

# Models for Estimating “Cure” from Cancer

Paul C Lambert  
Center for Biostatistics and Genetic Epidemiology  
University of Leicester  
paul.lambert@le.ac.uk

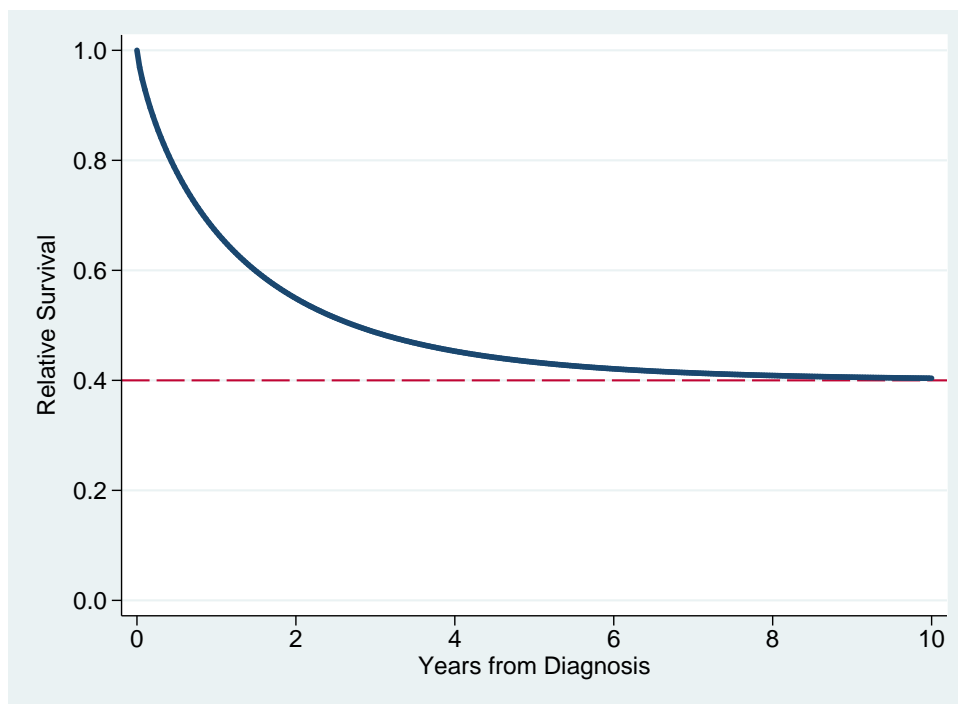
Workshop on Methods for Studying Cancer Patient Survival with  
Application in Stata  
Stockholm, 6th September 2007



## What is Cure?

- **Medical Cure** is defined as when all signs of cancer have been removed in a patient. This is usually when all malignant cells have been eliminated.
- This is an individual level definition of cure.
- It is difficult to assume that a patient is medically cured.
- **Population** or **Statistical Cure** occurs when the mortality (hazard) rate associated with the disease returns to the same level as that expected in the general population.
- Equivalently the excess mortality rate approaches zero.
- This is a population level definition of cure.
- When the excess mortality reaches (and stays) at zero, the relative survival curve is seen to reach a plateau.

## Plateau for Relative Survival



## Modelling Excess Mortality

- Usually model on the log excess hazard (mortality) scale

### Relative Survival Models

$$\lambda(t) = h^*(t) + \lambda_e(t)$$

$$\text{Observed Mortality Rate} = \text{Expected Mortality Rate} + \text{Excess Mortality Rate}$$

- where  $\lambda_e(t)$  denotes the excess mortality (hazard) rate.
- Cure Models assume that the excess mortality rate has an asymptote at zero.
- This implies that the relative survival function will also have an asymptote - at the cure fraction.

## Mixture and Non-Mixture Models

### Relative Survival Models

$$S(t) = S^*(t)R(t)$$
$$\lambda(t) = h^*(t) + \lambda_e(t)$$

- When modelling cure we define an asymptote at the cure fraction,  $\pi$ , for the relative survival function,  $R(t)$ .
- The excess mortality rate,  $\lambda_e(t)$ , has an asymptote at zero.
- Two main approaches

