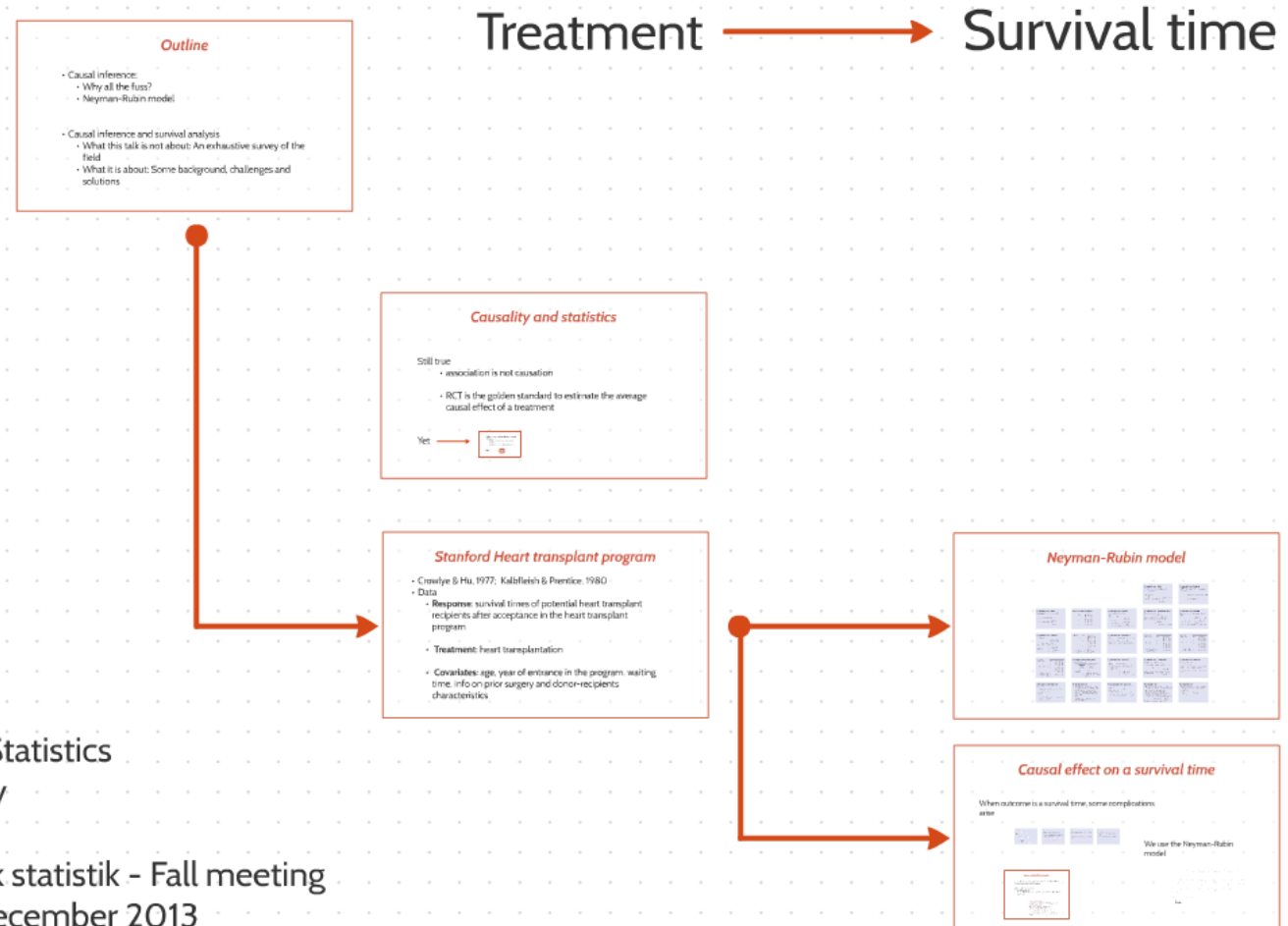


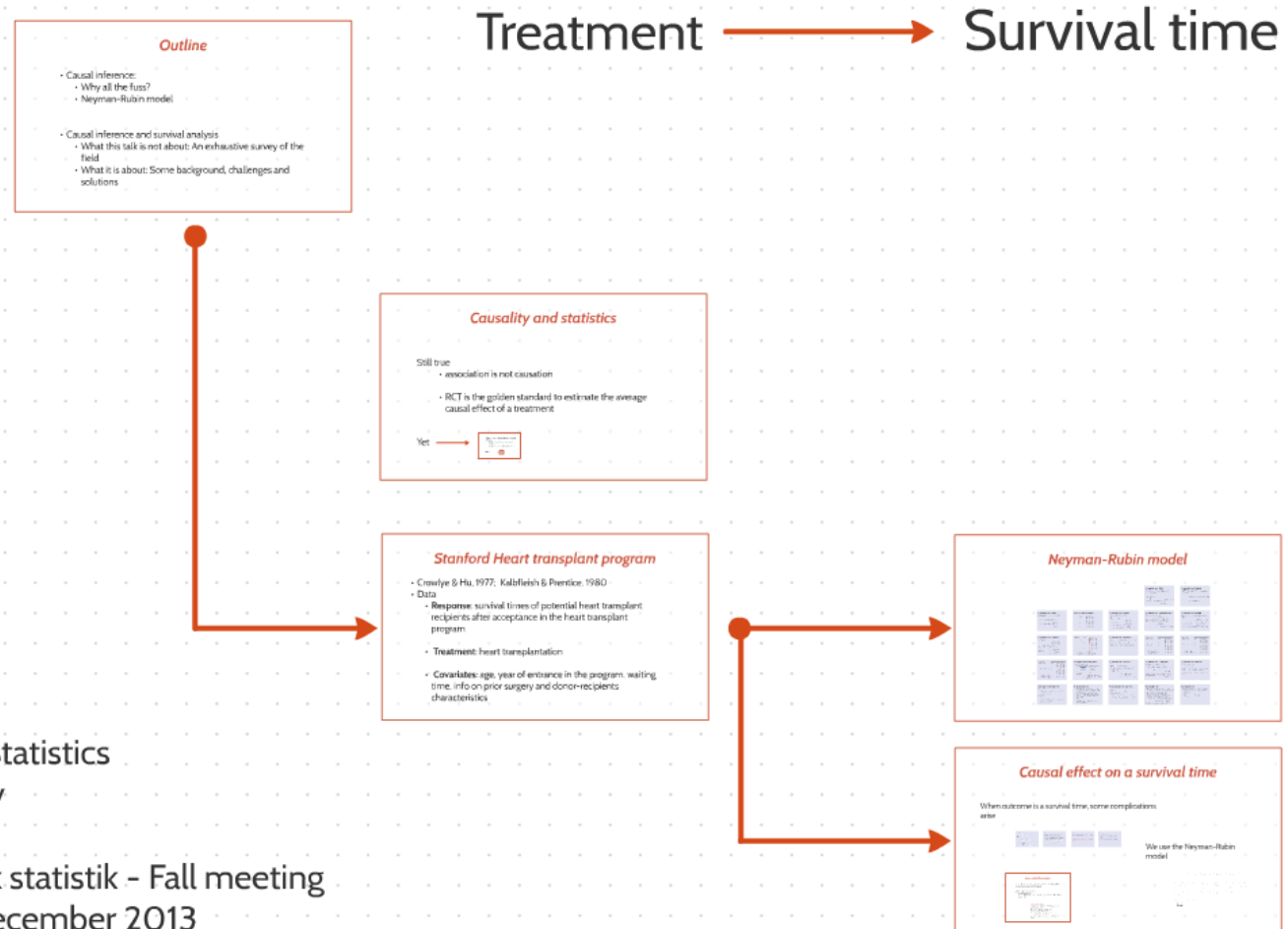
Causal inference and survival analysis



Xavier de Luna, Dept of Statistics
USBE @ Umeå University

Föreningen för medicinsk statistik - Fall meeting
AstraZeneca, Mölndal, December 2013

Causal inference and survival analysis



Xavier de Luna, Dept of Statistics
USBE @ Umeå University

Föreningen för medicinsk statistik - Fall meeting
AstraZeneca, Mölndal, December 2013

Outline

- Causal inference:
 - Why all the fuss?
 - Neyman-Rubin model
- Causal inference and survival analysis
 - What this talk is not about: An exhaustive survey of the field
 - What it is about: Some background, challenges and solutions



Causality and statistics


Still true

- association is not causation
- RCT is the golden standard to estimate the average causal effect of a treatment

Yet



• Explosion of scientific publications on causal inference:
• JSM-2002 had 13 papers on causal inference
• JSM-2012 had 73, JSM-2013 had 102

• WHY? → 

- Explosion of scientific publications on causal inference
 - JSM-2002 had 13 papers on causal inference
 - JSM-2012 had 73, JSM-2013 had 102

• WHY?



Why?

- Languages for causal reasoning have been developed, so association and causality can be disentangle
- RCT has its limitations (efficacy)
- Lots of observational data out there (efficiency)

Why?

- Languages for causal reasoning have been developed; so association and causality can be desintangle
