XV Italian Stata Users Group Meeting

Recurrent-event analysis with Stata: methods and applications

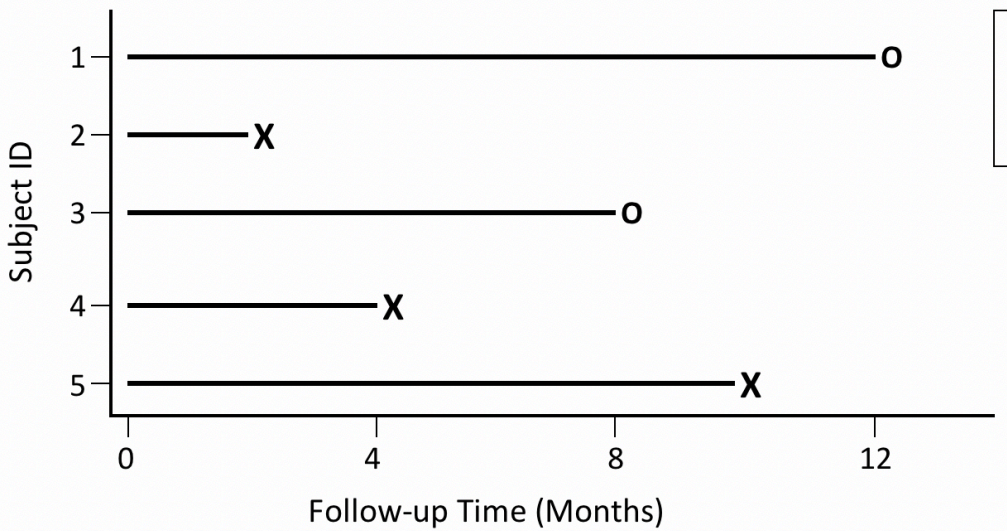
Francesca Ghilotti & Rino Bellocco

Bologna
November 15, 2018

Overview

Introduction to Survival Analysis

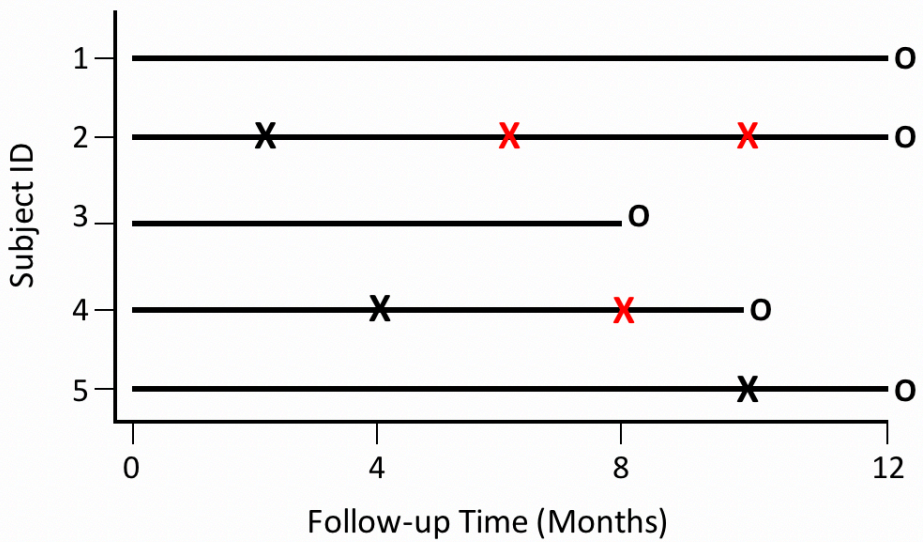
- The outcome variable is time until the occurrence of an event of interest
- Some observations might be censored, that is, the actual time until the event is not observed
- In Stata: `stset` *Time*, `failure(`*Event*`)`
- Cox proportional hazards model



ID	Time	Event
1	12	0
2	2	1
3	8	0
4	4	1
5	10	1

Introduction to Recurrent Events

- It is common in medical research that the event of interest can occur more than once in the same individual:
  `e.g. admissions to hospital, cardiovascular events, infections, cancer recurrences`



o Censoring
X First Event
X Recurrent Event

ID	Time	Event
1	12	0
2	2	1
2	6	1
2	10	1
2	12	0
3	8	0
4	4	1
4	8	1
4	10	0
5	10	1
5	12	0

