Multiple failure-time data

Multiple failure-time data or multivariate survival data are frequently encountered in biomedical and other investigations. These data arise from time-to-event studies when either of two or more events (failures) occur for the same subject, or from identical events occurring to related subjects. In these studies, failure times are correlated within subject, violating the independence of failure times assumption required in traditional survival analysis.

We follow Therneau's (1997) suggestion that for analysis purposes, failure events should be classified according to

- Whether they have a natural order
- Whether they are recurrences of the same types of events.

The counting process approach to survival analysis

A general approach to survival analysis was introduced by Andersen & Gill (1982) where each subject is considered as a counting process (counting events)

- $N_i^{(k)}(t)$ is the total number of events of type k for each subject i up to time t
- $Y_i^{(k)}(t)$ is an indicator function with $Y_{ik}(t) = 1$ if subject i is at risk at time t for event of type k

In this formulation the hazard is considered as an "intensity" process such that

$$\lambda_i^{(k)}(t) = Y_i^{(k)}(t)\lambda_0^{(k)}(t)\exp\{\beta' Z_i\}$$

By judicious choice of the various components of the process as defined above, the counting process approach can handle all kinds of survival data including

- Time updated covariates $Z_i(t)$
- Discontinuous risk sets
- Multiple failures of the different type (competing risks)
- Multiple failures of the same type (both ordered and unordered)

Unordered failures

Failures of the same type include, for example, repeated lung infections with pseudomonas in children with cystic fibrosis, or the development of breast cancer in genetically predisposed families.

Failures of different types include adverse reactions to therapy in cancer patients on a particular treatment protocol, or the development of connective tissue disease symptoms in a group of third graders exposed to hazardous waste.

Ordered failures

Ordered events may result from a study that records the time to first myocardial infarction (MI), second MI, and so on. These are ordered events in the sense that the second event cannot occur before the first event. Unordered events, on the other hand, can occur in any sequence. For example, in a study of liver disease patients, a panel of seven liver function laboratory tests can become abnormal in a specific order for one patient and in a different order for another patient. The order in which the tests become abnormal (fail) is random.

Two main approaches to modeling these data have gained popularity over the last few years:

- The frailty model method.
  In these models the association between failure times is explicitly modeled as a random-effect term, called the frailty shared by all members of the cluster and assumed to follow a known statistical distribution (often the gamma distribution), with mean equal to one and unknown variance.

- Variance-corrected models.
  In this approach the dependencies between failure times are not included in the models. Instead, the covariance matrix of the estimators is adjusted to account for the additional correlation. These models are easily estimated in Stata.

In this lecture we illustrate the main ideas for estimating these models using the Cox proportional hazard model.

Brief mathematical detail and definitions

Let $T_i^{(k)}$ and $U_i^{(k)}$ be the failure and censoring time of the kth failure type ($k = 1, \cdots, K$) in the ith subject ($i = 1, \cdots, m$), and let $\mathbf{Z}_i^{(k)}$ be a p-vector of possibly time-dependent covariates, for the ith subject with respect to the kth failure type.

"Failure type" is used here to mean both failures of different types and failures of the same type.

Assume that $T_i^{(k)}$ and $U_i^{(k)}$ are independent, conditional on the covariate vector $(\mathbf{Z}_i^{(k)})$.

Define $X_i^{(k)} = \min(T_i^{(k)}, U_i^{(k)})$ and $\delta_{ij} = I(T_i^{(j)} \leq U_i^{(j)})$ where $I(.)$ is the indicator function, and let β be a p-vector of unknown regression coefficients. Under the proportional hazard assumption, the hazard function of the ith subject for the kth failure type is

$$\lambda^{(k)}(t; \mathbf{Z}_i^{(k)}) = \lambda_0(t) e^{\mathbf{Z}_i^{(k)} \beta}$$

if the baseline hazard function is assumed to be equal for every failure type, or

$$\lambda^{(k)}(t; \mathbf{Z}_i^{(k)}) = \lambda_{0k}(t) e^{\mathbf{Z}_i^{(k)} \beta}$$

if the baseline hazard function is allowed to differ by failure type (Lin 1994).

Maximum likelihood estimates of for the above models are obtained from the Cox's partial likelihood function, $L(\beta)$, assuming independence of failure times. The estimator $\hat{\beta}$ has been shown to be a consistent estimator for β and is asymptotically normal as long as the marginal models are correctly specified (Lin 1994).