- RCT has its limitations (efficacy)
- Lots of observational data out there (efficiency)

Stanford Heart transplant program

- Crowlye & Hu, 1977; Kalbfleish & Prentice, 1980
- Data
 - **Response:** survival times of potential heart transplant recipients after acceptance in the heart transplant program
 - **Treatment:** heart transplantation
 - **Covariates:** age, year of entrance in the program, waiting time, info on prior surgery and donor-recipients characteristics

Neyman-Rubin model

Neyman inference: sample

We have a sample (observed and not observed) of n individuals

- Green: treated individuals (with $T=1$)
- Red: control individuals (with $T=0$)

Observed status of variables

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Observed status of variables

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Neyman inference: Notation

Observed: $y(0), y(1)$ and $y(1) - y(0)$

- Green: observed response
- Red: treated
- Green: control

Randomized assignment: $Y(0) = y(0) + Z \cdot (y(1) - y(0))$

$y(1) - y(0)$ is part of factor (observed response)

Neyman inference: Model

- Randomized assignment: $Y(0), Y(1)$ are i.i.d.
- Treatment assignment: $Z \sim \text{Bernoulli}(\tau)$
- Potential outcomes: $y(0), y(1)$ are i.i.d.
- Observed outcomes: $y(0) + Z \cdot (y(1) - y(0))$

Neyman inference: Estimand

What is the causal effect to be identified?

