# Relative survival – an introduction and recent developments

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

11 December 2008



## Outline

- 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.
- Is relative survival a useful measure for conditions other than cancer?
- Modelling relative survival.
  - Poisson regression.
  - Flexible parametric models.
- Cure models.
- Estimating survival in the presence of competing risks.



## Introduction to relative survival

- Interest is typically in net mortality (mortality associated with a diagnosis of cancer) [1].
- Cause-specific mortality is often used to estimate net mortality — only those deaths which can be attributed to the cancer in question are considered to be events.

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

## 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.* [2] 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.
- 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 [3].
- Although these tables include the effect of deaths due to the cancer being studied, Ederer *et al.* [4] 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 [5] 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.



## 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.

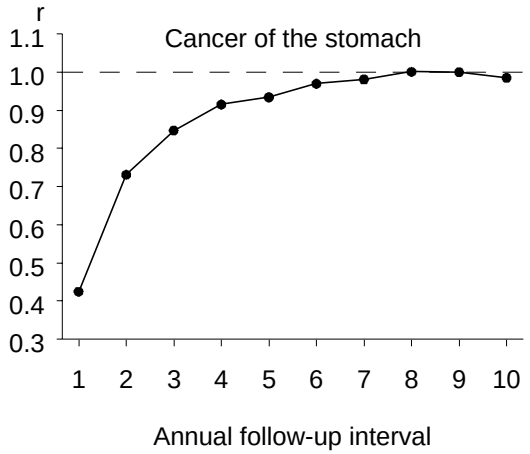



Figure 1: Plots of the annual (interval-specific) relative survival ratios (r) for males and females diagnosed with cancer of the stomach in Finland 1985–1994 and followed up to the end of 1995.



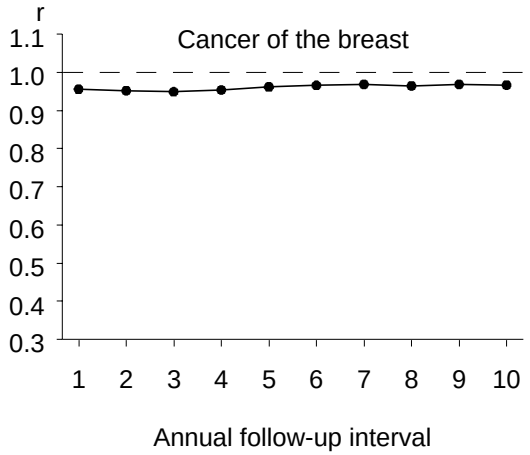



Figure 2: Plots of the annual (interval-specific) relative survival ratios (r) for females diagnosed with cancer of the breast in Finland 1985–1994 and followed up to the end of 1995.



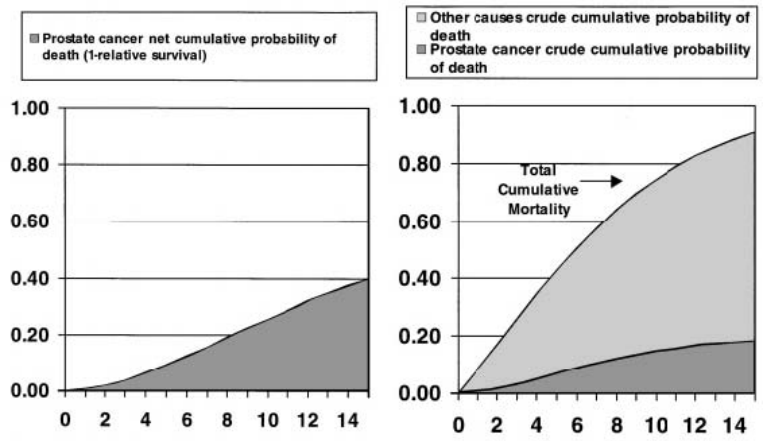
- Plots of the interval-specific RSR are also useful for assessing the quality of follow-up.
- If the interval-specific RSR levels out at a value greater than 1, this generally indicates that some deaths have been missed in the follow-up process.
- An interval-specific relative survival ratio of unity 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.

## Interpreting relative survival estimates

- 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. Estimates of the proportion of patients who will die of cancer in the presence of competing risks can also be made.
- Cronin and Feuer (2000) [6] extended the theory of competing risks to relative survival; their method is implemented in our Stata command `strs`.

