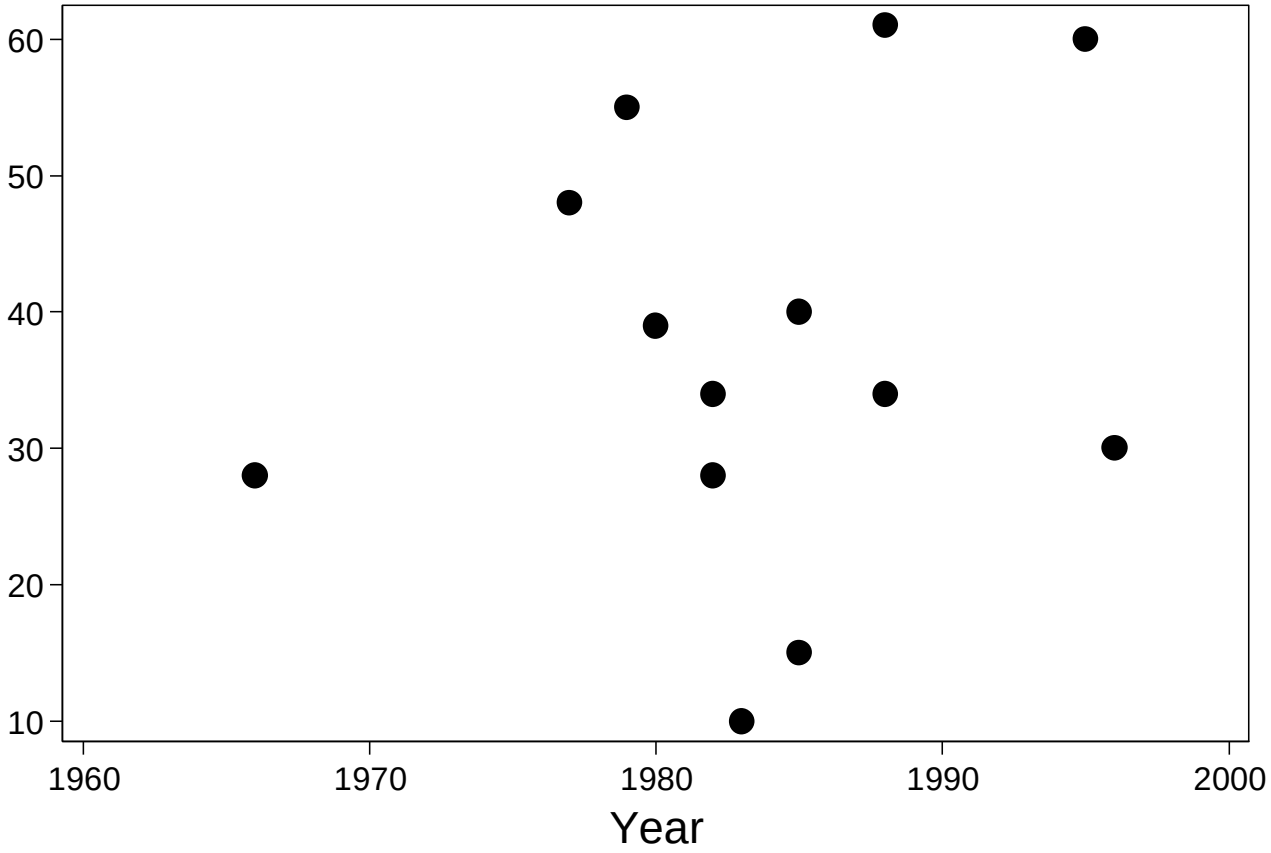
Erik Parner
Sektion for Biostatistik

Biostatistisk metode – et par eksempler

Kvantitativ data



Er der en trend i proportionerne?