Which causal quantities we can estimate the following randomized trial data?

Average causal effect (ACE)

by randomization

Neyman inference: estimand

Average causal effect

$$E[y(1) - y(0)]$$

Estimate of the causal effect

$$\frac{1}{n} \sum_{i=1}^n (y_i(1) - y_i(0))$$

Average causal effect τ estimated to be different

$$\tau = E[y(1) - y(0)]$$

What is different?

Answer:

Observed status of variables

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Neyman inference: random mess

- Consider the columns of $y(0)$ and $y(1)$ as given for each individual.
- Number of random messes (other than treatment assignment) τ .
- The randomization procedure of the estimator is applied to arbitrary randomization (not necessarily the same as the one used for randomization).

Estimation of causal effects

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Estimation of causal effects

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Source of randomness

Randomized assignment

Obs	$Y(0)$	$Y(1)$	Z
1	1	0	Obs
2	1	1	Obs
3	0	1	Obs
4	1	0	Obs
5	0	1	Obs
6	1	1	Obs
7	0	0	Obs
8	1	1	Obs
9	0	1	Obs
10	1	1	Obs

Reassign treatment many times!

Randomization

Use the $Y(0)$ and $Y(1)$ columns as given for each individual.

Neyman inference: properties

- We have τ level of randomization.
- First level randomization for each individual (observed).
- Randomization: $Z \sim \text{Bernoulli}(\tau)$.
- Random assignment: $Y(0) = y(0) + Z \cdot (y(1) - y(0))$.

Neyman inference: Assumptions

The randomization assumptions were made.

Assumptions (randomization)

$$E[Z] = \tau$$

Should we also assume that the values of $y(0)$ and $y(1)$ are given sufficient and not affected by the values taken by Z in our data analysis?

Neyman inference: comments

- Randomized assignment
- This is often easier to randomize on response. It is important that the randomization is done before the outcome is measured.
- How can we make it more efficient? Randomized trial randomization

Other frameworks of inference

- Frequentist inference
 - The causal effect is a fixed value (parameter to be estimated)
 - Randomized assignment
- Bayesian inference
 - Random assignment is not needed
 - Random assignment is needed, but not necessarily randomized
 - Random assignment is needed

Frequentist inference

- Random assignment is not needed
- Random assignment is needed, but not necessarily randomized
- Random assignment is needed

Frequentist inference: properties

- Randomization: $Z \sim \text{Bernoulli}(\tau)$.
- Random assignment: $Y(0) = y(0) + Z \cdot (y(1) - y(0))$.

Bayesian inference

- We do not have a prior distribution on the causal effect.
- Random assignment is not needed
- Random assignment is needed, but not necessarily randomized
- Random assignment is needed

Bayesian inference

- Random assignment is not needed
- Random assignment is needed, but not necessarily randomized
- Random assignment is needed

Neyman inference: Model

Potential outcomes: Neyman (1923), Rubin (1974).

★ Treatment assignment:


$z = 1$ for a treated individual,
 $z = 0$ when not treated.

★ Potential outcomes:

$y(1)$ outcome if treated,
 $y(0)$ outcome if not treated.

★ Causal effect at individual level:

Cannot be observed!


$$y(1) - y(0)$$

Neyman inference: Estimand

- ★ Which causal effect can be identified?
- ★ Under certain assumptions we may retrieve the following estimand from data:

$$\tau = E(y(1) - y(0))$$

Average Causal Effect (ACE)

for a given population

Neyman inference: sample

You have a sample (does not need to be random) of n individuals:

★ n_t treated individuals for which we observe:

$$y(1), z, x$$

★ n_c control individuals for which we observe:

$$y(0), z, x$$

Observed status of variables

Unit	z	$y(1)$	$y(0)$	x
1	1	Obs	Mis	Obs
2	1	Obs	Mis	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	1	Obs	Mis	Obs
1	0	Mis	Obs	Obs
2	0	Mis	Obs	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_c	0	Mis	Obs	Obs

Neyman inference: Notation

Denote: $y_i(1) = y_i^1$ and $y_i(0) = y_i^0$

We observe two groups:

★ treated:

$$y_1^1, y_2^1, \dots, y_{n_t}^1$$

★ controls:

$$y_1^0, y_2^0, \dots, y_{n_c}^0$$

Treatment assignment not random.

~~$y(1), y(0) \perp\!\!\!\perp z$~~

$\bar{y}^t - \bar{y}^c$ is not of interest (does not estimate τ)

Neyman inference: Unconfoundedness

If treatment z is randomized we have:

$$y(1), y(0) \perp\!\!\!\perp z$$

In an observational study this does typically not hold.

