An introduction and some recent developments in statistical methods for population-based cancer survival analysis

Paul W. Dickman
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden
paul.dickman@ki.se

http://www.pauldickman.com/

22 November 2010

Overview

- Introduction to relative survival and why it is often preferred over cause-specific survival for the study of cancer patient survival using data collected by population-based cancer registries.
- Period vs cohort estimation.
- Age standardised estimates of relative survival.
- Cure models
- Is net survival (e.g., relative survival) a useful measure for patients and clinicians?
  - estimating crude probabilities of death
- Estimating and modelling relative survival in Stata (strs command).

Measures used in cancer control

- The most common measures are incidence, mortality, and survival.
- By 'mortality' we typically mean mortality in the population, whereas 'survival' is nothing more than mortality among the patients (those diagnosed with cancer).
- In follow-up studies we can present estimates as either event rates or survival proportions.
- Population-based studies of patient survival provide a measure of the effectiveness of the health care system in diagnosing and treating cancer.

How might we measure the prognosis of cancer patients?

- Total mortality (among the patients).
- Our interest is typically in net mortality (mortality associated with a diagnosis of cancer).
- Cause-specific mortality provides an estimate of net mortality.
- When estimating cause-specific mortality only those deaths which can be attributed to the cancer in question are considered to be events.

$$\text{cause-specific mortality} = \frac{\text{number of deaths due to cancer}}{\text{person-time at risk}}$$

The survival times of patients who die of causes other than cancer are censored.

Potential disadvantages of cause-specific survival

- Using cause-specific mortality requires that reliably coded information on cause of death is available.
- Even when cause of death information is available to the cancer registry via death certificates, it is often vague and difficult to determine whether or not cancer is the primary cause of death.
- How do we classify, for example, deaths due to treatment complications?
- Consider a man diagnosed with prostate cancer and treated with estrogen who dies following a myocardial infarction. Do we classify this death as 'due entirely to prostate cancer' or 'due entirely to other causes'?
- Welch *et al.* [1] studied deaths among surgically treated cancer patients that occurred within one month of diagnosis. They found that 41% of deaths were not attributed to the coded cancer.

Relative survival

- Can instead estimate excess mortality: the difference between observed (all-cause) and expected mortality.

$$\begin{array}{ccccc} \text{excess} & = & \text{observed} & - & \text{expected} \\ \text{mortality} & & \text{mortality} & & \text{mortality} \end{array}$$

- Relative survival is the survival analog of excess mortality — the relative survival ratio is defined as the observed survival in the patient group divided by the expected survival of a comparable group from the general population.

$$\text{relative survival ratio} = \frac{\text{observed survival proportion}}{\text{expected survival proportion}}$$

- It is usual to estimate the expected survival proportion from nationwide (or statewide) population life tables stratified by age, sex, calendar time, and, where applicable, race [2].

- Although these tables include the effect of deaths due to the cancer being studied, Ederer *et al.* [3] showed that this does not, in practice, affect the estimated survival proportions.
- A major advantage of relative survival (excess mortality) is that information on cause of death is not required, thereby circumventing problems with the inaccuracy [4] or nonavailability of death certificates.
- We obtain a measure of the excess mortality experienced by patients diagnosed with cancer, irrespective of whether the excess mortality is directly or indirectly attributable to the cancer.
- Deaths due to treatment complications or suicide are examples of deaths which may be considered indirectly attributable to cancer.

Cervical cancer diagnosed in New Zealand 1994 – 2001
Life table estimates of patient survival

Women diagnosed 1994 - 2001 with follow-up to the end of 2002

I	N	D	W	Effective number at risk	Interval-specific observed survival	Cumulative observed survival	Cumulative expected survival	Interval-specific relative survival	Cumulative relative survival
1	1559	209	0	1559.0	0.86594	0.86594	0.98996	0.87472	0.87472
2	1350	125	177	1261.5	0.90091	0.78014	0.98192	0.90829	0.79450
3	1048	58	172	962.0	0.93971	0.73310	0.97362	0.94772	0.75296
4	818	32	155	740.5	0.95679	0.70142	0.96574	0.96459	0.72630
5	631	23	148	557.0	0.95871	0.67246	0.95766	0.96679	0.70218
6	460	10	130	395.0	0.97468	0.65543	0.94972	0.98284	0.69013
7	320	5	129	255.5	0.98043	0.64261	0.94198	0.98848	0.68219
8	186	3	134	119.0	0.97479	0.62641	0.93312	0.98405	0.67130
9	49	1	48	25.0	0.96000	0.60135	0.91869	0.97508	0.65457

Issues with relative survival

- The central issue in estimating relative survival is defining a 'comparable group from the general population' and estimating expected survival.
- If not all of the excess mortality is due to the cancer then the relative survival ratio will underestimate net survival (overestimate excess mortality).
- For example, patients diagnosed with smoking-related cancers will experience excess mortality, compared to the general population, due to both the cancer and other smoking related conditions.
- Should the patients be a selected group from the general population, for example, with respect to social class, the national population might not be an appropriate comparison group.

Relative survival example

Table 1: Number of cases (N) and 5-year observed (p), expected (p^*), and relative (r) survival for males diagnosed with localised skin melanoma in Finland during 1985–1994.

Age	N	p	p^*	r
15–29	67	0.947	0.993	0.954
30–44	273	0.856	0.982	0.872
45–59	503	0.824	0.943	0.874
60–74	449	0.679	0.815	0.833
75+	200	0.396	0.505	0.784

- Note that relative survival controls for the fact that expected mortality depends on demographic characteristics (age, sex, etc.).
- In addition, relative survival may, and usually does, depend on such factors.