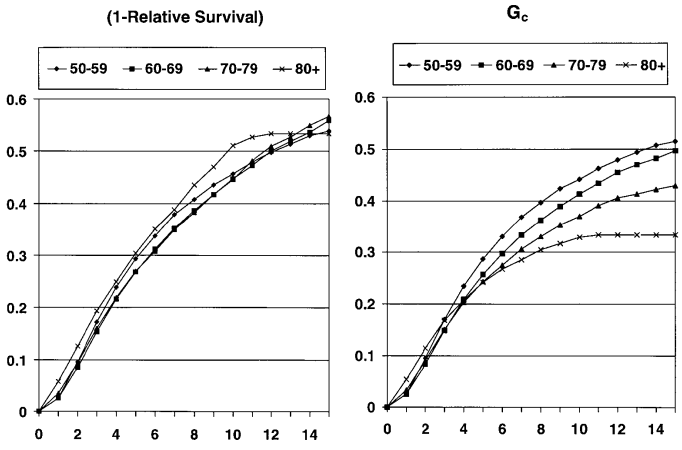


Cumulative probability of death in men with localized prostate cancer over the age of 70.





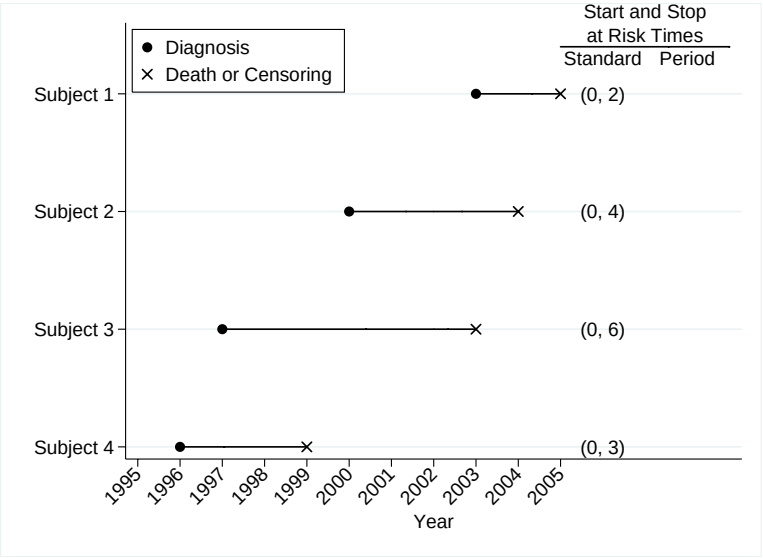

Cause-specific probability of death in women with regional breast cancer



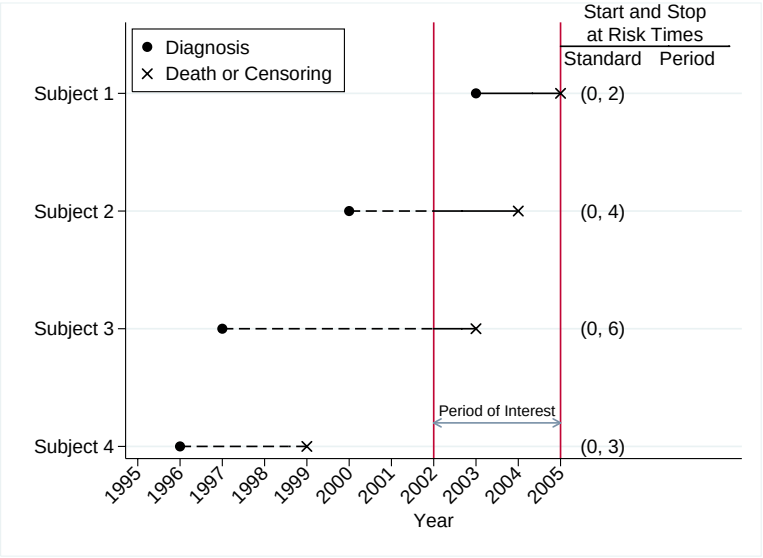
## Estimating relative survival using a period approach

