Estimating and Modelling the Proportion Cured of Disease in Population Based Cancer Studies

Paul C Lambert

Centre for Biostatistics and Genetic Epidemiology, University of Leicester, UK

12th UK Stata Users Group Meeting
London
11th-12th September 2006

Population Based Cancer Studies

- Using data from cancer registries.
- Attempt to obtain *all* diagnosed cancers.
- Information used for incidence and survival.
- Large sample sizes.
- Relative Survival methods used for survival analysis.
- Five year relative survival often reported.

Relative Survival

$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \qquad R(t) = S(t)/S^*(t)$$

- Expected survival obtained from national population life tables stratified by age, sex, year of diagnosis, other covariates.
- Estimate of mortality associated with a disease without requiring information on cause of death.
- On hazard scale

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

$$\begin{matrix}\text{Observed} \\ \text{Mortality Rate}\end{matrix} = \begin{matrix}\text{Expected} \\ \text{Mortality Rate}\end{matrix} + \begin{matrix}\textit{Excess} \\ \textit{Mortality Rate}\end{matrix}$$

Relative Survival

$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \qquad R(t) = S(t)/S^*(t)$$

- Expected survival obtained from national population life tables stratified by age, sex, year of diagnosis, other covariates.
- Estimate of mortality associated with a disease without requiring information on cause of death.
- On hazard scale

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

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

Relative Survival

$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \qquad R(t) = S(t)/S^*(t)$$

- Expected survival obtained from national population life tables stratified by age, sex, year of diagnosis, other covariates.
- Estimate of mortality associated with a disease without requiring information on cause of death.
- On hazard scale

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

$$\begin{array}{c}\text{Observed}\\ \text{Mortality Rate}\end{array} = \begin{array}{c}\text{Expected}\\ \text{Mortality Rate}\end{array} + \begin{array}{c}\textit{Excess}\\ \textit{Mortality Rate}\end{array}$$

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\beta X)$$

- Parameters are (log) excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[4].
- The models do not assume that a proportion of patients may be 'cured' of their disease.
- For details of Stata command `strs` for estimation and modelling of relative survival using piecewise methods see
  http://www.pauldickman.com/rsmodel/stata_colon/

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\beta X)$$

- Parameters are (log) excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[4].
- The models do not assume that a proportion of patients may be 'cured' of their disease.
- For details of Stata command `strs` for estimation and modelling of relative survival using piecewise methods see `http://www.pauldickman.com/rsmodel/stata_colon/`

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\beta X)$$

- Parameters are (log) excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[4].
- The models do not assume that a proportion of patients may be 'cured' of their disease.
- For details of Stata command `strs` for estimation and modelling of relative survival using piecewise methods see `http://www.pauldickman.com/rsmodel/stata_colon/`

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\beta X)$$

- Parameters are (log) excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[4].
- The models do not assume that a proportion of patients may be 'cured' of their disease.
- For details of Stata command `strs` for estimation and modelling of relative survival using piecewise methods see `http://www.pauldickman.com/rsmodel/stata_colon/`

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\beta X)$$

- Parameters are (log) excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[4].
- The models do not assume that a proportion of patients may be 'cured' of their disease.
- For details of Stata command `strs` for estimation and modelling of relative survival using piecewise methods see http://www.pauldickman.com/rsmodel/stata_colon/

Definition of Cure (1)

- For many cancers the hazard (mortality) rate returns to the same level as that in the general population.
- When this occurs the relative survival curve is seen to reach a plateau (or the excess hazard rate approaches zero).
- This is Population or Statististical Cure.
- Information of cure at the individual level not available.

Definition of Cure (1)

- For many cancers the hazard (mortality) rate returns to the same level as that in the general population.
- When this occurs the relative survival curve is seen to reach a plateau (or the excess hazard rate approaches zero).
- This is Population or Statististical Cure.
- Information of cure at the individual level not available.

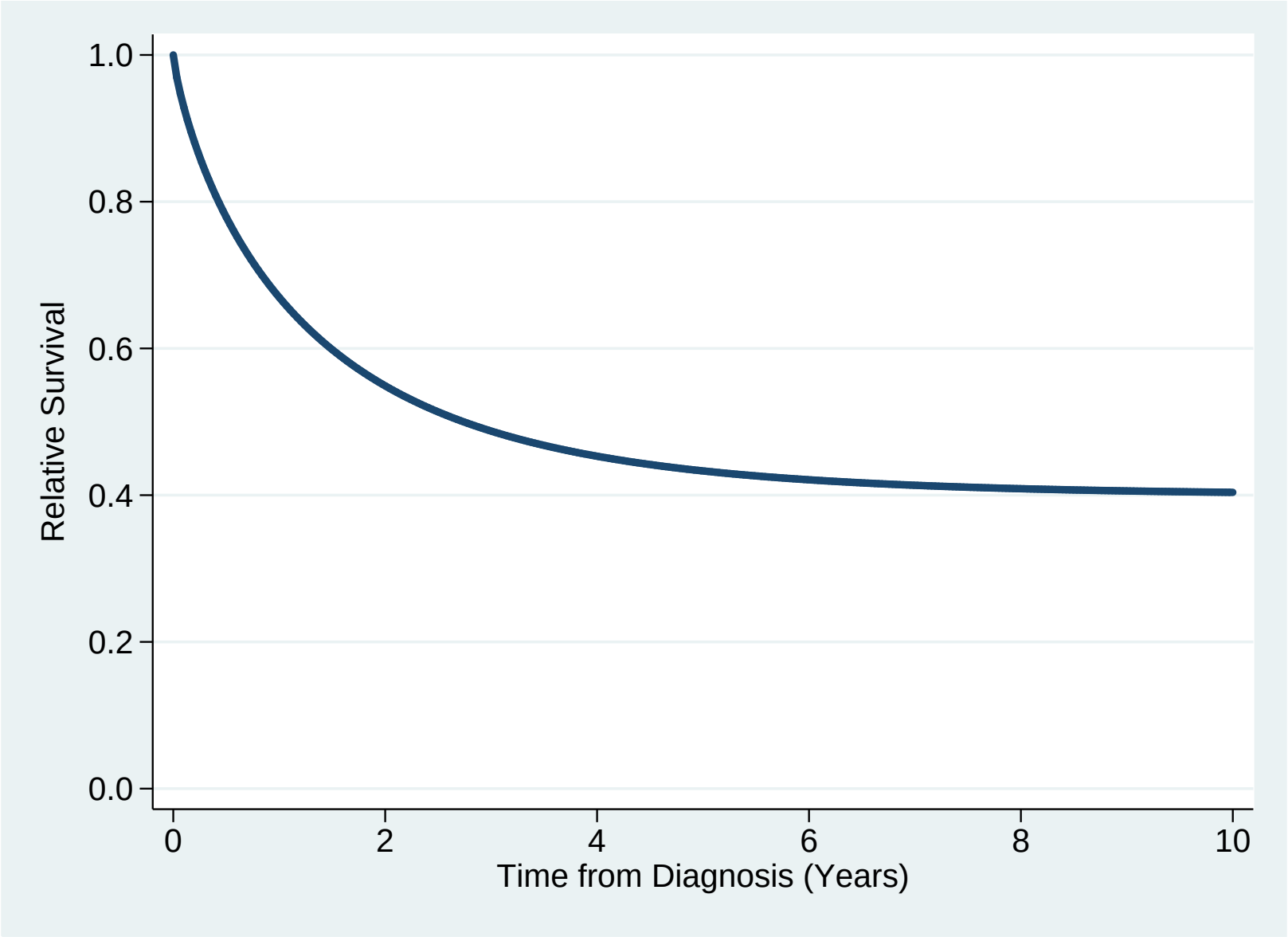
Definition of Cure (1)

- For many cancers the hazard (mortality) rate returns to the same level as that in the general population.
- When this occurs the relative survival curve is seen to reach a plateau (or the excess hazard rate approaches zero).
- This is Population or Statistical Cure.
- Information of cure at the individual level not available.

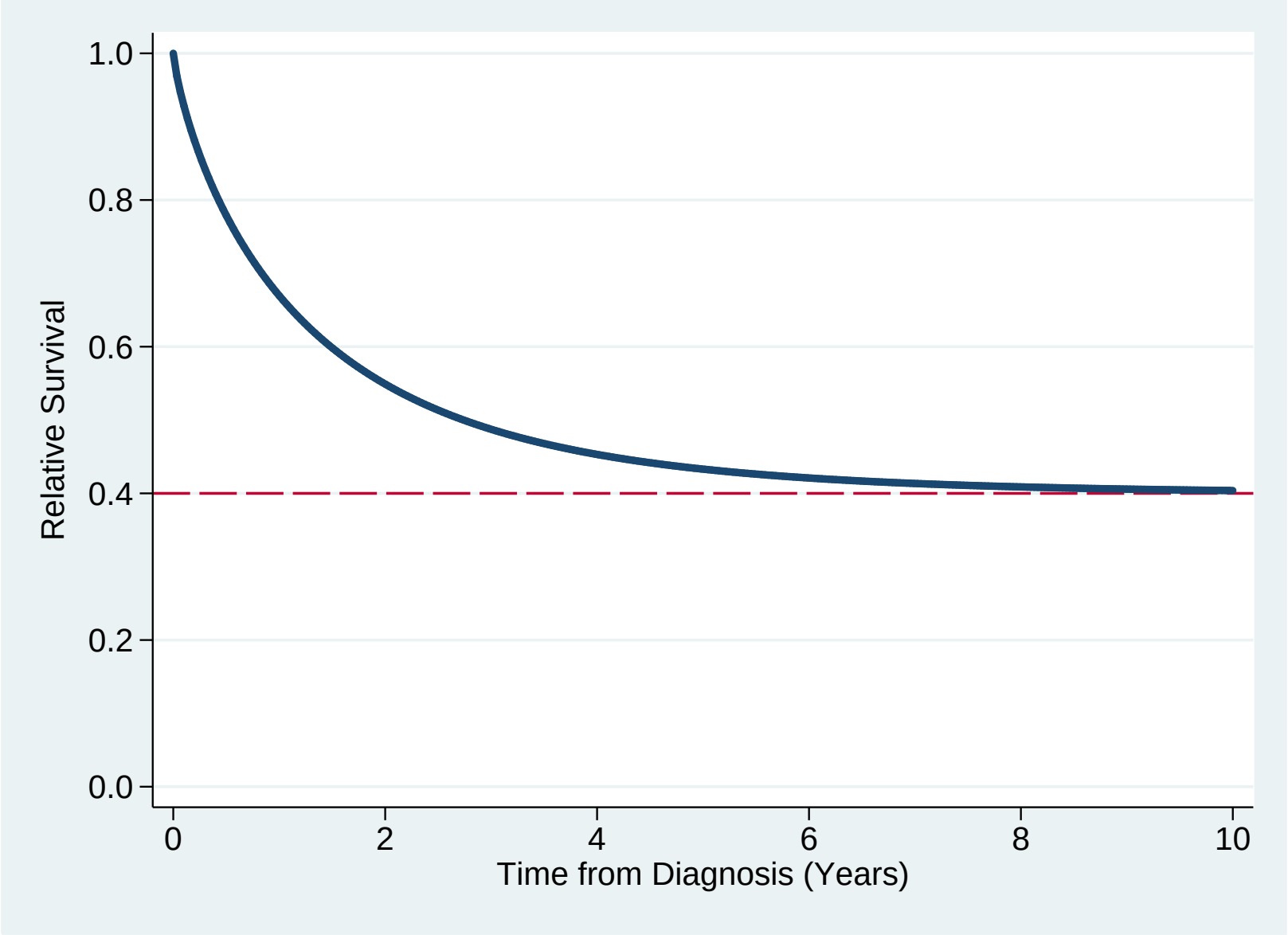
Definition of Cure (1)

- For many cancers the hazard (mortality) rate returns to the same level as that in the general population.
- When this occurs the relative survival curve is seen to reach a plateau (or the excess hazard rate approaches zero).
- This is Population or Statistical Cure.
- Information of cure at the individual level not available.