In some cases there may exist given a set of covariates \mathbf{x} s.t.:

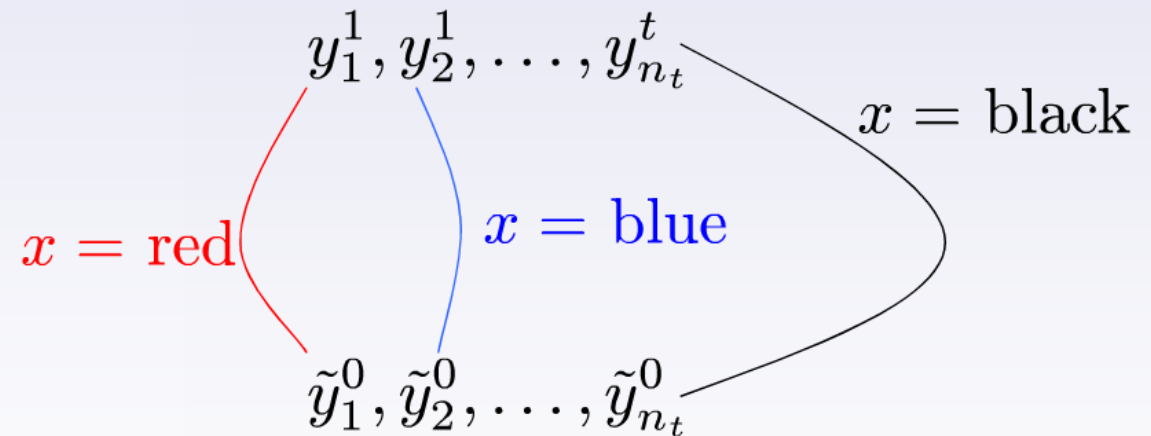
$$y(1), y(0) \perp\!\!\!\perp z | \mathbf{x}$$

[Unconfoundedness assumption]

Neyman inference: matching

Hence, construct a new control group which is comparable with the treated:

★ treated:



★ matched controls:

\tilde{y}_j^0 is a control individual which has same/similar \mathbf{x} than y_j^1 .

Neyman inference: estimand

A matching estimator:

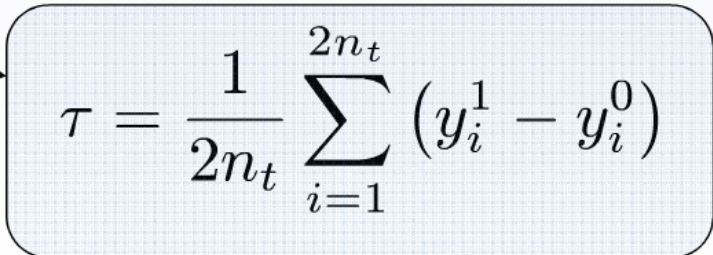
$$\hat{\tau} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_i^1 - \frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{y}_i^0$$

Estimator of what estimand?

Average causal effect τ , estimand to be defined:

$$\tau = E(y(1) - y(0))$$

What is that?


$$\tau = \frac{1}{2n_t} \sum_{i=1}^{2n_t} (y_i^1 - y_i^0)$$

Answer:

Unit	z	$y(1)$	$y(0)$	x
1	1	Obs	Mis	Obs
2	1	Obs	Mis	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	1	Obs	Mis	Obs
1	0	Mis	Obs	Obs
2	0	Mis	Obs	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	0	Mis	Obs	Obs

Matched pair

$$\tau = \frac{1}{2n_t} \sum_{i=1}^{2n_t} (y_i^1 - y_i^0) = \frac{1}{2n_t} \left(\sum_{i=1}^{2n_t} y_i^1 - \sum_{i=1}^{2n_t} y_i^0 \right)$$

Neyman inference: randomness

- Consider the outcomes $y(1)$ and $y(0)$ as given for each individuals.
- Source of randomness is then the treatment assignment z
- The sampling distribution of the estimator is obtained by randomly reassigning treatment with the constraint that within each matched pair both treatment ($z=1$) and non-treatment ($z=0$) arise.

Source of randomness

Observations and the resulting estimator

$$\hat{\tau} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_i^1 - \frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{y}_i^0$$

Unit	z	$y(1)$	$y(0)$	x
1	1	Obs	Mis	Obs
2	1	Obs	Mis	Obs
3	1	Obs	Mis	Obs
4	1	Obs	Mis	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	1	Obs	Mis	Obs
1	0	Mis	Obs	Obs
2	0	Mis	Obs	Obs
3	0	Mis	Obs	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	0	Mis	Obs	Obs

Source of randomness

Reassigning treatment randomly

$$\hat{\tau} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_i^1 - \frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{y}_i^0$$

Unit	z	$y(1)$	$y(0)$	x
1	0	Mis	Obs	Obs
2	1	Obs	Mis	Obs
3	1	Obs	Mis	Obs
4	0	Mis	Obs	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	0	Mis	Obs	Obs
1	1	Obs	Mis	Obs
2	0	Mis	Obs	Obs
3	0	Mis	Obs	Obs
\vdots	\vdots	\vdots	\vdots	\vdots
n_t	1	Obs	Mis	Obs