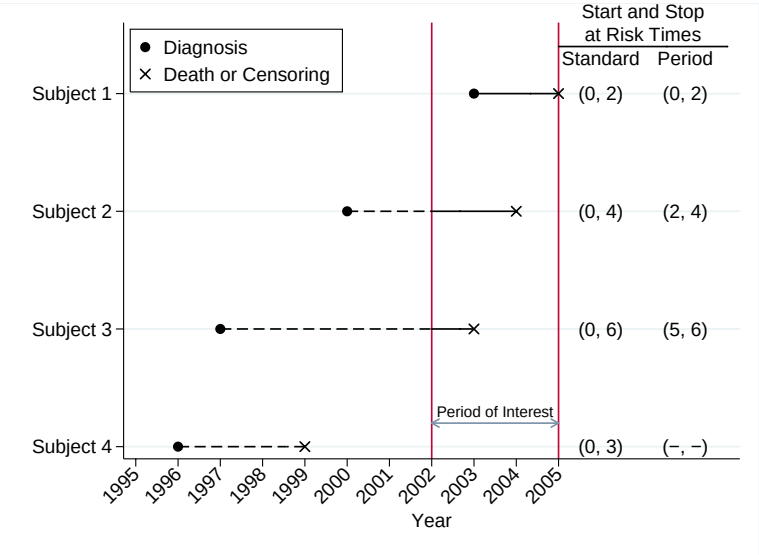
- In 1996 Hermann Brenner suggested estimating cancer patient survival using a period, rather than cohort, approach [7].
- Time at risk is left truncated at the start of the period window and right censored at the end.
- This suggestion was initially met with scepticism although studies based on historical data [8] 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.












## 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 |



## Age-standardisation of relative survival

- This is used primarily when interest is in obtaining estimates of relative survival for descriptive purposes and is of less interest when focus is on modelling.
- 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?
- See the papers by Brenner et al. [9, 10, 11]

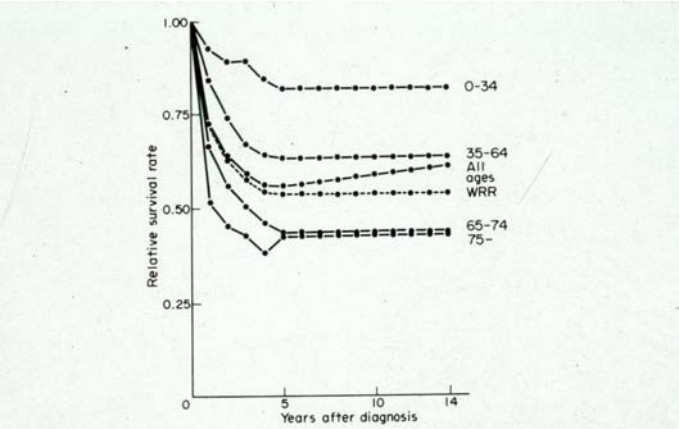



FIG. 1. A hypothetical example. The relative survival curves by age at diagnosis for female patients with localized cancer of the colon or small intestine in Finland 1953–1970, on the assumption that the 5-yr relative survival rate in each age group is the cure rate for that age group. (WRR: weighted average of the age-specific relative survival curves, with the age distribution of patients at diagnosis being used as constant weights.)



## 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[12] and HIV/AIDS[13].

  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.



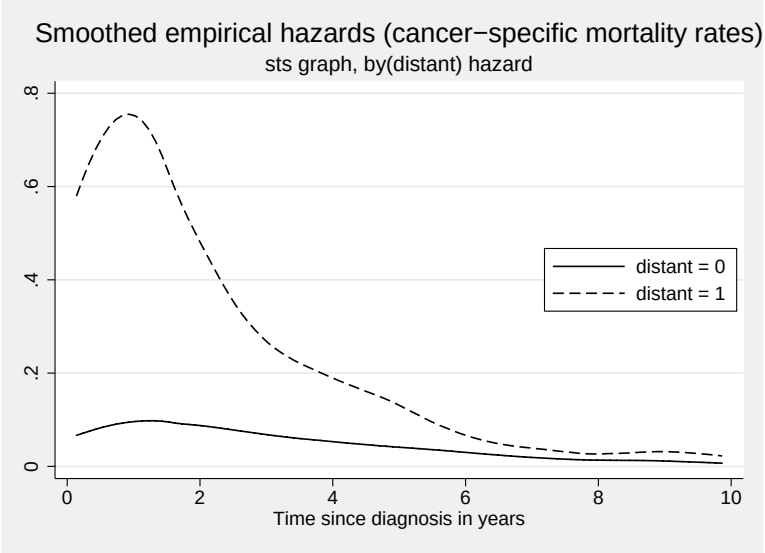
## Overview of approaches to modelling prognosis of cancer patients