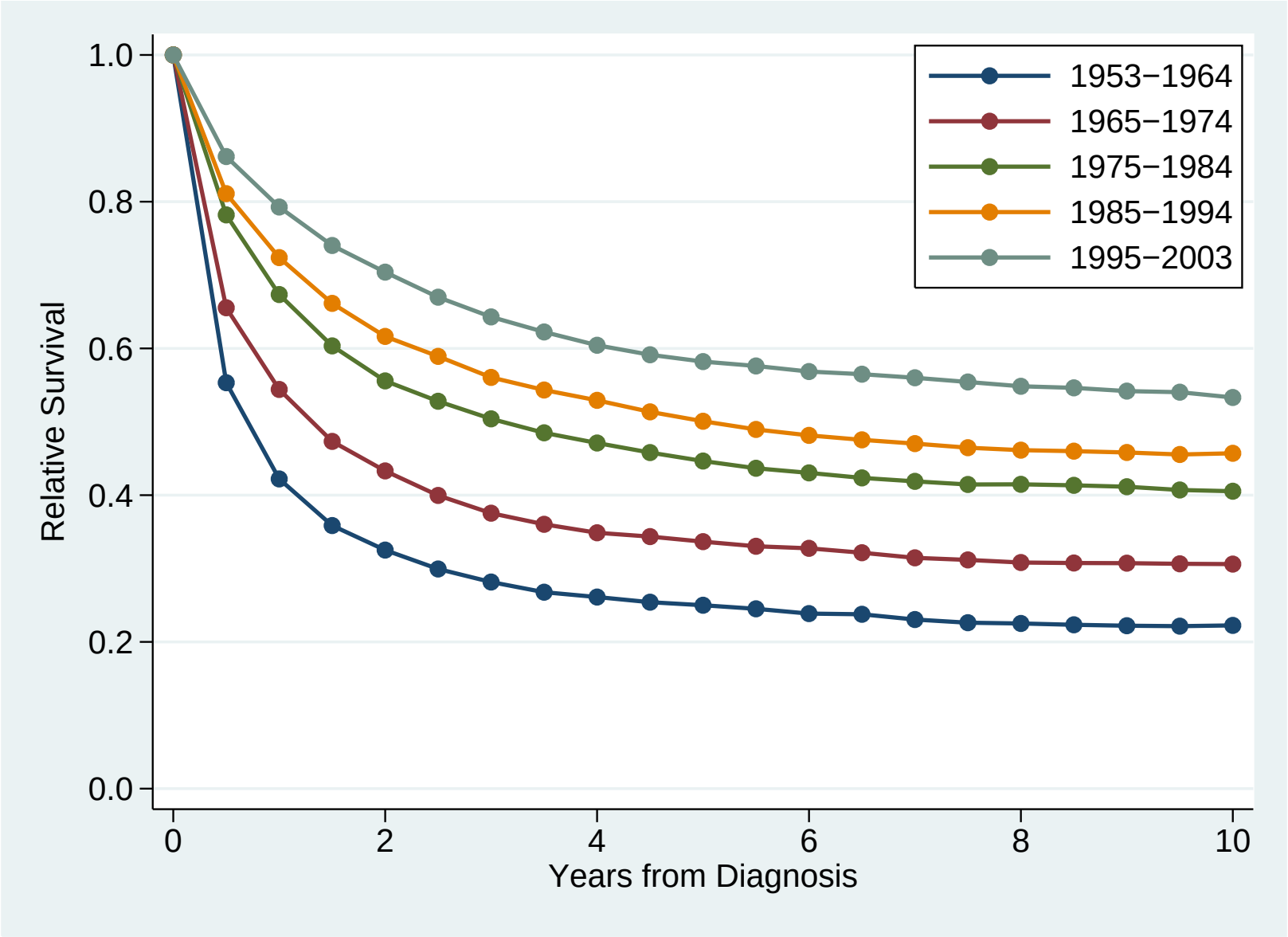
Definition of Cure (2)



Definition of Cure (2)



Relative Survival for Cancer of the Colon in Finland



Mixture and Non-Mixture Models

Relative Survival Models

$$S(t) = S^*(t)R(t)$$

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

- When modelling cure we define an asymptote at the cure fraction, π, for the relative survival function, $R(t)$.
- The excess hazard rate, $\lambda(t)$, has an asymptote at zero.
- Two main approaches
  - Mixture Model
  - Non-Mixture Model
  - Both of these models have been used in 'standard' survival analysis [9], i.e. not incorporating background mortality. Some of these models are implemented in Stata using the `cureregr` command.

Mixture and Non-Mixture Models

Relative Survival Models

$$S(t) = S^*(t)R(t)$$

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

- When modelling cure we define an asymptote at the cure fraction, π, for the relative survival function, $R(t)$.
- The excess hazard rate, $\lambda(t)$, has an asymptote at zero.
- Two main approaches
  - Mixture Model
  - Non-Mixture Model
  - Both of these models have been used in 'standard' survival analysis [9], i.e. not incorporating background mortality. Some of these models are implemented in Stata using the `cureregr` command.

Mixture Model

Mixture Model

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

- $S^*(t)$ is the expected survival.
- π 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.
- See De Angelis [2] and Verdecchia [10] for more details.

Non-Mixture Model

Non Mixture Model

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

- We have extended the non-mixture model to relative survival[7].
- If parameters in $f_z(t)$ do not vary by covariates then this is a proportional excess hazards model.
- The mixture model does not have proportional excess hazards as a special case.
- 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

Non Mixture Model

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

- We have extended the non-mixture model to relative survival[7].
- If parameters in $f_z(t)$ do not vary by covariates then this is a proportional excess hazards model.
- The mixture model does not have proportional excess hazards as a special case.
- 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.

Likelihood

Relative Survival Models

$$L_i = d_i \ln(h^*(t_i) + \lambda(t_i)) + \ln(S^*(t_i)) + \ln(R(t_i)) - \ln(S^*(t_{0i})) - \ln(R(t_{0i}))$$

- $S^*(t_i)$ and $S^*(t_{0i})$ do not depend on the model parameters and can be excluded from the likelihood.
- Merge in expected mortality rate at time of death, $h^*(t_i)$.
- Newton-Raphson algorithm implemented using Stata `ml` command (method `lf`).
- Incorporating delayed entry allows period analysis models to be fitted[8]. This is a method used to obtain *up-to-date* estimates of (relative) survival. Application in the cure models allows *up-to-date* estimates of cure to be obtained.

Likelihood

Relative Survival Models

$$L_i = d_i \ln(h^*(t_i) + \lambda(t_i)) + \ln(S^*(t_i)) + \ln(R(t_i)) - \ln(S^*(t_{0i})) - \ln(R(t_{0i}))$$

- $S^*(t_i)$ and $S^*(t_{0i})$ do not depend on the model parameters and can be excluded from the likelihood.
- Merge in expected mortality rate at time of death, $h^*(t_i)$.
- Newton-Raphson algorithm implemented using Stata `ml` command (method `lf`).
- Incorporating delayed entry allows period analysis models to be fitted[8]. This is a method used to obtain *up-to-date* estimates of (relative) survival. Application in the cure models allows *up-to-date* estimates of cure to be obtained.

Likelihood

Relative Survival Models

$$L_i = d_i \ln(h^*(t_i) + \lambda(t_i)) + \ln(S^*(t_i)) + \ln(R(t_i)) - \ln(S^*(t_{0i})) - \ln(R(t_{0i}))$$

- $S^*(t_i)$ and $S^*(t_{0i})$ do not depend on the model parameters and can be excluded from the likelihood.
- Merge in expected mortality rate at time of death, $h^*(t_i)$.
- Newton-Raphson algorithm implemented using Stata `ml` command (method `lf`).
- Incorporating delayed entry allows period analysis models to be fitted[8]. This is a method used to obtain *up-to-date* estimates of (relative) survival. Application in the cure models allows *up-to-date* estimates of cure to be obtained.

strsmix and strsnmix commands

strsmix [*varlist*] [*if*] [*in*] , distribution(*distribution*) link(*link function*) bhazard(*varname*) [k1(*varlist*) k2(*varlist*) k3(*varlist*) k4(*varlist*) pmix(*varlist*) noconstant noconsk1 noconsk2 noconsk3 noconsk4 noconspmix init(*matrix name*) skip inititer(#) stopconstraint valconstraint(#) eform]

strsnmix [*varlist*] [*if*] [*in*] , distribution(*distribution*) link(*link function*) bhazard(*varname*) [k1(*varlist*) k2(*varlist*) k3(*varlist*) k4(*varlist*) pmix(*varlist*) split(#) earlyk1(*varlist*) earlyk2(*varlist*) noconstant noconsk1 noconsk2 noconsk3 noconsk4 noconspmix earlynoconsk1 earlynoconsk2 init(*matrix name*) skip inititer(#) stopconstraint valconstraint(#) eform]

Stata

```
net from http://www.hs.le.ac.uk/personal/pl4/Software/Stata/strsnmix

install strsnmix
```

Some options for `strsnmix` and `strsmix`

- `distribution(`*distribution*`)` specifies the parametric distribution. Arguments for both `strsmix` and `strsnmix` are `weibull`, `lognomal` and `gamma`, `weibexp` and `weibweib`.
- `link(`*link function*`)` specifies the link function for the cure fraction. Options are `identity`, `logistic` and `loglog`. Note that `loglog` is $\ln(-\ln(\pi))$.
- `bhazard(`*varname*`)` gives the variable name for the baseline hazard at death/censoring. This option is compulsory, but standard cure models can be estimated by making *varname* a column of zeros.
- `k1-k4(`*varlist*`)` gives any covariates for the auxillary parameter. E.g. for the Weibull distribution `k1` refers to $\ln(\lambda)$ and `k2` refers to $\ln(\gamma)$.
- Commands submitted to *The Stata Journal*[5].

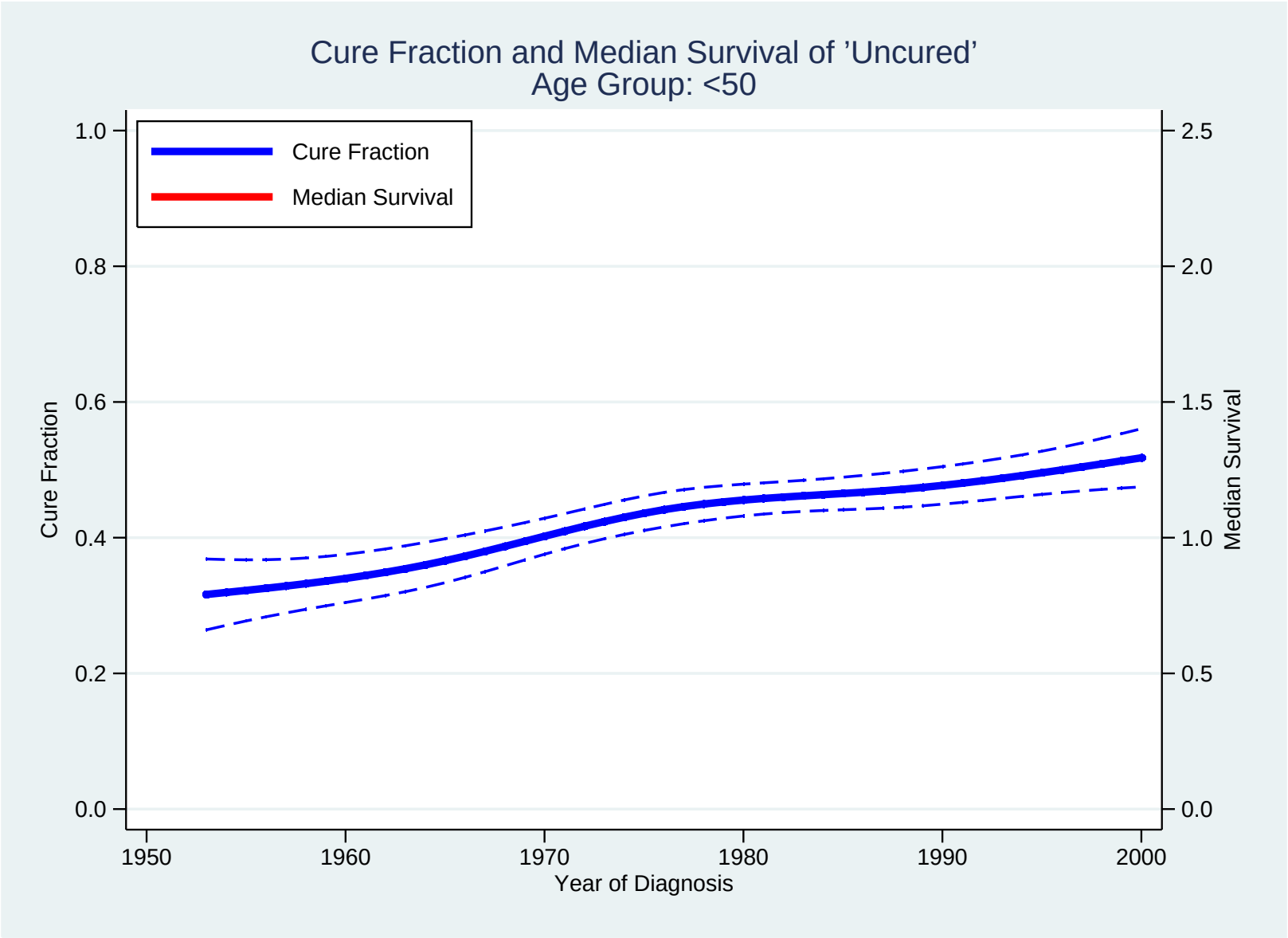
Cancer of the Colon in Finland

- Data from the Finnish Cancer Registry.
- 27,754 men and women diagnosed 1953-2003 with follow-up to 2004.
- Covariates age group and year of diagnosis.
- Exclude those aged 80 years and over.
- Use a mixture cure model with Weibull distribution for the 'uncured'.
- Year of diagnosis modelled using restricted cubic splines for cure fraction and both Weibull parameters.

