# Transformation survival models

Yulia Marchenko

Director of Biostatistics
StataCorp LP

2014 UK Stata Users Group meeting

1 Introduction

2 The igencox command

3 Examples: Single-event data

4 Examples: Single-event clustered data

5 Examples: Multiple-events data

6 Download

7 References

# Acknowledgements

This work was supported by the NIH Phase I SBIR (HHSN261201000090C) and the NIH Phase II SBIR (HHSN261201200096C) contracts titled "Software for Modern Extensions of the Cox model" to StataCorp LP.

Danyu Lin, Department of Biostatistics, University of North Carolina at Chapel Hill

Donglin Zeng, Department of Biostatistics, University of North Carolina at Chapel Hill

Rafal Raciborski, StataCorp

Xiao Yang, StataCorp

- The Cox proportional hazards (PH) regression model (Cox 1972) is the most common method for the analysis of survival data.
- Appeal: estimates of regression coefficients may be obtained without a parametric assumption about the baseline hazard function.
- Restriction: proportional hazards—covariates have multiplicative effects on the hazard function.
- Interpretation: a coefficient is a log-hazard ratio for a one-unit change in a covariate holding other covariates constant.

- The proportional-hazards assumption is often violated in applications. For example, sometimes it is reasonable to assume that the baseline and subject-specific hazard functions become more similar with time.
- The proportional odds (PO) regression model (Bennett 1983) may be used in this case because it assumes that the odds of survival are proportional and as a result that the ratio of hazards approaches unity with time.
- Appeal: estimates of regression coefficients may be obtained without a parametric assumption about the baseline hazard function.
- Restriction: proportional odds—covariates have multiplicative effect on the odds of survival beyond time $t$. The ratio of the hazards converges to unity as time increases.
- Interpretation: a coefficient is a log-odds ratio for a one-unit change in a covariate holding other covariates constant.

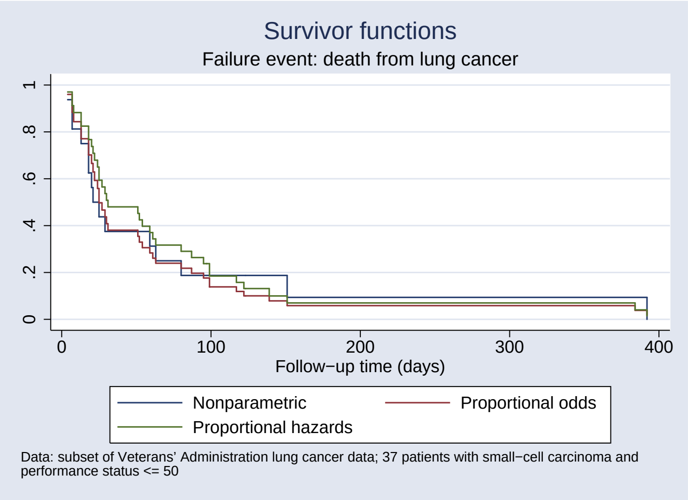


Figure: Survivor curves for lung-cancer patients with small-cell carcinoma who required frequent medical care: nonparametric, proportional-hazards, and proportional-odds estimators.

- Proportional hazards and proportional odds models are special cases of transformation survival models—a class of semiparametric linear transformation models, which relates an unknown transformation of the survival time linearly to covariates (Kalbfleisch and Prentice 2002, 241).
- Appeal: estimates of regression coefficients may still be obtained without a parametric assumption about the baseline hazard function as for the Cox model, but the assumption of proportional hazards is relaxed.
- Restriction: the shape of the subject-specific hazard is determined by the chosen transformation.
- Interpretation: coefficients no longer have intuitive interpretations except for the two special cases of PH and PO.