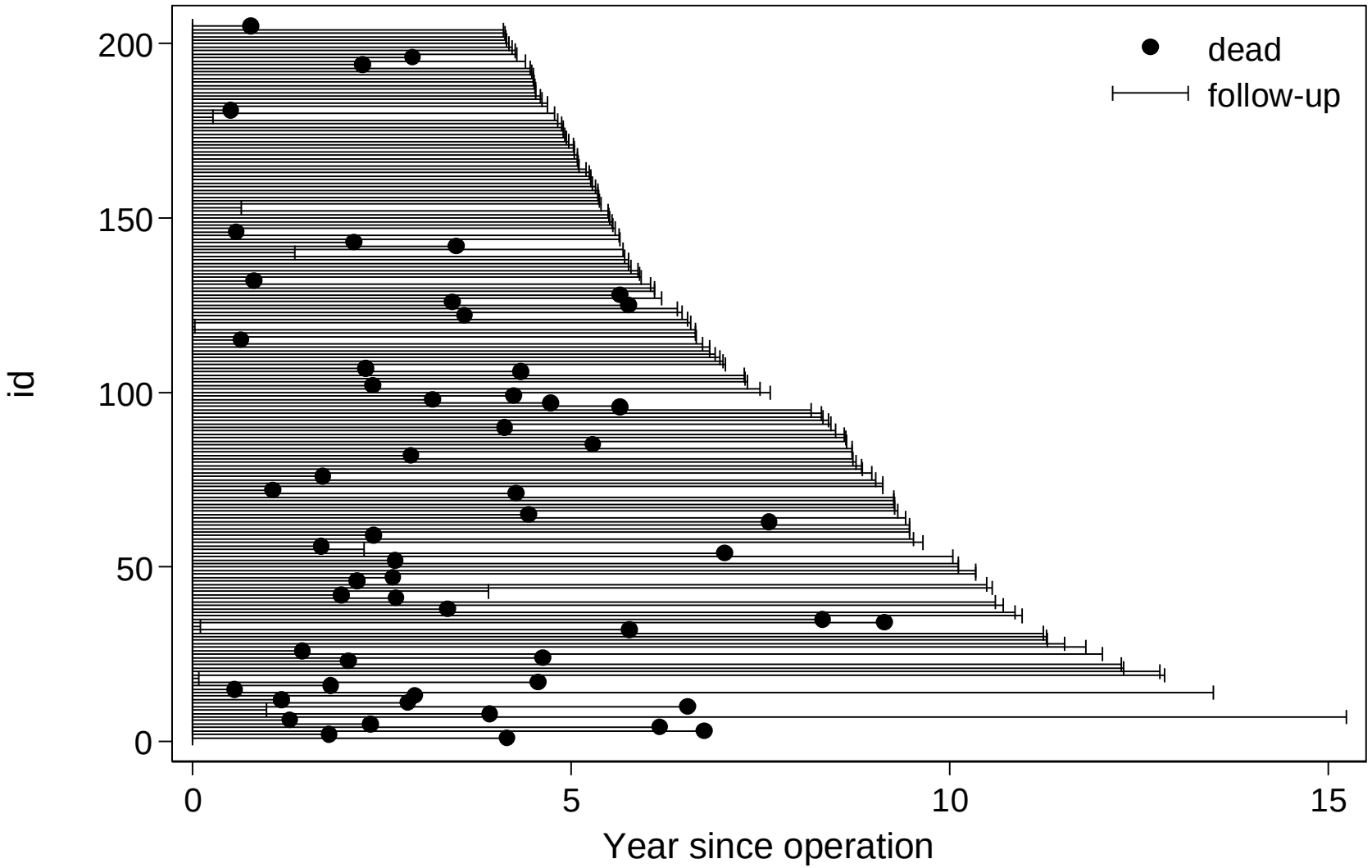
Mere informations

Table I. Summary of some reviews of the quality of statistics in medical journals, showing the percentage of 'acceptable' papers (of those using statistics)

Year published	First author	Number of papers	Number of Journals	% papers acceptable
1966	Schor[4]	295	10	28
1977	Gore[5]	77	1	48
1979	White[6]	139	1	55
1980	Glantz[7]	79	2	39
1982	Felson[8]	74	1	34
1982	MacArthur[9]	114	1	28
1983	Tyson[10]	86	4	10
1985	Avram[11]	243	2	15
1985	Thorn[12]	120	4	< 40
1988	Murray[13]	28	1	61
1988	Morris[14]	103	1	34
1995	McGuigan[15]	164	1	60
1996	Welch[16]	145	1	30

Altman (1998). Statistical reviewing in medical journals. Statist. Med. 17, 2661—2674.