The resulting estimated covariance matrix obtained as the inverse of the information matrix, however, $I^{-1} = -\partial^2 \log L(\beta)/\partial\beta\partial\beta'$ does not take into account the additional correlation in the data, and therefore, it is not appropriate for testing or constructing confidence intervals for multiple failure time data.

Sandwich estimators

Lin and Wei (1989) proposed a modification to this naive estimate, appropriate when the Cox model is misspecified. The resulting robust variance-covariance matrix is estimated as

$$V = I^{-1}U'UI^{-1} = D'D$$

where U is a $n \times p$ matrix of efficient score residuals and D is the $n \times p$ vector of leverage residuals resulting from differences in the estimation of β if each observation i is removed from the data set (this is called *dfbeta* by many software packages). The above formula assumes that the n observations are independent (i.e., there is a single observation per subject – no clustering).

Sandwich estimators with clustered survival data

When observations are not independent, but can be divided into m independent groups $(G_1, G_2, \cdots, G_m)$, then the robust covariance matrix takes the form

$$V = I^{-1} G' G I^{-1}$$

where G is a $m \times p$ matrix of the group efficient score residuals.

Implementation and examples

Implementation of all variance-adjusted models involves three steps: Setting up the data (mainly correctly specifying the time intervals), correct definition of the risk sets (by setting up $Y^{(k)}(t)$) and care in the estimation method. All of the following models can be handled:

1. Unordered failure events
   (a) Unordered failure events of the same type
   (b) Unordered failure events of different types (competing risk)
2. Ordered failure events
   (a) The Andersen-Gill model
   (b) The marginal risk set model
   (c) The conditional risk set model (time from entry)
   (d) The conditional risk set model (time from the previous event)

We will focus on the latter kind of models (i.e., ordered failure-time models):

1. *The Andersen & Gill approach*
   The simplest method to implement follows the counting process approach of Andersen and Gill (1982).

   The basic assumption is that all failure types are indistinguishable. This is a "conditional model" because the time interval for failure k starts at the conclusion of the interval when failure $k-1$ occurred.

   A major limitation of this approach is that it does not allow more than one event to occur at a given time. In addition, the A-G model assumes that all failures within the same subject are independent and models any clustering as explicit interactions included in the model. This assumption is usually untenable.

2. *The WLW model*

   A second model, proposed by Wei, Lin, and Weissfeld (1989), is based on the idea of marginal risk sets. For this analysis, the data are treated like a set of unordered failures, so each event has its own stratum and each patient appears in all strata.

   The marginal risk set at time t for event k is made up of all subjects under observation at time t regardless of whether they had experienced or not events $1, \cdots, k-1$.

3. *The PWP model*

   A third method proposed by Prentice, Williams, and Peterson (1981) is known as the conditional risk set model. The data are set up as for Andersen and Gill's counting processes method, except that the analysis is stratified by failure order. The assumption made is that a subject is not at risk of a second event until the first event has occurred and so on.

   Thus, the conditional risk set at time t for event k is made up of all subjects under observation at time t that have had event $k-1$.

There are two variations to this approach: Time from entry and time from previous event (the so-called "gap-time model").

In the first variation, time to each event is measured from entry time, and in the second variation, time to each event is measured from the previous event.

The above three approaches will be illustrated using the bladder cancer data presented by Wei, Lin, and Weissfeld (1989). These data were collected from a study of 85 subjects randomly assigned to either a treatment group receiving the drug thiotepa or to a group receiving a placebo control. For each patient, time for up to four tumor recurrences was recorded in months (`r1-r4`).

The bladder cancer data

The four models for ordered failures are illustrated by use of the bladder cancer data published in Wei, Lin & Weisfeld (1989).

```
. list in 1/9, noobs
```

id	group	futime	number	size	r1	r2	r3	r4
1	placebo	1	1	3	0	0	0	0
2	placebo	4	2	1	0	0	0	0
3	placebo	7	1	1	0	0	0	0
4	placebo	10	5	1	0	0	0	0
5	placebo	10	4	1	6	0	0	0
6	placebo	14	1	1	0	0	0	0
7	placebo	18	1	1	0	0	0	0
8	placebo	18	1	3	5	0	0	0
9	placebo	18	1	1	12	16	0	0

This dataset includes data on 86 subjects with bladder cancer with follow-up between 0 and 64 months. The data for the first subject that had zero follow-up have been excluded leaving data on 85 subjects. The following are the first nine observations in the data.

The `id` variable identifies the patients, group is the treatment group, `futime` is the total follow-up time for the patient, number is the number of initial tumors, `size` is the initial tumor size, `number` is the number of initial tumors and `r1` to `r4` are the times to first, second, third, and fourth recurrence of tumors.