Source of randomness

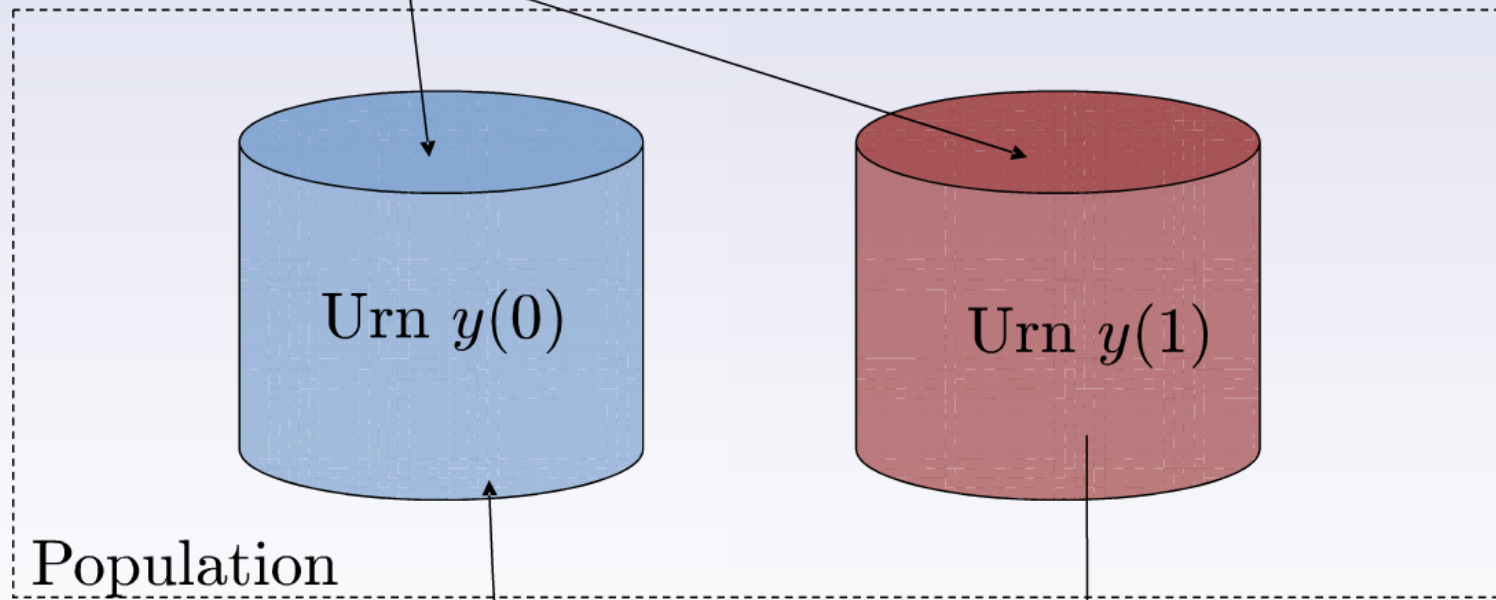
Reassigning treatment randomly and the resulting estimator

$$\hat{\tau} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_i^1 - \frac{1}{n_c} \sum_{i=1}^{n_c} \tilde{y}_i^0$$

Unit	z	$y(1)$	$y(0)$	x
1	0	Mis	Obs	Obs
2	1	Obs	Mis	Obs
3	1	Obs	Mis	Obs
4	0	Mis	Obs	Obs
⋮	⋮	⋮	⋮	⋮
n_t	0	Mis	Obs	Obs
1	1	Obs	Mis	Obs
2	0	Mis	Obs	Obs
3	0	Mis	Obs	Obs
⋮	⋮	⋮	⋮	⋮
n_c	1	Obs	Mis	Obs

Reassign treatment many times!

units $1, \dots, n_t$ and $1, \dots, n_t$



sample without replacement n_t
units and delete the corresponding
units from

Neyman inference: properties

★ We have 2^{n_t} possible randomizations.

★ Over these randomizations we have (Neyman, 1923):

▶ Unbiasedness:

$$E(\hat{\tau}) = \tau$$

▶ Variance estimator:

$$\widehat{Var}(\hat{\tau}) = \frac{1}{n_t} \sum_{i=1}^{n_t} \{(y_i^1 - y_{i+n_t}^0) - \hat{\tau}\}^2$$

(unbiased if additive constant treatment effect)

Neyman inference: Assumptions

Unconfoundedness assumption was made.

Another identifying assumption used in this framework is:

$$0 < \Pr(z = 1|\mathbf{x}) < 1$$

[common support]

Finally we also assume that the values $y(1)$ and $y(0)$ for a given individual are not affected by the values taken by z for any other individual.

[SUTVA]

Neyman inference: comments

- In this inferential framework:

Population = Sample

- This is often relevant in studies based on registries. In such cases it is often non-trivial to think of the sample as drawn randomly from super-population (often difficult to define).
- How can such results be generalized? Prediction? Only historical value?

