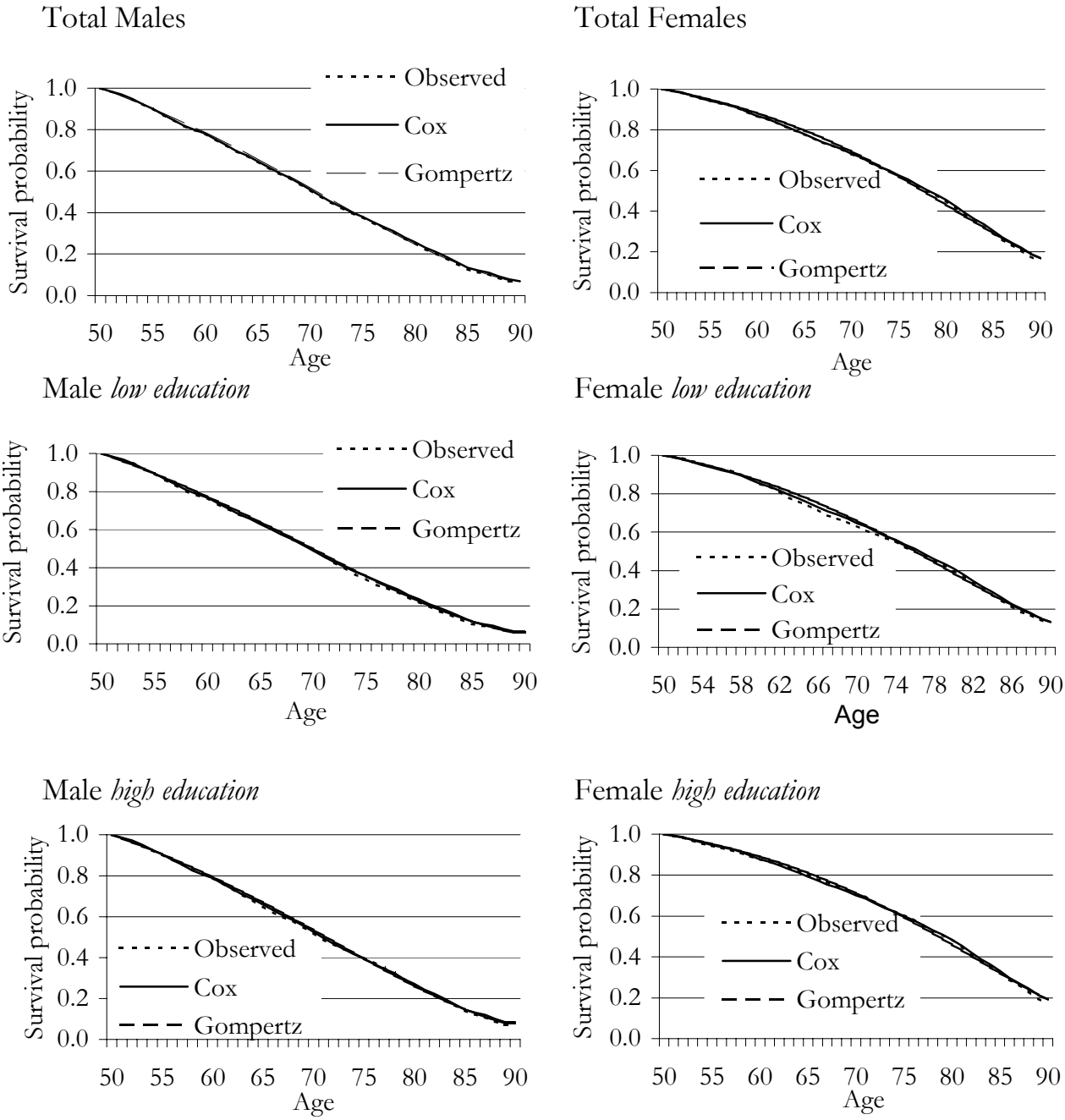
By comparing different transition rates, we were able to see how one curve fits better compared to another or how the outliers behaved over age. As was seen, both the Cox and the empirical rates had nearly the same outliers, while Gompertz gave smooth and exponentially increasing estimates of transitions rates. Here, we have constructed several life tables, using the age-specific transition rates obtained from the observed occurrence-exposure rates, Cox and Gompertz regression models. Some of the basic life table estimates such as survival curves, lifetime risk of developing cardiovascular disease and life expectancies are compared. Life tables are constructed following the procedure described in Section 3.3.4 of Chapter 3. Since CVD to death transition rates in the Cox model are replaced by the empirical rates, this might influence the life expectancy with CVD in the Cox model. We therefore used the MSLT to compare the life expectancy free of CVD, survival probability in NO-CVD state and lifetime probability of disease. Total life expectancies are presented from a single decrement (SDLT) life table.

