Session 9: Multiple failures

We will analyze the bladder data set (Wei et al., 1989). The data statement inputting the data into SAS is given below:

```
title 'Example 1: Multiple Failure Outcomes';

 data bladder(keep=id tstart tstop status trt number size visit);
    retain id tstart 0;
    array tt t1-t4;
    infile cards missover;
    input trt time number size @27 t1 @31 t2 @35 t3 @39 t4;
    id + 1;
    tstart=0;
    do over tt;
       visit=_i_;
       if tt = . then do;
          tstop=time;
          status=0;
       end;
       else do;
          tstop=tt;
          status=1;
       end;
       output;
       tstart=tstop;
    end;
    if (tstart < time) then do;
       tstop= time;
       status=0;
       visit=5;
       output;
    end;
    datalines;
 1       0       1     1
 1       1       1     3
 1       4       2     1
 1       7       1     1
 1       10      5     1
 .        .      .     .
 .        .      .     .
 .        .      .     .
 .        .      .     .
 2       49      3     3
 2       50      1     1
 2       50      4     1    4    24   47
 2       54      3     4
 2       54      2     1    38
 2       59      1     3
 ;
```

The data set is from a study in bladder cancer. The patients were followed for up to four recurrences (t1-t4). Some had less than four and some had none at all.

```
data bladder(keep=id tstart tstop status trt number size visit);
```

This snippet of code defines the first data set and tells SAS which variables will be ultimately kept in the resulting data set (regardless of any variables that will be created). This is very good style of programming.

```
    retain id tstart 0;
```

Since we will turn data with a single line per subject into data with multiple lines per subject (as many as the recurrences plus, in some cases, the residual time of follow-up) we need to keep some variables the same as we generate line after line of code. The `retain` statement does just that. Also, `id` and `tstart` are defined (initially) equal to zero. It will be changed from one subject to another.

```
    array tt t1-t4;
    infile cards missover;
    input trt time number size @27 t1 @31 t2 @35 t3 @39 t4;
```

The command `array` associates a bunch of variables in a variable list so that we can loop over them. It's the same as the command `foreach` in STATA. In this manner, we can loop over `t1`, `t2`, `t3` and `t4`.

```
    id + 1;
    tstart=0;
```

For each subject, variable id is incremented by one and tstart is initialized to zero.

```
    do over tt;
       visit=_i_;
       if tt = . then do;
          tstop=time;
          status=0;
       end;
       else do;
          tstop=tt;
          status=1;
       end;
       output;
       tstart=tstop;
    end;
```

The above loop is done over `t1-t4`. If any of these is missing, then there is no recurrence (this was set to zero in STATA). In that case, `status=0` (censored observation for the recurrence in question) and `tstop` is set to follow-up `time`. When `t1`, `t2`, `t3` or `t4` are non-missing, `status` is set to 1 (recurrence event) and `tstop` to the non-missing time of recurrence. A new observation is output. After this has been done (and we are at the next line of output), `tstart` is set equal to the previous `tstop` and we loop again. In this manner four observations will be created *for all subjects*. Notice that, by using the `output` command, we create new observations (lines of data) whereas, if we had not, SAS would by default only create as many lines as the original data set.

```
    if (tstart < time) then do;
       tstop= time;
       status=0;
       visit=5;
       output;
    end;
```

In a small number of cases, the last (`tstop`) observation (recall that after the output `tstart` is equal with the previous `tstop`), is still smaller than the total follow-up `time`. In that case there is

residual follow-up. We will create a fifth line for these subjects with `tstop` equal the total follow-up time.

A print out of the produced data set is as follows:

```
proc print data=bladder;
     by id;
     var  tstart tstop status  trt    number    size    visit    ;
run;
```