Data structure

ID	Time	Event	T_Start	T_End	Interval
1	12	0	0	12	1
2	2	1	0	2	1
2	6	1	2	6	2
2	10	1	6	10	3
2	12	0	10	12	4
3	8	0	0	8	1
4	4	1	0	4	1
4	8	1	4	8	2
4	10	0	8	10	3
5	10	1	0	10	1
5	12	0	10	12	2

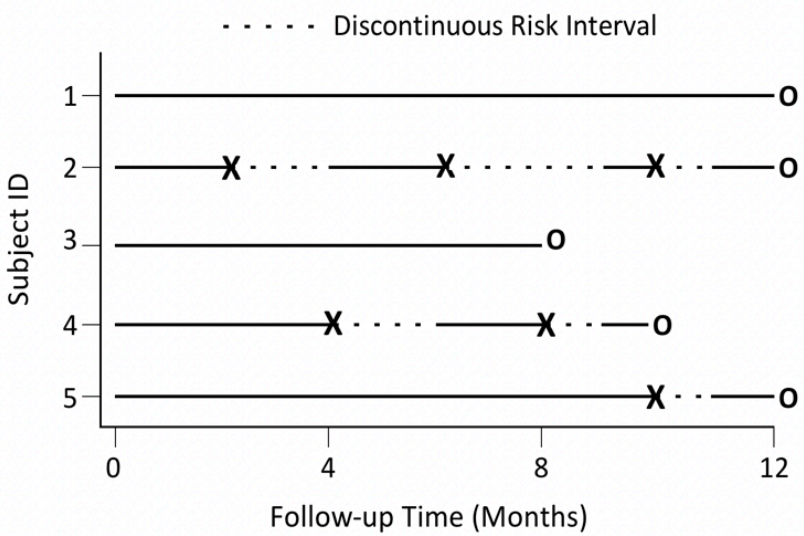
How to declare data in Stata

```
. stset t_end, failure(event) exit(time .) id(id)

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

--------------------------------------------------------------------------------
        11  total observations
         0  exclusions
--------------------------------------------------------------------------------
        11  observations remaining, representing
         5  subjects
         6  failures in multiple-failure-per-subject data
        54  total analysis time at risk and under observation
                                              at risk from t =         0
                                   earliest observed entry t =         0
                                        last observed exit t =        12
```

Discountinuos Risk Intervals



ID	Time	Event	T_Start*	T_End	Duration	Interval (K)
1	12	0	0	12	0	1
2	2	1	0	2	2	1
2	6	1	4	6	3	2
2	10	1	9	10	1	3
2	12	0	11	12	0	4
3	8	0	0	8	0	1
4	4	1	0	4	2	1
4	8	1	6	8	1	2
4	10	0	9	10	0	3
5	10	1	0	10	1	1
5	12	0	11	12	0	2

* T_Start[K] = T_End[K-1] + Duration[K-1]

How to declare data in Stata

```
. stset t_end, failure(event) exit(time .) id(id) enter(t_start)

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

--------------------------------------------------------------------------
        11  total observations
         0  exclusions
--------------------------------------------------------------------------
        11  observations remaining, representing
         5  subjects
         6  failures in multiple-failure-per-subject data
        54  total analysis time at risk and under observation
                                        at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        12
```

This stset is wrong!

How to declare data in Stata

```
. stset t_end, failure(event) exit(time .) id(id) time0(t_start)

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

--------------------------------------------------------------------------------
        11  total observations
         0  exclusions
--------------------------------------------------------------------------------
        11  observations remaining, representing
         5  subjects
         6  failures in multiple-failure-per-subject data
        44  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =        12
```

This is the correct stset for discontinuous risk intervals

How to analyze Recurrent-Event Data

Traditional methods are not wrong, but they imply an inefficient use of data.