- Cox proportional hazards model for cause-specific mortality
- Poisson regression model for cause-specific mortality
- Similarity of Poisson regression and Cox regression
- Poisson regression for excess mortality
- Alternative approaches to modelling excess mortality
  - Fine splitting and modelling the baseline hazard using splines or fractional polynomials
  - Flexible parametric models
- Cure models for relative survival

## Example: survival of patients diagnosed with colon carcinoma in Finland

- Patients diagnosed with colon carcinoma in Finland 1984–95. Potential follow-up to end of 1995; censored after 10 years.
- Outcome is death due to colon carcinoma (i.e., cause-specific mortality).
- Interest is in the effect of clinical stage at diagnosis (distant metastases vs no distant metastases).
- How might we specify a statistical model for these data?



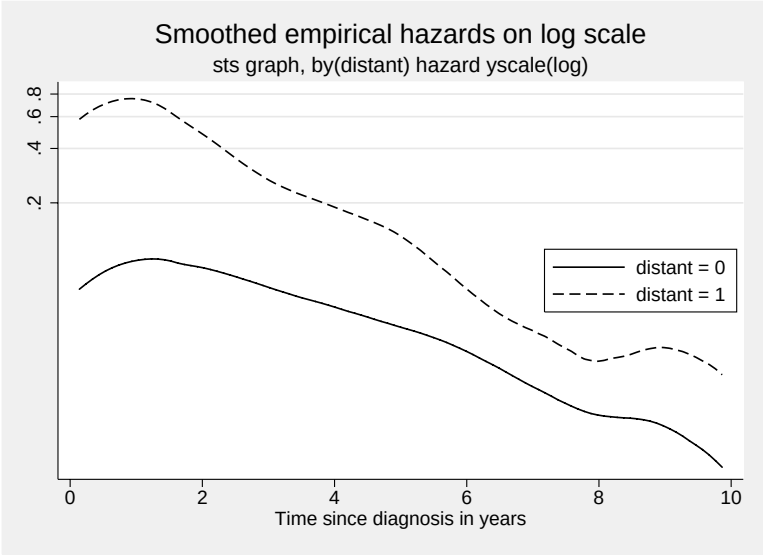





## The Cox proportional hazards model

- The 'intercept' in the Cox model [14], the hazard (event rate) for individuals with all covariates $z$ at the reference level, can be thought of as an arbitrary function of time[1], often called the baseline hazard and denoted by $\lambda_0(t)$.
- The hazard at time $t$ for individual with other covariate values is a multiple of the baseline
$$\lambda(t|x) = \lambda_0(t)\exp(\beta' x).$$
- Alternatively
$$\ln[\lambda(t|x)] = \ln[\lambda_0(t)] + \beta' x.$$

[1] time $t$ can be defined in many ways, e.g., attained age, time-on-study, calendar time, etc.