Other frameworks of inference

- Frequentist inference
 - The sample is randomly drawn from a population (often an ill-defined super-population)
 - Otherwise often practical
- Bayesian inference
 - Population concept is not needed
 - However, strong assumptions are needed: exchangeability and a parametric model for $f(y/x)$.
 - Computationally demanding

Stanford Heart transplant program

- Crowlye & Hu, 1977; Kalbfleish & Prentice, 1980
- Data
 - **Response:** survival times of potential heart transplant recipients after acceptance in the heart transplant program
 - **Treatment:** heart transplantation
 - **Covariates:** age, year of entrance in the program, waiting time, info on prior surgery and donor-recipients characteristics

Causal effect on a survival time

When outcome is a survival time, some complications arise

Note 1: Control group must include treated 	Case of randomized treatment - Assume that there is a random, common, unknown, cause of death, which is not related to the treatment or to the control group, and that the treatment is not related to the control group. 	Observed treatment - Assume that the treatment is not related to the control group, and that the treatment is not related to the control group. 	Censoring - We do not observe the survival time of the control group, and we observe the survival time of the treated group.
---	---	---	--

We use the Neyman-Rubin model

Some concluding remarks

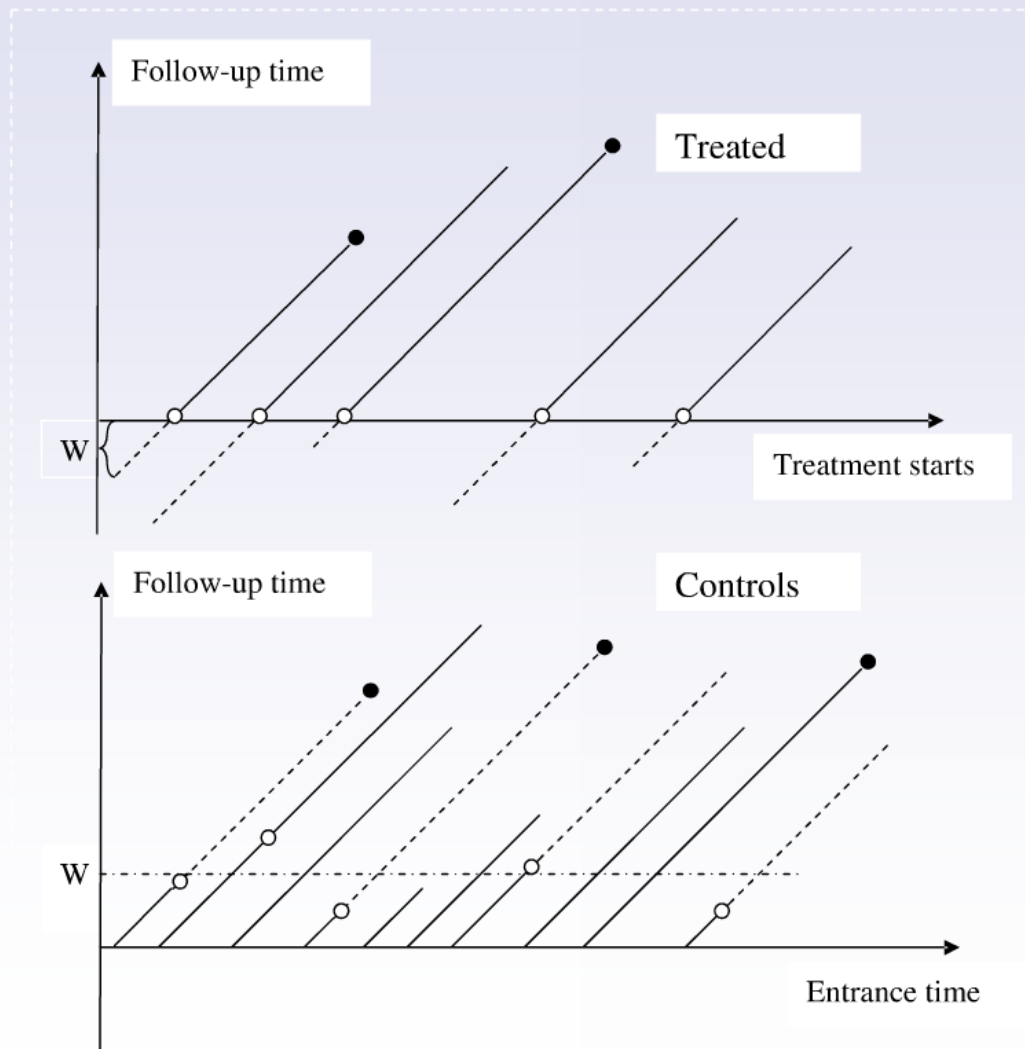
- Causal inference in observational studies: Protocols defining population, treatment assignment and control group
- With population wide registers
 - Sample is population
 - Large control groups and rich set of background characteristics allow for good designs

Some references

Robins JM, Greenland S, Hernan MA. Causal inference in epidemiology. *Journal of the Royal Society Series B*. 2004;66(2):102-130.
Robins JM, Greenland S, Hernan MA. Causal inference in epidemiology. *Journal of the Royal Society Series B*. 2004;66(2):102-130.
Robins JM, Greenland S, Hernan MA. Causal inference in epidemiology. *Journal of the Royal Society Series B*. 2004;66(2):102-130.