Survival curves

The age profiles of survival free of cardiovascular disease for the FHS male and female participants in the MSLT population are shown in Figure 7.7. This figure shows the survival free of cardiovascular disease in fifty-year-old men and women by educational status.

As expected, the more highly educated men and women led longer lives and survived longer free of any cardiovascular disease than less well educated individuals. Overall, the fitted survival curves using the observed rates and model estimated rates were very close to each other. For instance, a marginal difference between the fitted survival curves is displayed in Figure 7.7 (Female low education), as they did not overlap at every age. However, this small difference was to be expected as the Cox and empirical transaction rates are neither smooth nor equal. At older ages, the survival curves estimated by the Gompertz regression model turn slightly downwards compared to the Cox and observed rates.

Figure 7.7 Survival curves illustrating the probability of surviving free of cardiovascular disease by age and educational status

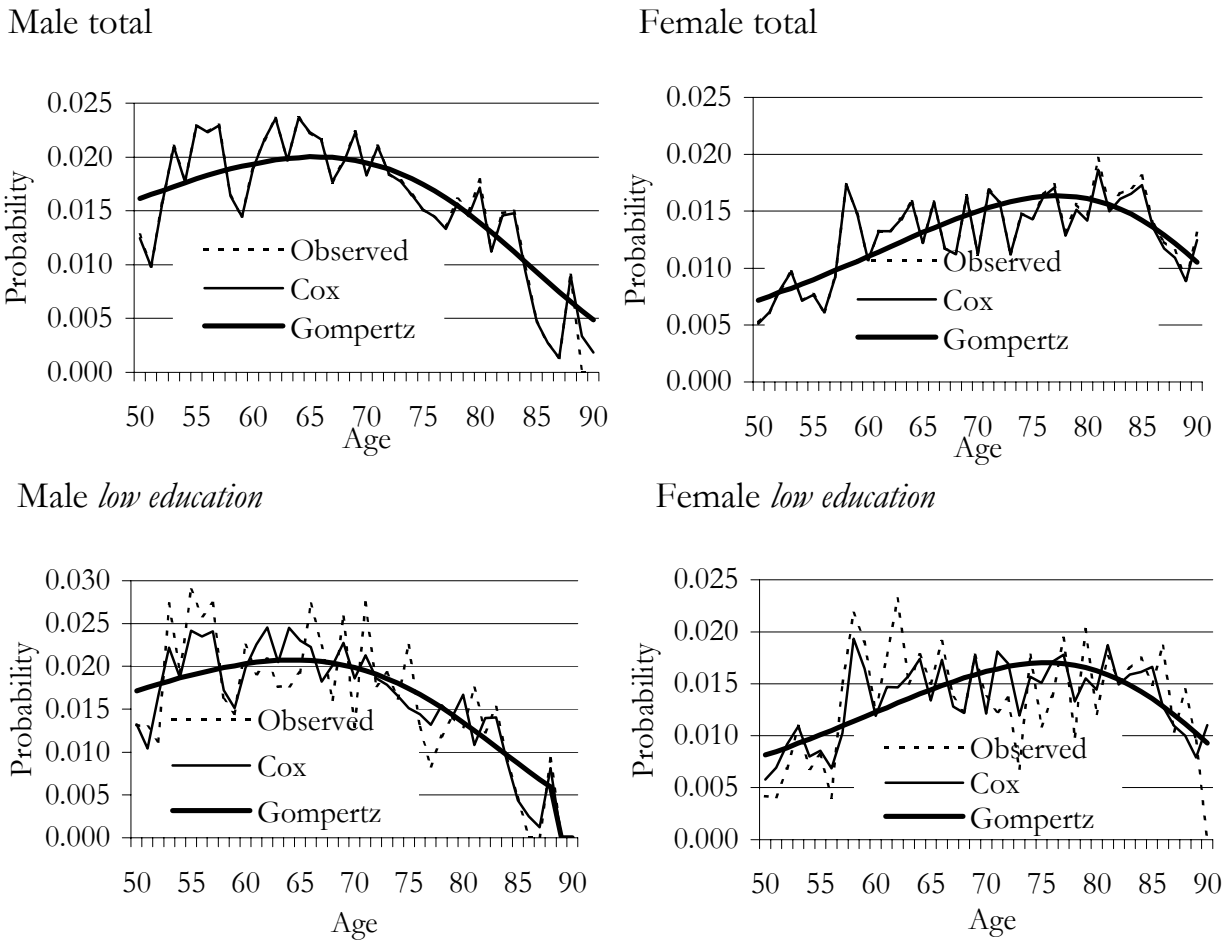


Lifetime probability

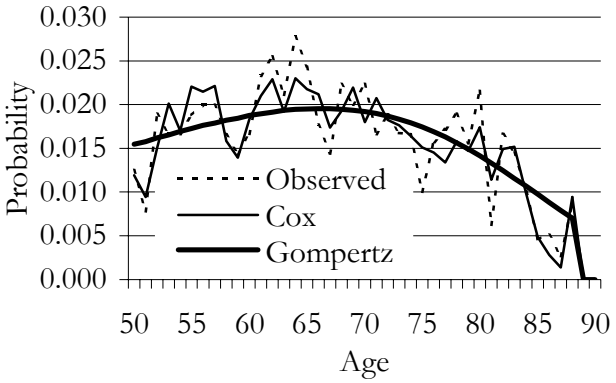
We compared the probability of cardiovascular disease-free people at age 50 developing cardiovascular disease by sex and education (Table 7.1 and Figure 7.8). The procedure to estimate the lifetime risk is the same as the procedure applied in Section 3.4.3 of Chapter 3.

The age-specific probability of developing cardiovascular disease is estimated by the Cox model to be the same as the observed estimate in the null model. Adding a covariate to the Cox model, causes the age-specific probability to jump less than the observed rates, because of the proportionality assumption of Cox model. The Gompertz estimate gives smooth age-specific probabilities of developing cardiovascular disease.

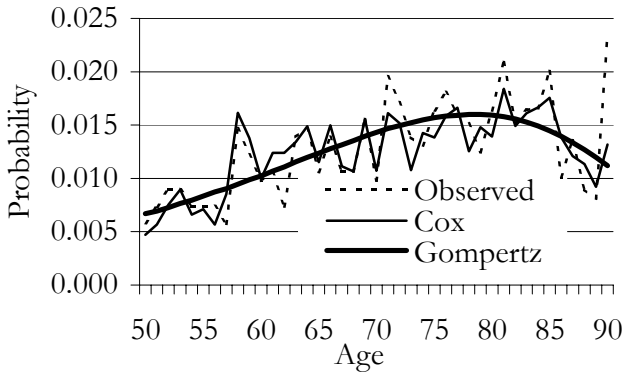
Figure 7.8 Age-specific probability of cardiovascular disease free people at age 50 developing cardiovascular disease, by sex and education