Statistical cure

- The life table is a useful tool for describing the survival experience of the patients over a long follow-up period.
- In particular, an interval-specific relative survival ratio equal to one indicates that, during the specified interval, mortality in the patient group was equivalent to that of the general population.
- The attainment and maintenance of an interval-specific RSR of one indicates that there is no excess mortality due to cancer and the patients are assumed to be 'statistically cured'.
- An individual is considered to be medically cured if he or she no longer displays symptoms of the disease.
- Statistical cure applies at a group, rather than individual, level.

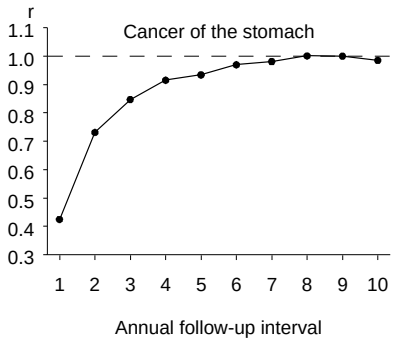
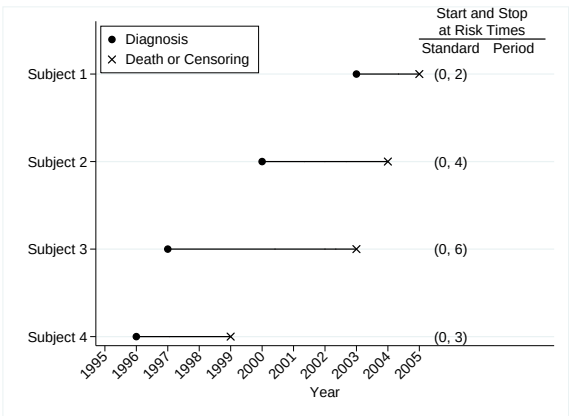


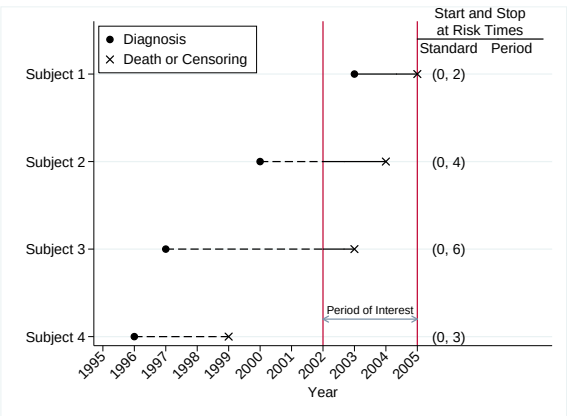
Figure 1: Plots of the annual (interval-specific) relative survival ratios (r) for males and females diagnosed with cancer of the stomach in Finland 1985–1994 and followed up to the end of 1995.

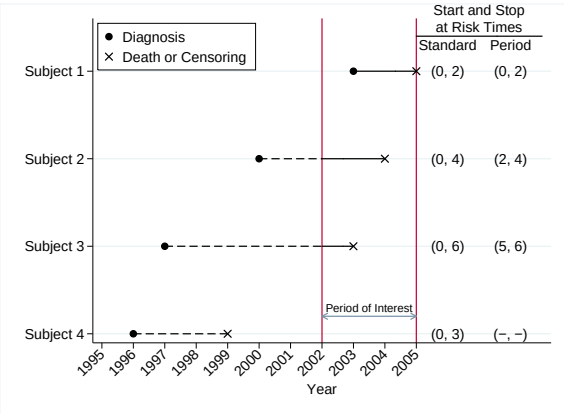
- Plots of the interval-specific RSR are also useful for assessing the quality of follow-up.
- If the interval-specific RSR levels out at a value greater than 1, this generally indicates that some deaths have been missed in the follow-up process.
- An interval-specific relative survival ratio of unity is generally not achieved for smoking-related cancers, such as cancer of the lung and kidney.
- Compared to the general population, these patients are subject to excess mortality due to the cancer in addition to excess mortality due to other conditions caused by smoking, such as cardiovascular disease.
- We'll return to these concepts later when we discuss cure models.

Estimating relative survival using a period approach

- In 1996 Hermann Brenner suggested estimating cancer patient survival using a period, rather than cohort, approach [5].
- Time at risk is left truncated at the start of the period window and right censored at the end.
- This suggestion was initially met with scepticism although studies based on historical data [6] have shown that
  - period analysis provides very good predictions of the prognosis of newly diagnosed patients; and
  - highlights temporal trends in patient survival sooner than cohort methods.







Age-standardisation of relative survival

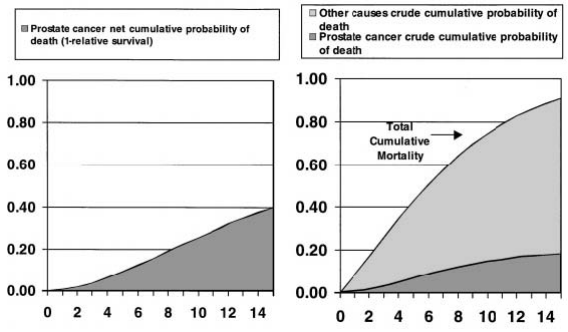
- The problem is more complex than age-standardisation of, for example, incidence rates since the age-distribution of the patients changes during follow-up.
- Which weights do we use and how does one interpret the resulting estimates?
- See the papers by Pokhrel et al and Brenner et al. [7, 8, 9, 10, 11].