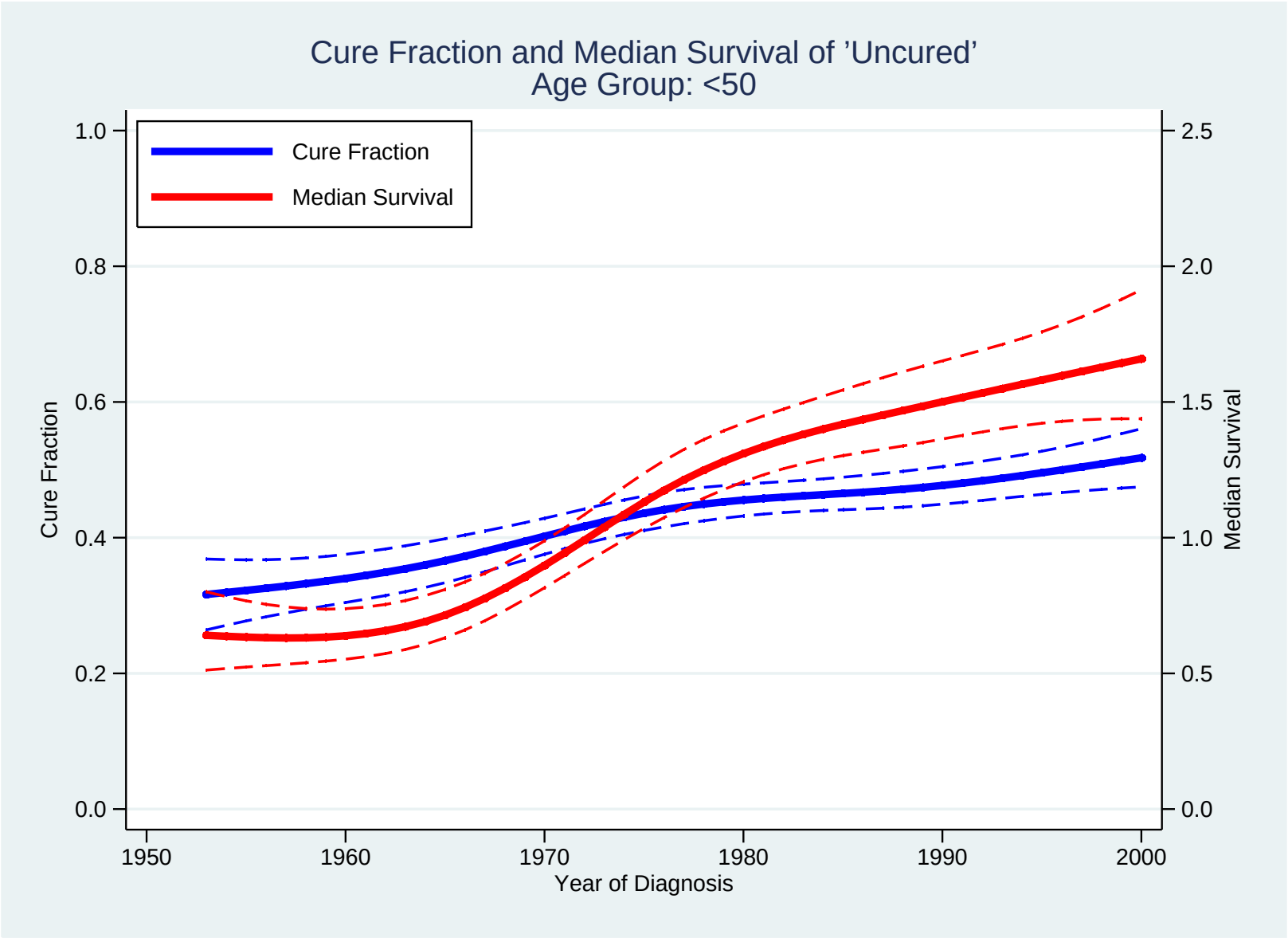
Stata Code

```
strsmix rcs1-rcs4 agegrp2 agegp3 agegrp4 age2rcs1 age3rcs1 age4rcs1, ///
        dist(weibull) link(identity) bhazard(brate) ///
        k1(rcs1-rcs4 agegrp2 agegrp3 agegrp4 age2rcs1 age3rcs1 age4rcs1) ///
        k2(rcs1-rcs4 agegrp2 agegrp3 agegrp4 age2rcs1 age3rcs1 age4rcs1)
predict cure, cure ci
predict rs, survival ci
predict rsu, survival uncured ci
predict exhaz, hazard ci
predict median, centile ci
```

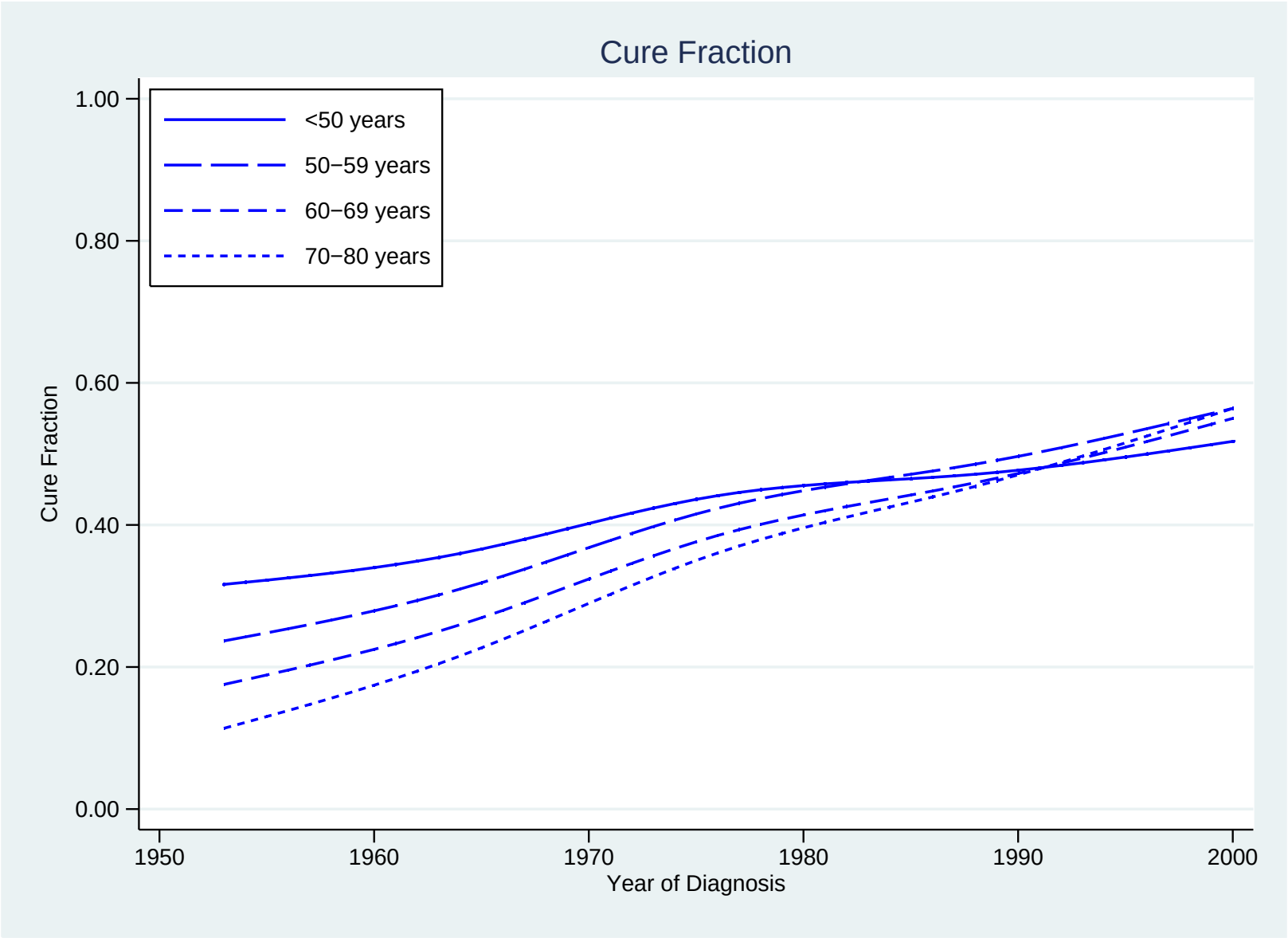
Time Trends for Cancer of the Colon Age <50



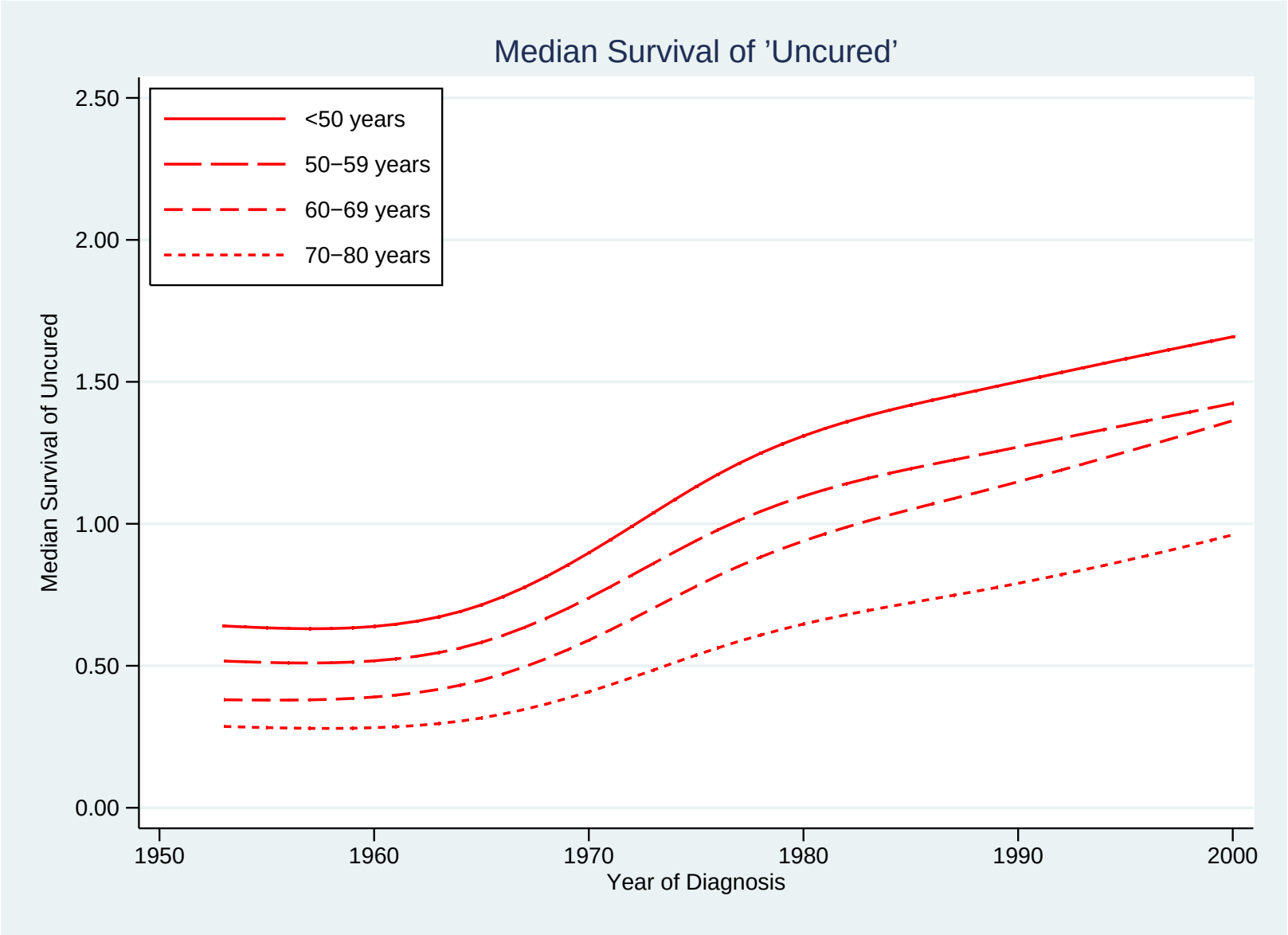
Time Trends for Cancer of the Colon Age <50