- Semiparametric linear transformation survival model:

$$H(T) = -\boldsymbol{\beta}^\top \mathbf{Z} + \epsilon, \tag{1}$$

  where $T$ is the failure time, $\mathbf{Z}$ is a vector of fixed (time-invariant) covariates, $\boldsymbol{\beta}$ is a vector of unknown coefficients, $\epsilon$ is an error term with a completely known distribution, and $H(\cdot)$ is an completely unspecified increasing function.
- When $\epsilon$ has an extreme-value distribution, (1) reduces to the proportional hazards model.
- When $\epsilon$ has a standard logistic distribution, (1) reduces to the proportional odds model.
- Various transformation models may be generated by specifying different distributions for $\epsilon$.

- There is a formulation equivalent to (1) based on the cumulative hazard function:

$$\Lambda(t|\mathbf{Z}) = G(\exp(\beta^\top \mathbf{Z})\Lambda_0(t)), \tag{2}$$

  where $\Lambda(t|\mathbf{Z})$ and $\Lambda_0(t)$ are the subject-specific and baseline cumulative hazard functions and $G(\cdot)$ is a completely specified continuous increasing function.

- Formula (2) can be extended to allow for time-varying covariates $\mathbf{Z}(t)$ and to multiple and recurrent events by considering a formulation similar to (2) for a cumulative intensity function of a counting process $N(t)$ recording the number of events occurred by time $t$.

- Zeng and Lin (2007) considered two functions for $G(\cdot)$ in (2).
- Box-Cox transformation, $BoxCox(\rho)$:

$$G(x) = \frac{(1+x)^{\rho} - 1}{\rho}, \rho \geq 0 \tag{3}$$

- Logarithmic transformations, $Log(r)$:

$$G(x) = \frac{\ln(1+rx)}{r}, r \geq 0 \tag{4}$$

- $G(x) = x$, with $\rho = 1$ in (3) or $r = 0$ in (4), leads to the proportional hazards model.
- $G(x) = \ln(1+x)$, with $\rho = 0$ in (3) or $r = 1$ in (4), leads to the proportional odds model.

- Nonparametric maximum likelihood: an unknown failure distribution is treated as an infinite-dimensional parameter.
- The likelihood is maximized over $\boldsymbol{\beta}$ and cumulative-hazard jump sizes $\Lambda_k$ at each observed failure time $t_k$ for $k = 1, 2, \ldots, K$
- Zeng and Lin (2007) showed that nonparametric maximum likelihood estimators (NPMLEs) of $\boldsymbol{\beta}$ and $\Lambda_k$s are consistent, asymptotically normal, and asymptotically efficient.
- The maximization is computationally intensive, especially for a large number of failures.
- A number of algorithms exist for the computation of NPMLEs of $\boldsymbol{\beta}$ and $\Lambda_k$s.
- One of them is the expectation-maximization (EM) algorithm.

The `igencox` command fits transformation survival models for failure-time data. Transformation survival models generalize the Cox model to allow for nonproportional hazards. The supported transformations are the class of Box-Cox transformations, $BoxCox(\rho)$, and the class of logarithmic transformations, $Log(r)$. Special cases include the Cox proportional hazards model with $BoxCox(1)$ or $Log(0)$, and the proportional odds model with $BoxCox(0)$ or $Log(1)$.

- Box-Cox and logarithmic transformations
- Proportional hazards models with $BoxCox(1)$ or $Log(0)$
- Proportional odds models with $BoxCox(0)$ or $Log(1)$
- Time-varying covariates
- Single-event clustered data with a Gaussian random effect or Gaussian frailties
- Multiple dependent events
- Recurrent data

`igenset` *timevar* ... [ , f̲ailure(*failvar*) ... ]

`igencox` [ *varlist* ] [ *if* ] [ *in* ]
[ , transform(boxcox | logarithmic [ # ]) ... ]

## Quick examples:

### Load and declare survival data

```
. use mydata
. igenset time, failure(died)
```

### Fit proportional hazards model:

```
. igencox age
. igencox age, transform(boxcox 1)
. igencox age, transform(logarithmic 0)
```