Interpreting relative survival estimates

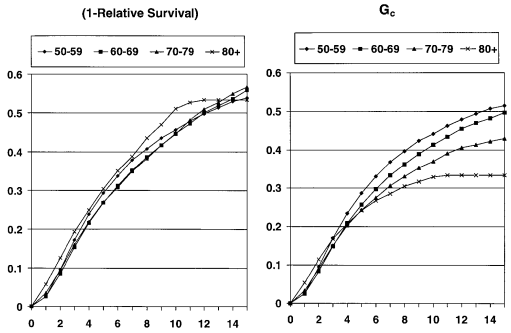
- Patient survival is the most important single measure of cancer patient care (the diagnosis and treatment of cancer) and is of considerable interest to clinicians, patients, researchers, politicians, health administrators, and public health professionals [12].
- However, relatively little attention has been paid to the fact that each of these consumers of survival statistics have quite different needs.
- The standard approach of estimating net survival (relative survival or cause-specific survival) is useful for comparing populations but not necessarily relevant to individual patients since such estimates are interpreted in the hypothetical scenario where cancer is the only possible cause of death.
- The cumulative relative survival ratio can be interpreted as the proportion of patients alive after i years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death.

- 1-RSR can be interpreted as the proportion of patients who will die of cancer within i years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death.
- We do not live in this hypothetical world (where we estimate what is called the net probability of death). Estimates of the proportion of patients who will die of cancer in the presence of competing risks can also be made (crude probabilities of death).
- Cronin and Feuer (2000) [13] extended the theory of competing risks to relative survival; their method is implemented in our Stata command `strs`.

Net (left) and crude (right) probabilities of death in men with localized prostate cancer aged 70+ at diagnosis.



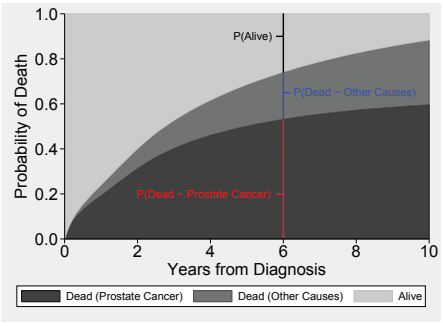
Net (left) and crude (right) probabilities of death due to cancer in women with regional breast cancer.



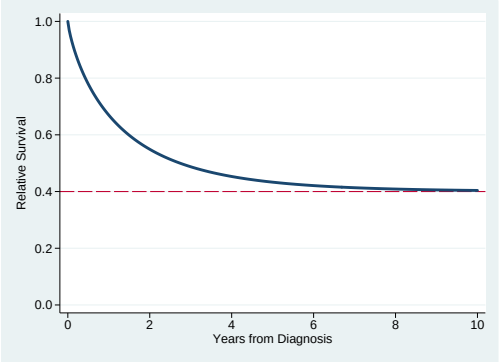
Estimating crude and net mortality for individual data

- Lambert et al [14] showed how to estimate crude and net probabilities based on flexible parametric models for relative survival.
- This approach avoids having to split the timescale and facilitates easy modelling of continuous covariates. That is, we can obtain predictions for individual values of age or other covariates.

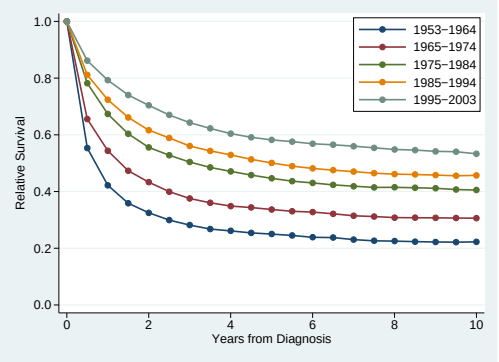
Crude probabilities of death for men diagnosed with prostate cancer in England and Wales, aged 75+ at diagnosis.



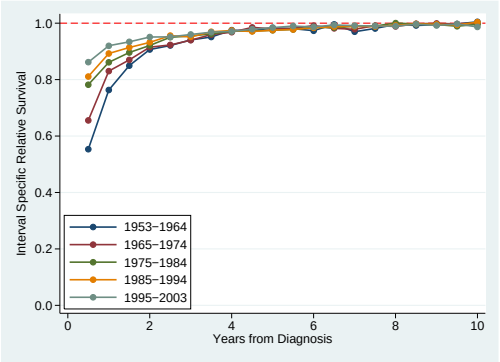
Cure models – definition of cure



Relative Survival for Cancer of the Colon in Finland



Relative Survival for Cancer of the Colon in Finland



Mixture cure models

- When modelling cure we define an asymptote at the cure fraction, π, for the relative survival function, $R(t)$[15, 16].

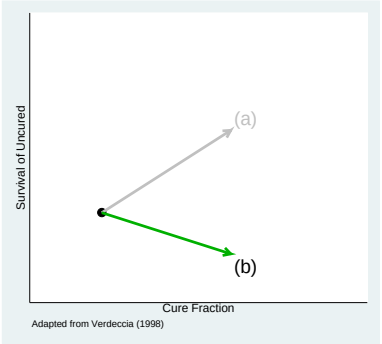
$$S(t) = S^*(t)(\pi + (1-\pi)S_u(t))$$

Survival for the 'Uncured'

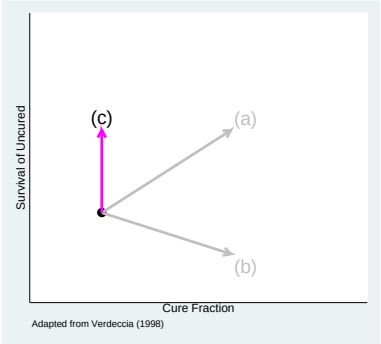
- As well as the cure fraction, summaries of the 'uncured' (those 'bound to die') are potentially useful.
- For example, mean or median survival or some other percentile of the survival distribution.
- We need to choose parametric form for $S(t)$.
- For many scenarios the Weibull distribution provides a good fit.