Male *high education*



Female *high education*



While Figure 7.8 only demonstrates the overall fit of the models compared to observed estimates, the probabilities of developing cardiovascular disease before age 70 and 90 are shown in Table 7.1. The probability of developing disease before age 90 according to both model estimates and the observed one (except high education females) was the same. Before age 70, the model estimate varied by one to two percent. The result also demonstrated that the probability of developing cardiovascular disease by educational status varied from one to three percent for males and one to seven percent for females.

Table 7.1 Lifetime probability that cardiovascular disease free people at age 50 develop cardiovascular disease, by sex and education (percentage)

	Before age 70			Before age 90		
	Observed	Cox	Gompertz	Observed	Cox	Gompertz
Male						
Total	41	40	39	64	64	65
Low education	42	42	41	65	65	65
High education	40	39	38	63	63	63
Female						
Total	24	24	23	53	53	53
Low education	28	26	26	55	55	55
High education	21	22	22	52	51	51

Life expectancy

In addition to survival probability and the risk of developing cardiovascular disease, the total life expectancy (from SDLT) and life expectancy free of cardiovascular disease (MSLT) at age 50 and 70 is presented in Table 7.2. Using the observed rates, a male at age 50 can expect to survive 26.90 additional years, while this was 27.01 years according to the model-derived rates. Similarly for females, the observed Cox and Gompertz estimates of life expectancy at age 50 were 32.03, 32.16 and 32.13 years respectively. As can be seen from this table, the model estimate of the total life expectancies and life expectancies free of cardiovascular disease at age 50 and 70 do not differ much. The results are comparable in the absence or presence of a covariate. A slight difference (0 to 0.5 years) continues mainly because of different assumptions of the models and sample size.