Logistic regression

- Binary outcome that indicates whether or not the event was ever experienced during follow-up
- Time at the event is not considered and it ignores all events after the first

Models for count data: Poisson and Negative Binomial

- Total number of events per a fixed period of time
- The time between repeated occurrences is ignored

Traditional Cox Model

- It considers time to the first event
- All events after the first are disregarded

How to analyze Recurrent-Event Data

Problems:

- Failure times are correlated within the same subject
- We need statistical methods that take into account the lack of independence

Solutions:

- Extensions of the traditional Cox model have been proposed:
  - a)* Andersen-Gill model (AG)
  - b)* Prentice, Williams and Peterson Total Time (PWP-TT)
  - c) Prentice, Williams and Peterson Gap Time (PWP-GT)
  - d) Wei, Lin and Weissfeld model (WLW)
  - e)* Frailty models
  - f) Multi-state models (MSM)

How to choose among the models

Some questions which are important to keep in mind:

- Is the order of the events important?
- Does the risk of recurrent event change as the number of previous events increases?
- Are we interested in the overall effect or in the effect for the k^{th} event?
- Are there many recurrences per subject?

Andersen-Gill model (AG)

$$\lambda_{ik}(t) = \lambda_0(t)\, e^{X_{ik}\beta}$$

$\lambda_{ik}(t)$ represents the hazard function for the k^{th} event of the i^{th} subject

- Simple extension of the Cox model
- It uses robust standard errors to account for correlation (variance-corrected method)
- It uses a common baseline hazard function for all events
- It estimates a global parameter
- It assumes that all failure types are equal (unordered)
- Subjects contribute to the risk-set for an event as long as they are under observation at the time the event occurs

Andersen-Gill model (AG)

How to implement it using Stata

```
. stcox var1 var2, robust
```

When to use it

- When the interest is on the overall effect of a covariate on the hazard of a recurrent event
- When the risk of recurrent events remains constant regardless of the number of previous events
- It is adequate for frequent events

Prentice, Williams and Peterson Total Time (PWP-TT)

$$\lambda_{ik}(t) = \lambda_{0k}(t)\, e^{X_{ik}\beta}$$

- Events are ordered and handled by stratification
- The PWP models are conditional models
- Everyone is at risk for the first stratum, but only who had an event in the previous stratum are at risk for the successive one
- It can estimate both overall and event-specific effects
- It uses robust standard errors to account for correlation (variance-corrected method)

Prentice, Williams and Peterson Total Time (PWP-TT)

How to implement it using Stata

```
. stcox var1 var2, robust strata(interval)

. stcox var1 var2 var1*interval, ///
  robust strata(interval)
```

When to use it

- When the effects of covariates are different in subsequent events
- When the occurrence of the first event increases the likelihood of a recurrence
- When there are few recurrent events per subject

Prentice, Williams and Peterson Total Time (PWP-TT)

Final Remarks

- Data should be restricted to a certain number of events if the risk set becomes very small as the number of strata increases
- PWP-TT models could significantly underestimate the overall effect if there is no strong biological relationship between events

Frailty Models

$$\lambda_i(t) = \lambda_0(t)\, \alpha_i\, e^{X_i\beta}$$

- α_i is the random effect that describes excess risk or frailty for distinct individuals and induces dependence among the recurrent events
- The random effect varies across subjects but it is constant over time within subject
- The baseline hazard function does not vary by event
- The event order is not taken into account

Frailty Models

How to implement it using Stata

```
. stcox var1 var2, shared(id)
```

frailties are assumed to be gamma-distributed

```
. streg var1, dist(weibull) shared(id) ///
  frailty(gamma|invgaussian)
```

When to use it

- When there is heterogeneous susceptibility to the risk of recurrent events

Example in Stata

Chronic Granulomatous Disease (CGD) Infection Data

The CGD data set in Fleming and Harrington (1991) is from a placebo-controlled randomized trial of gamma interferon in chronic granulomatous disease. In total, 128 patients were followed for about 1 year. Each patient may experience more than one infection.

Example in Stata

Number of infections per individual

Number of infections	Number of individuals
0	84
1	27
2	9
3	5
4	1
5	1
6	0
7	1
	128

Number of individuals per interval

Interval	Number of individuals
1	128
2	44
3	16
4	8
5	3
6	2
7	1
8	1

Example in Stata

	N events	HR	95% CI
Time to first infection			
Traditional Cox Model	44	0.31	0.16 - 0.61
Multiple-Failure Analysis			
AG	76	0.33	0.18 - 0.60
PWP-TT	69	0.41	0.23 - 0.73
Frailty	76	0.34	0.19 - 0.63
AG*	69	0.35	0.19 - 0.64
Frailty*	69	0.33	0.18 - 0.62

* Same number of events as in the PWP-TT model

References

1 Amorim, L. D., & Cai, J. (2015). Modelling recurrent events: a tutorial for analysis in epidemiology. International journal of epidemiology, 44(1), 324-333.