Interpreting changes over time – general improvement

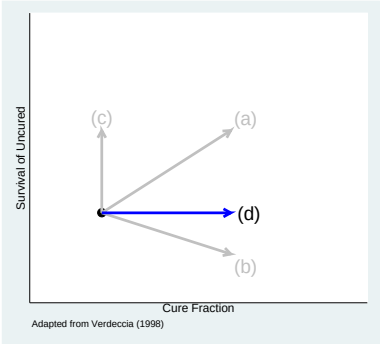
Interpreting changes over time – selective improvement



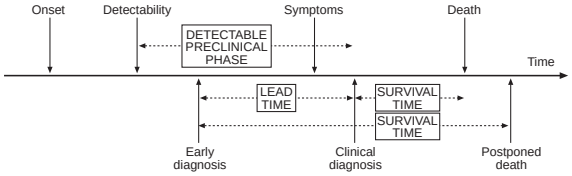
Improved palliative care or lead time



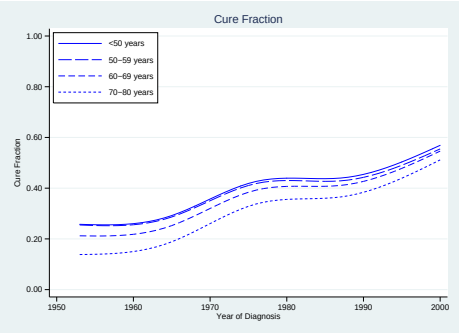
Inclusion of subjects with no excess risk



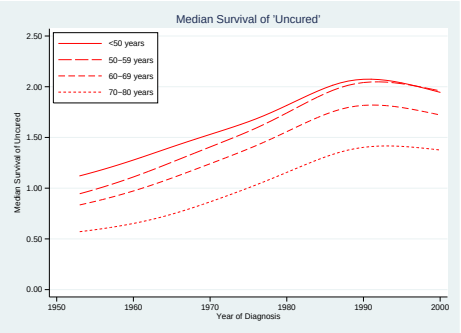
The cure proportion is not affected by lead time



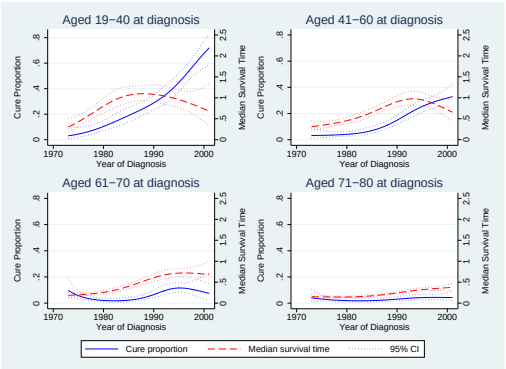
Time Trends for Cancer of the Rectum in Finland [15] cure proportion



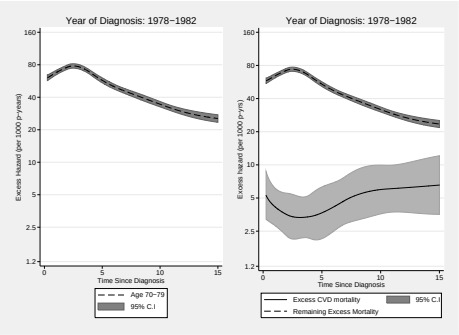
Time Trends for Cancer of the Rectum in Finland [15] median survival of the uncured



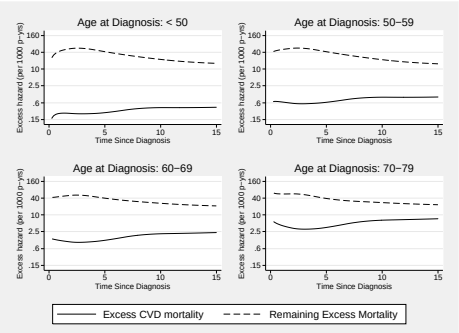
Time Trends for AML in Sweden [17]



Partitioning excess mortality for women diagnosed with breast cancer



Partitioning excess mortality for women diagnosed with breast cancer



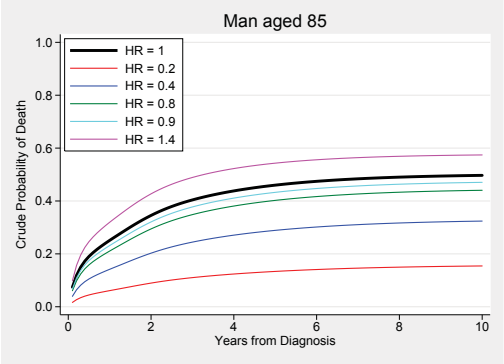
How do we compare treatments? What does a hazard ratio of 0.8 mean?

- The standard measure used for comparing treatments, the hazard ratio, is typically a ratio of net mortality rates.
- That is, a ratio of two rates each estimated in the hypothetical world where one cannot die of anything other than cancer.
- Are such estimates easily interpretable in the clinical setting?

Crude probability of death due to prostate cancer

Which line represents the crude probability for a HR of 0.8?

The answer



Treatment comparison and risk communication in the clinical setting

- Informal surveys we have conducted among colleagues suggest that not even epidemiologists and biostatisticians have a good intuitive feel for how a specified hazard ratio will impact survival probabilities.
- From a clinical perspective it may be more useful to compare crude probabilities of death, ideally as natural frequencies.

