XIII Italian Stata Users Group Meeting

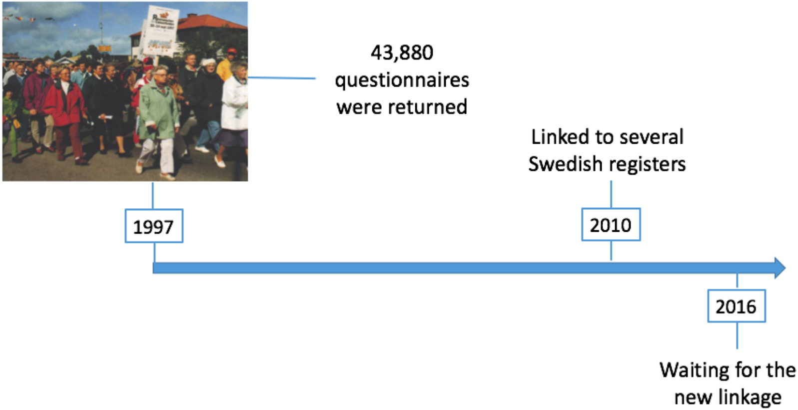
Time scales in Survival analysis: a review of definitions, methods and applications

Francesca Ghilotti, Sc.M
Rino Bellocco, Sc.D

November 17, 2016

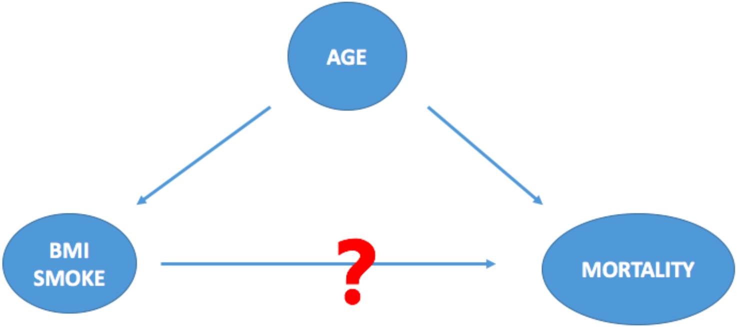
Swedish National March Cohort

Setting:



Aim

Study the relationship between BMI, smoke and overall Mortality



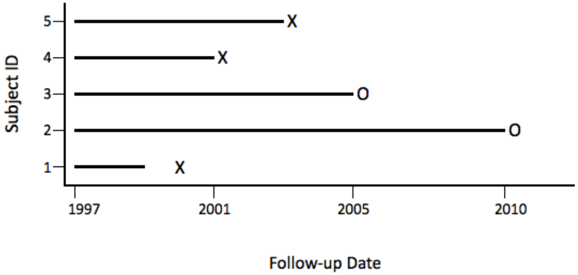
How to declare survival data

```
stset time_of_failure, failure(failure_var)
```

Some key concepts:

- **Time origin** defines time 0, when we start recording time
- **Entry time** defines when a subject starts being at risk
- **Exit time** defines when a subject stops being at risk
- **Analysis time** difference between exit and entry times
- **Failure indicator** defines if a subject has the event or is censored

Follow-up as time-scale



Follow-up as time-scale: stset

```
. stset endt_death, f(out_death) orig(beginf) scale(365.25)

     failure event:  out_death != 0 & out_death < .
obs. time interval:  (origin, endt_death]
 exit on or before:  failure
    t for analysis:  (time-origin)/365.25
            origin:  time beginf

------------------------------------------------------------------------------
      41647  total observations
          0  exclusions
------------------------------------------------------------------------------
      41647  observations remaining, representing
       4296  failures in single-record/single-failure data
 525471.316  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  13.24846
```

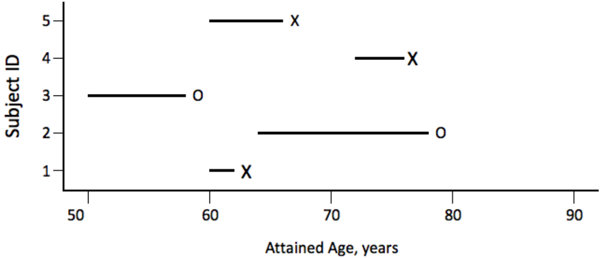
Follow-up as time-scale: Likelihood

ID	t	a_0	a
1	2	60	62
2	13+	64	77
3	8+	50	58
4	4	72	76
5	6	60	66

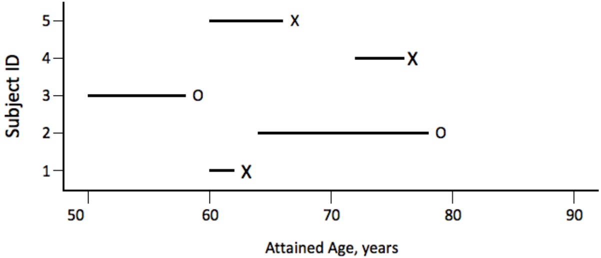
$$L(\beta) = \prod_{i=1}^{n} L_i = \left(\frac{\lambda_1(2)}{\lambda_1(2) + \lambda_2(2) + \lambda_3(2) + \lambda_4(2) + \lambda_5(2)} \right) \times$$