### Fit proportional odds model:

```
. igencox age, transform(boxcox 0)
. igencox age, transform(log 1)
```

### Fit a Box-Cox model with $\rho = 1.5$

```
. igencox age, transform(boxcox 1.5)
```

### Fit a logarithmic model with $r = 0.5$

```
. igencox age, transform(log 0.5)
```

```
igenset timevar ..., id(id) [failure(failvar) ...]

igencox [varlist] [if] [in], cluster(varname)
        [transform(boxcox|logarithmic [#]) ...]
```

## Quick examples:

Load and declare survival data

```
. use mydata
. igenset time, failure(died) id(id)
```

Fit proportional hazards model with a Gaussian frailty for `subject`:

```
. igencox age, cluster(subject)
. igencox age, cluster(subject) transform(boxcox 1)
. igencox age, cluster(subject) transform(logarithmic 0)
```

Fit proportional odds model with a Gaussian frailty for `subject`:

```
. igencox age, cluster(subject) transform(boxcox 0)
. igencox age, cluster(subject) transform(log 1)
```

Fit a random-effects Box-Cox model with $\rho = 1.5$

```
. igencox age, cluster(subject) transform(boxcox 1.5)
```

Fit a random-effects logarithmic model with $r = 0.5$

```
. igencox age, cluster(subject) transform(log 0.5)
```

```
igenset timevar ..., id(id) failtype(failtypevar) [failure(failvar)
      ...]
igencox [varlist] [if] [in],
        [transform#(boxcox|logarithmic [#])]
        [failcov#(varlist[, add]) ...]
```

**Quick examples:**

Consider a study investigating the effect of a treatment on a disease relapse (`type==1` event) and death (`type==2` event) of patients.

Load and declare survival data

```
. use mydata
. igenset time, failure(died) id(id) failtype(type)
```

Fit proportional hazards model for both types of events:

```
. igencox treat
. igencox treat, transform(boxcox 1)
. igencox treat, transform(logarithmic 0)
```

Fit a proportional hazards model for relapses and proportional odds model for deaths:

```
. igencox treat, transform1(boxcox 1) transform2(log 1)
```

Above, include an additional covariate `x` to model relapses:

```
. igencox treat, transform1(boxcox 1) transform2(log 1) failcov1(x, add)
```

Use `predict` after `igencox` to compute

- baseline and covariate-adjusted survivor functions:

```
. igencox ..., baseq(varname) ...
. predict varname , basesurv
. predict varname , survival [at(varname=# [...])]
```

- baseline and covariate-adjusted cumulative hazard functions:

```
. igencox ..., baseq(varname) ...
. predict varname , basechazard
. predict varname , cumhaz [at(varname=# [...])]
```

- standard errors of the survivor or cumulative hazard functions:

```
. igencox ..., baseq(varname) savesigma(filename) ...
. predict varname , basesurv|basechazard se(newvarname)
. predict varname , survival|cumhaz se(newvarname) [at(varname=# [...])]
```

- unconditional or conditional (on random-effects) survivor or cumulative hazard functions for single-event clustered data:

```
. igencox ..., cluster(varname) baseq(varname) ...
. predict varname , ...
. predict varname , conditional [ebayes(varname | #)] ...
```

- failure-specific survivor or cumulative hazard functions for multiple-failure data:

```
. igencox ..., baseq(varname) ...
. predict varname , failtype(failtypevar=#) ...
```

- survivor or cumulative hazard functions conditional on another failure occurring at a specific time for multiple-failure data:

```
. igencox ..., baseq(varname) ...
. predict varname , failtype(failtypevar=#) condition(failtypevar=# time=#) ...
```

## Example

- Veterans' Administration lung cancer trial (Prentice 1973)
- 97 patients without prior therapy
- Covariates: performance status and tumor type

```
. use lungcancer

. describe

Contains data from lungcancer.dta
  obs:            97
 vars:             5                        23 Jul 2014 11:24
 size:           582
------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
id              byte    %9.0g                 Observation identifier
time            int     %9.0g                 Survival time (days)
died            byte    %9.0g                 Death indicator
performance     byte    %9.0g                 Performance status measured
                                                between 0 and 100 (%)
tumor           byte    %10.0g     tumorlab   Lung-cancer tumors: large-cell,
                                                adeno, small-cell, and
                                                squamous-cell carcinomas
------------------------------------------------------------------------------
Sorted by:
```

- Zeng and Lin (2006) fit logarithmic $Log(r)$ and $BoxCox(\rho)$ transformation models for different values of $r$ and $\rho$ with `performance` and `tumor` as covariates.
- The likelihood is maximized for the logarithmic model at $r = 0.83$, and is very close to that for $r = 1$.
- I reproduce their results for $r = 0$ (proportional hazards model) and $r = 1$ (proportional odds model) using `igencox`.
- I also fit the Cox PH model using `stcox` for comparison.