### Mixture Model

$$S(t) = S^*(t)(\pi + (1 - \pi)S_u(t))$$

### Non Mixture Model

$$S(t) = S^*(t)\pi^{F_z(t)}$$

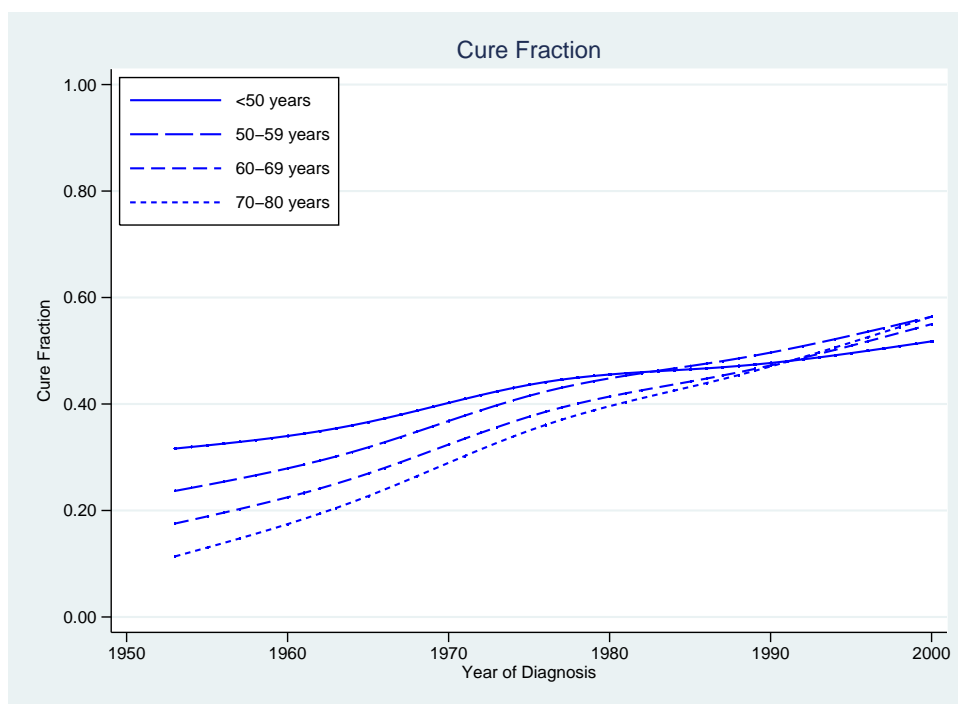
## Mixture Model

### Mixture Model

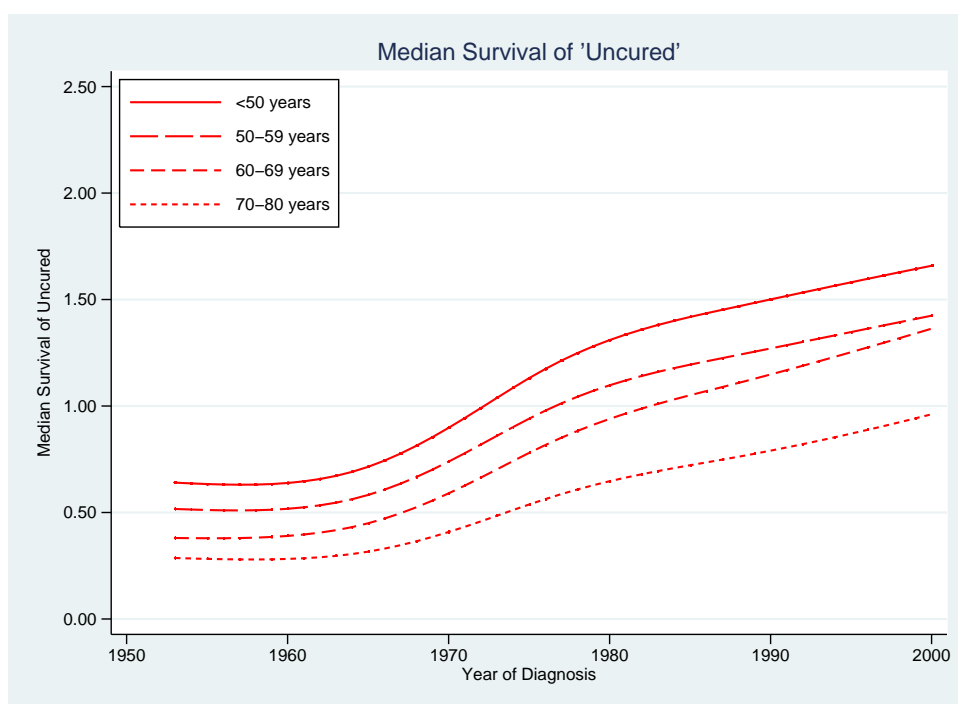
$$S(t) = S^*(t)(\pi + (1 - \pi)S_u(t)) \quad \lambda(t) = h^*(t) + \frac{(1-\pi)f_u(t)}{\pi+(1-\pi)S_u(t)}$$

- $S^*(t)$  is the expected survival.
- $\pi$  is the proportion cured (the cure fraction).
- $(1 - \pi)$  is the proportion 'uncured' (those 'bound to die').
- $S_u(t)$  is the survival for the 'uncured' group.
- The excess mortality rate has an asymptote at zero.
- See (De Angelis et al., 1999), (Verdecchia et al., 1998) and (Lambert et al., 2007b) for more details.
- see (Lambert et al., 2007a) for application of this model to Finnish colon cancer data

## Trends in the Proportion Cured of Colon Cancer in Finland



## Trends in Median Survival of the Uncured



## Non-Mixture Model 1

- The non-mixture model has been used for 'standard' cure models (not incorporating background mortality) Sposto (2002).
- We have extended the non-mixture model to incorporate background mortality and thus be fitted to relative survival data (Lambert et al., 2007b).

### Non Mixture Model

$$S(t) = S^*(t)\pi^{F_z(t)} \quad \lambda(t) = h^*(t) - \ln(\pi)f_z(t)$$

- If parameters in  $f_z(t)$  do not vary by covariates then this is a proportional excess hazards model.
- The non-mixture model can also be written as;

$$S(t) = S^*(t) \left( \pi + (1 - \pi) \left( \frac{\pi^{F_z(t)} - \pi}{1 - \pi} \right) \right)$$

- This is a mixture cure fraction model and thus the survival function of 'uncured' patients can also be obtained from a non-mixture model by a simple transformation of the model parameters

## Non-Mixture Model 2

### Non Mixture Model

$$S(t) = S^*(t)\pi^{F_z(t)} \quad \lambda(t) = h^*(t) - \ln(\pi)f_z(t)$$

- At time zero  $F_z(t) = 0$  and thus the relative survival,  $R(t) = 1$ .
- At time  $\infty$   $F_z(t) = 1$  and thus the relative survival,  $R(t) = \pi$ .
- The non-mixture model (without incorporation of background mortality) was developed for the the modelling of tumour recurrence where the cure fraction is the probability no cancerous cells remain.
- However, it can be considered as a useful mathematical function with an asymptote that can be used to estimate the cure fraction even if the data do not 'fit' the biological definition as long as it is reasonable to assume cure (Ibrahim et al., 2001).

## The strsmix and strsnmix commands

- See (Lambert, 2007) for details.
- strsmix mixture cure fraction models.
- strsnmix non-mixture cure fraction models.
- Options include;
  - distribution - weibull, lognormal, (generalized) gamma.
  - link - identity, logistic or loglog
  - bhazard(*varname*) - baseline hazard rate at event time.
- Various prediction options including
  - Proportion cured.
  - Overall (relative) survival function.
  - (Relative) survival function of 'uncured' group.
  - Overall excess hazard function.
  - Overall excess hazard function for 'uncured' group.
  - Centiles (e.g. median) of survival distribution of 'uncured' group.
  - Probability of cure as a function of time.
  - etc.