Illustration of the use of natural frequencies to communicate expected 10-year prognosis under two alternative treatments. Treatment B has a 20% lower net mortality than treatment A (i.e., the hazard ratio is 0.8).

	Age 55		*Age 85*	
Outcome	*A*	*B*	*A*	*B*
Die of cancer	55	48	46	40
Die of other causes	2	2	32	35
Alive	43	50	22	25
Total	100	100	100	100

Applying relative survival to diseases other than cancer

- In order to interpret excess mortality as 'mortality due to the disease of interest' we need to accurately estimate expected mortality (the mortality that would have been observed in the absence of the disease).
- General population mortality rates may not satisfy this criteria.
- Excess mortality (compared to the general population) may nevertheless still be of interest.
- Recent applications in cardiovascular disease[18] and HIV/AIDS[19, 20].

  Nelson et al. Relative survival: what can cardiovascular disease learn from cancer? *Eur Heart J.* 2008;29:941-7.

  Bhaskaran et al. Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *JAMA* 2008;300:51-9.

Modelling excess mortality (relative survival)

- Instead of cause-specific mortality we estimate excess mortality: the difference between observed (all-cause) and expected mortality.

$$\begin{array}{ccccc} \text{excess} & = & \text{observed} & - & \text{expected} \\ \text{mortality} & & \text{mortality} & & \text{mortality} \end{array}$$

- Relative survival is the survival analog of excess mortality.
- A major advantage of relative survival (excess mortality) is that information on cause of death is not required.
- We obtain a measure of the excess mortality experienced by patients diagnosed with cancer, irrespective of whether the excess mortality is directly or indirectly attributable to the cancer.
- Both cause-specific survival and relative survival estimate the same underlying quantity (net survival) and the estimates should be similar.

Modelling excess mortality (relative survival)

- The hazard at time since diagnosis t for persons diagnosed with cancer is modelled as the sum of the known baseline hazard, $\lambda^*(t)$, and the excess hazard due to a diagnosis of cancer, $\nu(t)$ [21, 22, 23, 24, 25].

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

- Follow-up time is partitioned into bands corresponding to life table intervals and indicator variables included in the design matrix. The model is written as

$$\lambda(\mathbf{x}) = \lambda^*(\mathbf{x}) + \exp(\mathbf{x}\beta) \quad (1)$$

or

$$\ln\left[\lambda(\mathbf{x}) - \lambda^*(\mathbf{x})\right] = \mathbf{x}\beta.$$

- The exponentiated parameter estimates have an interpretation as excess hazard ratios, also known as relative excess risks.
- An excess hazard ratio of, for example, 1.5 for males compared to females implies that the excess hazard associated with a diagnosis of cancer is 50% higher for males than females.
- Non-proportional excess hazards are common but can be incorporated by introducing follow-up time by covariate interaction terms.

Modelling excess mortality using Poisson regression

- The model assumes piecewise constant hazards which implies a Poisson process for the number of deaths in each interval. We can therefore estimate the model in the framework of generalised linear models.
- We assume that the total number of deaths, d_j, for observation j can be described by a Poisson distribution, $d_j \sim \text{Poisson}(\mu_j)$ where $\mu_j = \lambda_j y_j$ and y_j is person-time at risk for the observation. Equation 1 is then written as

$$\ln(\mu_j - d_j^*) = \ln(y_j) + \mathbf{x}\beta, \tag{2}$$

  where d_j^* is the expected number of deaths (due to causes other than the cancer of interest and estimated from general population mortality rates).
- This implies a generalised linear model with outcome d_j, Poisson error structure, link $\ln(\mu_j - d_j^*)$, and offset $\ln(y_j)$. This is not a standard link function so the link is defined in `rs.ado`.

rs.ado

```
program define rs
   version 7
   args todo eta mu return
   if `todo' == -1 {
      global SGLM_lt "Relative survival"
      global SGLM_lf "log(u-d*)"
      exit
   }
   if `todo' == 0 {          /* eta = g(mu) */
      gen double `eta' = ln(`mu'-$SGLM_p)
      exit
   }
   if `todo' == 1 {          /* mu = g^-1(eta) */
      gen double `mu' = exp(`eta')+$SGLM_p
      exit
   }
   if `todo' == 2 {          /* (d mu)/(d eta) */
      gen double `return' = exp(`eta')
      exit
   }
   if `todo' == 3 {          /* (d^2 mu)(d eta^2) */
      gen double `return' = exp(`eta')
      exit
   }
   di as error "Unknown call to glm link function"
   exit 198
end
```


Poisson regression for the colon carcinoma data

- When we `stset` the data we specify all deaths as events.

```
. stset exit, fail(status==1 2) origin(dx) scale(365.25) id(id)
```

- We use `strs` to estimate relative survival for each combination of relevant predictor variables and save the results to a file.