- Declare survival data using `stset`:

```
. stset time, failure(died)

     failure event:  died != 0 & died < .
obs. time interval:  (0, time]
 exit on or before:  failure

------------------------------------------------------------------------------
       97  total observations
        0  exclusions
------------------------------------------------------------------------------
       97  observations remaining, representing
       91  failures in single-record/single-failure data
    10879  total analysis time at risk and under observation
                                              at risk from t =         0
                                   earliest observed entry t =         0
                                        last observed exit t =       587
```

- Fit a Cox model using `stcox`:

```
. stcox performance i.tumor, nolog

         failure _d:  died
   analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           97                 Number of obs    =        97
No. of failures =           91
Time at risk    =        10879
                                               LR chi2(4)       =     33.65
Log likelihood  =   -312.35474                 Prob > chi2      =    0.0000

-------------------------------------------------------------------------------
           _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
  performance |   .9759256   .0057701    -4.12   0.000     .9646817    .9873006
              |
        tumor |
       Adeno  |   2.342992   .8149609     2.45   0.014     1.184945    4.632799
  Small-cell  |   1.729192   .5550104     1.71   0.088     .9218008    3.243764
       Squam  |    .807025   .2803153    -0.62   0.537      .408533    1.594215
-------------------------------------------------------------------------------

. estimates store stcox
```

- Declare survival data using `igenset`:

```
. igenset time, failure(died)

     failure event:  died != 0 & died < .
obs. time interval:  (0, time]
 exit on or before:  failure

------------------------------------------------------------------------------
       97  total observations
        0  exclusions
------------------------------------------------------------------------------
       97  observations remaining, representing
       91  failures in single-record/single-failure data
    10879  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =       587
```

- Fit proportional odds and proportional hazards models using `igencox`:

```
. igencox performance i.tumor, transform(logarithmic 1) nolog
         failure _d:  died
   analysis time _t:  time

Generalized Cox regression -- Breslow method for ties

Transformation: Logarithmic(1)
No. of subjects =           97                  No. of obs    =         97
No. of failures =           91
Time at risk    =        10879
                                                Wald chi2(4)  =      41.35
Log likelihood  =  -399.164425                  Prob > chi2   =     0.0000
```

| _t | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| performance | -.0531533 | .0101397 | -5.24 | 0.000 | -.0730267 | -.0332799 |
| tumor | | | | | | |
| 1 | 1.313782 | .5542236 | 2.37 | 0.018 | .2275237 | 2.40004 |
| 2 | 1.382667 | .5237982 | 2.64 | 0.008 | .3560419 | 2.409293 |
| 3 | -.1813802 | .5876395 | -0.31 | 0.758 | -1.333132 | .970372 |

```
. estimates store igencoxPO
. qui igencox performance i.tumor, transform(logarithmic 0)
. estimates store igencoxPH
```

- Compare coefficients:

```
. estimates table stcox igencoxPH igencoxPO, b(%9.3f) se(%9.3f) p(%9.3f)
```

| Variable | stcox | igencoxPH | igencoxPO |
|---|---|---|---|
| performance | -0.024 | -0.024 | -0.053 |
| | 0.006 | 0.006 | 0.010 |
| | 0.000 | 0.000 | 0.000 |
| tumor | | | |
| 1 | 0.851 | 0.851 | 1.314 |
| | 0.348 | 0.348 | 0.554 |
| | 0.014 | 0.014 | 0.018 |
| 2 | 0.548 | 0.548 | 1.383 |
| | 0.321 | 0.321 | 0.524 |
| | 0.088 | 0.088 | 0.008 |
| 3 | -0.214 | -0.214 | -0.181 |
| | 0.347 | 0.347 | 0.588 |
| | 0.537 | 0.537 | 0.758 |

legend: b/se/p

- Compare models:

```
. estimates stats igencoxPH igencoxPO
Akaike´s information criterion and Bayesian information criterion
```

