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

\[ \text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \quad R(t) = \frac{S(t)}{S^*(t)} \]

- 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

\[ h(t) = h^*(t) + \lambda(t) \]

\[ \text{Observed Mortality Rate} = \text{Expected Mortality Rate} + \text{Excess Mortality Rate} \]
Relative Survival

$$R(t) = \frac{S(t)}{S^*(t)}$$

- 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

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

| Observed Mortality Rate | = | Expected Mortality Rate | + | Excess Mortality Rate |
Relative Survival

Relative Survival = \frac{\text{Observed Survival}}{\text{Expected Survival}} \quad R(t) = \frac{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

\[ h(t) = h^*(t) + \lambda(t) \]

Observed Mortality Rate = Expected Mortality Rate + Excess Mortality Rate
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/](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/
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.
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.
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.
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)

![Graph showing relative survival over time from diagnosis.]

- **Relative Survival** vs. **Time from Diagnosis (Years)**

- The graph illustrates the decline in relative survival over time from diagnosis, with a gradual decrease approaching zero after approximately 10 years.
Definition of Cure (2)
### 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, \( \pi \), 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.
When modelling cure we define an asymptote at the cure fraction, $\pi$, 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

\[ S(t) = S^*(t) \left( \pi + (1 - \pi) S_u(t) \right) \]
\[ h(t) = h^*(t) + \frac{(1 - \pi) f_u(t)}{\pi + (1 - \pi) S_u(t)} \]

- \( S^*(t) \) 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.
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;

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

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.
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;

\[
S(t) = S^*(t) \pi F_z(t) \quad h(t) = h^*(t) - \ln(\pi)f_z(t)
\]

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. 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.
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

\[
\text{strsmix [ } \varlist \text{ ] [ } \textit{if} \text{ ] [ } \textit{in} \text{ ] , \textit{distribution}(\textit{distribution}) \ \textit{link}(\textit{link function}) \ \text{bhazard(\varname)} \ \text{[} \text{k}1(\varlist) \ k2(\varlist) \ k3(\varlist) \ k4(\varlist) \ pmix(\varlist) \ ] \ \text{noconstant} \ \text{noconsk}1 \ \text{noconsk}2 \ \text{noconsk}3 \ \text{noconsk}4 \ \text{noconspmix} \ \text{init}(\textit{matrix \ name}) \ \text{skip \ inititer(#)} \ \text{stopconstraint} \ \textit{valconstraint}(\#) \ \text{eform} \ ]
\]

\[
\text{strsnmix [ } \varlist \text{ ] [ } \textit{if} \text{ ] [ } \textit{in} \text{ ] , \textit{distribution}(\textit{distribution}) \ \textit{link}(\textit{link function}) \ \text{bhazard(\varname)} \ \text{[} \text{k}1(\varlist) \ k2(\varlist) \ k3(\varlist) \ k4(\varlist) \ pmix(\varlist) \ ] \ \text{split(\#)} \ earlyk1(\varlist) \ earlyk2(\varlist) \ \text{noconstant} \ \text{noconsk}1 \ \text{noconsk}2 \ \text{noconsk}3 \ \text{noconsk}4 \ \text{noconspmix} \ \text{earlynoconsk}1 \ \text{earlynoconsk}2 \ \text{init}(\textit{matrix \ name}) \ \text{skip \ inititer(#)} \ \text{stopconstraint} \ \textit{valconstraint}(\#) \ \text{eform} \ ]
\]