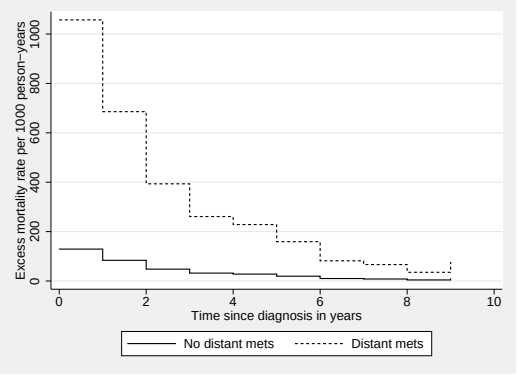
```
. strs using popmort, br(0(1)10) mergeby(_year sex _age)
> by(sex distant agegrp year8594) notables save(replace)
```

- We then fit the Poisson regression model using the resulting output file (which contains the observed (`d`) and expected (`d_star`) numbers of deaths for each life table interval along with person-time at risk (`y`)).


```
. use grouped, clear
. xi: glm d i.end distant, fam(pois) link(rs d_star) lnoffset(y) eform
------------------------------------------------------------------------------
           d |       ExpB   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iend_2 |   .6485116   .0222687   -12.61   0.000      .606302    .6936598
    _Iend_3 |    .372179   .0202302   -18.18   0.000     .3345676    .4140186
    _Iend_4 |   .2468263   .0196903   -17.54   0.000     .2110997    .2885992
    _Iend_5 |   .2160604   .0210312   -15.74   0.000     .1785334    .2614753
    _Iend_6 |   .1505581   .0215428   -13.23   0.000     .1137389    .1992964
    _Iend_7 |   .0773745   .0191536   -10.34   0.000     .0476308    .1256921
    _Iend_8 |   .0628595   .0191633    -9.08   0.000     .0345839     .114253
    _Iend_9 |   .0333979    .018285    -6.21   0.000     .0114208    .0976658
   _Iend_10 |   .0728102   .0235883    -8.09   0.000     .0385859    .1373902
    distant |    8.18794   .2588833    66.50   0.000      7.69594    8.711393
------------------------------------------------------------------------------
```

- We estimate that excess mortality is 8.2 times higher for patients with distant metastases at diagnosis compared to patients without distant metastases at diagnosis.

Fitted values from the excess mortality model



- Can adjust for additional variables.

```
. xi: glm d i.end sex i.agegrp year8594 distant, fam(pois)
> link(rs d_star) lnoffset(y) eform
i.end          _Iend_1-10       (naturally coded; _Iend_1 omitted)
i.agegrp       _Iagegrp_0-3     (naturally coded; _Iagegrp_0 omitted)
------------------------------------------------------------------------------
           d |       ExpB   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iend_2 |   .6582263     .022388   -12.30   0.000     .6157772    .7036016
[output omitted]
   _Iend_10 |    .07477   .0243443    -7.97   0.000     .0394989    .1415367
        sex |   .9878062   .0272241    -0.45   0.656     .9358634    1.042632
_Iagegrp_1 |   1.046824   .0680002     0.70   0.481     .9216818    1.188959
_Iagegrp_2 |    1.17649    .070505     2.71   0.007     1.046109     1.32312
_Iagegrp_3 |   1.549778   .0950402     7.14   0.000     1.374262     1.74771
  year8594 |   .8909376   .0238832    -4.31   0.000     .8453358    .9389994
   distant |   8.008541   .2490144    66.91   0.000     7.535056    8.511779
------------------------------------------------------------------------------
```


- The variable `year8594` is coded as 1 for patients diagnosed 1985–1994 and 0 for patients diagnosed 1975–1984.
- We see that patients diagnosed in the recent period are estimated to experience 11% lower excess mortality compared to those diagnosed in the earlier period.
- There is evidence that excess mortality decreases with follow-up time, higher excess mortality in the older age groups, and no evidence of a difference between males and females.
- No evidence that the effect of distant metastases at diagnosis is confounded by sex, age at diagnosis, or period of diagnosis.

Other (better) approaches to modelling excess mortality

- Flexible parametric models; developed by Royston & Parmar [26] and extended to relative survival by Nelson *et al.* [27]. Stata command `stpm2` developed by Lambert & Royston.
- See exercises 28 (cause-specific mortality) and 29 (excess mortality) in `http://www.pauldickman.com/survival/labs.pdf` for exercises on flexible parametric models.
- Perme *et al.* [28] describe an approach to modelling excess mortality that is analogous to the Cox model; it makes no assumption about the form of the baseline hazard and the estimation approach is easy to implement in standard software.
- Cure models for relative survival.

The `strs` command for estimating and modelling relative survival using Stata

- Estimating relative survival.
  - cohort, period, or hybrid approach
  - choice of three methods for estimating expected survival (Ederer I, Ederer II, Hakulinen)
  - estimation in the presence of competing risks (Cronin and Feuer (2000) [13]).
  - estimates can be standardised (by age for example)
  - saves estimates for subsequent modelling (or presentation in tables or graphs)
- Modelling excess mortality (relative survival)
  - several alternative approaches to estimating the model
- See Dickman et al. [29] for details.

An example: localised colon carcinoma

```
. use colon if stage==1, clear
. stset surv_mm, fail(status==1 2) id(id) scale(12)
. strs using popmort, br(0 0.5 1(1)9) mergeby(_year sex _age) by(sex)
-> sex = Male
+-------------------------------------------------------------------------+
|interval    n    d    w       p  p_star       r      cp   cp_e2   cr_e2 |
|-------------------------------------------------------------------------|
|  0 .5   2620  229    0  0.9126  0.9728  0.9381  0.9126  0.9728  0.9381 |
| .5  1   2391   99    0  0.9586  0.9749  0.9833  0.8748  0.9484  0.9224 |
|  1  2   2292  229  166  0.8963  0.9483  0.9452  0.7841  0.8993  0.8719 |
|  2  3   1897  180  139  0.9015  0.9470  0.9519  0.7069  0.8517  0.8300 |
|  3  4   1578  140  119  0.9078  0.9449  0.9607  0.6417  0.8048  0.7974 |
|-------------------------------------------------------------------------|
|  4  5   1319  113  104  0.9108  0.9428  0.9660  0.5845  0.7588  0.7703 |
|  5  6   1102  102   81  0.9039  0.9414  0.9601  0.5283  0.7143  0.7396 |
|  6  7    919   71   71  0.9196  0.9409  0.9774  0.4859  0.6721  0.7229 |
|  7  8    777   59   72  0.9204  0.9391  0.9800  0.4472  0.6312  0.7084 |
|  8  9    646   49   62  0.9203  0.9380  0.9811  0.4115  0.5921  0.6950 |
+-------------------------------------------------------------------------+
```