## Example of strsmix using Finnish Colon Data

### Stata Code

```
. strsmix if agegrp == 3, dist(weibull) link(identity) bhazard(brate)
initial:      log likelihood = -13734.336
alternative:  log likelihood = -7779.3205
rescale:     log likelihood = -7779.3205
rescale eq:  log likelihood = -7023.9247
Iteration 0: log likelihood = -7023.9247
Iteration 1: log likelihood = -7018.4665
Iteration 2: log likelihood = -7018.4619
Iteration 3: log likelihood = -7018.4619

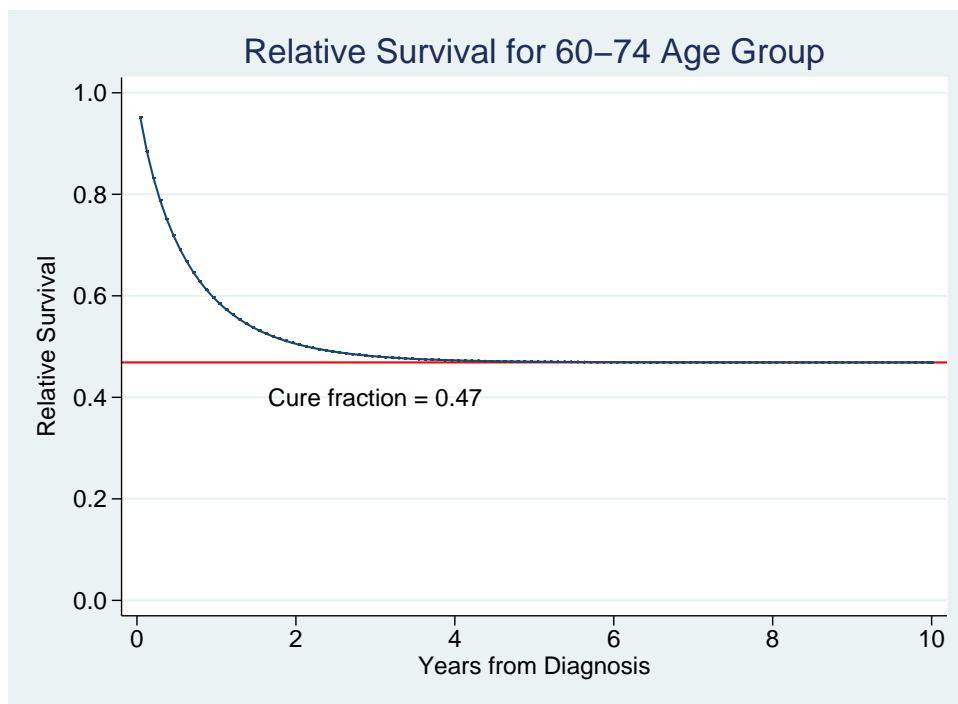
                    Number of obs   =      5868
                    Wald chi2(0)     =          .
                    Prob > chi2      =          .

Log likelihood = -7018.4619
```

	_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pi						
	_cons	.4687827	.0102975	45.52	0.000	.4485999 .4889655
ln_lambda						
	_cons	.386464	.0423715	9.12	0.000	.3034174 .4695107
ln_gamma						
	_cons	-.148694	.0206601	-7.20	0.000	-.1891871 -.108201

```
. predict rs if e(sample), survival
. predict rs_unc if e(sample), survival uncured
```

## Predicted Relative Survival



## Median Survival of the 'Uncured'

- A useful summary of the survival of the 'uncured' is the median survival of the survival distribution of the 'uncured'.
- In the mixture model the median survival can be obtained by setting the relative survival function for the 'uncured' to 0.5 and solving for  $t$ .
- When using the Weibull Distribution

### Median Survival of the Uncured

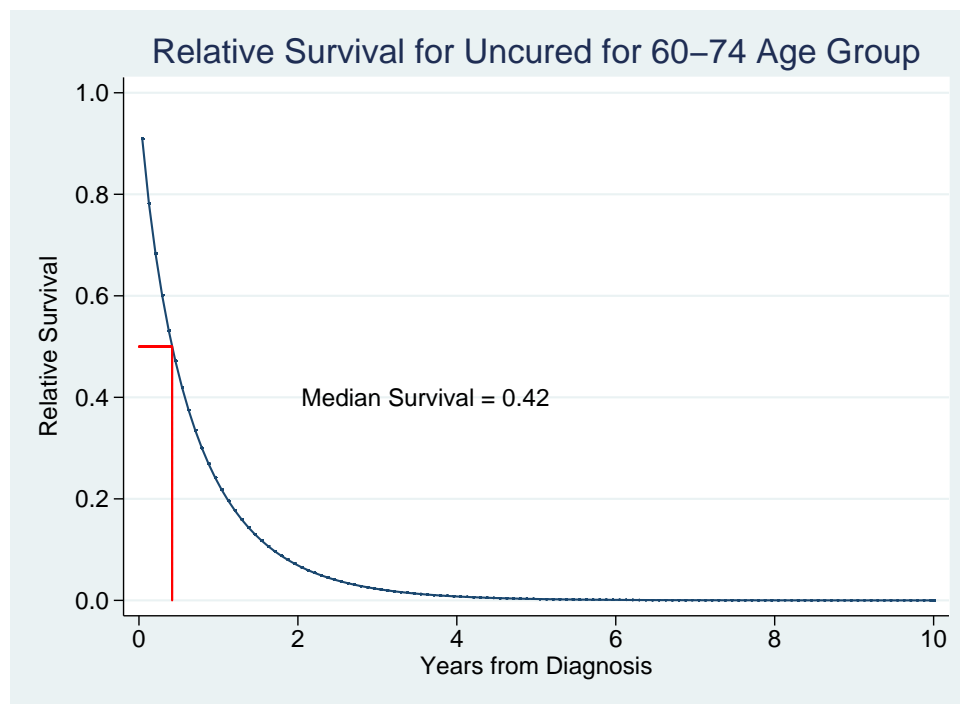
$$R(t) = \exp(-\lambda t^\gamma)$$

$$0.5 = \exp(-\lambda t^\gamma)$$

$$t_{Median} = \left( \frac{-\ln(0.5)}{\lambda} \right)^{\frac{1}{\gamma}}$$

- Centiles other than the median are easily obtained.
- Standard errors can be obtained using the delta-method.