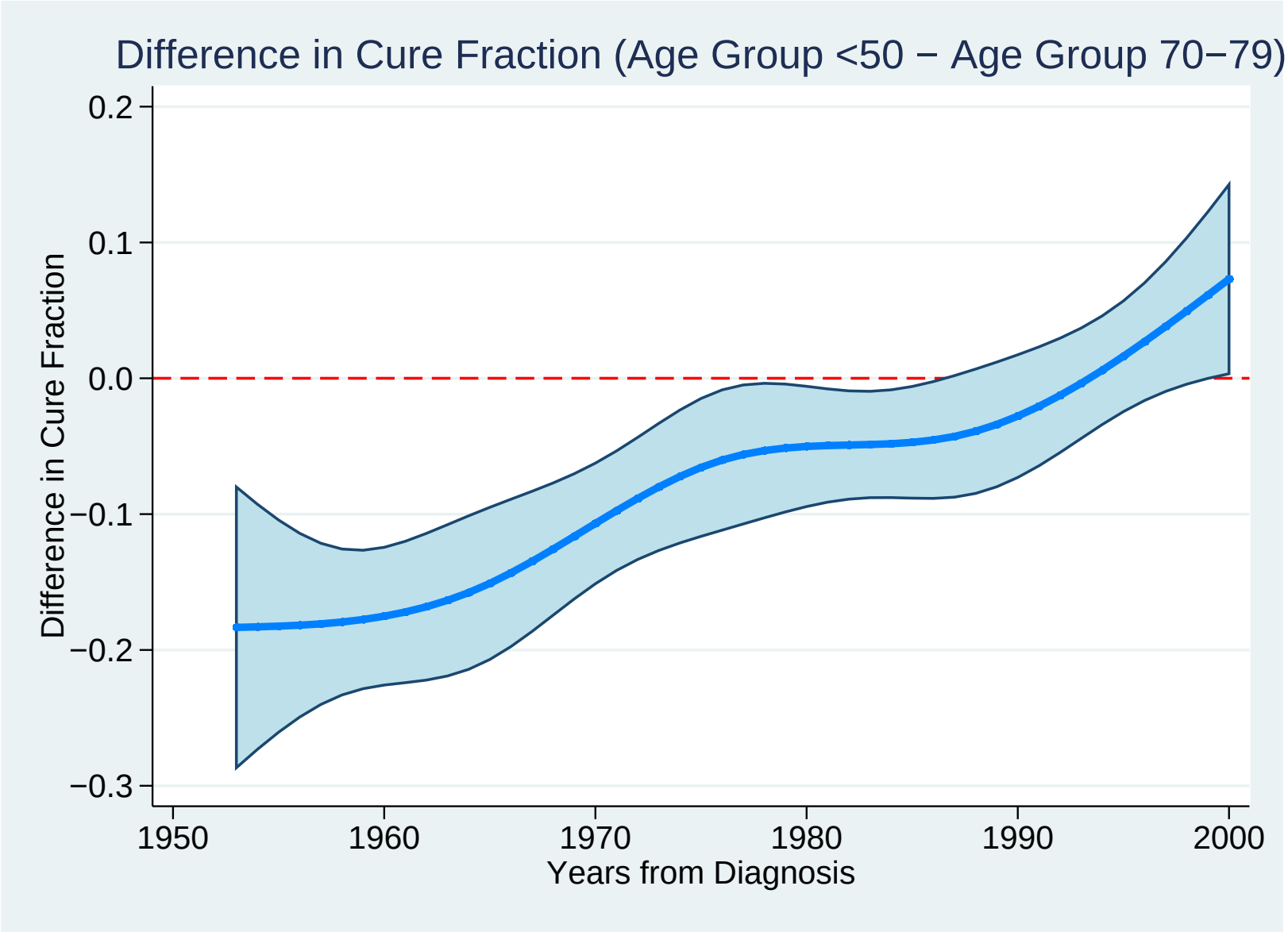
Time Trends for Cancer of the Colon



Time Trends for Cancer of the Colon



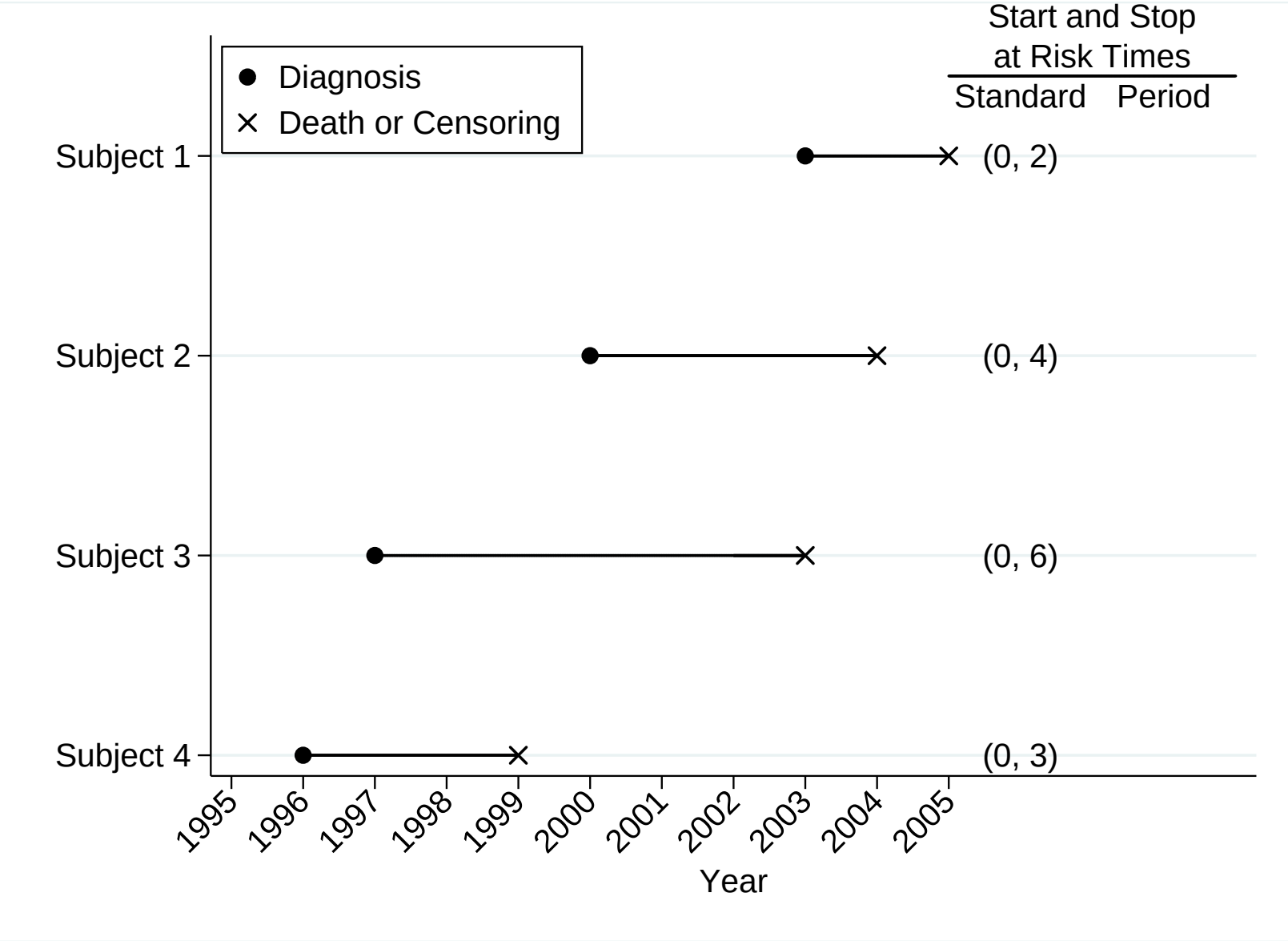
Quantifying Differences



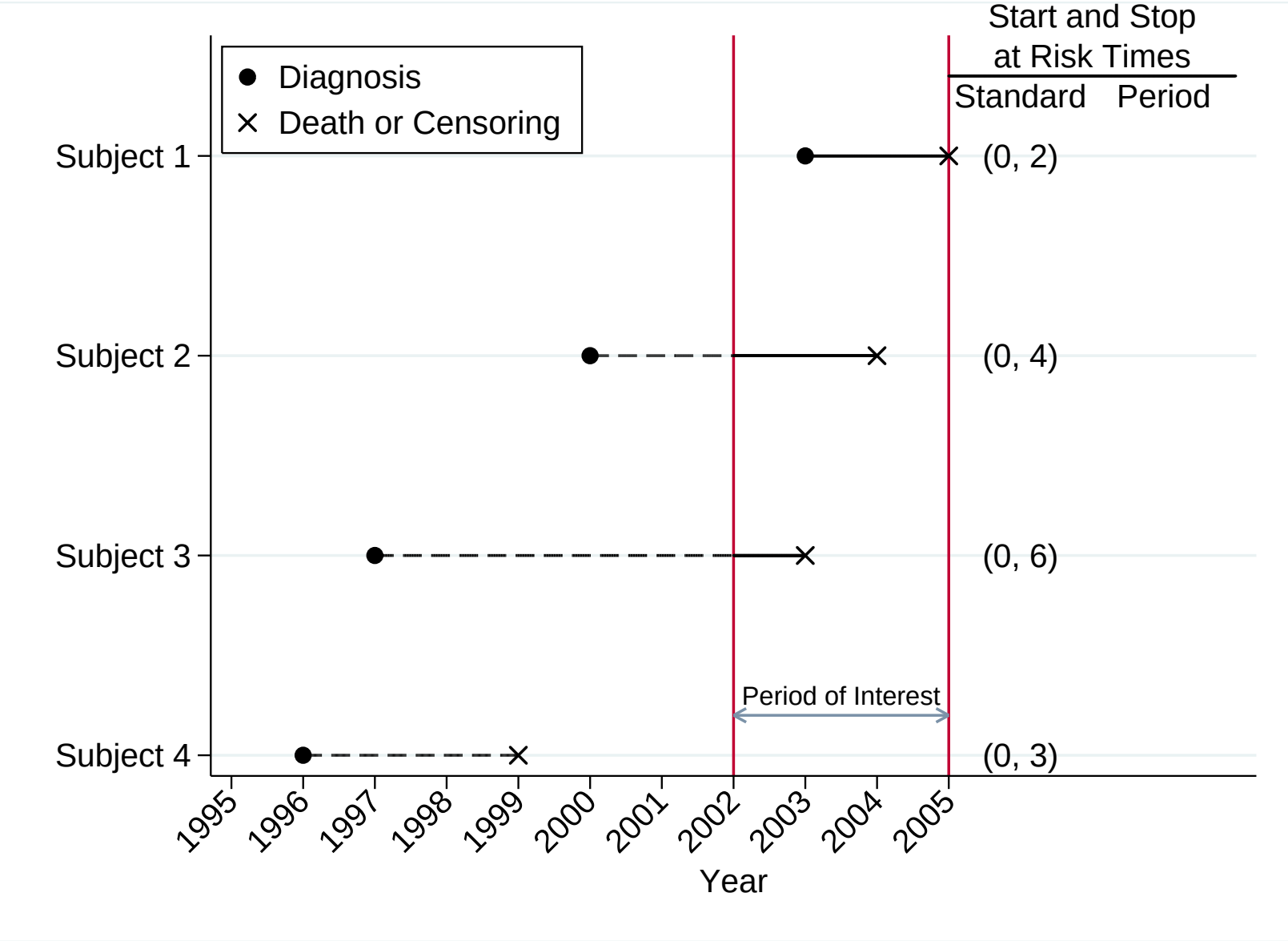
Period Analysis

- Long-term estimates of survival may be out-of-date.
- Period Analysis estimates (relative) survival by only incorporating survival experience in a recent time window[1].
- Period Analysis generally estimated in lifetables, but simple to incorporate in to modelling environment[8].
- In survival models period analysis can be incorporated using delayed entry techniques.