Table 7.2 Total life expectancy (SDLT) and life expectancy free of cardiovascular disease (MSLT)

		Age 50			Age 70		
		Total (null)	Low education	High education	Total	Low education	High education
Male							
Total (SDLT)	Observed	26.90	26.33	27.29	12.55	12.19	12.80
	Cox	27.01	26.40	27.43	12.69	12.30	12.98
	Gompertz	27.01	26.41	27.43	12.41	12.02	12.71
Free of CVD (MSLT)	Observed	20.83	20.12	21.32	7.40	6.80	7.80
	Cox	21.01	20.34	21.47	7.59	7.13	7.89
	Gompertz	20.99	20.37	21.52	7.45	7.07	7.84
Female							
Total (SDLT)	Observed	32.03	30.96	32.69	15.86	15.21	16.25
	Cox	32.16	31.28	32.47	16.02	15.41	16.12
	Gompertz	32.13	31.08	32.44	15.71	14.88	15.84
Free of CVD (MSLT)	Observed	26.60	25.13	27.51	11.16	10.10	11.78
	Cox	26.76	25.46	27.61	11.21	10.38	11.95
	Gompertz	26.82	25.49	27.63	11.10	10.10	11.70

7.5 Discussion

The aim of this chapter was to estimate the age-specific transition rates applying Cox and Gompertz regression models, and to assess the model estimates by comparing them to the observed occurrence-exposure rates. The different estimates of the rates were subsequently transferred into life table outcomes. In the null model, both the Cox and the observed rates were similar. Adding covariates to the Cox model did not cause the estimated age-specific rates to overlap with the observed rates, because of the proportionality assumption. The Gompertz estimates behaved the same way in both the null model and the model with covariates.

When the available data on the exact timing of events and occurrences of events followed a specific shape, a parametric model could be used instead of a semi-parametric Cox model. The Cox null model could be applied instead of observed rates when the number of occurrences at each age group was sufficiently large. In that situation, the Cox null model and the observed rates were exactly the same. The estimation of the Cox null model was simple, easy and software is available.

The Gompertz model smoothed the rates exponentially. When applying the Cox model and estimating observed rates, we could only predict the rates within the follow-up time. For the Cox estimate, we needed a baseline hazard to delineate the predicted rates. Gompertz could be used both for macro data and micro data. Cox could be used mainly for the micro data. Gompertz had fewer parameters. Cox provided partial likelihood estimates, which is theoretically weaker. The precision of the partial likelihood estimates of parameters can be much less than that for the maximum likelihood estimates when the sample size is small (Coleman, 1981). When we estimated the age-specific transition rates at older ages the sample population was small. Because of the small sample size and partial likelihood method, the transition rates at older ages were not same as the observed rates.

The Cox proportional hazard model depends on the proportionally assumption. It does not have any untenable distributional assumptions. The main advantages of the Gompertz estimate are the smooth estimates that are more interpretable and easy to apply for the life table construction. Since Gompertz provides smooth estimates, the variability of the results will be less compared to the Cox method and the observed rates. However, to estimate the transition rates and accordingly construct a life table for the cardiovascular disease history in the FHS, a Gompertz model with or without covariate fits well. Using Gompertz regression, the rates could be predicted for any ages. Since the variability of the estimated transition rates in the Gompertz model is less than the Cox or empirical rates, life table estimates using Gompertz in the presence or absence of covariate is recommended.

References

Blossfeld HP, Rohwer G, (2002). *Techniques of Event History Modeling. New Approaches to Causal Analysis*. Second edition, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Coleman JS, (1981). *Longitudinal Data Analysis*. Basic Books, New York.

Cox DR, (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, 34:187-220.

Cox DR, (1975). Partial likelihood. *Biometrika*, 62:269-276.

Cox DR, Oakes DO, (1984). *Analysis of Survival Data*. Chapman and Hall, London.

Efron B, (1977). The efficiency of Cox's likelihood function for censored analysis. *Journal of the American Statistical Association*, 72: 557-565.

Fox J, (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, New York.

Gompertz B, (1825). On the nature of the function expressive of the law of human mortality, and on a new method of determining the value of life contingencies. *Philosophical Transactions of the Royal Society*, 115:513-585.