Example 1: Multiple Failure Outcomes

id=1

Obs	tstart	tstop	status	trt	number	size	visit
1	0	0	0	1	1	1	1
2	0	0	0	1	1	1	2
3	0	0	0	1	1	1	3
4	0	0	0	1	1	1	4

id=2

Obs	tstart	tstop	status	trt	number	size	visit
5	0	1	0	1	1	3	1
6	1	1	0	1	1	3	2
7	1	1	0	1	1	3	3
8	1	1	0	1	1	3	4

id=3

Obs	tstart	tstop	status	trt	number	size	visit
9	0	4	0	1	2	1	1
10	4	4	0	1	2	1	2
11	4	4	0	1	2	1	3
12	4	4	0	1	2	1	4

id=4

Obs	tstart	tstop	status	trt	number	size	visit
13	0	7	0	1	1	1	1
14	7	7	0	1	1	1	2
15	7	7	0	1	1	1	3
16	7	7	0	1	1	1	4

id=5

Obs	tstart	tstop	status	trt	number	size	visit
17	0	10	0	1	5	1	1
18	10	10	0	1	5	1	2
19	10	10	0	1	5	1	3
20	10	10	0	1	5	1	4

id=6

Obs	tstart	tstop	status	trt	number	size	visit
21	0	6	1	1	4	1	1
22	6	10	0	1	4	1	2
23	10	10	0	1	4	1	3
24	10	10	0	1	4	1	4

Note that, for the first five subjects, there were no recurrences, so only the first line of data has a different `tstart` and `tstop` (in fact the first subject had zero follow-up and will be excluded from all analyses). By contrast, subject 6 has a recurrence at six months and is followed to 10 months. So that subject has two lines of data with different `tstart` and `tstop`.

The log output of this data step is as follows:

```
NOTE: The data set WORK.BLADDER has 356 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
```

There are four ways to analyze these data that we will show below. These are:

- The Andersen-Gill (conditional model)
- The marginal (Wei-Lin-Weisfeld or WLW model)
- The conditional Prentice-Williams-Peterson (PWP) model. This has two verions:
  - The time from start model
  - The gap-time model

All of these models have in common that they attempt to describe the risk set (i.e., which subjects are at risk for which type of failure, first, second, third or fourth) and estimating the variance.

The Andersen-Gill model

This model (Andersen & Gill, 1981), assumes that the failures are ordered and each subject is at risk for failure *k* only after he or she has had failure *k*-1. That is, you cannot be at risk for the second failure before you have experienced the first failure. While this is a reasonable assumption, the model also assumes that the failures are *independent* from each other, that is, the model does not account for clustering of failures within the same subject.

The analysis of the A-G model is given as follows:

```
title2 'Andersen-Gill Multiplicative Hazards Model';

proc phreg data=bladder;
     model (tstart, tstop) * status(0) = trt number size;
     where tstart < tstop;
 run;
```

Notice that we specify that the model is to only be executed in cases where `tstart < tstop`. This will exclude all the superfluous observations generated by the code above. The output is as follows:

```
                          Example 1: Multiple Failure Outcomes                 37
                       Andersen-Gill Multiplicative Hazards Model
                                                 17:49 Tuesday, January 22, 2008

                                The PHREG Procedure

                                 Model Information

                           Data Set                 WORK.BLADDER
                           Dependent Variable       tstart
                           Dependent Variable       tstop
                           Censoring Variable       status
                           Censoring Value(s)       0
                           Ties Handling            BRESLOW


                          Number of Observations Read         190
                          Number of Observations Used         190
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
190	112	78	41.05

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	934.210	920.159
AIC	934.210	926.159
SBC	934.210	934.315

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.0509	3	0.0028
Score	15.4173	3	0.0015
Wald	15.1736	3	0.0017

We note that, out of the 356 observations produced, only 190 are used in this analysis (including all 112 recurrence events).

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
trt	1	-0.40710	0.20007	4.1402	0.0419	0.666
number	1	0.16065	0.04801	11.1980	0.0008	1.174
size	1	-0.04009	0.07026	0.3256	0.5683	0.961

The result is that treatment reduces recurrences significantly. In addition, the number of tumors at baseline is predictive of subsequent recurrences.

The Wei-Lin-Weisfeld marginal model

The WLW model assumes that each tumor is a separate tumor type. Thus, the first tumor recurrence is a failure of type 1, the second of type 2 and so on. In addition, each subject is eligible for all recurrences (since they are simply failures of different types) *simultaneously*. While this is a mathematical approach (it is not logical in our setting of ordered failures) it makes sense in that, by setting the data in this manner, the approach allows construction of the correct matrices for calculation of the standard errors of the point estimates of the regression coefficients. The WLW approach uses a "sandwich estimator" of the variance of the type

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

where $I = \partial^2 \log L(\beta) / \partial\beta\partial\beta'$ is the usual information matrix and U is an $n \times p$ matrix of the score residuals. Matrix $D = UI^{-1}$ (is the matrix of leverage residuals – also called *dfbeta* by some

packages) with elements d_{ij} that are the differences in the estimate of $\hat{\beta}_j$ if observation i is removed from the dataset. The WLW data set is constructed from the original bladder data set as follows:

```
data bladder2;
    set bladder;
    if visit < 5;
    trt1= trt * (visit=1);
    trt2= trt * (visit=2);
    trt3= trt * (visit=3);
    trt4= trt * (visit=4);
    number1= number * (visit=1);
    number2= number * (visit=2);
    number3= number * (visit=3);
    number4= number * (visit=4);
    size1= size * (visit=1);
    size2= size * (visit=2);
    size3= size * (visit=3);
    size4= size * (visit=4);
 run;
```

The above code simply excludes all observations past the fourth failure as subjects that have experienced all four failures cannot be at risk for anything else. The log output is as follows:

```
NOTE: There were 356 observations read from the data set WORK.BLADDER.
NOTE: The data set WORK.BLADDER2 has 344 observations and 20 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

Since there are 86 subjects with four observations each, there is a total of 344 observations in the `bladder2` data set. In addition, the data step has created four interaction terms between `trt`, `size` and `number` with `visit`.

The WLW model is fit as follows:

```
title2 'Marginal Proportional Hazards Models';

proc phreg data=bladder2 covs(aggregate);
     model tstop*status(0)=trt1-trt4 number1-number4 size1-size4;
     strata visit;
       id ID;
       TREATMENT: test trt1, trt2, trt3, trt4/average e;
run;
```

Note some new features of the `PHREG` procedure. The WLW model essentially fits four separate models, one each for the first, second, third and fourth failure. Then it literally aggregates the matrices of the score residuals from the four models in a new matrix of the form

$$D'D = \begin{pmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{11} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{pmatrix}$$

where each $D_{ij} = I_i^{-1} G_i' G_j I_j^{-1}$, i.e., the information matrix for each model i and j and G is the $g \times p$ score residual matrix (g=4). The `strata` command performs a stratification over each visit (type of failure).

In addition, an overall test for treatment with an estimate of an average effect is requested. The analysis of the WLW model with stata is as follows:

Example 1: Multiple Failure Outcomes

Marginal Proportional Hazards Models

The PHREG Procedure

Model Information

Data Set	WORK.BLADDER2
Dependent Variable	tstop
Censoring Variable	status
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	344
Number of Observations Used	344

Summary of the Number of Event and Censored Values

Stratum	visit	Total	Event	Censored	Percent Censored
1	1	86	47	39	45.35
2	2	86	29	57	66.28
3	3	86	22	64	74.42
4	4	86	14	72	83.72
Total		344	112	232	67.44

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	880.828	851.435
AIC	880.828	875.435
SBC	880.828	908.057

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	29.3932	12	0.0034
Score (Model-Based)	33.0747	12	0.0009
Score (Sandwich)	17.7990	12	0.1219
Wald (Model-Based)	31.0544	12	0.0019

Example 1: Multiple Failure Outcomes

Marginal Proportional Hazards Models

The PHREG Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Wald (Sandwich)	34.8311	12	0.0005

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq	Hazard Ratio
trt1	1	-0.51762	0.30750	0.974	2.8336	0.0923	0.596
trt2	1	-0.61944	0.36391	0.926	2.8975	0.0887	0.538
trt3	1	-0.69988	0.41516	0.903	2.8419	0.0918	0.497
trt4	1	-0.65079	0.48971	0.848	1.7661	0.1839	0.522
number1	1	0.23599	0.07208	0.947	10.7204	0.0011	1.266
number2	1	0.13756	0.08690	0.946	2.5059	0.1134	1.147
number3	1	0.16984	0.10356	0.984	2.6896	0.1010	1.185
number4	1	0.32880	0.11382	0.909	8.3453	0.0039	1.389
size1	1	0.06789	0.08529	0.842	0.6336	0.4260	1.070
size2	1	-0.07612	0.11812	0.881	0.4153	0.5193	0.927
size3	1	-0.21131	0.17198	0.943	1.5097	0.2192	0.810
size4	1	-0.20317	0.19106	0.830	1.1308	0.2876	0.816