Event data


dead
follow-up
id
200
150
100
50
0
0
5
10
15
Year since operation

Traditionelt har man i analysen af event data ofte brugt

- Kaplan-Meier estimat til at sig noget om risiko (Aalen-Johansen)
- Cox regression.

Der er i disse år en rivende udvikling inden for event analyser. Vi arbejder meget med pseudo observation metoden.

Et par eksempler på samarbejde:

Rasmus Østergaard Nielsen, Sektion for Idræt

Kausale analyser

Mogens Vestergaard, Sektion for Almen Medicin

Instrument variable analyser

Kim Overvad, Sektion for Epidemiologi

Risiko analyser på case-cohort data
(biologisk interaktion)

Cecilia Ramlau-Hansen, Sektion for Almen Medicin

Interval censurerede data

Jiong Li, Sektion for Epidemiologi,

Diana Schendel, Sektion for Epidemiologi

Estimation af en kumuleret incidens under kalendertids-trends

Stefan Nygaard Hansen
Sektion for Biostatistik, Aarhus Universitet

1. december, 2016

Nyhedshistorie fra 2014

Hver tredje dansker bliver psykisk syg i løbet af livet

Ny forskning kortlægger det reelle omfang af psykisk sygdom i Danmark. Psykiatrifond håber på opprioritering.

En ny undersøgelse viser, at op mod hver 3. får en psykisk lidelse i løbet af livet. (© Colourbox)

PRINT

DEL ARTIKLEN:

MAIL

TWITTER

Op mod hver tredje får i løbet af livet en psykisk lidelse, og det er betydeligt flere end end hidtil antaget.

Det viser nyt opsigtsvækkende forskningsstudie fra Aarhus Universitet af danskernes såkaldte livstidsrisiko for psykiske lidelser.

- Blev bl.a. dækket af DR and Politiken.

Nyhedshistorie fra 2014

- Hvilken population knytter dette udsagn sig til?
- Lad os kigge på den videnskabelige artikel bag:

Original Investigation

A Comprehensive Nationwide Study of the Incidence Rate and Lifetime Risk for Treated Mental Disorders

Carsten Bøcker Pedersen, DrMedSc; Ole Mors, PhD; Aksel Bertelsen, MD; Berit Lindum Waltoft, MSc; Esben Agerbo, DrMedSc; John J. McGrath, MD; Preben Bo Mortensen, DrMedSc; William W. Eaton, PhD

IMPORTANCE Understanding the epidemiologic profile of the life course of mental disorders is fundamental for research and planning for health care. Although previous studies have used population surveys, informative and complementary estimates can be derived from population-based registers.

OBJECTIVE To derive comprehensive and precise estimates of the incidence rate of and lifetime risk for any mental disorder and a range of specific mental disorders.

DESIGN, SETTING, AND PARTICIPANTS We conducted a follow-up study of all Danish residents (5.6 million persons), to whom all treatment is provided by the government health care system without charge to the patient, from January 1, 2000, through December 31, 2012 (total follow-up, 59.5 million person-years). During the study period, 320 543 persons received first lifetime treatment in a psychiatric setting for any mental disorder; 489 006 persons were censored owing to death; and 69 987 persons were censored owing to emigration. Specific categories of mental disorders investigated included organic mental disorders, substance abuse disorders, schizophrenia, mood disorders, anxiety, eating disorders, personality disorders, mental retardation, pervasive developmental disorders, and behavioral and emotional disorders.