Gompertz B, (1827). On one union law of mortality from birth to extreme old age, and on the law of sickness. *Journal of the Institute of Actuaries*, 16:329-344

Kalbfleisch JD, Prentice RL, (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Kleinbaum DG, (1996). *Survival Analysis. A Self-Learning Text*. Springer, New York.

Manton KG, Stallard E, Woodbury MA, Dowd JED, (1994). Time-varying covariates in models of human mortality and aging: multidimensional generalizations of the Gompertz. *Journal of Gerontology: Biological Sciences*, 49(4):169-90.

Manton KG, Stallard E, Corder L, (1997). The limits of longevity and their implications for health and mortality in developed countries. Paper presented at the United Nations Symposium on Health and Mortality, Brussels, 19-22 November 1997.

Nusselder W, (1998). *Compression of Morbidity? A Life-table Approach*. Ph.D. Dissertation, Erasmus University, Rotterdam, The Netherlands.

Oakes D, (1977). The asymptotic information in censored data. *Biometrika*, 64:441-448.

Olshansky SJ, Carners BA, (1997). Ever since Gompertz. *Demography*, 34(1):1-15.

Rohwer G, Pötter Y, (1999). TDA User's manual. Version 1. Ruhr-universität Bochum. http://www.stat.ruhr-uni-bochum.de/

Stamler J, Stamler R, Neaton JD, et al., (1999). Low risk-factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy. *Journal of American Medical Association,* 282(21):2012-2018.

Therneau TM, Grambsch PM, (2000). *Modeling Survival Data: Extending The Cox Model*. Springer, New York.

Willekens F, (2001). Gompertz in context: the Gompertz and related distributions. In: Tabeau E, Anneke van den B Jeths, Heathcote C, (eds.), *Forecasting Mortality in Developed Countries- Insights from a Statistical, Demographic and Epidemiological Perspective*, Kluwer Academic Publishers, Dordrecht.

Wong WH, (1986). Theory of partial likelihood. *Annals of Statistics*, 14:88-123.

Yamaguchi K, (1991). *Event History Analysis*. Applied social series methods series volume 28. Sage Publications, Newbury Park.

7.1 Introduction

Survival analysis examines and models the duration of events of interest. It focuses on the distribution of survival times. While there are well known methods for estimating survival distributions (e.g. Kaplan-Meier estimate), the most interesting method to model survival examines the relationship between survival and one or more predictors, usually termed covariates (Fox, 1997). In this chapter, we have focused on two survival models: the semi-parametric *Cox* regression model (Cox, 1972; 75) and the parametric *Gompertz* regression model (Gompertz, 1825; 1827). The Cox regression model is the model on which modern survival analysis is founded and is widely used in numerous fields. The parametric Gompertz regression model has dominated mortality analysis for over 100 years and has been applied in a range of disciplines from botany to sociology. To construct a life table (with or without covariates), we needed age-specific transition rates. Transition rates can be derived directly from the data set (e.g. occurrence-exposure rates) or can be estimated by fitting regression models to the data. However, the age-specific transition rates obtained from the Cox and Gompertz models in the presence or absence of covariates and the consequences for the life table estimates have not been assessed. In this chapter, an assessment of these two widely used models is made by comparing the estimated rates with empirical age-specific *occurrence-exposure* rates.

The Cox proportional hazard model is the most general of the regression models, because it is not based on any assumptions concerning the nature or shape of the underlying distribution. The basic assumption of this model is that the underlying transition rate (rather than the survival time) is a function of the independent variables (covariates). No assumptions are made about the shape of the hazard function (Kleinbaum, 1996; Blossfeld and Rohwer, 2002; Therneau and Grambsch, 2000). Proportional hazard models consider the transition rates at each time among those subjects who have not failed. The predicted transition rates using the Cox model gives an efficient estimate, if proportionality exists, as compared to a parametric proportional hazards model, such as the Weibull, even when the data actually come from the parametric model (Kleinbaum, 1996). To know the effect of covariates on the occurrences of event of interest, researchers mostly estimate the hazard ratios. However, the estimated age-specific rates, and accordingly the life table outcomes using Cox model with age as time scale, in the presence or absence of covariates, has not been compared with empirical age-specific rates.

While the Cox model does not have any pre-specified form of the baseline hazard, the Gompertz function follows an exponential form (increasing or decreasing). The Gompertz model has fewer parameters and some other facilities. Both the Cox and Gompertz models have advantages and disadvantages (described in Section 7.2). One advantage of the Gompertz model is that it produces smooth