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| igencoxPH | 97 | . | -403.3547 | 4 | 814.7095 | 825.0083 |
| igencoxPO | 97 | . | -399.1644 | 4 | 806.3288 | 816.6277 |

```
Note:  N=Obs used in calculating BIC; see [R] BIC note
```

- Compute covariate-adjusted survivor functions and their standard errors using `predict`.
- Option `baseq()` must be specified with `igencox` to compute a survivor function (or cumulative hazard).
- Option `savesigma()` must be specified with `igencox` to compute standard errors.

```
. qui igencox performance i.tumor, transform(log 1) baseq(q) savesigma(sigma)
. predict double surv80, survival at(performance=80 tumor=0) se(se80)
. predict double surv40, survival at(performance=40 tumor=2) se(se40)
```

- Compute 95% pointwise confidence intervals:

```
. qui gen double tmp = 1.96*se80/(surv80*log(surv80))
. qui gen cil80 = surv80^exp(-tmp)
. qui gen ciu80 = surv80^exp(tmp)
. qui replace tmp = 1.96*se40/(surv40*log(surv40))
. qui gen cil40 = surv40^exp(-tmp)
. qui gen ciu40 = surv40^exp(tmp)
```

- Plot covariate-adjusted survivor functions and their confidence intervals:

```
. sort _t
. twoway (line surv80 surv40 _t, connect(J J))  ///
>        (rline cil80 ciu80 _t,  connect(J))    ///
>        (rline cil40 ciu40 _t,  connect(J)),   ///
>        ytitle(Survival probabilities) xtitle(Follow-up time (days)) ///
>        scheme(s2mono) legend(off)
```

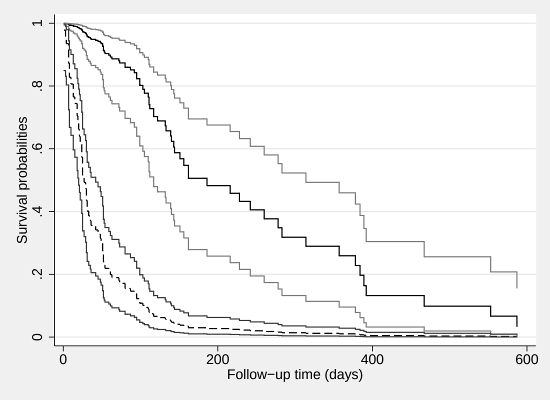


Figure: Estimated survivor curves for the lung-cancer patients: the upper three curves correspond to the point estimates and 95% confidence limits for a subject with a large-cell carcinoma and performance status of 80%, and the lower curves to those of a patient with a small-cell carcinoma and performance status of 40%.

### Example

- Diabetic Retinopathy Study (Huster et al. 1989) evaluating the ability of laser photocoagulation to delay visual loss among patients with diabetic retinopathy
- 197 high-risk patients, 394 observations on times to visual loss (visual acuity $< 5/200$)
- Covariates: laser treatment and onset of diabetes
- Treatment is randomly applied to one of the eyes

```
. use drs
(Visual-Loss Data from Diabetic Retinopathy Study)

. describe

Contains data from drs.dta
  obs:           394                          Visual-Loss Data from Diabetic
                                                Retinopathy Study
 vars:             6                          21 Jul 2014 14:16
 size:         5,122                          (_dta has notes)
-------------------------------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
-------------------------------------------------------------------------------
id              int    %9.0g                 Subject identifier
obsid           float  %9.0g                 Observation identifier
time            float  %9.0g                 Time to visual loss measured as
                                               visual activity less than 5/200
failure         byte   %9.0g                 Failure indicator
treat           byte   %9.0g                 Laser treatment (0: not received,
                                               1: received)
onset           byte   %9.0g                 Onset of diabetics (0: juvenile
                                               onset, 1: adult onset)
-------------------------------------------------------------------------------
Sorted by:
```