A recurrence time of zero indicates no tumor.

1. *The Andersen-Gill model*
   To illustrate the bladder cancer data and how each of the four models creates a different data set we consider the data for subject #25 under the four models.

   Under the A-G model, the data from this subject are as follows:

id	group	number	size	rec	status	tstart	tstop
25	1	2	1	1	1	0	3
25	1	2	1	2	1	3	6
25	1	2	1	3	1	6	8
25	1	2	1	4	1	8	12
25	1	2	1	5	0	12	30

   This subject has experienced four recurrences at times 3, 6, 8 and 12 and was followed until time 30.

2. *The Wei, Lin & Weisfeld model*
   Under the WLW model, each patient is simultaneously at risk for all failures (thus the clock starts at time zero).

   Once the fourth failure has been experienced the subject is no longer at risk for another failure (unlike the A-G model above) so the data for subject # 25 above become, under the WLW model:

id	group	number	size	rec	status	tstart	tstop
25	1	2	1	1	1	0	3
25	1	2	1	2	1	0	6
25	1	2	1	3	1	0	8
25	1	2	1	4	1	0	12

3. *The Prentice, Williams and Peterson model*

In the time since entry PWP model, data are set up similarly with the A-G model but the ordering of the failure is considered by the model. In addition, time starts from entry for each interval.

(a) <u>The total time model</u>

Under the PWP total time model the above data will be given as

id	group	number	size	rec	status	tstart	tstop
25	1	2	1	1	1	0	3
25	1	2	1	2	1	0	6
25	1	2	1	3	1	0	8
25	1	2	1	4	1	0	12

(b) The gap-time model

Under the gap-time model the clock starts at the end of the previous failure, so the data for the same subject are given by

id	group	number	size	rec	status	gap
25	1	2	1	1	1	3
25	1	2	1	2	1	3
25	1	2	1	3	1	2
25	1	2	1	4	1	4

Implementing the Andersen-Gill model

To implement the Andersen and Gill model using the results from the bladder cancer study, the data are set up as follows: for each patient there must be one observation per event or time interval.

In general, if a subject has one event, then there will be two observations. The first observation will cover the time from entry until the time of the event, and the second observation the time from the event to the end of follow-up.

The data for the nine subjects listed above are

```
. list if id<=10, noobs
```

id	group	tstart	tstop	status	number	size
1	placebo	0	1	0	1	3
2	placebo	0	4	0	2	0
3	placebo	0	7	0	1	0
4	placebo	0	10	0	5	0
5	placebo	0	6	1	4	0
5	placebo	6	10	0	4	0
6	placebo	0	14	0	1	0
7	placebo	0	18	0	1	0
8	placebo	0	5	1	1	3
8	placebo	5	18	0	1	3
9	placebo	0	12	1	1	1
9	placebo	12	16	1	1	1
9	placebo	16	18	0	1	1

In the original data, subjects 1 through 4 had no tumors recur, thus, each of these 4 patients has only one censored (`status==0`) observation spanning from `tstart`=0 to end of follow-up.

Patient 5 (`id==5`) had one tumor recur at 6 months and was followed until month 14. This patient has two observations in the final dataset; one from `tstart` to tumor recurrence (`tstop==6`), ending in an event (`status==1`), and another from `tstart==6` to end of follow-up (`tstop==10`), ending as censored (`status==0`).

The data are set-up as follows:

```
. stset tstop , fail(status) exit(time .) id(id) enter(tstart)

                id:  id
     failure event:  status != 0 & status < .
obs. time interval:  (tstop[_n-1], tstop]
 enter on or after:  time tstart
 exit on or before:  time .

------------------------------------------------------------------------------
      190  total obs.
        0  exclusions
------------------------------------------------------------------------------
      190  obs. remaining, representing
       85  subjects
      112  failures in multiple failure-per-subject data
     2711  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        64
```

A critical component of the data set-up is to specify the start of each interval by the `tstart` variable.