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 auxiliary 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.
- 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**

```stata
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

Cure Fraction and Median Survival of 'Uncured'
Age Group: <50

Year of Diagnosis

Cure Fraction
Median Survival

P. Lambert
Cure Models and Relative Survival
12th UK Stata Users Group
Time Trends for Cancer of the Colon Age <50

Cure Fraction and Median Survival of 'Uncured' Age Group: <50

Cure Fraction and Median Survival of 'Uncured' Age Group: <50

- Cure Fraction
- Median Survival

Year of Diagnosis


Cure Fraction

Median Survival
Time Trends for Cancer of the Colon

Cure Fraction

- <50 years
- 50–59 years
- 60–69 years
- 70–80 years

Year of Diagnosis


Cure Fraction

0.00 0.20 0.40 0.60 0.80 1.00
Time Trends for Cancer of the Colon

Median Survival of 'Uncured'

- <50 years
- 50–59 years
- 60–69 years
- 70–80 years

Year of Diagnosis:
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000
Quantifying Differences

Difference in Cure Fraction (Age Group <50 – Age Group 70–79)

Years from Diagnosis

Difference in Cure Fraction

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 into modelling environment[8].
- In survival models period analysis can be incorporated using delayed entry techniques.
Period Analysis

Start and Stop at Risk Times

Standard Period

(0, 2)

(0, 4)

(0, 6)

(0, 3)

Subject 1

Subject 2

Subject 3

Subject 4

Diagnosis

Death or Censoring

Year


Paul C Lambert  Cure Models and Relative Survival  12th UK Stata Users Group

19/27
Period Analysis

- **Period of Interest**
- **Start and Stop at Risk Times**
  - **Standard Period**
    - Subject 1: (0, 2)
    - Subject 2: (0, 4)
    - Subject 3: (0, 6)
    - Subject 4: (0, 3)

**Year**

**Diagnosis**
- Subject 1
- Subject 2
- Subject 3
- Subject 4

**Death or Censoring**
- Subject 4
Period Analysis

<table>
<thead>
<tr>
<th>Subject</th>
<th>Start and Stop at Risk Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
</tr>
<tr>
<td>Subject 1</td>
<td>(0, 2)</td>
</tr>
<tr>
<td>Subject 2</td>
<td>(0, 4)</td>
</tr>
<tr>
<td>Subject 3</td>
<td>(0, 6)</td>
</tr>
<tr>
<td>Subject 4</td>
<td>(0, 3)</td>
</tr>
</tbody>
</table>

- **Diagnosis**
- **Death or Censoring**
Cancer of the Colon: Cure Fraction

- Standard
- Lag 10 years
- Period Analysis
- Model 1
- Model 2
- Model 3
Period Analysis: Cancer of the Colon

Cancer of the Colon: Cure Fraction

Year


Cure Fraction

0.0 0.1 0.2 0.3 0.4 0.5 0.6

Standard

Lag 10 years

Period Analysis

Model 1

Model 2

Model 3
Period Analysis: Cancer of the Colon

Cancer of the Colon: Cure Fraction

Year

Cure Fraction


Standard
Lag 10 years
Model 1
Model 2
Model 3

Cure Models and Relative Survival
12th UK Stata Users Group
Period Analysis: Cancer of the Colon

Cancer of the Colon: Cure Fraction

Year

Cure Fraction

Standard
Lag 10 years
Period Analysis
Model 1
Model 2
Model 3

Period Analysis: Cancer of the Colon

Cancer of the Colon: Cure Fraction

Year

Cure Fraction

Standard
Lag 10 years
Period Analysis
Model 1
Model 2
Model 3

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 parametric 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 parametric model for the relative survival and then fit a cure fraction model condition on survival to time $k$.
- Use a Finite Mixture of Distributions.
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 parametric 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

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

- 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

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

- For mixture models on relative survival scale.
- Mixture of two Weibull distributions generally works well.
Cancer of the Colon: Weibull and Mixture of Weibulls

Age Group 70–79

- Ederer II
- Weibull
- Mixture of Weibulls
Cancer of the Colon: Weibull and Mixture of Weibulls

Age Group 70–79

- Ederer II
- Weibull
- Mixture of Weibulls

Relative Survival vs. Years from Diagnosis

0.0 0.2 0.4 0.6 0.8 1.0

0 2 4 6 8 10

Years from Diagnosis
Cancer of the Colon: Weibull and Mixture of Weibulls

Age Group 70–79

Relative Survival

Years from Diagnosis

- Ederer II
- Weibull
- Mixture of Weibulls
Cancer of the Colon: Weibull and Mixture of Weibulls

Age Group 80+

Relative Survival vs. Years from Diagnosis

- Ederer II
- Weibull
- Mixture of Weibulls
Cancer of the Colon: Weibull and Mixture of Weibulls

Age Group 80+

Relative Survival vs. Years from Diagnosis for different models:
- Ederer II
- Weibull
- Mixture of Weibulls

Paul C Lambert  Cure Models and Relative Survival  12th UK Stata Users Group
Cancer of the Colon: Excess Hazard Rate

Age Group 80+

- Survival of Uncured (Weibull)
- Survival of Uncured (Mixture of Weibulls)
Cancer of the Colon: Excess Hazard Rate

**Age Group 80+**

- Excess Hazard Rate (Mixture of Weibulls)
- Mixture Component 1 (37%)
- Mixture Component 2 (63%)

Years from Diagnosis:

- 0
- 2
- 4
- 6
- 8
- 10

Excess hazard rate:

- 0.0
- 1.0
- 2.0
- 3.0

Paul C Lambert

Cure Models and Relative Survival

12th UK Stata Users Group
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.
\texttt{strsmix} and \texttt{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%).