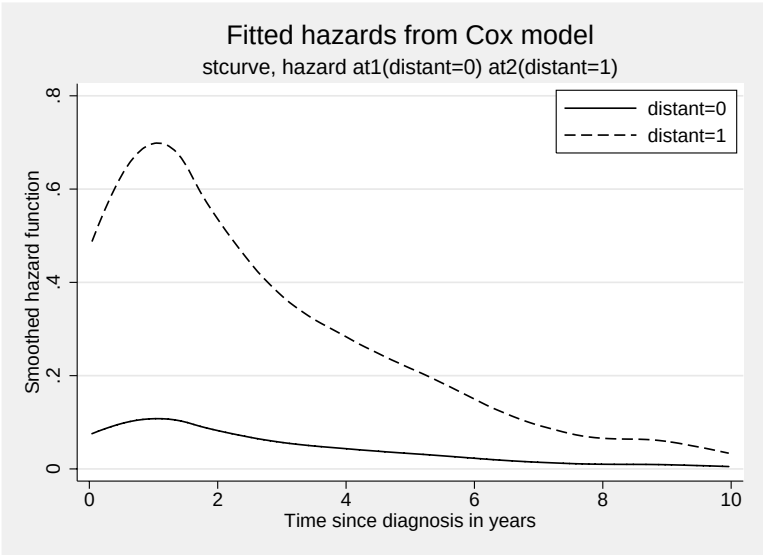


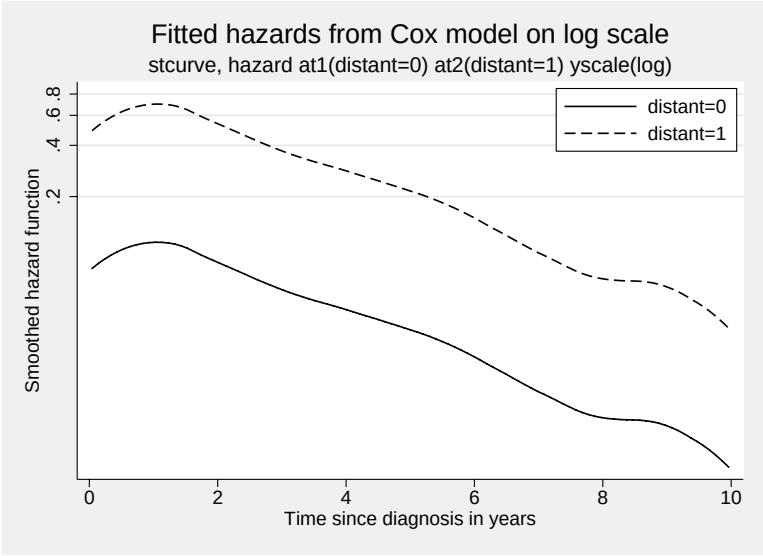

## Cox model to estimate the cause-specific mortality rate ratio

```
. stcox distant
        failure _d:  status == 1
  analysis time _t:  (exit-origin)/365.25
            origin:  time dx
              note:  trim>10 trimmed

No. of subjects =        13208          Number of obs  =   50666
No. of failures =         7122
Time at risk    =  44013.26215
                                        LR chi2(1)     = 5544.65
Log likelihood  =   -61651.446          Prob > chi2    =  0.0000
------------------------------------------------------------------------------
     _t | Haz. Ratio   Std. Err.     z    P>|z|   [95% Conf. Interval]
--------+---------------------------------------------------------------------
distant |   6.557777   .1689328   73.00   0.000   6.234895    6.897381
------------------------------------------------------------------------------
```










## Regression models commonly applied in epidemiology

- Linear regression

$$\mu = \beta_0 + \beta_1 X$$

- Logistic regression

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X$$

- Poisson regression

$$\ln(\lambda) = \beta_0 + \beta_1 X$$

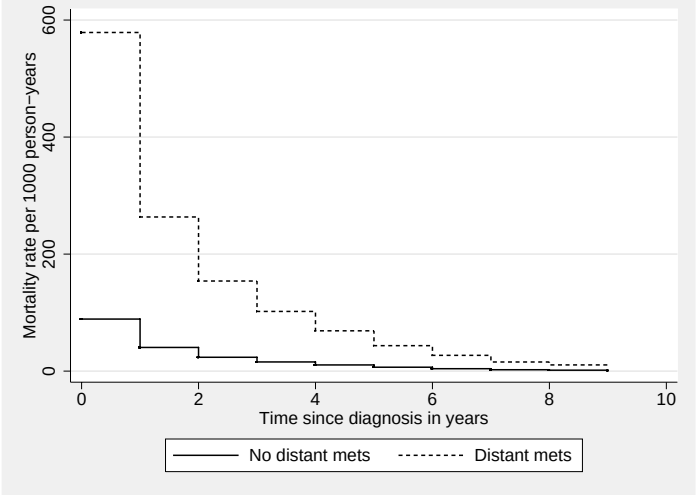
- In each case $\beta_1$ is the effect per unit of $X$, measured as a change in the mean (linear regression); the change in the log odds (logistic regression); the change in the log rate (Poisson regression).