Supplemental content at jamapsychiatry.com

Hovedresultatet

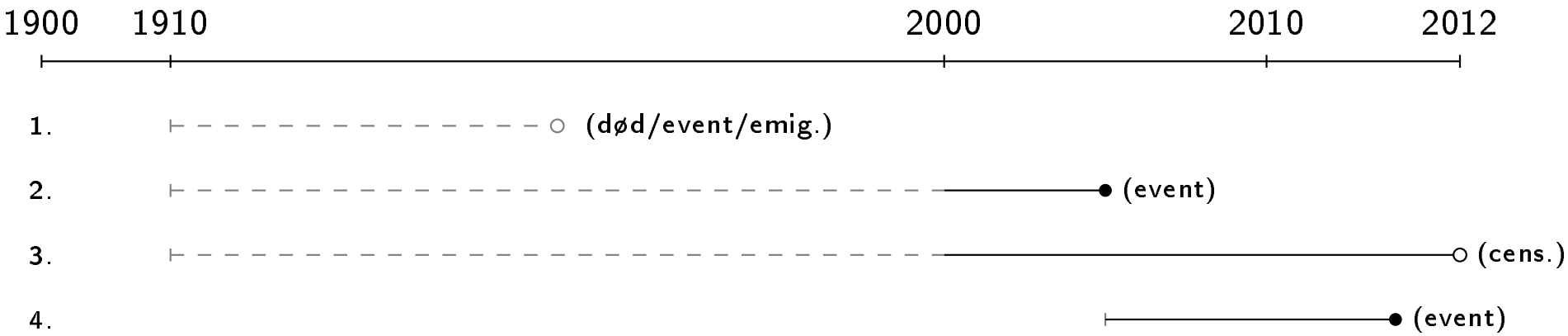
Table 2. Lifetime Risk and Cumulative Incidence at 50 Years of Age for All Psychiatric Disorders

Psychiatric Disorder	*ICD-10-DCR* Code	Cumulative Incidence (95% CI), %[a]		Lifetime Risk (95% CI), %[b]	
		Males	Females	Males	Females
Any psychiatric disorder	F00-F99	22.60 (22.49-22.73)	25.26 (25.13-25.39)	32.05 (31.91-32.19)	37.66 (37.52-37.80)
Organic, including symptomatic, mental disorders	F00-F09	0.41 (0.40-0.43)	0.26 (0.24-0.27)	8.84 (8.74-8.95)	11.98 (11.87-12.09)
Dementia in Alzheimer disease	F00	0.01 (0.01-0.01)	0.01 (0.01-0.02)	3.20 (3.13-3.27)	5.14 (5.06-5.22)
Vascular dementia	F01	0.01 (0.01-0.01)	0.01 (0.00-0.01)	1.85 (1.80-1.90)	2.01 (1.96-2.06)
Mental and behavioral disorders due to psychoactive substance abuse	F10-F19	5.96 (5.89-6.03)	2.95 (2.90-3.00)	7.79 (7.71-7.87)	4.49 (4.43-4.55)
Mental and behavioral disorders due to alcohol use	F10	3.52 (3.47-3.58)	1.76 (1.72-1.80)	5.28 (5.22-5.35)	2.96 (2.91-3.01)
Mental and behavioral disorders due to cannabis use	F12	2.12 (2.08-2.17)	0.77 (0.74-0.79)	2.20 (2.15-2.24)	0.79 (0.76-0.82)

Studiedesign

Deres studiedesign kan kort beskrives ved følgende:

- Event/outcome: første psyk. diagnose (i register).
- Population: individer født 1900–2010.
- Delayed entry: i risiko ved fødsel eller i 2000.
- Fast end of follow-up dato i slut 2012.
- Competing events: død, emigration.
- Tidsskala: alder.



Den statistiske analyse

Per design har vi altså time-to-event data, dvs. hver observation er et par (T, Δ), hvor T er alderen hvor et event indtræffer og Δ indikerer hvilket.