Example 1: Multiple Failure Outcomes

Marginal Proportional Hazards Models

The PHREG Procedure

Linear Coefficients for Test TREATMENT

Parameter	Row1	Row2	Row3	Row4	Average Effect
trt1	1	0	0	0	0.67684
trt2	0	1	0	0	0.25723
trt3	0	0	1	0	-0.07547
trt4	0	0	0	1	0.14140
number1	0	0	0	0	0.00000
number2	0	0	0	0	0.00000
number3	0	0	0	0	0.00000
number4	0	0	0	0	0.00000
size1	0	0	0	0	0.00000
size2	0	0	0	0	0.00000
size3	0	0	0	0	0.00000
size4	0	0	0	0	0.00000
CONSTANT	0	0	0	0	0.00000

Test TREATMENT Results

Wald Chi-Square	DF	Pr > ChiSq
3.9668	4	0.4105

Average Effect for Test TREATMENT

Estimate	Standard Error	z-Score	Pr > |z|
-0.5489	0.2853	-1.9240	0.0543

The overall average treatment effect has $\hat{\beta} = -0.5489$, which corresponds to a hazard rate of 0.578, that is, the treatment is associated with about half of the risk for recurrence compared to the control.

To understand the workings of this model, fit a separate model for the first failure. This is done as follows:

```
proc phreg data=bladder2;
     model tstop*status(0)=trt1 number1 size1;
     where visit=1;
     title3 'Model for first visit';
run;
```

The relevant output is as follows:

```
                          Example 1: Multiple Failure Outcomes

                          Marginal Proportional Hazards Models

                                  Model for first visit

                                   The PHREG Procedure

                              Analysis of Maximum Likelihood Estimates

                          Parameter      Standard                                 Hazard
       Variable    DF      Estimate         Error    Chi-Square    Pr > ChiSq       Ratio

       trt1         1      -0.51757       0.31576        2.6868        0.1012       0.596
       number1      1       0.23605       0.07607        9.6287        0.0019       1.266
       size1        1       0.06790       0.10125        0.4498        0.5024       1.070
```

We notice that the point estimate for $\hat{\beta} = -0.518$ is almost identical to the point estimate produced above for `trt1`. The standard error estimate is slightly different because, in the previous case, all standard errors were produced by a sandwich estimator, while the standard error in this case follows the naïve approach.

The Prentice-Williams-Peterson model

There are two types of PWP models: The gap time model and the total time model. In both cases, the setup of the data set is identical to the A-G model, with the exception that time of observation past the last failure is not considered (i.e., once the fourth failure has occurred the patient is not considered further).

a) <u>The gap time model</u>
   In this case, the PWP approach is a version of the A-G conditional model where each subject is considered at risk for each failure *conditional* on having experienced the previous failure. The differentiation of the model is in the fact that the variance estimation proceeds by a stratified analysis according to each failure (i.e., just as in the WLW model, the first failure is considered as failure of type 1, the second of type 2 and so on). In the gap-time model the length of the interval (i.e., (`tstart`, `tstop`]) is considered, where the start of the interval, just as in the A-G case, is past the occurrence of the previous failure (i.e., the subject cannot be eligible to experience a subsequent failure prior to having experienced all previous failures.

   The setup of the data are similar to the A-G model, but the clock starts from the occurrence of the previous model.

We will define variable `gap=tstop-tstart` and we will set up the data as follows:

```
* Fitting the models of Prentice, Williams and Peterson;
 data bladder3(drop=lstatus);
    retain lstatus;
    set bladder2;
    by id;
    if first.id then lstatus=1;
    if (status=0 and lstatus=0) then delete;
    lstatus=status;
    gaptime=tstop-tstart;
 run;
```

In this formulation, all observations following a censored observartion, other than the first observation when lstatus is set to 1, are discarded from the model. This in effect discards all duplicate censored visits but keeps the last censored visit (since the subject would be at risk for a subsequent failure then). A print out of the data is as follows:

```
title2 'PWP Total Time Model with Noncommon Effects';
proc print data=bladder3;
     by id;
     var  tstart tstop status  trt    number    size    visit   gaptime ;
run;
```