## Predicted Relative Survival for Uncured



## Some Issues with Cure Models

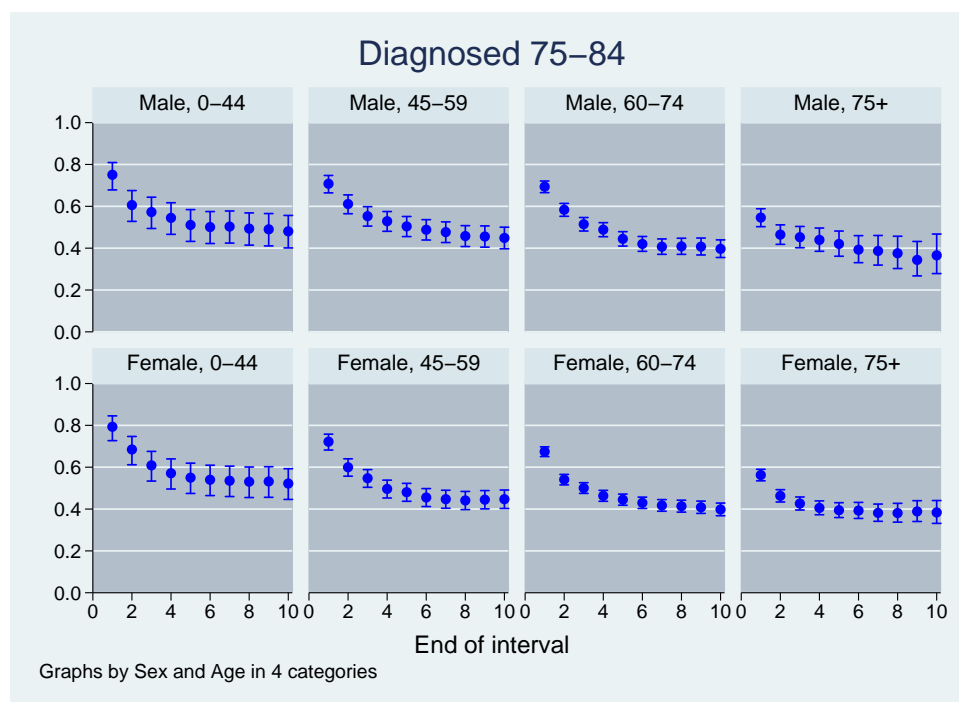
- Is cure a reasonable assumption?
- How do define time to 'cure' ?
- Modelling the Weibull parameters.
- Modelling continuous variables continuously.
- Mixture or Non-Mixture Models?
- More complex parametric distributions - when should they be used?
- Up-to-Date estimates of cure.



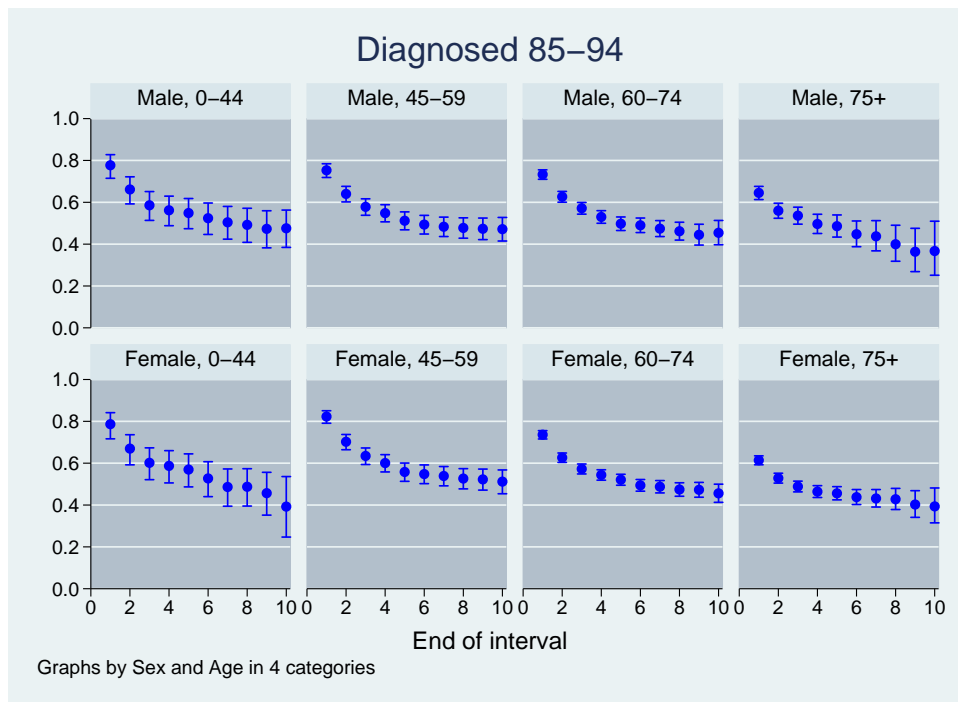
## Is Cure a Reasonable Assumption?

- You can try to fit cure models to any survival data, but if cure is not a reasonable assumption then the results may be meaningless.
- If the (relative) survival curve has not reached a plateau then the estimate of cure (and the survival distribution of the 'uncured' group) will be based on extrapolation beyond the range of the data.
- Before fitting a cure model investigate life table estimates of relative survival.