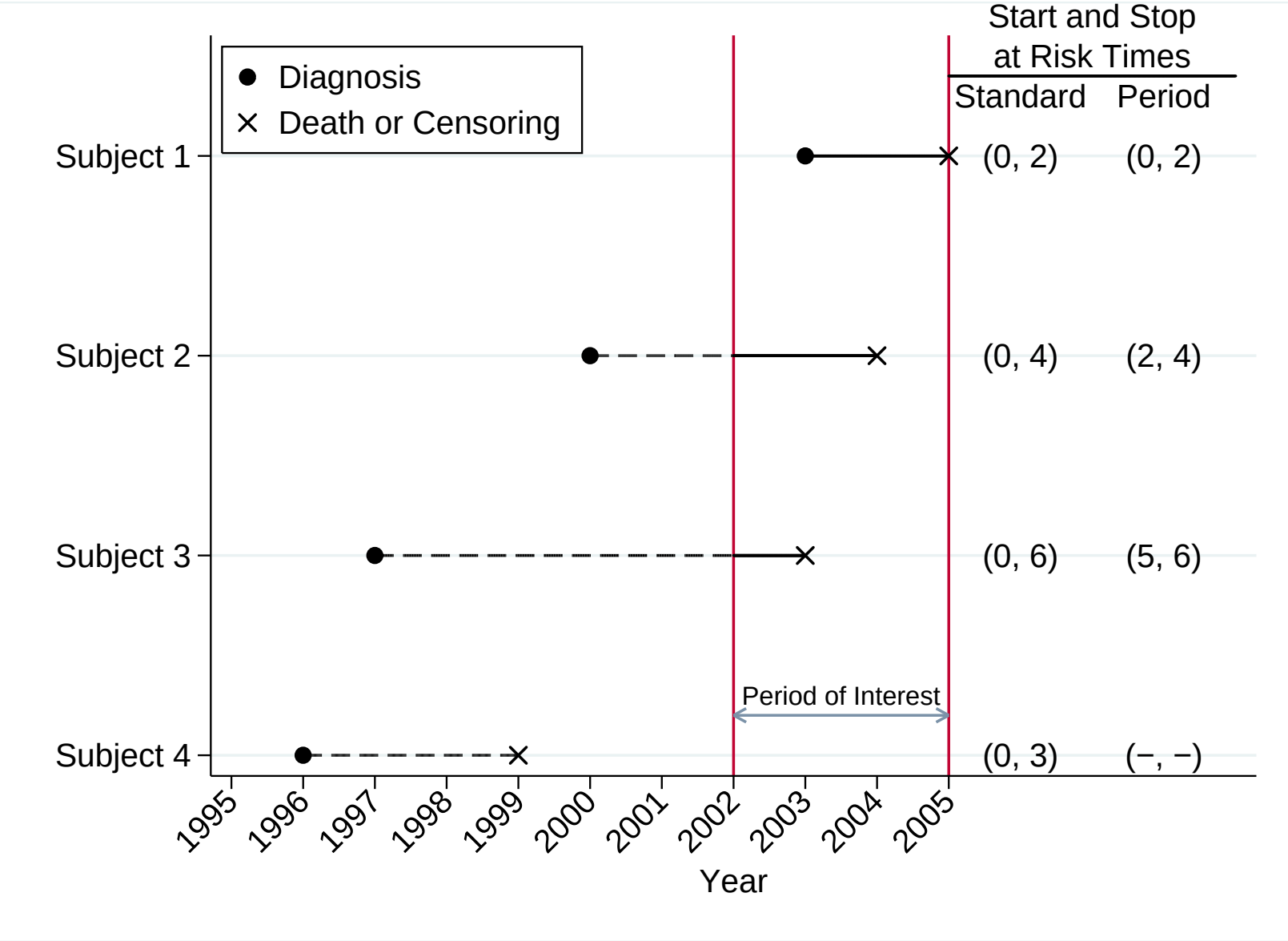
Period Analysis



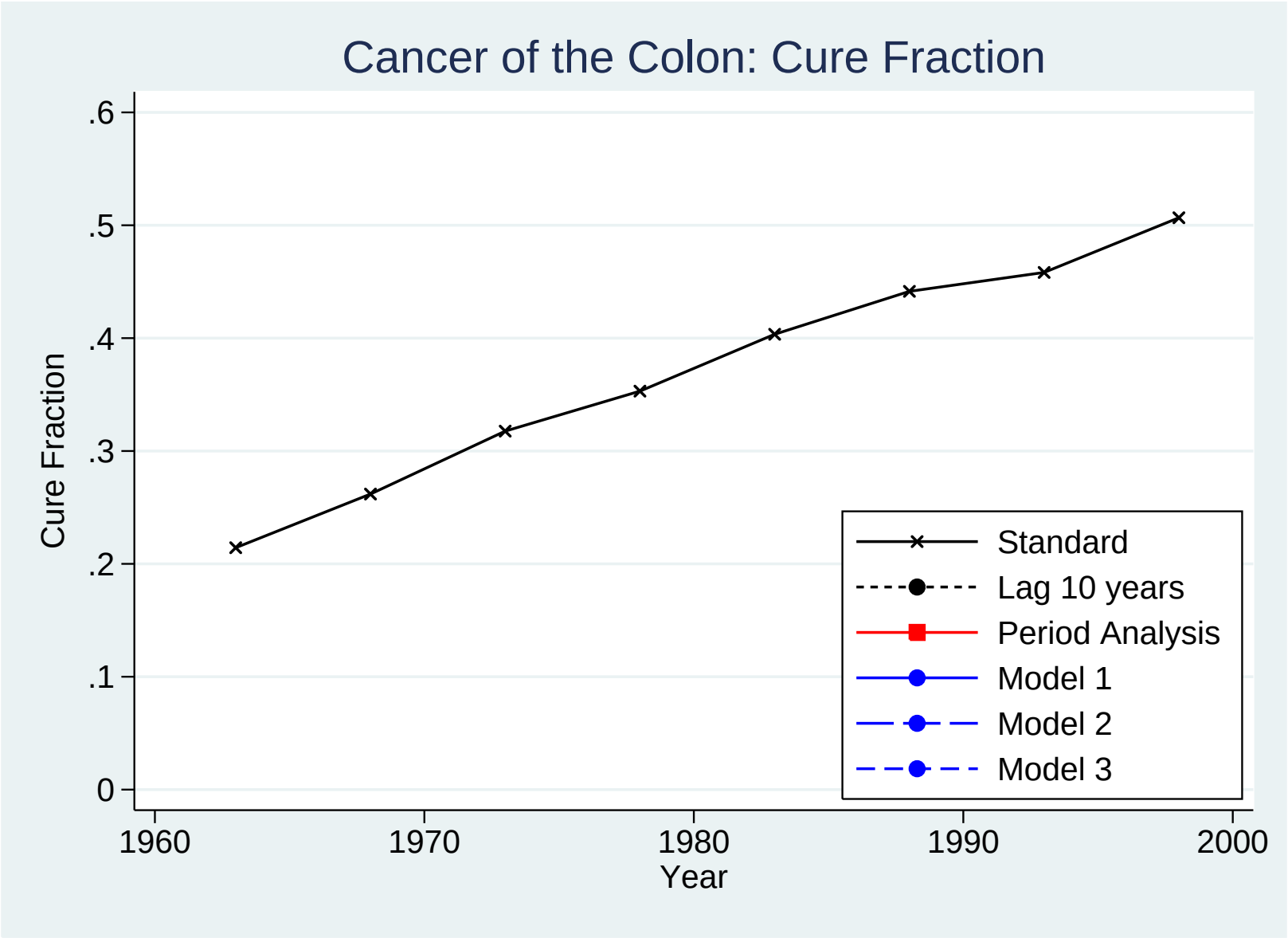
Period Analysis



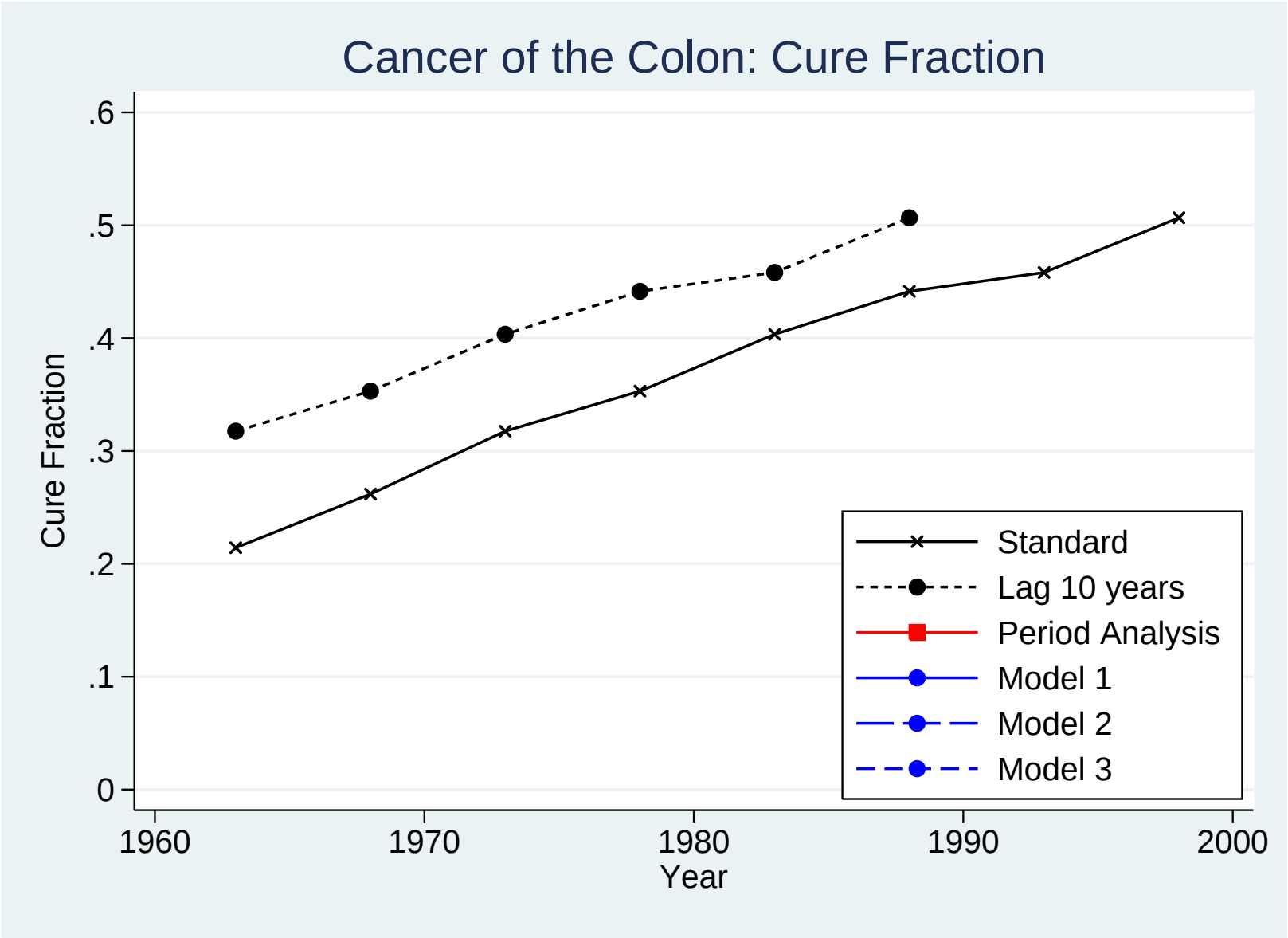
Period Analysis



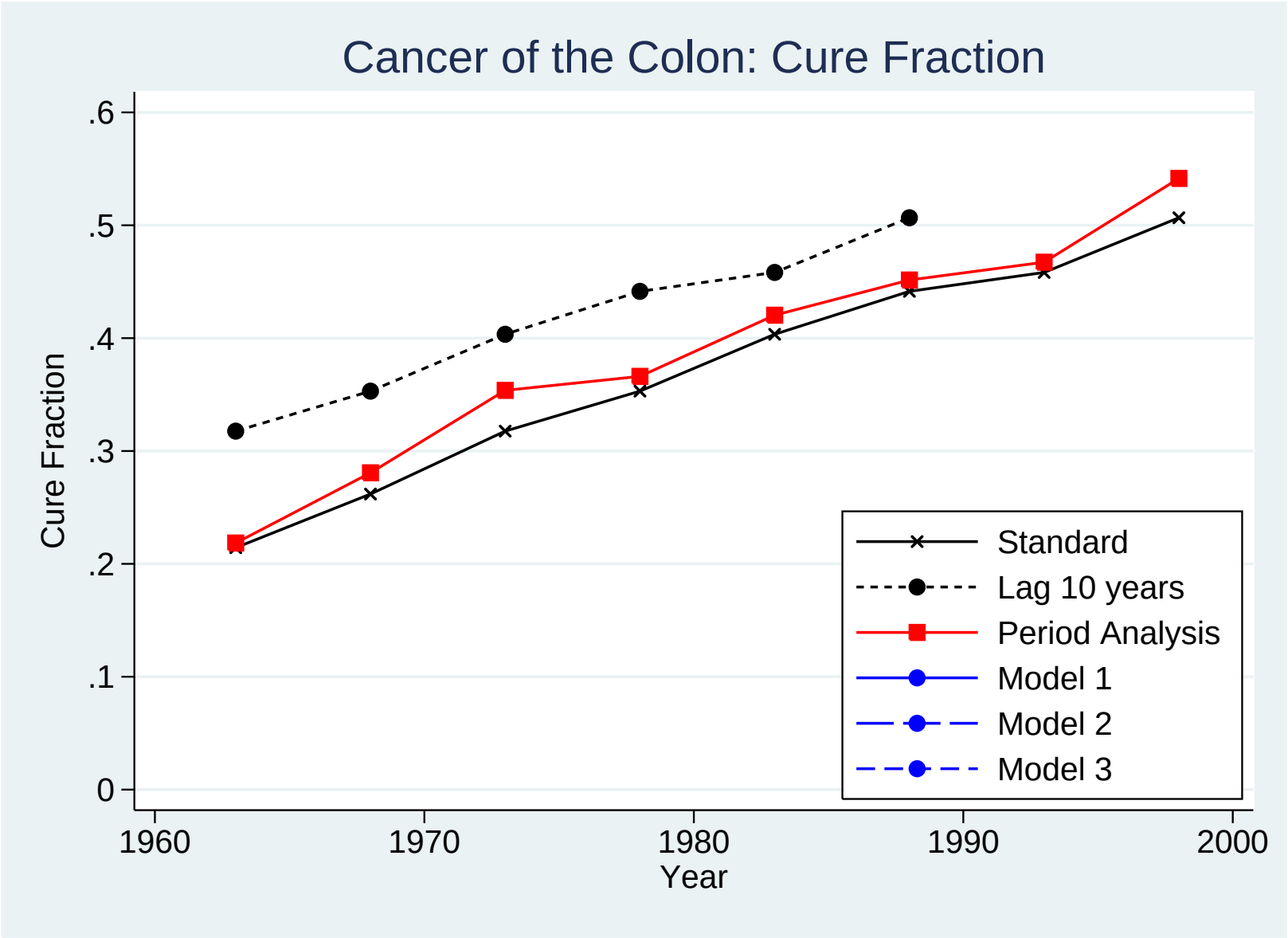
Period Analysis: Cancer of the Colon



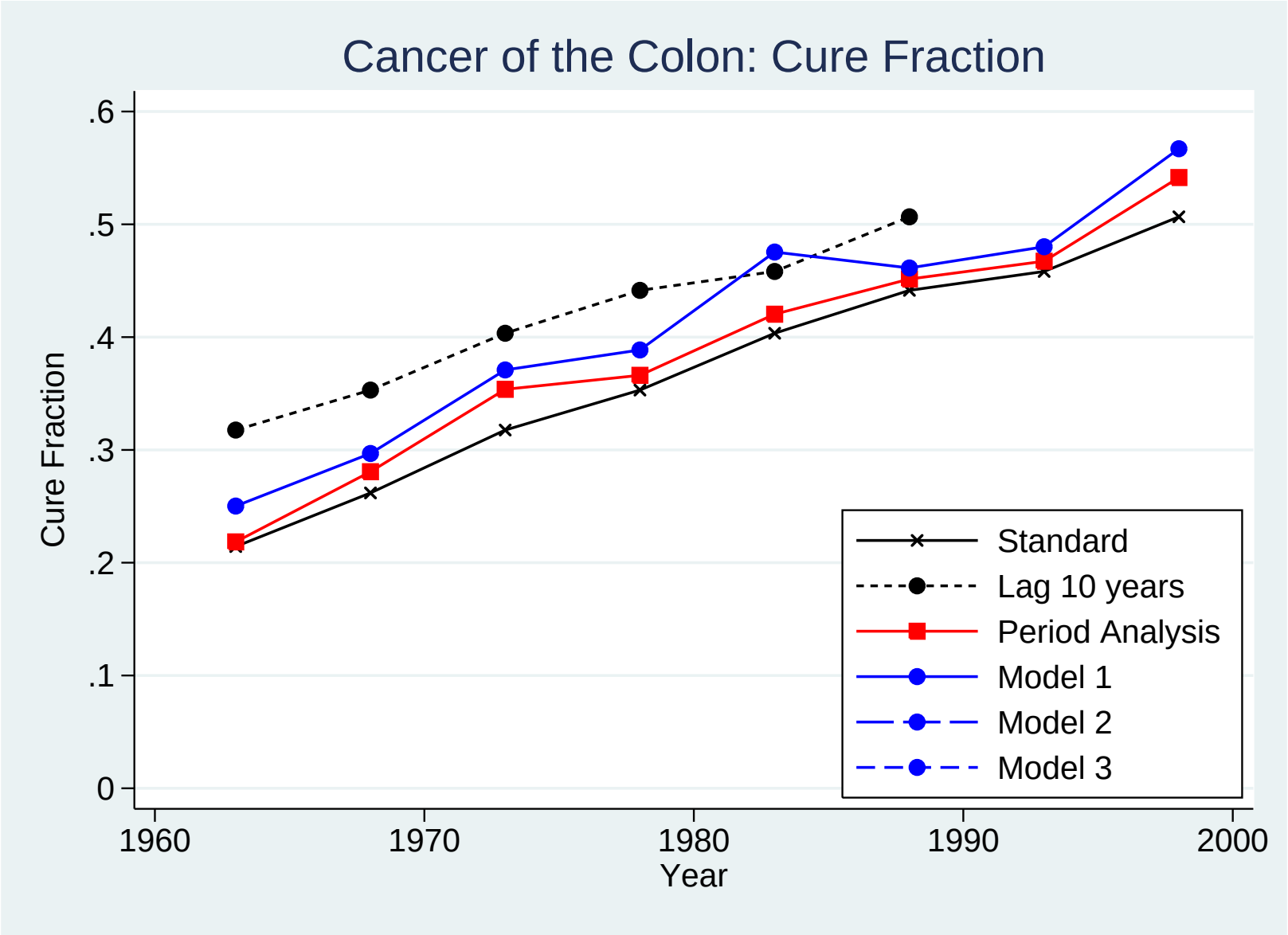
Period Analysis: Cancer of the Colon



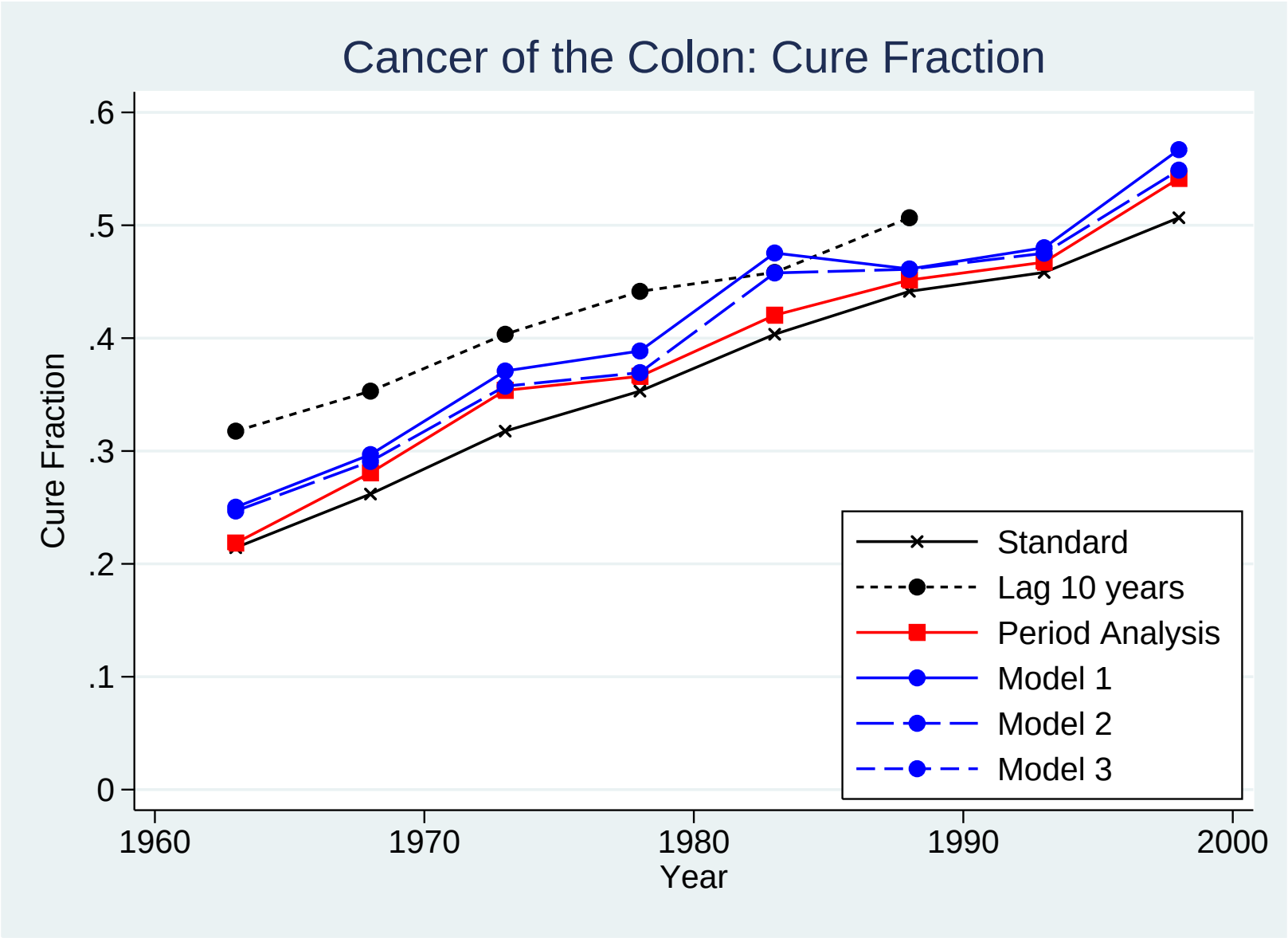
Period Analysis: Cancer of the Colon



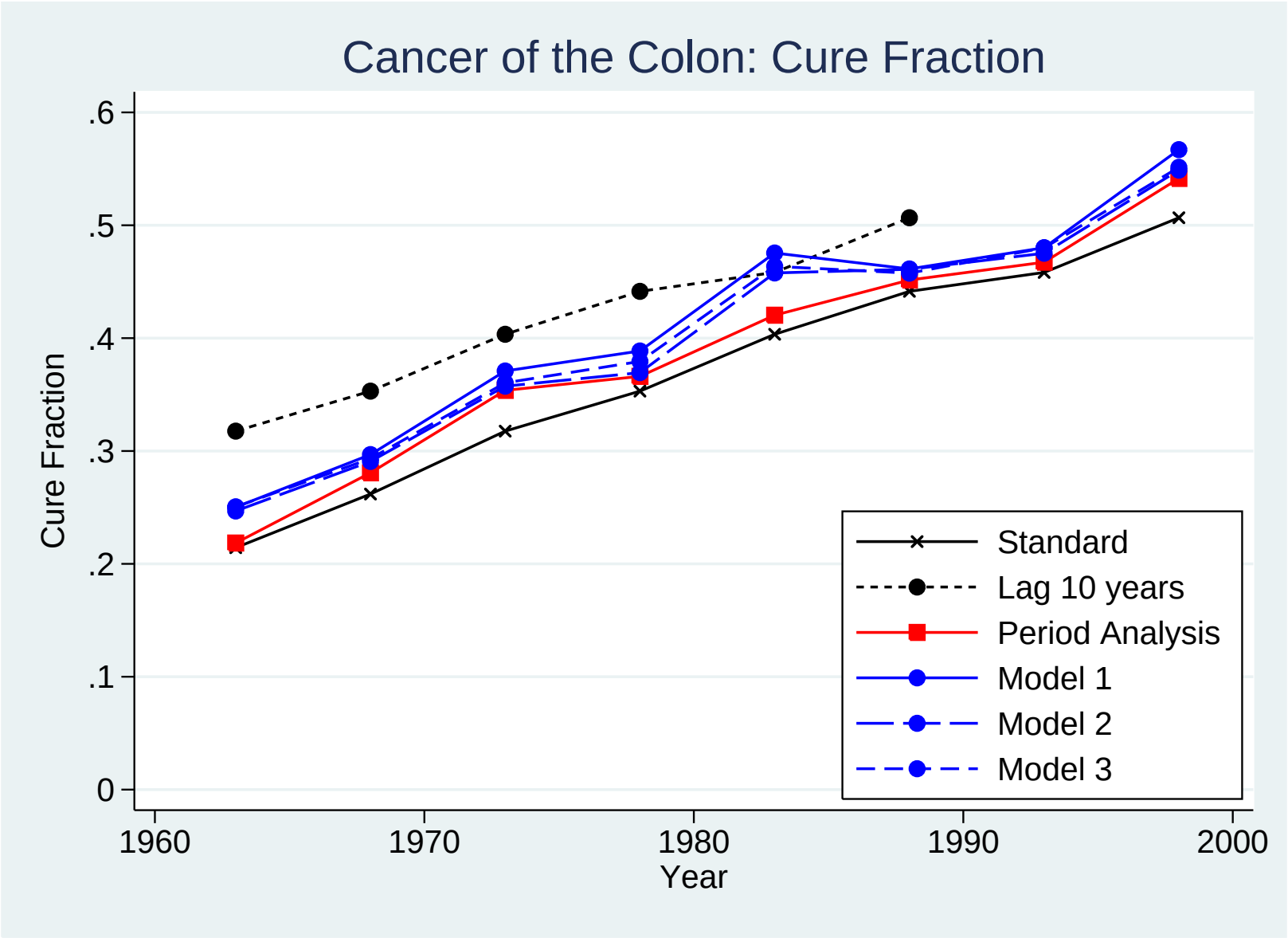
Period Analysis: Cancer of the Colon



Period Analysis: Cancer of the Colon



Period Analysis: Cancer of the Colon



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
  - LogNormal and Generalized Gamma are implemented
  - LogNormal fits poorly due to Long tail
  - Some Convergence problems with Generalized Gamma
- Two Extensions
  - Split-time models. These split the time scale into two. Within the first time interval (up to time k) use simple parametic model for the relative survival and then fit a cure fraction model condition on survival to time k.
  - Use a Finite Mixture of Distributions.

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
  - LogNormal and Generalized Gamma are implemented
  - LogNormal fits poorly due to Long tail