Note 1: Control group must include treated

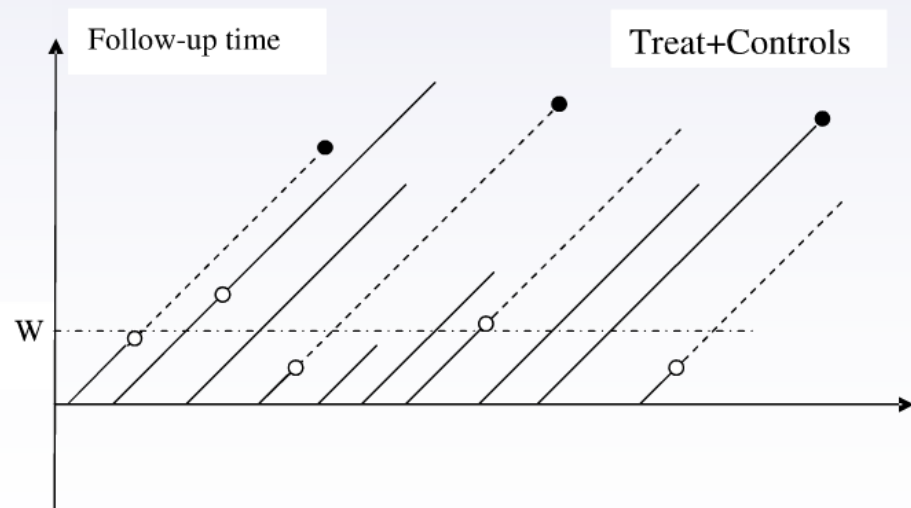
Lexis diagrams



Case of randomized treatment

- Assume each time a heart is available, a patient is randomly chosen.
- In contrast with usual studies, treated and controls cannot be directly compared: on average, survival time of a treated after transplantation is shorter than survival time of a control

Note 2: Inference must be conditioned on waiting time.



Observed treatment

- Among those having a given waiting time, match for covariates affecting response and treatment.

Note 3: For a given waiting time, conditional on the covariates, the treatment can be considered as randomized. (unconfoundedness assumption)

Censoring

- Note 4: Patient's survival is censored in two ways:
 - end of study, drop out, etc. ; independent mechanism (assumption)
 - controls may receive treatment (need of an extra assumption)

We use the Neyman-Rubin model

Theoretical framework

Randomization: Neyman (1934), Fisher (1935)

Randomized in our context: for an individual within treatment or control, the randomization is the mechanism that determines the treatment assignment.

$Z^*(N)$ – randomization for the whole N if treated or not

$Z^*(N)$ – randomization for N if not treated or not

Further, consider

$$Z^*(N) = \begin{cases} 1 & \text{if treated in group } N \\ 0 & \text{if not treated in group } N \end{cases}$$

Observed status of the variables

As an example, let $N = 0$, then:

patient id	$Z(0)$	$T(0)$	$Y(0)$
111	0	0.0	0.000
112	0	0.0	0.0
113	0	0.0	0.000
114	1	0.0	0.0
115	1	0.0	0.0
116	1	0.0	0.0
117	1	0.0	0.0
118	1	0.0	0.0
119	1	0.0	0.0

Note: NA for non-observable. (Can be observed at time t , but not treated at time t)

Assumptions

A $Z^*(N)$ The value $Z^*(N)$ and $T^*(N)$ for a given individual are not affected by the value taken by $Z^*(N)$ or any other individual.

B $Z^*(N)$ is independent of the potential outcomes, conditional on the set of potential characteristics X , i.e.,

$$Z^*(N) \perp\!\!\!\perp (Y(0), Y(1)) \mid X, N$$

C $Z^*(N)$ is a function of (X, N) only.

D $Z^*(N)$ is a function of (X, N) only.

Estimands

We can now define a few estimands:

- $\frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$, where i is the number of patients treated in waiting time N . This is the average treatment effect for the treated patients in the study (waiting time N).
- $\frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$, i.e. the average treatment effect for treated patients and the impact in the control group for a given waiting time.

Results

The average treatment effect for a given patient exposed to treatment, when treatment is assigned with randomization, is equal to the difference in the average treatment effect for a given patient in each treatment in a given (waiting) waiting time N .

We go for B.

Inference (if no censoring...)

Without censoring we could estimate δ with

$$\hat{\delta}(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$$

where $Z^*(N)$ is a vector for Z , and $Z = 1$ means...
 Note that in the usual $Z = Z^*$

Sampling distribution

Sampling distribution of the estimator can be traced under random assignment, for fixed values of $Z^*(N)$ and $T^*(N)$, and conditioning on treatment and its effect (strong, strong treatment assignment).

This (Theorem 1.0.2) is the estimator **unbiased** and its variance can be computed.

Censoring due to treatment

Let $T^*(N)$ denote time to treatment for an individual not treated in N .

Conditional $T^*(N)$ is constant value $T^*(N) = T^*(N)$.

Assumption 3:

$$T^*(N) \perp\!\!\!\perp (Y(0), Y(1)) \mid X, N$$

End of study, drop out, etc.

Let $T^*(N)$ be the time to censoring. Let us assume that treatment value individual has been censored at $T^*(N)$.

Conditional $T^*(N)$ is constant value $T^*(N) = T^*(N)$.

Assumption 4: $T^*(N)$ is independent of $T^*(N)$ and $T^*(N)$ when conditioning on X .

Note: censoring $T^*(N) = 0.1$ denotes time to death (not censoring).

Hazards

New estimator: $\hat{\delta}_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$

$$\hat{\delta}_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$$

for $j = 0, 1$.