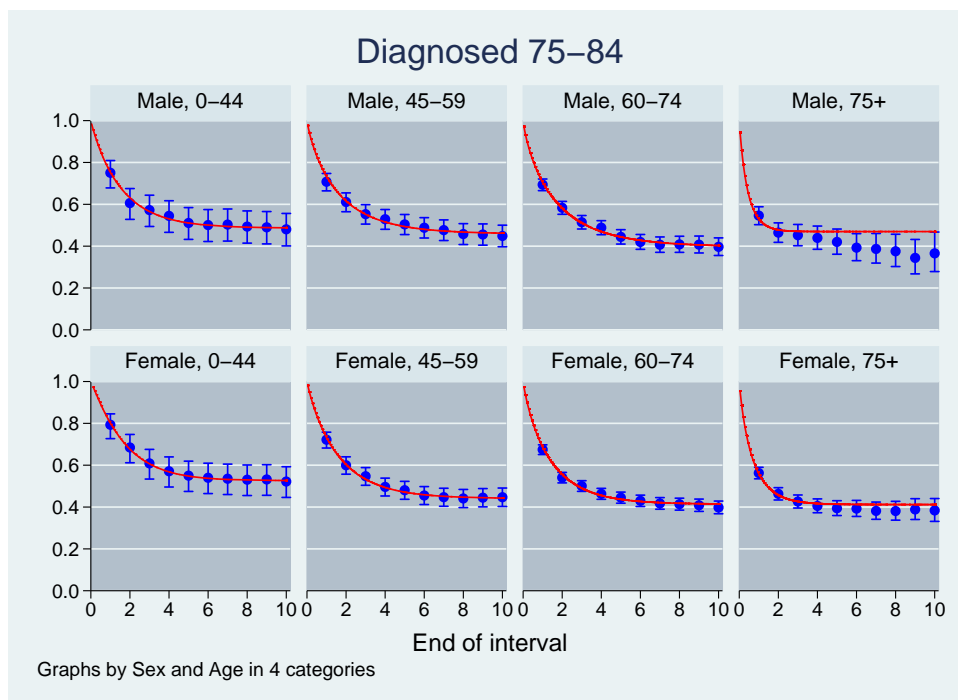
## Colon Cancer: Diagnosed 1975-1984



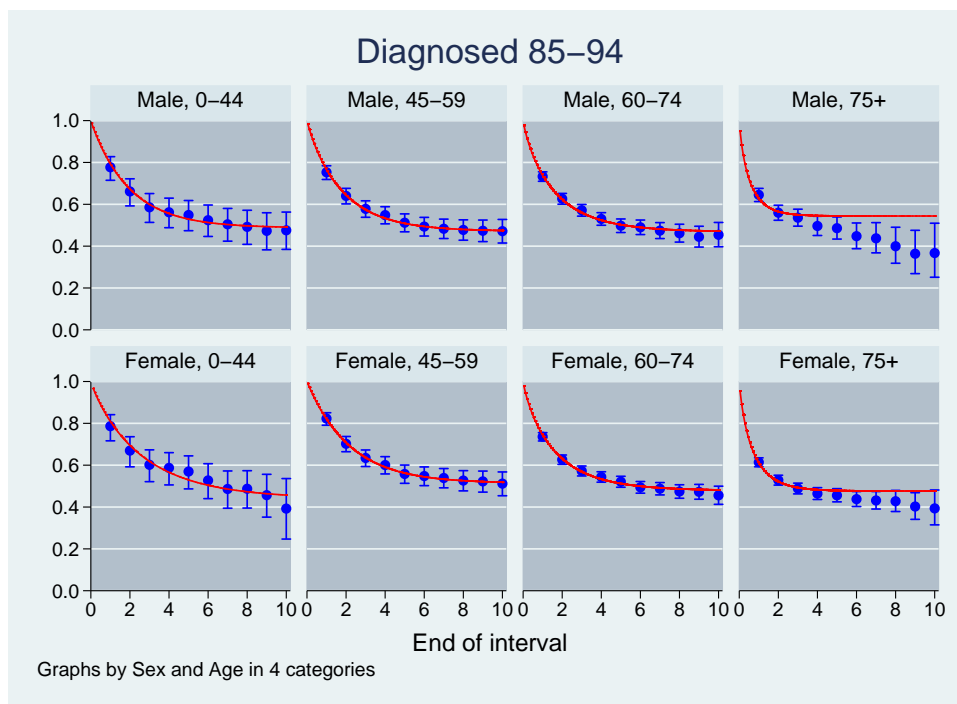
# Colon Cancer: Diagnosed 1985-1994



# Colon Cancer: Diagnosed 1975-1984



## Colon Cancer: Diagnosed 1985-1994



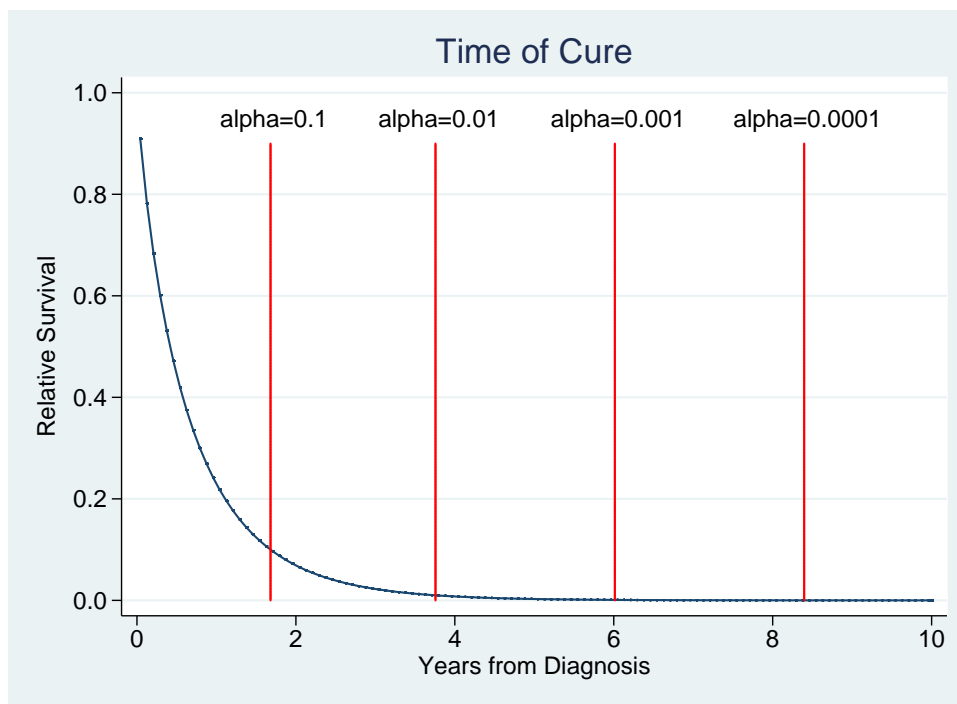
## Estimating Time to Cure 1

- A sensible question is to ask “When does cure occur”.
- However, this is a difficult question answer.
- This is because in the cure models define cure as occurring at time  $t = \infty$ .
- This is not useful!
- However, we can ask when is the Relative Survival function of the ‘uncured’ group below a stated amount (e.g. 0.1, 0.01, 0.001).
- This is essentially the same calculation as the for the median survival of the uncured where we were interested in when the Relative Survival function of the ‘uncured’ group was 0.5.
- If  $\alpha$  is the proportion of the ‘uncured’ group still alive then the time to ‘cure’ is

### Time to Cure

$$t_{Cure} = \left( \frac{-\ln(\alpha)}{\lambda} \right)^{\frac{1}{\gamma}}$$

## Estimating Time to Cure 2



## Modelling the Survival Distribution Parameters

- Failure to model the parameters in the parametric (Weibull) distribution can result in biased estimates of the cure fraction.
- For the Weibull, one ( $\lambda$ ) or both ( $\lambda$  or  $\gamma$ ) can vary by covariates.
- The following two slides show the differences in the estimate of the cure fraction and median survival of the 'uncured' for the mixture cure fraction models.
  - No modelling of  $\lambda$  or  $\gamma$ .
  - Modelling of  $\lambda$  but not  $\gamma$ .
  - Modelling of both  $\lambda$  and  $\gamma$ .

```