  - Some Convergence problems with Generalized Gamma
- Two Extensions
  - Split-time models. These split the time scale into two. Within the first time interval (up to time k) use simple parametic model for the relative survival and then fit a cure fraction model condition on survival to time k.
  - Use a Finite Mixture of Distributions.

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
  - LogNormal and Generalized Gamma are implemented
  - LogNormal fits poorly due to Long tail
  - Some Convergence problems with Generalized Gamma
- Two Extensions
  - Split-time models. These split the time scale into two. Within the first time interval (up to time k) use simple parametic model for the relative survival and then fit a cure fraction model condition on survival to time k.
  - Use a Finite Mixture of Distributions.

Mixture of Distributions

Non-Mixture Model

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

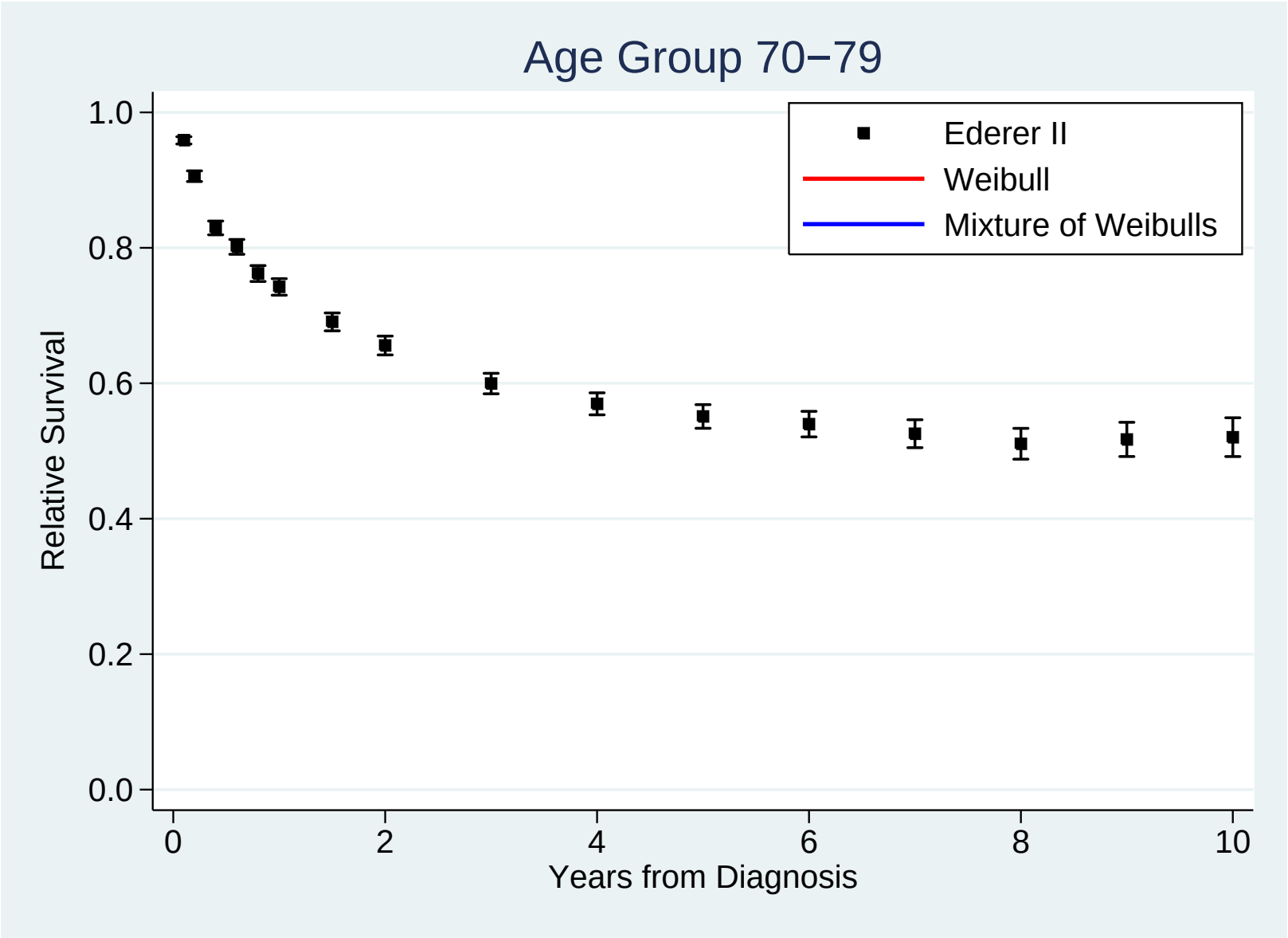
- This allows a much more flexible shape for the excess hazard and relative survival function[11].
- Mixture of two Weibull distributions generally works well.
- Can also think of two groups of individuals, those who die after a short time and those who die after a longer time.