$$\times \left(\frac{\lambda_4(4)}{\lambda_2(4) + \lambda_3(4) + \lambda_4(4) + \lambda_5(4)} \right) \times \left(\frac{\lambda_5(6)}{\lambda_2(6) + \lambda_3(6) + \lambda_5(6)} \right)$$

Age as time-scale



Age as time-scale



We have to deal with delayed entries and left truncation

Age as time-scale: stset

```
. stset endt_death, f(out_death) orig(birthd) entry(beginf) scale(365.25)

     failure event:  out_death != 0 & out_death < .
obs. time interval:  (origin, endt_death]
 enter on or after:  time beginf
 exit on or before:  failure
    t for analysis:  (time-origin)/365.25
            origin:  time birthd

------------------------------------------------------------------------------
      41647  total observations
          0  exclusions
------------------------------------------------------------------------------
      41647  observations remaining, representing
       4296  failures in single-record/single-failure data
 525471.316  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =  18.00137
                                          last observed exit t =  107.5099
```

Age as time-scale: Likelihood

ID	t	a_0	a
1	2	60	62
2	13+	64	77
3	8+	50	58
4	4	72	76
5	6	60	66

$$L(\beta) = \prod_{i=1}^{n} L_i = \left(\frac{\lambda_1(62)}{\lambda_1(62) + \lambda_5(62)}\right) \times \left(\frac{\lambda_2(66)}{\lambda_2(66) + \lambda_5(66)}\right) \times \left(\frac{\lambda_4(76)}{\lambda_2(76) + \lambda_4(76)}\right)$$

The hazard function

The hazard function is the instantaneous rate at which the event occurs:

$$\lambda_T(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P(T \in [t, t + \Delta) | T \geq t)$$

Choice of the Time-Scale

Follow-up as Time-Scale: $\lambda_T(t|a_0, z)$

Age as Time-Scale: $\lambda_A(a|a_0, z)$

Which time scale should be used and how do we make such a decision in general?

Clinical trial:

- Subjects start to be followed for the outcome after random allocation
- Reasonable to assume subjects start to be at risk upon study entry
- Time-on-Study typically used as the time-scale

Observational study:

- Subjects already at risk prior to study entry
- Unknown time or age when first at risk
- Example: Subjects with diabetes enter study, but unknown date or age when first diagnosed (prior to study entry)
- Attained age typically used as the time-scale

Characteristics of the two time-scales

Follow-up as Time-Scale

1. Closed cohort
2. Risk-set decreases over time
3. Censoring occurs at the end of follow-up
4. Assumptions about the relationship between age and the outcome are needed
5. Preferable if time-on-study is a stronger determinant of the outcome than age at the event

Age as Time-Scale

1. Open cohort
2. Risk-sets are not nested
3. Censored observations are spread across the entire range of age
4. Adjustment for age is handled nonparametrically
5. Preferable if age is a much stronger determinant of the outcome than time-on-study

Follow-up as Time-Scale

Model 0: $\lambda(t, Z) = \lambda_0(t) \cdot e^{\beta z}$
Unadjusted for a_0

Model 1: $\lambda(t, Z, a_0) = \lambda_0(t) \cdot e^{\beta z + \gamma_1 a_0}$
Adjusted for a_0 as a linear covariate

Model 2: $\lambda(t, Z, a_0) = \lambda_0(t) \cdot e^{\beta z + \gamma_1 a_0 + \gamma_2 a_0^2}$
Adjusted for a_0 with quadratic covariate

Model 3: $\lambda(t, Z, a_0) = \lambda_0(t) \cdot e^{\beta z + \gamma_1 a_{65}}$
Adjusted for a_0 as binary $(\geq 65 years)$

Model 4: $\lambda_g(t, Z, a_0) = \lambda_{0g}(t) \cdot e^{\beta z}$
Stratified by a_0 or birth cohort, g=1,...,s