- Zeng, Lin, and Lin (2008) fit a logarithmic $Log(r)$ transformation model for different values of $r = 0, 0.1, \ldots, 1$ with `treat`, `onset`, and their interaction as covariates.
- The likelihood is maximized at $r = 0.3$.
- Below I reproduce their results for a logarithmic model with $r = 0.3$.

- Declare survival data using igenset

```
. igenset time, failure(failure) id(obsid)
                id:  obsid
     failure event:  failure != 0 & failure < .
obs. time interval:  (time[_n-1], time]
 exit on or before:  failure

------------------------------------------------------------------------------
      394  total observations
        0  exclusions
------------------------------------------------------------------------------
      394  observations remaining, representing
      394  subjects
      155  failures in single-failure-per-subject data
 14018.24  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =     74.97
```

- Fit a logarithmic model with $r = 0.3$ :

```
. igencox treat##onset, cluster(id) transform(log 0.3) baseq(bq) nolog
         failure _d:  failure
   analysis time _t:  time
                 id:  obsid

Generalized Cox regression -- Breslow method for ties

Cluster: id

Transformation: Logarithmic(.3)

No. of subjects =          394                      No. of obs   =       394
No. of failures =          155
Time at risk    =  14018.24001
                                                    Wald chi2(4)   =     31.25
Log likelihood  = -1002.024549                      Prob > chi2    =    0.0000
```

| _t | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 1.treat | -.5640272 | .2379078 | -2.37 | 0.018 | -1.030318 | -.0977365 |
| 1.onset | .446635 | .256513 | 1.74 | 0.082 | -.0561213 | .9493913 |
| treat#onset | | | | | | |
| 1 1 | -1.073225 | .3850929 | -2.79 | 0.005 | -1.827993 | -.3184567 |
| /sigma2 | 1.24072 | .4326958 | | | .3926516 | 2.088788 |

- Compute covariate-adjusted survivor functions:

```
. predict s00, basesurv
. predict s10, survival at(treat=1 onset=0)
. predict s01, survival at(treat=0 onset=1)
. predict s11, survival at(treat=1 onset=1)
```

- Plot covariate-adjusted survivor functions:

```
. twoway line s00 s10 s01 s11 _t, sort connect(J J J J) title("Survivor functions")
```

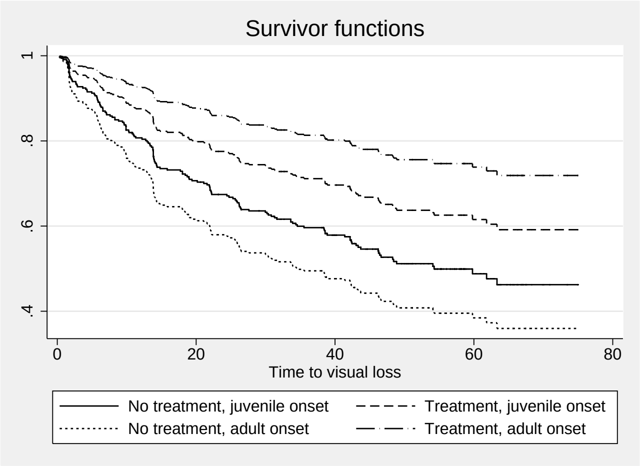


Figure: Survivor functions for treated and untreated eyes for patients with juvenile-onset and adult-onset diabetes.