2 Guo, Z., Gill, T. M., & Allore, H. G. (2008). Modeling repeated time-to-event health conditions with discontinuous risk intervals. Methods of information in medicine, 47(02), 107-116.

3 Westbury, L. D., Syddall, H. E., Simmonds, S. J., Cooper, C., & Sayer, A. A. (2016). Identification of risk factors for hospital admission using multiple-failure survival models: a toolkit for researchers. BMC medical research methodology, 16(1), 46.

References

4 Kelly, P. J., & Lim, L. L. Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. Statistics in medicine, 19(1), 13-33.

5 Cleves, M. (2000). Analysis of multiple failure-time data with Stata. Stata Technical Bulletin, 9(49).

6 Fleming, T. R., & Harrington, D. P. (2011). Counting processes and survival analysis (Vol. 169). John Wiley & Sons.

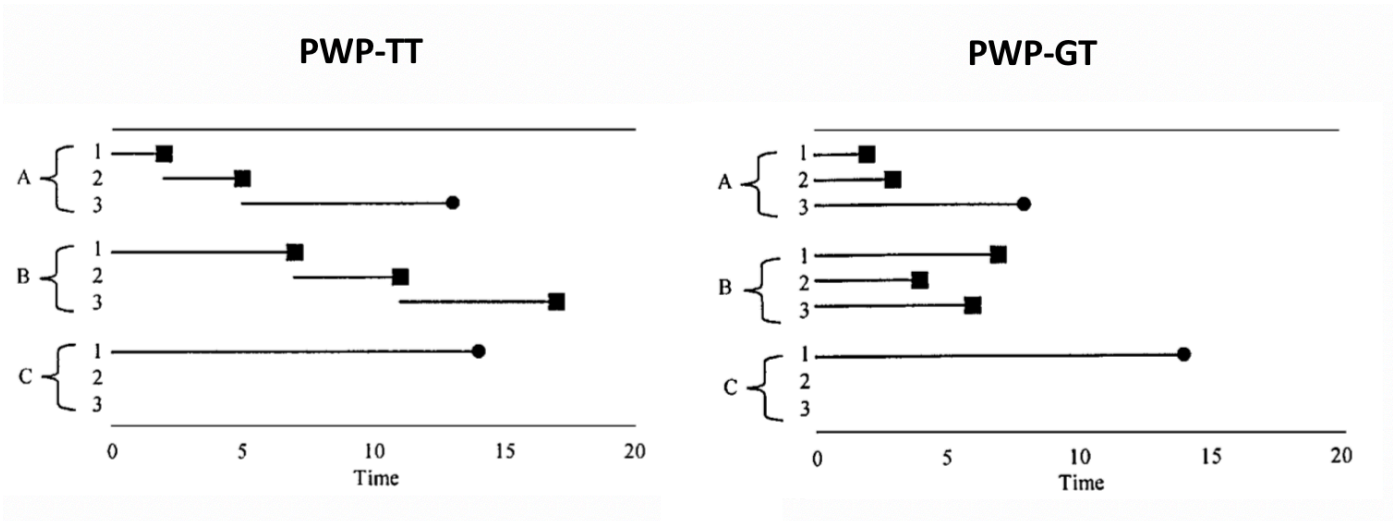
7 Kleinbaum, D. G., & Klein, M. (2001). Survival Analysis A Self-Learning Text.

Thank You!

PWP-TT and PWP-GT

Depending on how the starting point of the risk interval is set, there are two variations of PWP models:

- In the PWP-TT model the time scale is time t, from beginning of study
- In the PWP-GT model the time scale is time t, from the previous event



Proportional Hazard (PH) Assumption

- Hazards have to be proportional over time
- With AG model the PH assumption may be too strong in practice:
  hazard ratio assumed to be constant through time and common across recurrent events

Interpretation of the estimates

The interpretation of the estimates in multiple-failure survival models is unchanged compared to the traditional Cox model.

The individual likelihood L_i gives the conditional probability of failing at time $t(f)$ given that the subject is remaining in the risk set at $t(f)$, i.e. not have failed since the last event.

AG: HR=0.33
Treated patients have a 67% lower hazard of recurrent infections

Frailty models: HR=0.34
Conditional on unmeasured heterogeneity, treatment is associated with a 66% reduction in the recurrent risk