Example 1: Multiple Failure Outcomes
PWP Total Time Model with Noncommon Effects

id=1

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
1	0	0	0	1	1	1	1	0

id=2

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
2	0	1	0	1	1	3	1	1

id=3

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
3	0	4	0	1	2	1	1	4

id=4

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
4	0	7	0	1	1	1	1	7

id=5

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
5	0	10	0	1	5	1	1	10

id=6

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
6	0	6	1	1	4	1	1	6
7	6	10	0	1	4	1	2	4

id=7

Obs	tstart	tstop	status	trt	number	size	visit	gaptime
8	0	14	0	1	1	1	1	14

The analysis proceeds as in the case of single-observation per subject data.

```
title2 'PWP Gap Time Model with Common Effects';
 proc phreg data=bladder3;
    model gaptime * status(0) = trt number size;
    strata visit;
 run;
```

```
                        Example 1: Multiple Failure Outcomes

                      PWP Gap Time Model with Common Effects

                                The PHREG Procedure

                                Model Information

                      Data Set                  WORK.BLADDER3
                      Dependent Variable        gaptime
                      Censoring Variable        status
                      Censoring Value(s)        0
                      Ties Handling             BRESLOW


                      Number of Observations Read         184
                      Number of Observations Used         184


                 Summary of the Number of Event and Censored Values

                                                                        Percent
          Stratum    visit          Total       Event    Censored     Censored

                1    1                 86          47          39        45.35
                2    2                 47          29          18        38.30
                3    3                 29          22           7        24.14
                4    4                 22          14           8        36.36
          ---------------------------------------------------------------------
            Total                     184         112          72        39.13


                                Convergence Status

                 Convergence criterion (GCONV=1E-8) satisfied.


                           Model Fit Statistics

                                       Without          With
                      Criterion     Covariates     Covariates

                      -2 LOG L         735.076        726.320
                      AIC              735.076        732.320
                      SBC              735.076        740.476


                      Testing Global Null Hypothesis: BETA=0

                Test                 Chi-Square       DF     Pr > ChiSq

                Likelihood Ratio         8.7559        3         0.0327
                Score                    9.5977        3         0.0223
                Wald                     9.4570        3         0.0238
```

```
                    Example 1: Multiple Failure Outcomes

                  PWP Gap Time Model with Common Effects

                          The PHREG Procedure

                 Analysis of Maximum Likelihood Estimates

                         Parameter      Standard                                Hazard
Variable      DF          Estimate         Error    Chi-Square    Pr > ChiSq     Ratio

trt            1          -0.26952       0.20766        1.6845        0.1943     0.764
number         1           0.15353       0.05211        8.6823        0.0032     1.166
size           1           0.00684       0.07001        0.0095        0.9222     1.007
```

Note that this is equivalent to a model with `tstart=0` and `tstop=gaptime`.

b) <u>The total time conditional model</u>

In this model, `tstart` is set to zero, i.e., the time at risk for each failure is the total time from entry until the occurrence of the failure. The analysis of the PWP model proceeds as follows:

```
title2 'PWP Total Time Model with Common Effects';
 proc phreg data=bladder3;
    model tstop * status(0) = trt number size;
    strata visit;
 run;
```

The analysis by the Cox model is given by the following output:

```
                    Example 1: Multiple Failure Outcomes

                 PWP Total Time Model with Common Effects

                          The PHREG Procedure

                          Model Information

                Data Set                     WORK.BLADDER3
                Dependent Variable           tstop
                Censoring Variable           status
                Censoring Value(s)           0
                Ties Handling                BRESLOW


                Number of Observations Read          184
                Number of Observations Used          184


            Summary of the Number of Event and Censored Values

                                                                 Percent
        Stratum    visit          Total        Event    Censored    Censored

              1    1                 86           47          39       45.35
              2    2                 47           29          18       38.30
              3    3                 29           22           7       24.14
              4    4                 22           14           8       36.36
        ------------------------------------------------------------------
          Total                     184          112          72       39.13
```

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	743.098	734.347
AIC	743.098	740.347
SBC	743.098	748.502

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.7512	3	0.0328
Score	8.8795	3	0.0309
Wald	8.7957	3	0.0321

Example 1: Multiple Failure Outcomes

PWP Total Time Model with Common Effects

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
trt	1	-0.48972	0.20925	5.4775	0.0193	0.613
number	1	0.11027	0.05105	4.6659	0.0308	1.117
size	1	-0.03773	0.06754	0.3121	0.5764	0.963