Syntax of the `strs` command

`strs using` *filename* [*if*] [*in*] [`iweight=`*varname*], `breaks(`*numlist ascending*`)` `mergeby(`*varlist*`)` [`by(`*varlist*`)` `diagage(`*varname*`)` `diagyear(`*varname*`)` `attage(`*newvarname*`)` `attyear(`*newvarname*`)` `survprob(`*varname*`)` `maxage(`*int 99*`)` `standstrata(`*varname*`)` `brenner` `list(`*varlist*`)` `potfu(`*varname*`)` `format(`*%fmt*`)` `ederer1` `notables` `level(`*int*`)` `save[ (replace) ]` `savind(`*filename*`[ , replace ])` `savgroup(`*filename*`[ , replace ])`]

the patient data file must be `stset` using the `id()` option with time since entry in years as the timescale before using `strs`

`using` *filename* specifies a file containing general population survival probabilities sorted by the variables specified in `mergeby()`.

Life table quantities calculated by `strs`

start	Start of life table interval
end	End of life table interval
n	Number alive at start
d	Number of deaths during the interval
d_star	Expected number of deaths
ns	Number of survivors
w	Withdrawals (censorings) during the interval
n_prime	Effective number at risk
y	Person-time at risk
p	Interval-specific observed survival
se_p	Standard error of P
lo_p	Lower 95% CI for P
hi_p	Upper 95% CI for P
p_star	Interval-specific expected survival (Ederer II)
r	Interval-specific relative survival (Ederer II)
se_r	Standard error of R
lo_r	Lower 95% CI for R

hi_r	Upper 95% CI for R
cp	Cumulative observed survival
se_cp	Standard error of CP
lo_cp	Lower 95% CI for CP
hi_cp	Upper 95% CI for CP
nu	Estimated excess mortality rate, (d-d_star)/y
cp_e1	Cumulative expected survival (Ederer I)
cr_e1	Cumulative relative survival (Ederer I)
lo_cr_e1	Lower 95% CI for CR (Ederer I)
hi_cr_e1	Upper 95% CI for CR (Ederer I)
cp_e2	Cumulative expected survival (Ederer II)
cr_e2	Cumulative relative survival (Ederer II)
lo_cr_e2	Lower 95% CI for CR (Ederer II)
hi_cr_e2	Upper 95% CI for CR (Ederer II)
cp_hak	Cumulative expected survival (Hakulinen)
cr_hak	Cumulative relative survival (Hakulinen)
lo_cr_hak	Lower 95% CI for CR (Hakulinen)
hi_cr_hak	Upper 95% CI for CR (Hakulinen)

Estimates can be saved to a file

```
. use colon if stage==1, clear
. stset surv_mm, fail(status==1 2) id(id) scale(12)
. strs using popmort, br(0(1)10) mergeby(_year sex _age) by(sex agegrp) save
. use grouped, clear
. gen n0=n[_n-4]
. list sex agegrp n0 cp cr_e2 lo_cr_e2 hi_cr_e2 if end==5, sepby(sex) noobs
  +-------------------------------------------------------------+
  |    sex   agegrp     n0       cp    cr_e2   lo_cr_e2   hi_cr_e2 |
  |-------------------------------------------------------------|
  |   Male     0-44    161   0.7737   0.7881     0.7102     0.8486 |
  |   Male    45-59    462   0.7686   0.8233     0.7766     0.8636 |
  |   Male    60-74   1228   0.5945   0.7512     0.7128     0.7878 |
  |   Male      75+    769   0.4131   0.7777     0.7067     0.8479 |
  |-------------------------------------------------------------|
  | Female     0-44    136   0.7657   0.7709     0.6866     0.8358 |
  | Female    45-59    531   0.7765   0.7953     0.7536     0.8314 |
  | Female    60-74   1488   0.6993   0.7873     0.7588     0.8141 |
  | Female      75+   1499   0.4854   0.7816     0.7374     0.8249 |
  +-------------------------------------------------------------+
```


Period estimates of relative survival with strs

- The approach is to first, using `stset`, restrict the person-time at risk for each individual to the person-time in the 'period window' and then estimate the life table in the usual manner.

```
use melanoma if stage==1, clear

stset exit, enter(time mdy(1,1,1994)) exit(time mdy(12,31,1995)) ///
  origin(dx) f(status==1 2) id(id) scale(365.24)

strs using popmort, br(0(1)10) mergeby(_year sex _age) ///
  by(sex) list(n d w p r cr_e2 se_cp)
```

- The above code is available in `survival_period.do`.
- See the Stata Journal article for details of hybrid estimation.