strsmix agegrp2-agegrp4 female year8594, dist(weibull) link(logit) bhaz(brate)
```

```
strsmix agegrp2-agegrp4 female year8594, dist(weibull) link(logit) bhaz(brate) ///  
k1(agegrp2-agegrp4 female year8594)
```

```
strsmix agegrp2-agegrp4 female year8594, dist(weibull) link(logit) bhaz(brate) ///  
k1(agegrp2-agegrp4 female year8594) k2(agegrp2-agegrp4 female year8594)
```

## Comparison of Proportion Cured

### Compare Proportion Cured

Age in 4 categories	Year of diagnosis 1985-94 and Sex			
	— Diagnosed 75-84 —		— Diagnosed 85-94 —	
	Male	Female	Male	Female
0-44	0.475	0.474	0.545	0.544
	0.468	0.456	0.523	0.511
	0.475	0.468	0.527	0.520
45-59	0.455	0.453	0.525	0.524
	0.447	0.435	0.502	0.490
	0.450	0.442	0.502	0.494
60-74	0.419	0.418	0.489	0.488
	0.425	0.413	0.480	0.467
	0.420	0.413	0.471	0.464
75+	0.331	0.330	0.396	0.395
	0.448	0.435	0.502	0.490
	0.441	0.434	0.493	0.486

Line 1 -  $\lambda$  and  $\gamma$  not modelled,

Line 2 -  $\lambda$  modelled,

Line 3 -  $\lambda$  and  $\gamma$  modelled .

## Comparison of Median Survival for 'Uncured' Group

### Compare Median Survival

Age in 4 categories	Year of diagnosis 1985-94 and Sex			
	— Diagnosed 75-84 —		— Diagnosed 85-94 —	
	Male	Female	Male	Female
0-44	0.846	0.846	0.846	0.846
	1.115	1.225	1.257	1.381
	1.160	1.236	1.311	1.393
45-59	0.846	0.846	0.846	0.846
	1.044	1.147	1.176	1.292
	1.066	1.144	1.212	1.296
60-74	0.846	0.846	0.846	0.846
	0.870	0.957	0.981	1.078
	0.878	0.954	1.003	1.087
75+	0.846	0.846	0.846	0.846
	0.360	0.395	0.405	0.445
	0.360	0.405	0.403	0.452

Line 1 -  $\lambda$  and  $\gamma$  not modelled,

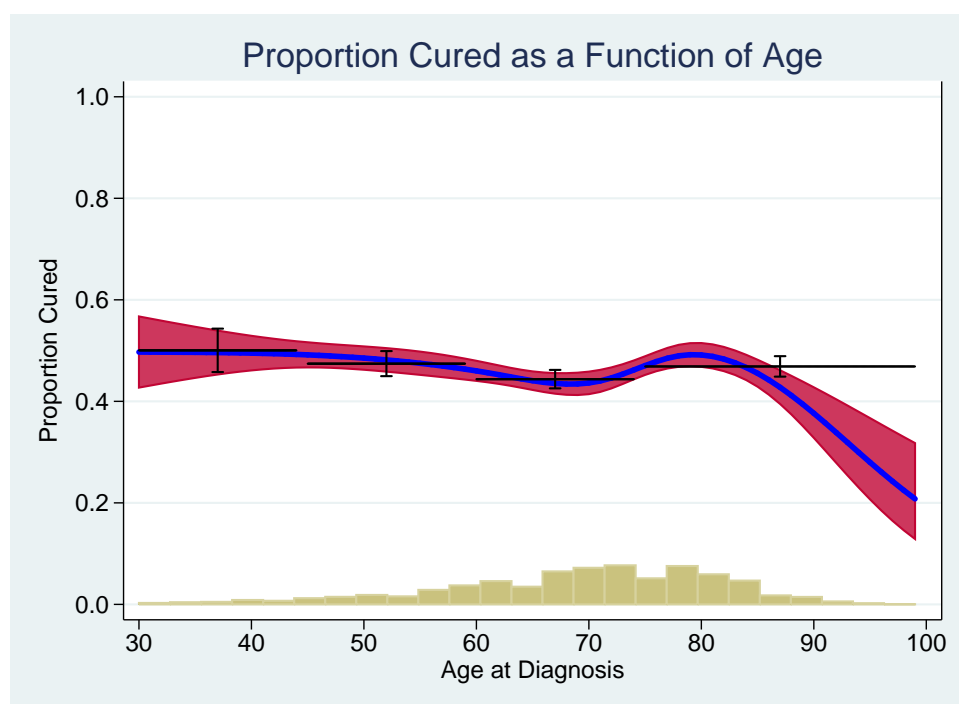
Line 2 -  $\lambda$  modelled,

Line 3 -  $\lambda$  and  $\gamma$  modelled .