The Andersen-Gill Cox model is fit as follows:

```
. stcox group size number, nohr nolog

        failure _d:  status
  analysis time _t:  tstop
 enter on or after:  time tstart
 exit on or before:  time .
               id:  id

Cox regression -- Breslow method for ties

No. of subjects =           85                     Number of obs   =       190
No. of failures =          112
Time at risk    =         2711
                                                   LR chi2(3)      =     14.05
Log likelihood  =   -460.07958                     Prob > chi2     =    0.0028

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.4070966   .2000726    -2.03   0.042    -.7992317   -.0149615
        size |  -.0400877   .0702575    -0.57   0.568    -.1777899    .0976146
      number |   .1606478   .0480081     3.35   0.001     .0665536    .2547419
------------------------------------------------------------------------------
```

The marginal risk set model (Wei, Lin, and Weissfeld)

The marginal risk model ignores the ordering of events and treats each failure as differen type of failure $i = 1, \cdots, 4$. The resulting data for the first five subjects are given as follows:

```
. list if id<=2, noobs


. list if id <2
   +-------------------------------------------------+
   | id   group   futime   number   size   rec   status |
   |-------------------------------------------------|
   |  1       1        1        1      3     1        0 |
   |  1       1        1        1      3     2        0 |
   |  1       1        1        1      3     3        0 |
   |  1       1        1        1      3     4        0 |
   |  2       1        4        2      1     1        0 |
   |-------------------------------------------------|
   |  2       1        4        2      1     2        0 |
   |  2       1        4        2      1     3        0 |
   |  2       1        4        2      1     4        0 |
   |-------------------------------------------------|
```

The data are set up as follows:

```
. stset futime, failure(status)

     failure event:  status != 0 & status != .
obs. time interval:  (0, time]
 exit on or before:  failure
--------------------------------------------------------------------------
      340  total obs.
        0  exclusions
--------------------------------------------------------------------------
      340  obs. remaining, representing
      112  failures in single record/single failure data
     8522  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59
```

Conspicuous is the fact that there is no accounting for clustering of these data by subject. Each observation is considered independent of the others.

The Cox model is fitted with the sandwich estimator, clustering on each subject and stratifying on each failure type.

```
. stcox group size number, nohr strata(rec) cluster(id) nolog

         failure _d:  status
   analysis time _t:  futime
Stratified Cox regr. -- Breslow method for ties

No. of subjects      =          340                 Number of obs   =       340
No. of failures      =          105
Time at risk         =         8522
                                                    Wald chi2(3)    =     15.19
Log pseudolikelihood =   -402.74353                 Prob > chi2     =    0.0017

                                  (Std. Err. adjusted for 85 clusters in id)
------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.5575149   .3096526    -1.80   0.072    -1.164423    .0493931
        size |  -.0663418   .0973908    -0.68   0.496    -.2572242    .1245407
      number |   .2082931   .0660605     3.15   0.002     .0788169    .3377692
------------------------------------------------------------------------------
                                                              Stratified by rec
```

The conditional risk set model (time from entry)

As previously mentioned, there are two variations of the conditional risk set model. The first variation in which time to each event is measured from entry is illustrated in this section.

The data are set up as for Andersen and Gill's method, however, a variable indicating the failure order is included. The analysis is then stratified by this variable. The resulting observations for the first five subjects are

```
. list if id<=5, noobs
```

id	group	tstart	tstop	status	number	size	rec
1	1	0	1	0	1	3	1
2	1	0	4	0	2	1	1
3	1	0	7	0	1	1	1
4	1	0	10	0	5	1	1
5	1	0	6	1	4	1	1
5	1	0	10	0	4	1	2

The resulting dataset is identical to that used to fit Andersen and Gill's model except that the `rec` variable identifies the failure risk group for each time span.

For the first 4 individuals, who have not had a recurrence, the `rec` value is one so that they are at risk for a first recurrence the whole follow-up time. The last individual, `id==5`, was at risk for a first recurrence for 6 months (`rec==1`) and at risk of a second recurrence (`rec==2`) from 6 months to the end of follow-up at 10 months.

The data are set up as follows:

```
. stset tstop, fail(status) exit(time .) enter(tstart)

     failure event:  status != 0 & status < .
obs. time interval:  (0, tstop]
 enter on or after:  time tstart
 exit on or before:  time .

------------------------------------------------------------------------------
      183  total obs.
        0  exclusions
------------------------------------------------------------------------------
      183  obs. remaining, representing
      112  failures in single record/single failure data
     3907  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59
```

Note that there is no clustering by subject as the time for all intervals starts at zero.