For at opnå et estimat af den kumulerede incidens anvendes typisk Kaplan–Meier estimatoren

$$\widehat{\mathrm{CI}}_{\mathrm{KM}}(t) = 1 - \prod_{j:s_j \leqslant t} \frac{Y(s_j) - d(s_j)}{Y(s_j)},$$

eller dets competing risk pendant; Aalen–Johansen estimatoren

$$\widehat{\mathrm{CI}}_{\mathrm{AJ}}(t) = \sum_{j:s_j \leqslant t} \frac{\widehat{S}(s_j-)}{Y(s_j)} \mathbf{1}_{D_j=1}.$$

Begge disse estimatorer er unbiased for den kumulurede incidens $P(T \leqslant t)$ **hvis** censurerings- og delayed-entry mekanismerne er *uafhængige*.

Interpretation

- For "any psychiatric disorder" estimerede de

$$\text{livstidsrisiko} = \widehat{\mathrm{CI}}_{\mathrm{AJ}}(100 \text{ år}) \approx 35\%.$$

- Betyder det, at den gennemsnitlige risiko for en psykiatrisk diagnose gennem livet for individer født 1900–2010 er cirka 35%?

- Ja, **hvis** vi tror på at censurerings- og delayed-entry mekanismerne virkeligt er uafhængige. Hvis ikke, så siger dette estimat intet brugbart eller interessant om den sande kumulerede incidens i vores population.

Uafhængighedsantagelsen

Vi kan tænke på den føromtalte uafhængighedsantagelse som en repræsentativitetsantagelse:

Antagelse

Individer som er i risiko skal være repræsentative for den uncensurerede stikprøve med hensyn til deres instantane (tænk hazard rate) sygdomsrisiko til enhver given alder.

I dette eksempel, kan vi ikke rigtig tjekke denne antagelse da det vil kræve information vi ikke har.

Vi kan og bør dog stadig diskutere hvor plausibel antagelsen er.

Uafhængighedsantagelsen

Lad os derfor fokusere på alderen 80 år:

- Det er kun individer født i perioden 1920–1932 der er i risiko ved deres 80 års fødselsdag.
- Tror vi, at individer født i denne periode er repræsentative for eksempelvis din og min risiko for en psykiatrisk diagnose omkring vores 80 års fødselsdag?
- Mon ikke et individ født i perioden 1920–1932 har en stor risiko for en psykiatrisk diagnose omkring deres 80 års fødselsdag ift. den tilsvarende risiko for du og jeg?
- Det virker meget usandsynligt grundet kalendertids-trends:
  - diagnostik har ændret sig betydeligt over tid
  - nationale registre har kun været i brug siden 60'erne.

Take home message

- I studier med delayed-entry og/eller censoring samt kalendertids-trends vil uafhængighedsantagelsen ofte ikke være opfyldt.
- Det betyder specielt, at de sædvanlige Kaplan–Meier og Aalen–Johansen estimatorer ikke er gyldige estimatorer for den kumulerede incidens.
- Hvad kan man så estimere i et sådant setup?
  - Man kan i hvert fald ikke estimere eksempelvis livstidsrisiko medmindre man er villige til at gøre sig nogle kraftige antagelser omkring tids-trenden.
- Se kommende artikel i BMC Medical Research Methodology af Hansen et al. for muligheder.

Tak.