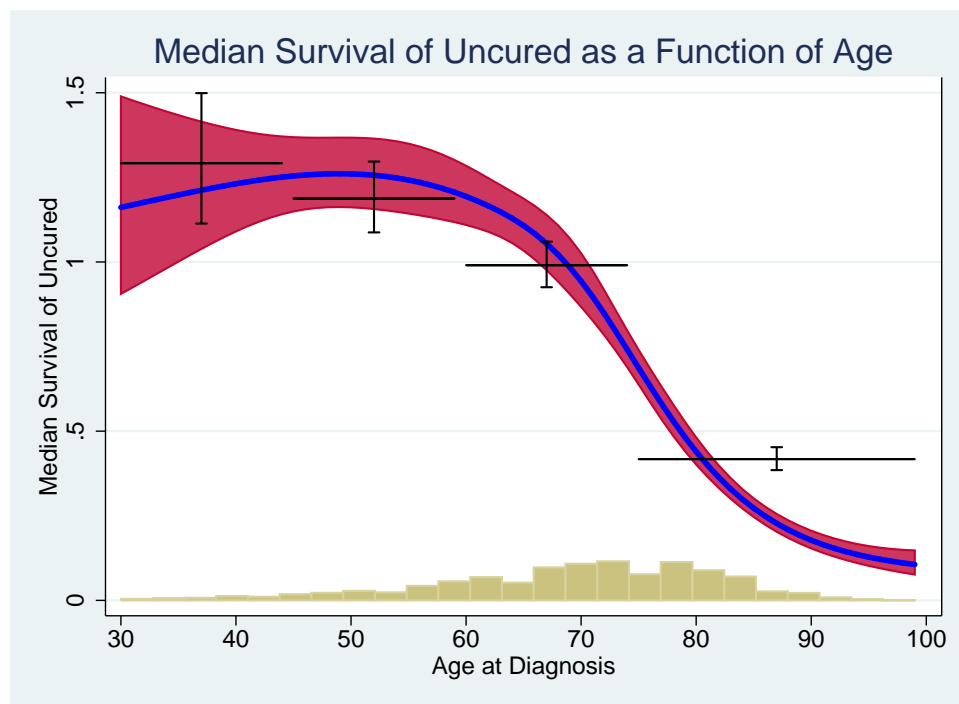
## Modelling Continuous Variables as a Continuous Function

- It is fairly standard to split continuous variables into a number of groups.
- This is generally agreed to be bad practice.
- It is fairly easy to model functions using splines or fractional polynomials.
- Balance between ease of interpretation and better/more appropriate modelling.
- As an example I have used splines (4 internal knots) to model the relationship between age and the proportion cured.

## Proportion Cured as Function of Age



## Median Survival of Uncured as Function of Age



## Mixture or Non-Mixture Models?

- Does it matter whether we use a mixture or non-mixture model?
- Probably not in most cases.
- The mixture model is easier to explain and provides a direct estimate of the survival distribution of the survival function of the 'uncured' group.
- The non-mixture model usually gives a slightly better fit. For example, for the colon cancer data
  - Mixture Model AIC = 42206.93
  - Non-Mixture Model AIC = 42117.05
- There are usually fewer convergence problems with the non-mixture model.

## More Flexible Models

- In some situations the Weibull distribution is not flexible enough and results in a poor fit.
- Usually when very high excess mortality rate in first few weeks after diagnosis.
- Other, more flexible, distributions can be considered
  - Generalized Gamma
  - Generalized F
- Often convergence problems.
- Alternative is to use a mixture of 2 (or more) distributions

## Mixture of Distributions

### Non-Mixture Model

$$\lambda(t) = h^*(t) - \ln(\pi) (p f_{z1}(t) + (1 - p) f_{z2}(t))$$

- This allows a much more flexible shape for the excess hazard and relative survival function (Yakovlev et al., 1999).
- Mixture of two Weibull distributions generally works well.
- Can also think of two groups of individuals, those who die after a short time and those who die after a longer time.