## Example

- Colon-cancer study evaluating adjuvant therapy on cancer relapse and death for patients with resected colon cancer (Zeng and Lin 2007)
- 315 patients in the observation group and 304 patients in Lev+5-FU group
- 268 cancer relapses and 192 deaths
- Covariates: Lev+5-FU treatment, time of surgery >20, depth of invasion (serosa or not), number of nodes $> 4$

```
. use coloncancer
(Colon-Cancer Data)

. describe

Contains data from coloncancer.dta
  obs:         1,238                          Colon-Cancer Data
 vars:             8                          23 Jul 2014 12:30
 size:        14,856                          (_dta has notes)
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
id              float   %9.0g                 Subject identifier
type            byte    %9.0g                 Failure type (1: cancer
                                                recurrence, 2: death)
time            int     %9.0g                 Time to event: cancer recurrence
                                                or death, in days
failure         byte    %9.0g                 Failure indicator
treat           byte    %9.0g                 Lev+5-FU treatment (0: no
                                                treatment, 1: treatment)
surgery20       byte    %9.0g                 Surgery took place >20 days prior
                                                to randomization (0: <=20, 1:
                                                >20)
serosa          byte    %9.0g                 Depth of invasion (0: submucosa
                                                or muscular layer, 1: serosa)
nodes4          byte    %9.0g                 Number of nodes greater than 4
                                                (0: <=4, 1: > 4)
-------------------------------------------------------------------------------
```

- Zeng and Lin (2007) fit a joint model for the two endpoints by characterizing the dependence between cancer relapse and death through a Gaussian random effect.
- They consider a set of different transformation models.
- Below I use `igencox` to fit the joint model for two types of events assuming proportional hazards.

- Declare multiple-failure survival data using `igenset`:

```
. igenset time, failure(failure) id(id) failtype(type)
-------------------------------------------------------------------------------
type = 1
                id:  id
     failure event:  failure != 0 & failure < .
obs. time interval:  (time[_n-1], time]
 exit on or before:  failure

------------------------------------------------------------------------------
      619  total observations
        0  exclusions
------------------------------------------------------------------------------
      619  observations remaining, representing
      619  subjects
      269  failures in single-failure-per-subject data
   472256  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =      1925
-------------------------------------------------------------------------------
type = 2
                id:  id
     failure event:  failure != 0 & failure < .
obs. time interval:  (time[_n-1], time]
 exit on or before:  failure

------------------------------------------------------------------------------
      619  total observations
        0  exclusions
------------------------------------------------------------------------------
      619  observations remaining, representing
      619  subjects
      192  failures in single-failure-per-subject data
   554878  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =      1925
-------------------------------------------------------------------------------
```

- Fit a joint proportional hazards model for cancer relapse and death:

```
. igencox treat surgery20 serosa nodes4, transform(log 0) baseq(bq) nolog
         failure _d:  failure
   analysis time _t:  time
                 id:  id
 failure type:  type
Generalized Cox regression -- Breslow method for ties
Failure types:
        fail1: type = 1
        fail2: type = 2
Transformation:
        fail1: Logarithmic(0)
        fail2: Logarithmic(0)
No. of subjects =          619                    No. of obs    =       1238
No. of failures =          461              No. of failure types    =          2
Time at risk    =      1027134
                                                   Wald chi2(9)     =     467.00
Log likelihood  = -2895.093543                     Prob > chi2      =     0.0000
```

| _t | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| fail1 | | | | | | |
| treat | -1.481461 | .234295 | -6.32 | 0.000 | -1.940671 | -1.022251 |
| surgery20 | -.6888368 | .2179453 | -3.16 | 0.002 | -1.116002 | -.2616718 |
| serosa | 2.244092 | .2219512 | 10.11 | 0.000 | 1.809076 | 2.679108 |
| nodes4 | 2.893161 | .2053575 | 14.09 | 0.000 | 2.490667 | 3.295654 |
| fail2 | | | | | | |
| treat | -.7216884 | .2803086 | -2.57 | 0.010 | -1.271083 | -.1722937 |
| surgery20 | -.6427658 | .2510799 | -2.56 | 0.010 | -1.134873 | -.1506583 |
| serosa | 1.937907 | .2851385 | 6.80 | 0.000 | 1.379046 | 2.496768 |
| nodes4 | 3.096585 | .2401568 | 12.89 | 0.000 | 2.625886 | 3.567284 |
| /sigma2 | 11.63789 | 1.170385 | | | 9.343973 | 13.9318 |