- Cox model

$$\ln[\lambda(t)] = \ln[\lambda_0(t)] + \beta_1 X$$



## Fitted values: Poisson regression, cause-specific mortality



## Estimated effect of distant metastases while controlling for time since diagnosis

```
. xi: glm dead i.fu distant, family(poisson) lnoff(risktime) eform
------------------------------------------------------------------------------
             |                 OIM
        dead |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Ifu_1 |   .4551995   .0100359   -35.70   0.000     .4359484    .4753006
      _Ifu_2 |   .2660856   .0076698   -45.93   0.000     .2514698    .2815508
      _Ifu_3 |   .1763302   .0063845   -47.93   0.000     .1642505    .1892983
      _Ifu_4 |   .1190439   .0054356   -46.61   0.000      .108853    .1301888
      _Ifu_5 |   .0751727   .0045158   -43.08   0.000     .0668231    .0845656
      _Ifu_6 |   .0466152   .0037402   -38.21   0.000     .0398319    .0545538
      _Ifu_7 |   .0269519   .0030126   -32.33   0.000     .0216493    .0335532
      _Ifu_8 |   .0183221    .002654   -27.61   0.000     .0137936    .0243373
      _Ifu_9 |   .0116515   .0022895   -22.66   0.000     .0079273    .0171253
     distant |   6.504972    .112855   107.93   0.000     6.287499    6.729967
    risktime | (exposure)
------------------------------------------------------------------------------
```



## Recall the estimate from Cox model

```
. stcox distant

No. of subjects =        13208          Number of obs  =   50666
No. of failures =         7122
Time at risk    =  44013.26215
                                        LR chi2(1)     = 5544.65
Log likelihood  =   -61651.446          Prob > chi2    =  0.0000
------------------------------------------------------------------------------
         _t | Haz. Ratio   Std. Err.     z    P>|z|   [95% Conf. Interval]
------------+-----------------------------------------------------------------
    distant |   6.557777   .1689328   73.00   0.000   6.234895    6.897381
------------------------------------------------------------------------------
```



## 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, $\nu(t)$ [15, 16, 17, 18, 19].

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

- It is common to assume that the excess hazards are piecewise constant and proportional. Provides estimates of relative excess risk.
- The model can be estimated in the framework of generalised linear models using standard statistical software (e.g., SAS, Stata, R) [15].
- 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 can be written as

$$\ln(\mu_j - d_j^*) = \ln(y_j) + \mathbf{x}\beta, \tag{1}$$

  where $\mu_j = E(d_j)$, $d_j^*$ the expected number of deaths, and $y_j$ person-time.

- This implies a generalised linear model with outcome $d_j$, Poisson error structure, link $\ln(\mu_j - d_j^*)$, and offset $\ln(y_j)$.

- Such models have previously been described by Breslow and Day (1987) [20, pp. 173–176] and Berry (1983) [18].

- The usual regression diagnostics (residuals, influence statistics) and method for assessing model fit for generalised linear models can be utilised.



## 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)
```

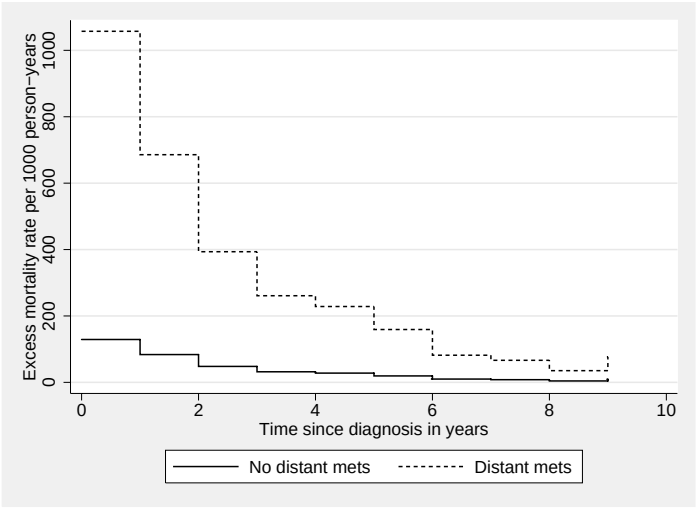
- We then fit the Poisson regression model using the resulting output file (which contains the observed (`d`) and expected (`d_star`) numbers of deaths for each life table interval along with person-time at risk (`y`)).



```
. 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
------------------------------------------------------------------------------
```



## References


[1] Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *Journal of Internal Medicine* 2006;**260**:103–17.

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

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

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

[5] 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.

[6] 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* Jul 2000;**19**:1729–40.

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

[8] 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.

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

[13] 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* Jul 2008;**300**:51–59.

[14] Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B* 1972;**34**:187–220.

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

[16] 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.

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



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

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

[20] Breslow NE, Day NE. *Statistical Methods in Cancer Research: Volume II - The Design and Analysis of Cohort Studies*. IARC Scientific Publications No. 82. Lyon: IARC, 1987.