Modelling excess mortality
First estimate relative survival and save the estimates

- Produces lifetable estimates of relative survival and stores the estimates in the data file `colon_grouped.dta`. The `notables` option supresses printing of the lifetables in the outpput window.

```
use colon if stage==1, clear

stset surv_mm, fail(status==1 2) id(id) scale(12)

strs using popmort, br(0(1)10) mergeby(_year sex _age) ///
   by(sex year8594 agegrp) notables ///
   savind(colon_individ, replace) savgroup(colon_grouped, replace)
```


Now fit the model to the first 5 years of follow-up

```
. use colon_grouped if end < 6, clear
(Collapsed (or grouped) survival data)

. xi: glm d i.end i.sex i.year8594 i.agegrp , fam(pois) ///
      link(rs d_star) lnoffset(y) eform

Generalized linear models                 No. of obs      =        80
Optimization     : ML                     Residual df     =        70
                                          Scale parameter =         1
Deviance         =  131.4342128           (1/df) Deviance =  1.877632
Pearson          =  130.1530694           (1/df) Pearson  =   1.85933

Variance function: V(u) = u               [Poisson]
Link function    : g(u) = log(u-d*)       [Relative survival]

                                          AIC             =   6.39959
Log likelihood   = -245.9836017           BIC             = -175.3077
```



```
------------------------------------------------------------------------------
             |                 OIM
           d |      exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iend_2 |    .7984084   .0730515    -2.46   0.014     .6673339      .955228
     _Iend_3 |    .6230213   .0671961    -4.39   0.000     .5043086     .7696785
     _Iend_4 |    .4969433   .0645561    -5.38   0.000     .3852391     .6410374
     _Iend_5 |    .4334347    .065147    -5.56   0.000      .322838     .5819191
     _Isex_2 |    .9564493   .0729823    -0.58   0.560     .8235891     1.110742
_Iyear8594_1 |    .7308044   .0539291    -4.25   0.000     .6323935     .8445296
  _Iagegrp_1 |    .8642841   .1353083    -0.93   0.352      .635911     1.174672
  _Iagegrp_2 |    1.071568   .1534869     0.48   0.629     .8092774     1.418869
  _Iagegrp_3 |    1.436319   .2146593     2.42   0.015     1.071613     1.925147
------------------------------------------------------------------------------
```


Would you like to learn more?

- We are holding a one-week course on 'Statistical methods for population-based cancer survival analysis' in Stockholm next July.
- See `http://www.cansurv.net/`

References

[1] Welch HG, Black WC. Are deaths within 1 month of cancer-directed surgery attributed to cancer? *J Natl Cancer Inst* 2002;**94**:1066–70.

[2] Berkson J, Gage RP. Calculation of survival rates for cancer. *Proceedings of Staff Meetings of the Mayo Clinic* 1950;**25**:270–286.

[3] Ederer F, Axtell LM, Cutler SJ. The relative survival rate: A statistical methodology. *National Cancer Institute Monograph* 1961;**6**:101–121.

[4] Percy CL, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American Journal of Public Health* 1981;**71**:242–250.

[5] Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996;**78**:2004–2010.

[6] Brenner H, Gefeller O, Hakulinen T. Period analysis for 'up-to-date' cancer survival data: theory, empirical evaluation, computational realisation and applications. *European Journal of Cancer* 2004;**40**:326–35.

[7] Pokhrel A, Hakulinen T. Age-standardisation of relative survival ratios of cancer patients in a comparison between countries, genders and time periods. *Eur J Cancer* 2009;**45**:642–647.

[8] Pokhrel A, Hakulinen T. How to interpret the relative survival ratios of cancer patients. *European Journal of Cancer* 2008;**00**:000–000.

[9] Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age adjustment of cancer survival rates. *Eur J Cancer* 2004;**40**:2317–2322.

[10] Brenner H, Hakulinen T. Age adjustment of cancer survival rates: methods, point estimates and standard errors. *Br J Cancer* 2005;**93**:372–375.

[11] Brenner H, Hakulinen T. On crude and age-adjusted relative survival rates. *J Clin Epidemiol* 2003;**56**:1185–91.

[12] Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006;**260**:103–117.

[13] Cronin K, Feuer E. Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Stat Med* 2000;**19**:1729–40.

[14] Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med* 2010; **29**:885–895.

[15] Lambert PC, Dickman PW, Osterlund P, Andersson T, Sankila R, Glimelius B. Temporal trends in the proportion cured for cancer of the colon and rectum: A population-based study using data from the finnish cancer registry. *Int J Cancer* 2007;**121**:2052–2059.

[16] Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007;**8**:576–594.

[17] Andersson TML, Lambert PC, Derolf AR, Kristinsson SY, Eloranta S, Landgren O, *et al.*. Temporal trends in the proportion cured among adults diagnosed with acute myeloid leukaemia in sweden 1973-2001, a population-based study. *Br J Haematol* 2009;.

[18] Nelson CP, Lambert PC, Squire IB, Jones DR. Relative survival: what can cardiovascular disease learn from cancer? *Eur Heart J* 2008;**29**:941–947.

[19] Bhaskaran K, Hamouda O, Sannes M, Boufassa F, Johnson AM, Lambert PC, *et al.*. Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *JAMA* 2008;**300**:51–59.

[20] Harrison KM, Ling Q, Song R, Hall HI. County-level socioeconomic status and survival after HIV diagnosis, United States. *Ann Epidemiol* 2008;**18**:919–927.

[21] Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Stat Med* 2004;**23**:51–64.

[22] Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: Elements for further discussion. *Statistics in Medicine* 1990;**9**:529–538.

[23] Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Applied Statistics* 1987;**36**:309–317.

[24] Berry G. The analysis of mortality by the subject-years method. *Biometrics* 1983; **39**:173–184.

[25] Pocock S, Gore S, Kerr G. Long term survival analysis: the curability of breast cancer. *Stat Med* 1982;**1**:93–104.

[26] Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;**21**:2175–2197.

[27] Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. *Stat Med* 2007;**26**:5486–5498.

[28] Perme MP, Henderson R, Stare J. An approach to estimation in relative survival regression. *Biostatistics* 2009;**10**:136–146.

[29] Dickman PW, Coviello E, Hills M. Estimating and modelling relative survival. *The Stata Journal* 2010;(in press).