Mixture Model

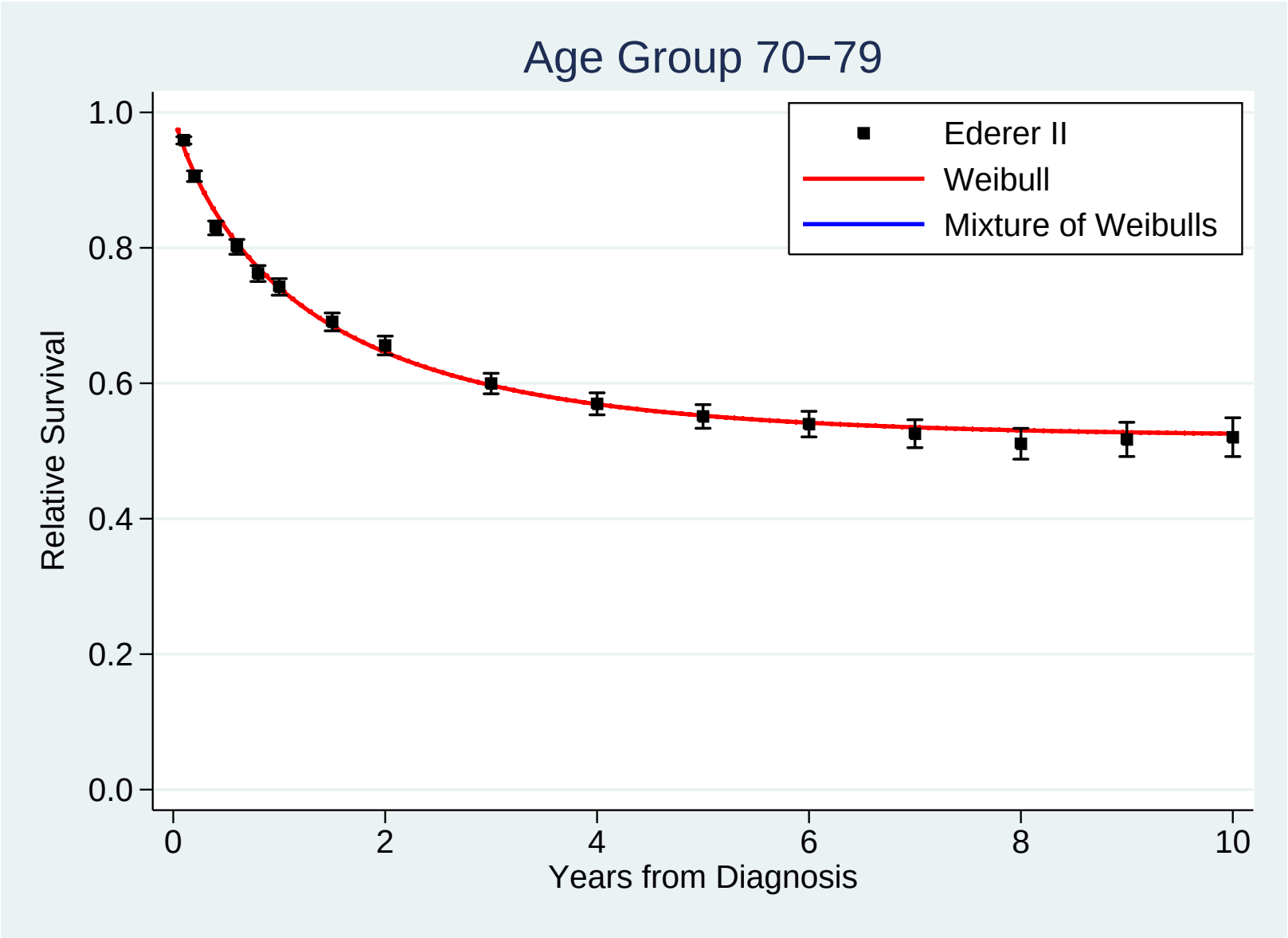
$$S(t) = S^*(t)\left(\pi + (1-\pi)\left(pS_{u1}(t) + (1-p)S_{u2}(t)\right)\right)$$

- For mixture models on relative survival scale.
- Mixture of two Weibull distributions generally works well.

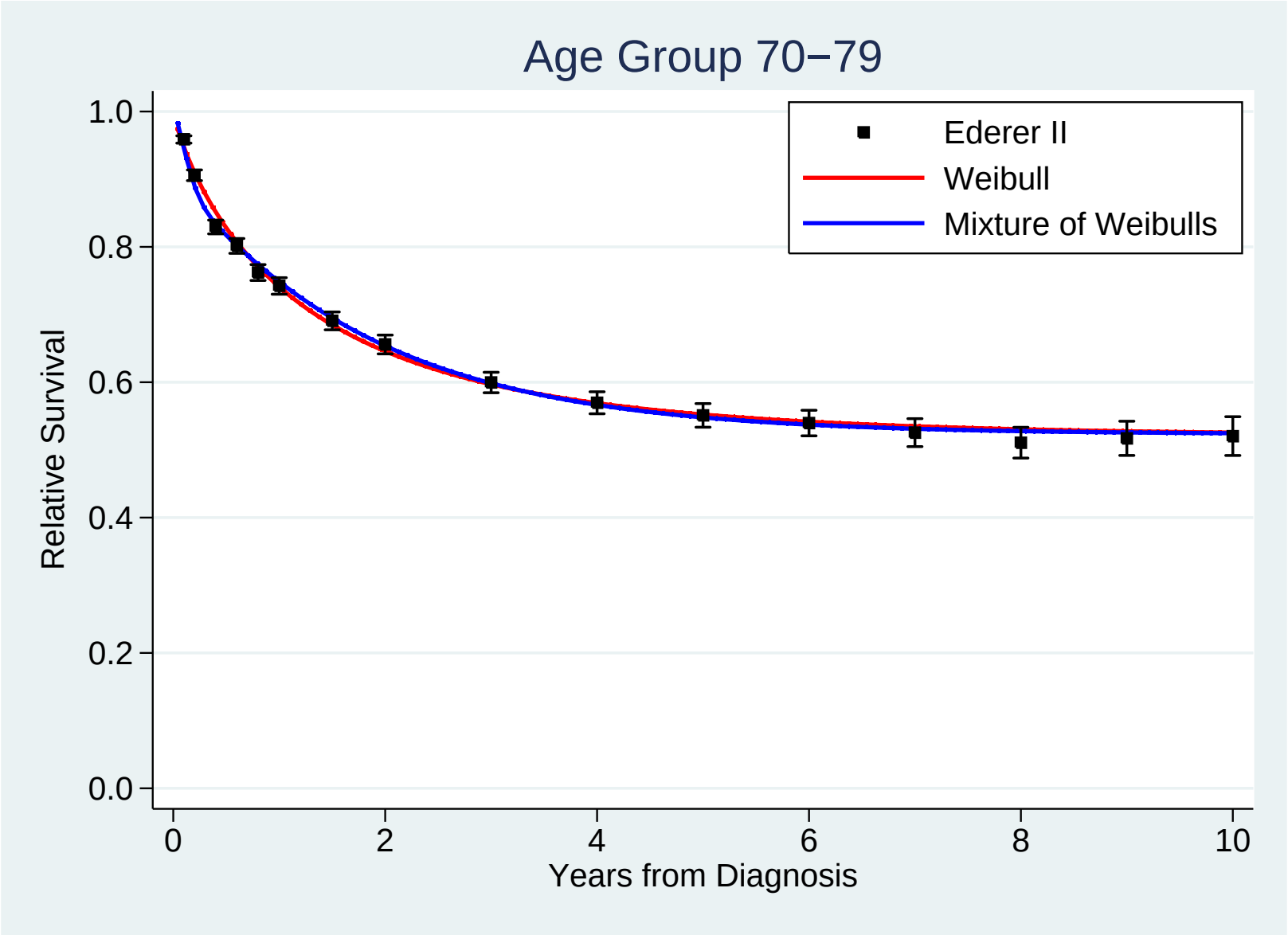
Cancer of the Colon: Weibull and Mixture of Weibulls



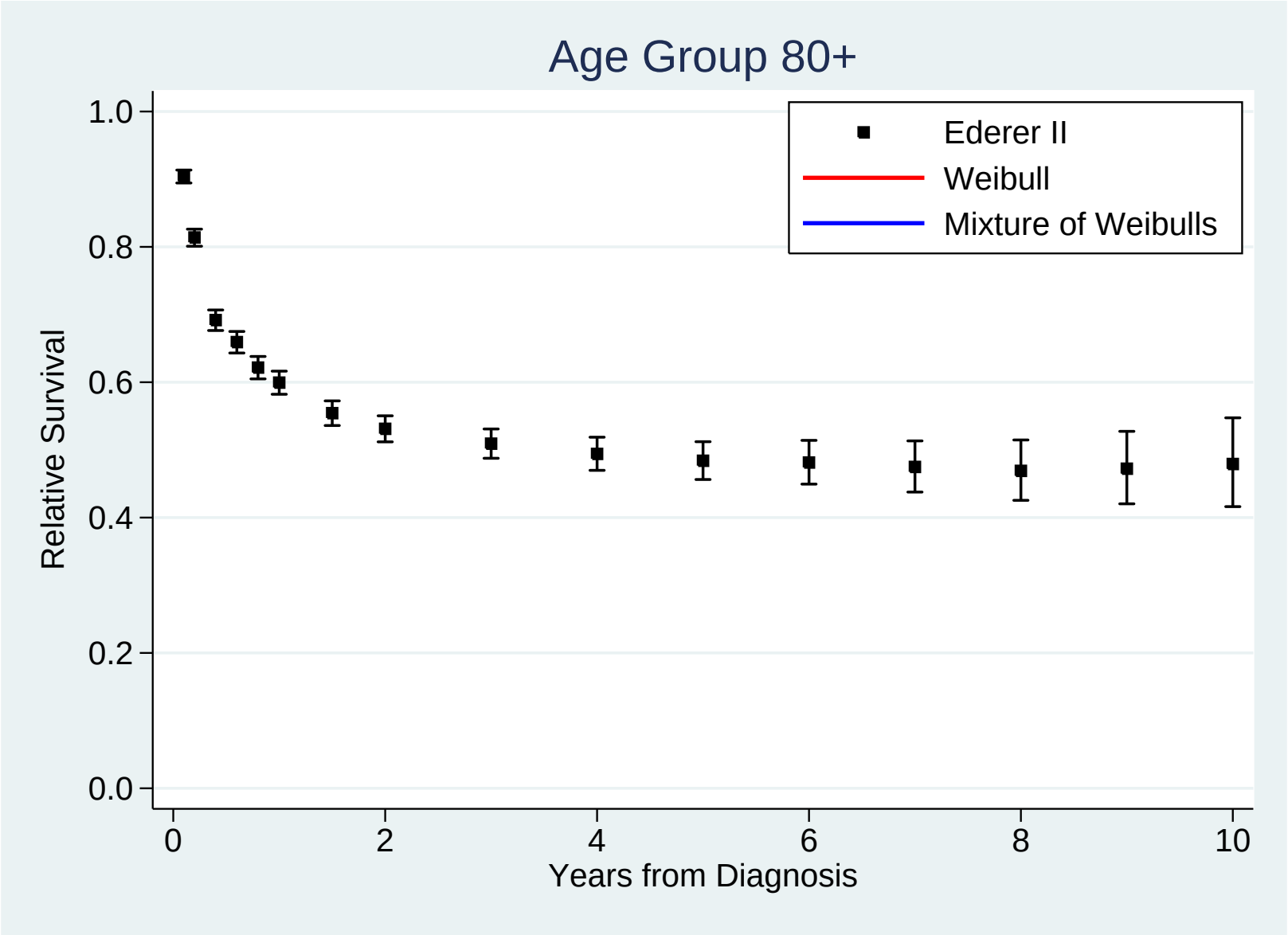
Cancer of the Colon: Weibull and Mixture of Weibulls



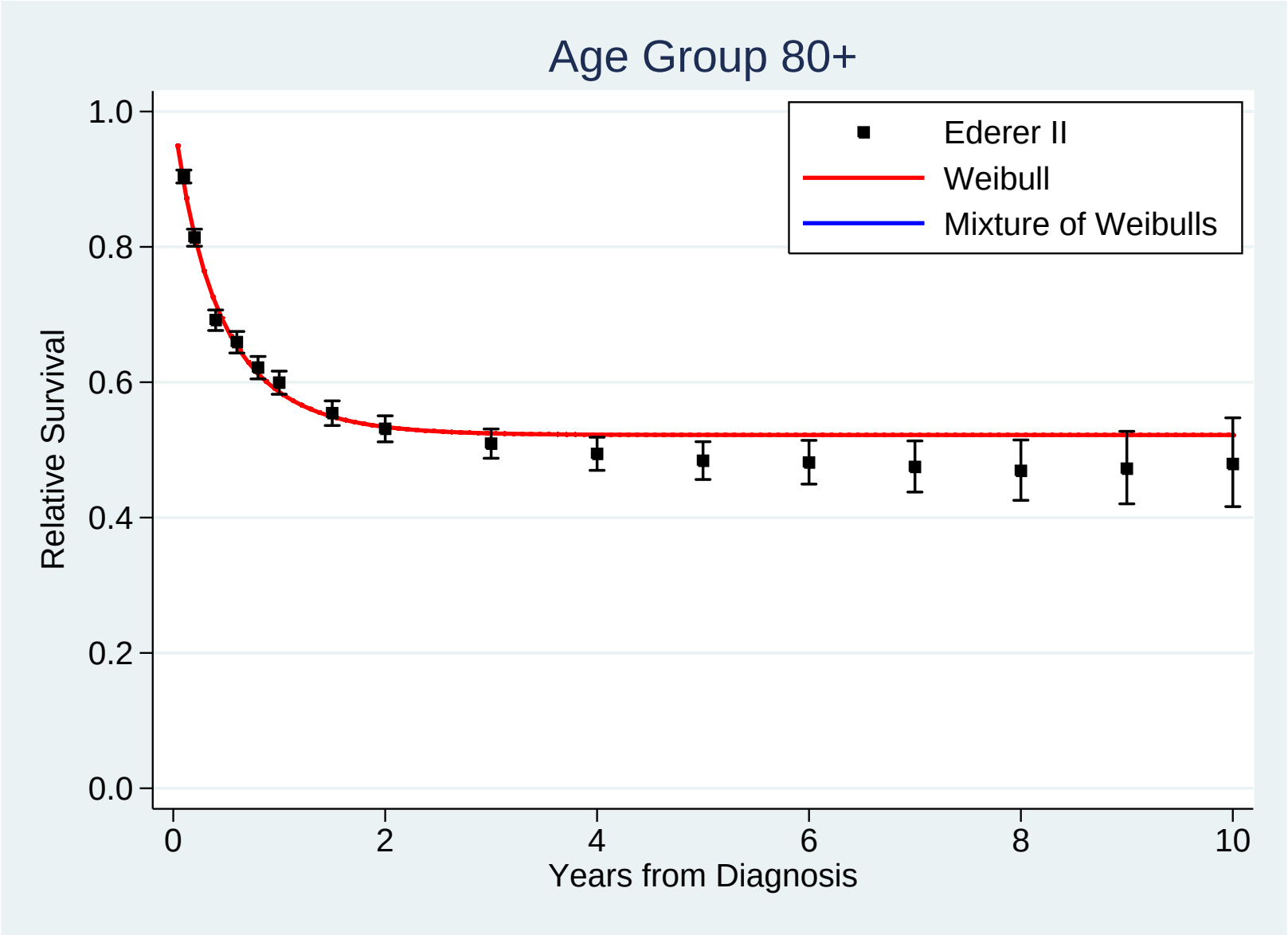
Cancer of the Colon: Weibull and Mixture of Weibulls