- Compute event-specific survivor functions:

```
. predict basesurv1, basesurv failtype(type=1)
. predict basesurv2, basesurv failtype(type=2)
```

- Plot event-specific survivor functions:

```
. twoway line basesurv1 basesurv2 _t, sort connect(J J) ///
>   title(Baseline survivor functions)
```

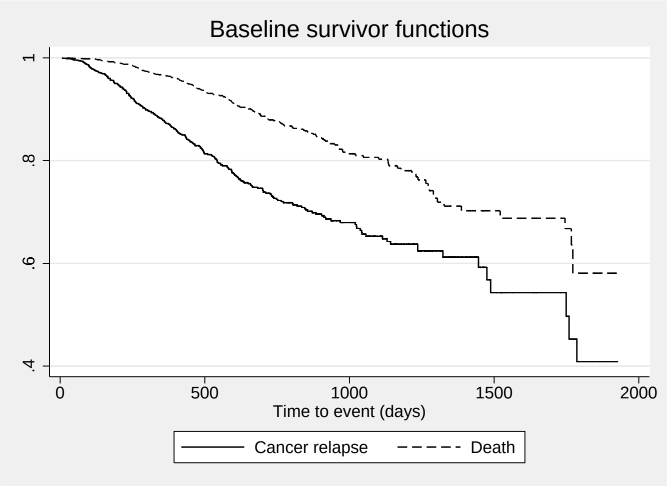


Figure: Baseline survivor functions for cancer relapse and death.

- Compute treatment-group survivor functions for death from colon cancer given relapses of cancer at 200 and 500 days:

```
. predict surv0, survival failtype(type=2) condition(type=1 time=200) ///
>                at(treat=1 surgery20=0 serosa=0 nodes4=0)
. predict surv1, survival failtype(type=2) condition(type=1 time=500) ///
>                at(treat=1 surgery20=0 serosa=0 nodes4=0)
```

- Plot the survivor functions:

```
. twoway line surv0 surv1 _t, sort connect(J J) ///
>   title("Colon-cancer survivor functions" "for treated patients with cancer relapse
```

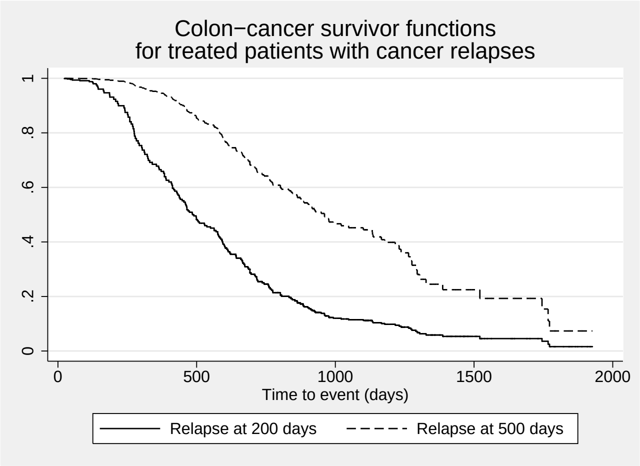


Figure: Colon-cancer survivor functions of treated patients given relapses at 200 and 500 days.

- The first version of `igencox` is available at:

```
. net describe igencox, from(http://fmwww.bc.edu/RePEc/bocode/i)
. net install igencox, from(http://fmwww.bc.edu/RePEc/bocode/i)
```

- The version of `igencox` above fits transformation models of type (1)—single-failure models with iid observations and fixed covariates.
- Extensions to time-varying covariates, clustered data, and multiple and recurrent failures are forthcoming.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.*, **2**, 273–277.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.

Huster, W. J, Brookmeyer, R., and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, **45** 145–156.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd Ed. Hoboken, Wiley.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, **60**, 278–288.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, **93**, 627–640.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Statist. Soc. B*, **69**, 507–564.

Zeng, D., Lin, D. Y. and Lin, X. (2008). Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica*, **18**, 355–377.