The total-time PWP model is

```
. stcox group size number, nohr nolog strata(rec)

        failure _d:  status
  analysis time _t:  tstop
 enter on or after:  time tstart
 exit on or before:  time .

Stratified Cox regr. -- Breslow method for ties

No. of subjects =          183                     Number of obs   =       183
No. of failures =          112
Time at risk    =         3907
                                                   LR chi2(3)      =      8.75
Log likelihood  =   -367.17326                     Prob > chi2     =    0.0328

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.4897246   .2092469    -2.34   0.019    -.8998411   -.0796082
        size |  -.0377304   .0675414    -0.56   0.576    -.1701092    .0946484
      number |   .1102692   .0510491     2.16   0.031     .0102149    .2103235
------------------------------------------------------------------------------
                                                              Stratified by rec
```

A robust (sandwich) estimate of the variance can be added:

```
. stcox group size number, nohr nolog robust strata(rec)

         failure _d:  status
   analysis time _t:  tstop
  enter on or after:  time tstart
  exit on or before:  time .

Stratified Cox regr. -- Breslow method for ties

No. of subjects      =          183                 Number of obs   =       183
No. of failures      =          112
Time at risk         =         3907
                                                    Wald chi2(3)    =      9.32
Log pseudolikelihood =   -367.17326                 Prob > chi2     =    0.0254
------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.4897246   .1979229    -2.47   0.013    -.8776464   -.1018029
        size |  -.0377304   .0652989    -0.58   0.563    -.1657138     .090253
      number |   .1102692   .0501329     2.20   0.028     .0120105    .2085278
------------------------------------------------------------------------------
                                                              Stratified by rec
```

Gap time model

The gap time PWP model measures time to each event from the time of the previous event. Time is measured from zero to the gap between each failure.

```
. list if id<=5, noobs
```

id	group	status	number	size	rec	gap
1	1	0	1	3	1	1
2	1	0	2	1	1	4
3	1	0	1	1	1	7
4	1	0	5	1	1	10
5	1	0	4	1	2	6
5	1	1	4	1	1	4

The gap reflects the time between failures. The first four subjects had no recurrences so the gap is the total time of follow-up. Subject 5 experienced a recurrence at 6 months and then was followed up to 10 months, so the gap between the first failure and the end of follow-up is 4 months.

The data are set up as follows:

```
. stset gap status

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

--------------------------------------------------------------------------
      183  total obs.
        5  obs. end on or before enter()
--------------------------------------------------------------------------
      178  obs. remaining, representing
      112  failures in single record/single failure data
     2480  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59
```

The analysis proceeds as in the case of data with single observations per subject, i.e., we do not include the `id()` option.

The corresponding gap-time model is

```
. stcox group size number, nohr nolog strata(rec)

         failure _d:  status
   analysis time _t:  gap

Stratified Cox regr. -- Breslow method for ties

No. of subjects =          178                     Number of obs   =       178
No. of failures =          112
Time at risk    =         2480
                                                   LR chi2(3)      =      8.76
Log likelihood  =   -363.16022                     Prob > chi2     =    0.0327

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.2695213   .2076622    -1.30   0.194    -.6765318    .1374892
        size |   .0068402   .0700105     0.10   0.922    -.1303777    .1440582
      number |   .1535334   .0521059     2.95   0.003     .0514077     .255659
------------------------------------------------------------------------------
                                                              Stratified by rec
```

Clustering by `id` and robust variance estimation is done as follows:

```
. stcox group size number, nohr nolog robust strata(rec) cluster(id)

         failure _d:  status
   analysis time _t:  gap

Stratified Cox regr. -- Breslow method for ties

No. of subjects      =          178                 Number of obs   =       178
No. of failures      =          112
Time at risk         =         2480
                                                    Wald chi2(3)    =     11.99
Log pseudolikelihood =   -363.16022                 Prob > chi2     =    0.0074

                                    (Std. Err. adjusted for 85 clusters in id)
------------------------------------------------------------------------------
             |               Robust
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -.2695213   .2093108    -1.29   0.198    -.6797628    .1407203
        size |   .0068402   .0625862     0.11   0.913    -.1158265     .129507
      number |   .1535334   .0491803     3.12   0.002     .0571418    .2499249
------------------------------------------------------------------------------
                                                          Stratified by rec
```

References

Andersen, P. K. and R. D. Gill. 1982. Cox's regression model for counting processes: A large sample study. Ann Statist 10: 1100-1120. Lee, E. W., L. J. Wei, and D. Amato. 1992.

Cox-type regression analysis for large number of small groups of correlated failure time observations. In Survival Analysis, State of the Art, 237-247. Netherlands: Kluwer.

Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data: The marginal approach. Stat Med 13: 2233-2247.

Lin, D. Y. and L. J. Wei. 1989. The robust inference for the Cox proportional hazards model. Journal of the American Statistical Association 84: 1074-1078.

Therneau, T. M. 1997. Extending the Cox model. Proceedings of the First Seattle Symposium in Biostatistics. New York: Springer.

Wei, L. J., D. Y. Lin, and L. Weissfeld. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the JASA 84: 1065-1073.