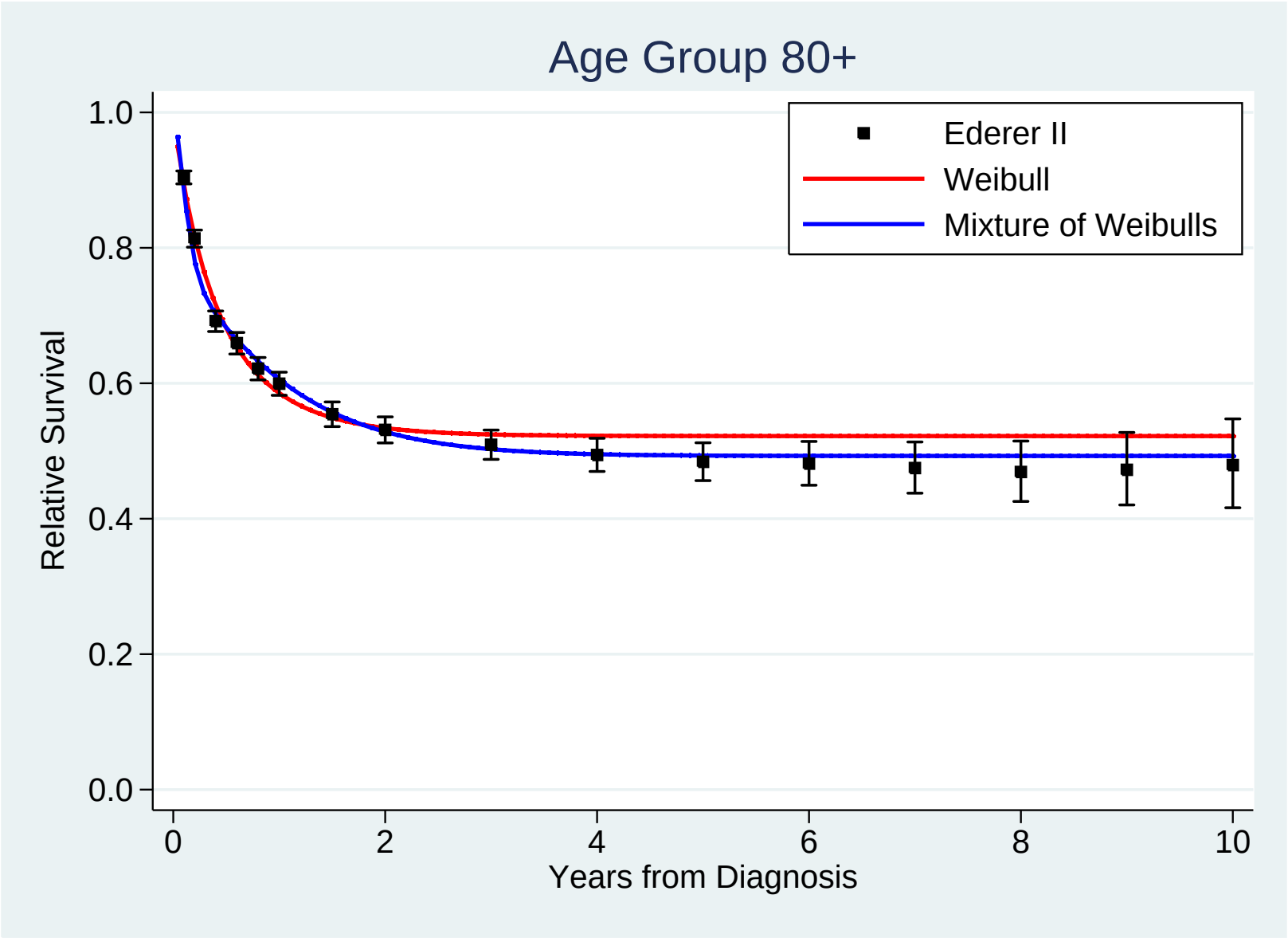
Cancer of the Colon: Weibull and Mixture of Weibulls



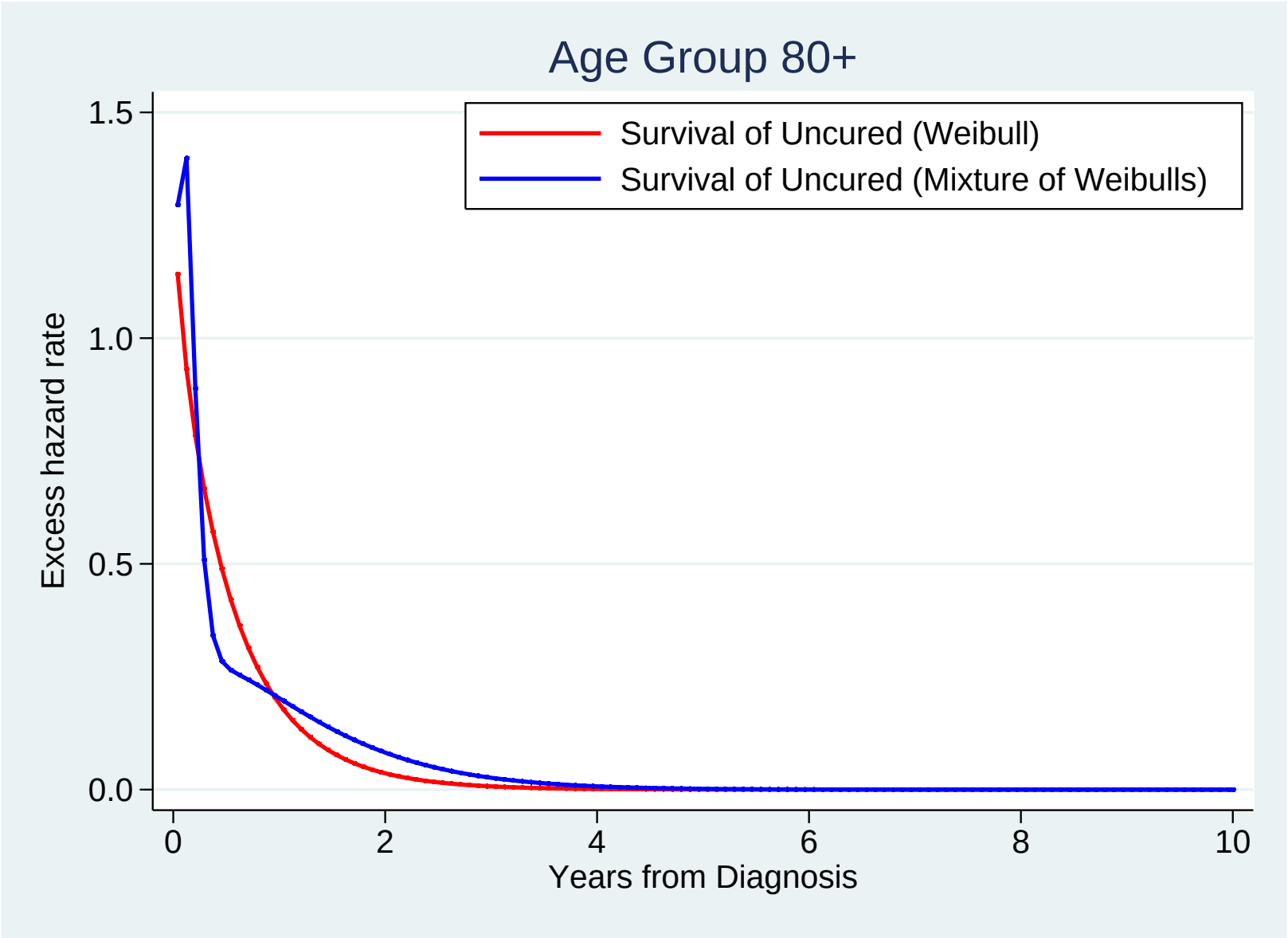
Cancer of the Colon: Weibull and Mixture of Weibulls



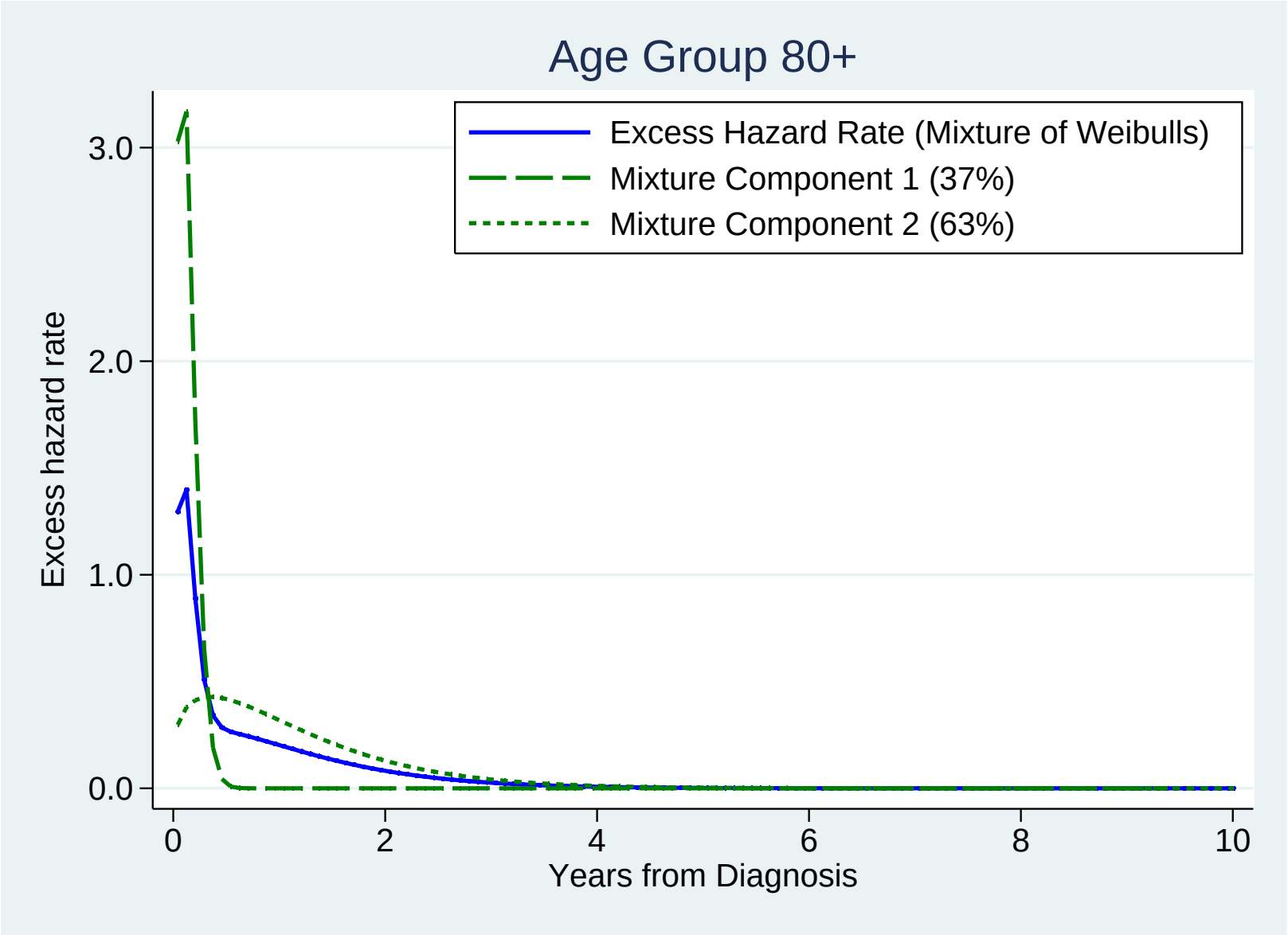
Cancer of the Colon: Weibull and Mixture of Weibulls



Cancer of the Colon: Excess Hazard Rate



Cancer of the Colon: Excess Hazard Rate



Summary

- In population based cancer studies 'cure' is often observed.
- Relative survival models that explicitly allow for 'cure' are useful for monitoring trends and differences in (relative) survival.
- `strsnmix` and strsmix fit a wide range of models.
- Incorporation of delayed entry models allows up-to-date estimates of cure to be obtained.
- Still needs to be a degree of caution
  - When 'cure' is not a reasonable assumption.
  - Follow-up not long enough.
  - Simple models may not fit the data well, but alternatives are available.
  - When the cure fraction is high (over 75-80%).

[1] H. Brenner and O. Gefeller. Deriving more up-to-date estimates of long-term patient survival. *Journal of Clinical Epidemiology*, 50(2):211–216, 1997.

[2] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–54, 1999.

[3] P.W. Dickman, A. Sloggett, M. Hills, and T. Hakulinen. Regression models for relative survival. *Statistics in Medicine*, 23:41–64, 2004.

[4] R. Giorgi, M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet, and J. Faivre. A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22(17):2767–84, 2003.

[5] P.C. Lambert. Estimating and modelling the cure fraction in population-based cancer survival analysis. *Submitted to The Stata Journal*, 2006.

[6] P.C. Lambert, L. K. Smith, D. R. Jones, and J.L. Botha. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine*, 24:3871–3885, 2005.

[7] P.C. Lambert, J.R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modelling the cure fraction in population-based cancer survival analysis. *Submitted to Biostatistics*, 2006.

[8] L. K. Smith, P. C. Lambert, J. L. Botha, and D. R. Jones. 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, 2004.

[9] R. Sposto. Cure model analysis in cancer: an application to data from the children's cancer group. *Statistics in Medicine*, 21(2):293–312, 2002.

[10] A. Verdecchia, R. De Angelis, R. Capocaccia, M. Sant, A. Micheli, G. Gatta, and F. Berrino. The cure for colon cancer: results from the eurocare study. *International Journal of Cancer*, 77(3):322–9, 1998.

[11] A.Y. Yakovlev, A.D. Tsodikov, K. Boucher, and R. Kerber. The shape of the hazard function in breast carcinoma: Curability of the disease revisited. *Cancer*, 85:1789–98, 1999.