Age as Time-Scale

Model 5: $\lambda(a, Z) = \lambda_0(a) \cdot e^{\beta z}$

Unadjusted for left truncation at a_0

Model 6: $\lambda(a, Z) = \lambda_0(a|a_0) \cdot e^{\beta z}$

Adjusted for left truncation at a_0

Model 7: $\lambda_g(a, Z) = \lambda_{0g}(a|a_0) \cdot e^{\beta z}$

Adjusted for left truncation at a_0 and stratified by birth cohort, g=1,...,s

Comparisons among the estimated HR

HR estimates (BMI $\geq$ 25 vs BMI $<$ 25) adjusted for age

Follow-up as Time-Scale		**Age as Time-Scale**	
Model	HR	Model	HR
(0) Unadjusted	1.286	(5) Unadjusted truncation	1.137
(1) Linear age	1.082	(6) Adjusted truncation	1.117
(2) Quadratic age	1.102	(7) Adjusted and stratified	1.109
(3) Binary age	1.112		
(4) Stratified on age	1.107		

Comparisons among the estimated HR

HR estimates (Smoke vs Non-Smoke) adjusted for age

Follow-up as Time-Scale		**Age as Time-Scale**	
Model	HR	Model	HR
(0) Unadjusted	1.036	(5) Unadjusted truncation	1.513
(1) Linear age	1.374	(6) Adjusted truncation	1.401
(2) Quadratic age	1.391	(7) Adjusted and stratified	1.400
(3) Binary age	1.299		
(4) Stratified on age	1.385		

What if age is an effect modifier?

- **Follow-up as time scale (Stratified analysis):**

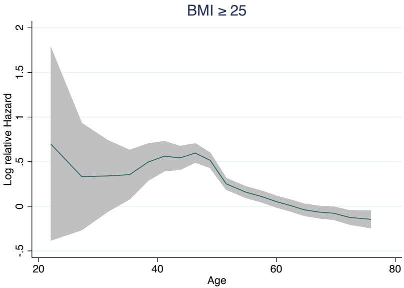
	Age < 65			Age ≥ 65		
	HR	se	$p - value$	HR	se	$p - value$
BMI $\geq$ 25	1.81	0.10	< 0.001	0.90	0.03	0.005

- **Follow-up as time scale (Interaction term):**

Var	HR	se	$p - value$
BMI $\geq$ 25	1.35	0.085	< 0.001
Centered Age	1.12	0.002	< 0.001
Interaction	0.99	0.003	< 0.001

What if age is an effect modifier?

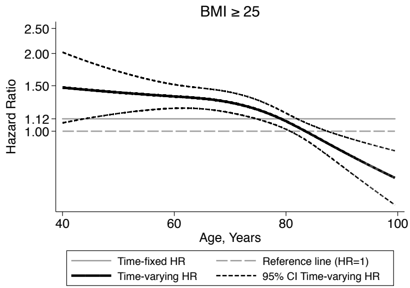
STEPP procedure:



What if age is an effect modifier?

- **Age as time scale:**

If the effect of a risk factor changes with age the hazards would not be proportional.



Conclusions

- When analysing epidemiologic cohort data is better to use age as time-scale
- If adjustment for age at entry is made correctly the difference among the estimated regression coefficients depending on the time-scale chosen is minimal
- Pay attention to the calendar-period and/or birth cohort effects (stratified models)


THANK
YOU

References

1 Kleinbaum, D. G., & Klein, M. (2012). Survival analysis: a self-learning text. Springer Science & Business Media.

2 Klein, J. P., & Moeschberger, M. L. (2005). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

3 Korn, E. L., Graubard, B. I., & Midthune, D. (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. American journal of epidemiology, 145(1), 72-80.

4 Ingram, D. D., Makuc, D. M., & Feldman, J. J. (1997). Re:'Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale'. American journal of epidemiology, 146(6), 528-529.

5 Thiebaut, A., & Benichou, J. (2004). Choice of timescale in Cox's model analysis of epidemiologic cohort data: a simulation study. Statistics in medicine, 23(24), 3803-3820.

6 Cologne, J., Hsu, W. L., Abbott, R. D., Ohishi, W., Grant, E. J., Fujiwara, S., & Cullings, H. M. (2012). Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. Epidemiology, 23(4), 565-573.

7 Pencina, M. J., Larson, M. G., & D'Agostino, R. B. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. Statistics in medicine, 26(6), 1343-1359.

8 Gail, M. H., Graubard, B., Williamson, D. F., & Flegal, K. M. (2009). Comment on 'Choice of time scale and its effect on significance of predictors in longitudinal studies'. Statistics in medicine, 28(8), 1315.

9 Lamarca, R., Alonso, J., Gmez, G., & Muoz, A. (1998). Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 53(5), M337-M343.