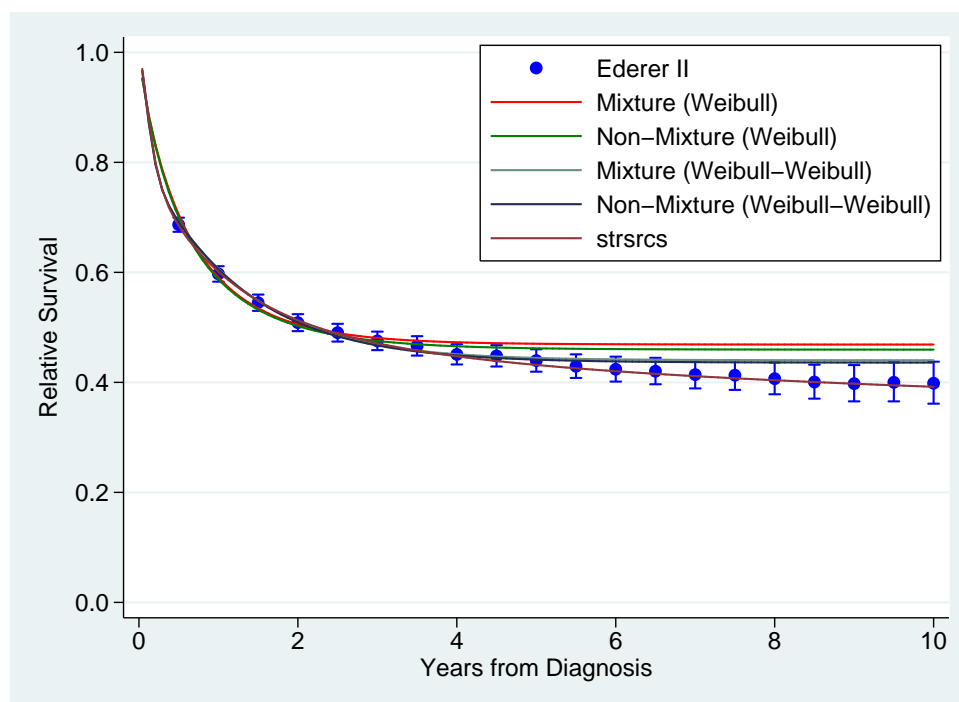
# Mixture of Distributions

## Mixture Model

$$S(t) = S^*(t) (\pi + (1 - \pi) (pS_{u1}(t) + (1 - p)S_{u2}(t)))$$

- For mixture models on relative survival scale.
- Mixture of two Weibull distributions generally works well.
- However, if we allow all parameters to vary by parameters  $(\pi, \lambda_1, \gamma_1, \lambda_2, \gamma_2, p)$  we end with a very complex model.

## Colon Cancer Age 75+



- The cure fraction is a measure of long-term survival.
- Measures of long-term survival may be out of date.
- Recent interest in obtaining up-to-date estimates of relative survival using **Period Analysis** or **Modelling Trends**.

## Period Analysis

- Estimates of (relative) survival obtained by only incorporating survival experience after a certain date (?).
- Period Analysis generally estimated in lifetables, but simple to incorporate in to modelling environment.
- In survival models period analysis can be incorporated using delayed entry techniques (?).
- Incorporation of delayed entry is easy in parametric models

### Likelihood with delayed entry

$$\ln L = d_i \ln(h^*(t_i) + \lambda_e(t_i)) + \ln(R(t_i)) - \ln(R(t_{0i}))$$

- $t_{0i}$  is the time when the subject becomes at risk.
- $t_i$  is the time when the subject dies or is censored.
- $d_i$  is the censoring indicator.

## Example of Using Period Analysis

### Standard Analysis

```
. keep if agegrp == 2
(8971 observations deleted)
. strsmix , dist(weibull) link(identity) bhazard(brate)
initial:      log likelihood = -23736.295
alternative:  log likelihood = -10677.286
rescale:     log likelihood = -10677.286
rescale eq:  log likelihood = -9736.3655
Iteration 0:  log likelihood = -9736.3655
Iteration 1:  log likelihood = -9716.5446
Iteration 2:  log likelihood = -9716.4648
Iteration 3:  log likelihood = -9716.4648
```

Number of obs = 6593  
Wald chi2(0) = .  
Prob > chi2 = .

Log likelihood = -9716.4648

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pi					
_cons	.4437026	.0092847	47.79	0.000	.425505 .4619002
ln_lambda					
_cons	-.3581668	.0300444	-11.92	0.000	-.4170528 -.2992808
ln_gamma					
_cons	-.1444423	.0182711	-7.91	0.000	-.180253 -.1086316

## Example of Using Period Analysis

### Period Analysis

```
. stset exit, failure(status = 1,2) origin(dx) entry(time mdy(1,1,1993)) scale(
> 365.25) exit(time dx+10*365.25)
(output omitted)
. strsmix , dist(weibull) link(identity) bhazard(brate)
note: delayed entry models are being fitted
initial:      log likelihood = -5124.498
alternative:  log likelihood = -1995.6225
rescale:     log likelihood = -1995.6225
rescale eq:  log likelihood = -1778.5844
Iteration 0:  log likelihood = -1778.5844
Iteration 1:  log likelihood = -1777.6456
Iteration 2:  log likelihood = -1777.6423
Iteration 3:  log likelihood = -1777.6423
```

Number of obs = 2496  
Wald chi2(0) = .  
Prob > chi2 = .

Log likelihood = -1777.6423

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
pi					
_cons	.4704467	.0227293	20.70	0.000	.4258981 .5149953
ln_lambda					
_cons	-.5524744	.076701	-7.20	0.000	-.7028055 -.4021432
ln_gamma					
_cons	-.1601728	.0507684	-3.15	0.002	-.259677 -.0606685

## Potential Points for Discussion

- Criteria for using cure models
- Definition of time to cure.
- Continuous covariates.
- Mixture or non-mixture models.
- More complex models (mixture of distributions etc).
- Up-to-date estimates of cure.

## References I

- Brenner, H. and Gefeller, O. (1997). Deriving more up-to-date estimates of long-term patient survival. *Journal of Clinical Epidemiology*, 50(2):211–216.
- De Angelis, R., Capocaccia, R., Hakulinen, T., Soderman, B., and Verdecchia, A. (1999). Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–54.
- Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer.
- Lambert, P. (2007). Modelling the cure fraction in survival studies. *The Stata Journal (in press)*.
- Lambert, P., Dickman, P. W., Osterlund, P., Andersson, T., Sankila, R., and Glimelius, B. (2007a). Temporal trends in the proportion cured for cancer of the colon and rectum: A population-based study using data from the finnish cancer registry. *International Journal of Cancer*, 121:2052–2059.
- Lambert, P., Thompson, J., Weston, C. L., and Dickman, P. W. (2007b). Estimating and modelling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8:576–94.
- Smith, L. K., Lambert, P. C., Botha, J. L., and Jones, D. R. (2004). Providing more up-to-date estimates of patient survival: a comparison of standard survival analysis with period analysis using life-table methods and proportional hazards models. *Journal of Clinical Epidemiology*, 57(1):14–20.
- Spoto, R. (2002). Cure model analysis in cancer: an application to data from the children's cancer group. *Statistics in Medicine*, 21(2):293–312.
- Verdecchia, A., De Angelis, R., Capocaccia, R., Sant, M., Micheli, A., Gatta, G., and Berrino, F. (1998). The cure for colon cancer: results from the eurocare study. *International Journal of Cancer*, 77(3):322–9.
- Yakovlev, A., Tsodikov, A., Boucher, K., and Kerber, R. (1999). The shape of the hazard function in breast carcinoma: Curability of the disease revisited. *Cancer*, 85:1789–98.