The pseudo-observation method

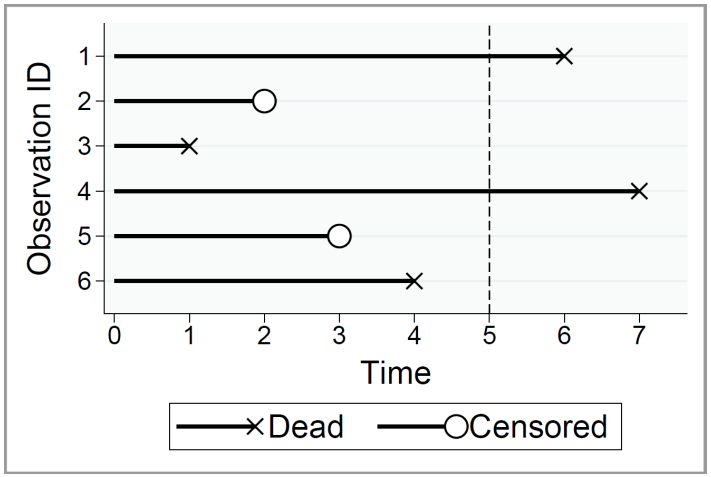
Morten Overgaard

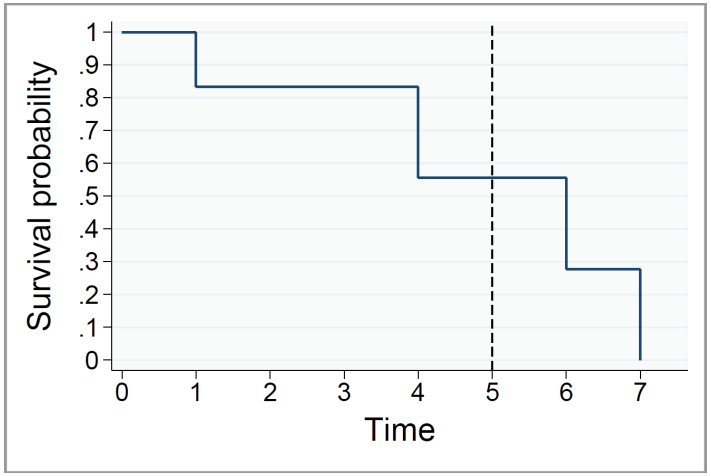
Section for Biostatistics

December 1, 2016

- Allows for a regression of possibly missing outcomes on the covariates of interest.
- Designed with time-to-event analysis and right-censored outcomes in mind.
- Can be used for estimation of e.g. risk difference, relative risk, and odds ratio.

Survival data and the Kaplan–Meier estimator





ID	Status at time 5
1	1
2	???
3	0
4	1
5	???
6	0

Time	Survival probability
1	0.83
2	0.83
3	0.83
4	0.56
6	0.28
7	0.00

The Kaplan–Meier estimate estimates the overall probability of survival. **How does the probability depend on covariates?**

- Most often a Cox regression is used to answer such questions.
- It relies on an assumption of proportional hazards.
- It estimates the hazard ratio of dying related to a certain comparison (e.g. male vs. female).

The hazard ratio may be difficult to interpret, and we are probably more interested in an estimate of the relative risk or the risk difference!

The pseudo-observations and how to use them

If $\hat{\theta}_n$ is the Kaplan–Meier estimate (at time 5) on the entire sample of size n and $\hat{\theta}_n^{(i)}$ is the estimate on the sample excluding the ith observation, **the jack-knife pseudo-observation** for observation i is defined as

$$\hat{\theta}_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_n^{(i)}.$$

- Related to observation i's influence on the estimate.
- Available for any observation in the dataset.

Observation ID	1	2	3	4	5	6
Status at time 5	1	???	0	1	???	0
Pseudo-observation	1.33	0.67	0.00	1.33	0.67	-0.67

The pseudo-observations and how to use them

The claim is:
That the ith pseudo-observation holds detailed information about observation i's probability of survival.

The main idea of the pseudo-observation method:
In estimating procedures, use the pseudo-observations as if they were the outcome.

What my project is about

- Looking at other outcomes than survival and other estimators than the Kaplan–Meier.
  - E.g. competing risks.
- Studying why and when the method makes sense.
  - Rather strict conditions on the censoring mechanism must be imposed when the method is based on an estimator such as the Kaplan–Meier.
- Estimating the variance of the parameter estimates.
  - Often a bit smaller than what was initially thought to be the case.
- Implementation of the method in Stata.
  - Check out `stpsurv` and related commands.