Letting estimator $\hat{\delta}_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$

where $\hat{\delta}_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$

for $j = 0, 1$.

Results

Missing estimator of the hazards are unbiased under the sampling distribution defined earlier. The variance can be estimated with:

$$V(\hat{\delta}_j(N)) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))^2$$

The variance of the estimator is generally biased (Coxeter and Hertzog, 2003, p. 11, 12, 13).

Survival functions

Denote by $S_j(t) = P(Y_j(t) = 1)$ the survival function of the survival function $S_j(t) = P(Y_j(t) = 1)$, where t is the time to event and $S_j(t)$ is the survival function.

$$S_j(t) = \prod_{s=0}^{t-1} (1 - H_j(s))$$

The estimator of interest is the difference in survival function:

$$\hat{\delta}_j(N) = S_j(t) - S_j(t)$$

Inference

An estimator of $\delta_j(N)$ is obtained by replacing the $Y_j(t)$ by their estimator, and the Fisher and Fisher (1935) type estimator.

The asymptotic variance of the Kaplan-Meier estimator is obtained with the Greenwood (1926) formula.

Estimator expressed by the estimator when the treatment effect is not easy to interpret (stability needed).

A Kaplan-Meier estimator is a **MLE** (maximum likelihood estimator) based on the Greenwood estimator (Kaplan-Meier estimator).

Averaging over waiting times

What two estimators are the central point estimate conditional on N ?

This can be seen as the **observed waiting times**, however, interpretation of survival function problematic, where N is the number of waiting times and N is the number of waiting times.

Heart transplant program

Estimating $\delta_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$

Employment subsidy program

Forward, labor market and welfare (2004)

- Swedish employment subsidy for the long-term unemployed – 50% of total wage costs to job for 6 months.
- Requires unemployment benefits (paid to unemployed).
- Condition: applicants must be registered unemployed.

Results

Results of the registered unemployed at least 12 months in unemployment:

- Program area: 40,000, 400,000, 1% reduction in unemployment.
- 400,000 applicants, 2000 individuals in the control group.

Exact one-to-one matching

Estimating $\delta_j(N) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$

Concluding slide

- Large observational studies show us to relate with the intervention, e.g. propensity score matching, probability of treated, propensity of treated with waiting time.
- To show that the effect is generalizable, estimator of a treatment effect on a random outcome.
- Randomized trials.
- Fisher (1935)
- Fisher and Neyman (1936)
- Fisher (1935), Fisher, Bradford Hill, and Rubin (2002), Rubin and van der Vaart (2003), Rubin and van der Vaart (2003)



Theoretical framework

Potential outcomes: Neyman (1923,1990), Rubin (1974).

Adaptation to our context: For an individual which has spent at least time W in the study without being treated, we define two potential outcomes:

$T^1(W)$ = survival time after time W if treated at W ,

$T^0(W)$ = survival time after W if neither treated at W nor later.

Further, consider

$$D(W) = \begin{cases} 1 & \text{if treated at time } W, \\ 0 & \text{if not treated at time } W. \end{cases}$$



Observed status of the variables

As an example, let $W = 21$ days:

patient ident.	$D(21)$	$T^1(21)$	$T^0(21)$
101	0	NA	$C@10$
66	0	NA	21
4	0	NA	$T@15$
47	1	51	NA
97	1	$C@110$	NA
58	1	321	NA

Note: NA for non-available; $C@t$ for censored at time t ; $T@t$ for treated at time t .



Censoring due to treatment

- Let $C^T(W)$ denote time to treatment for an individual not treated at W
- Convention: $T^0(W)$ is censored when $C^T(W) < T^0(W)$
- **Assumption D:**

For $i < t_0$,

$$\Pr(C^T(W) = i | \mathbf{X}, T^0(W) = t^0) = g(\mathbf{X}, W)$$



End of study, drop out, etc.

- Let $C^E(W)$ be the time to censoring (other reasons than treatment) when individual has survived until W
- Convention: Survival time censored when $C^E(W) < T^0(W)$ or $C^E(W) < T^1(W)$.
- **Assumption E:** $C^E(W)$ is independent of $T^0(W)$ and $T^1(W)$ when conditioning on \mathbf{X} .
- New notation: $T^j(W)$, $j = 0, 1$ denotes time to death OR censoring.



Hazards

New estimand: $\Delta_h(t; W) = h^1(t; W) - h^0(t; W)$,
where

$$h^j(t; W) = \frac{\sum_{i=1}^{2n_1} I(T_i^j(W) = t)}{\sum_{i=1}^{2n_1} I(T_i^j(W) \geq t)}$$

for $j = 0, 1$.

Matching estimator: $\widehat{\Delta}_h(t; W) = \widehat{h}^1(t; W) - \widehat{h}^0(t; W)$,
where

$$\widehat{h}^j(t; W) = \frac{\sum_{i:D=1} I(T_i^j(W) = t)}{\sum_{i:D=1} I(T_i^j(W) \geq t)},$$

for $j = 0, 1$.



Results

Matching estimator of the hazards are **unbiased** under the sampling distribution defined earlier. The **variance** can be estimated with

$$\begin{aligned} \widehat{Var} \left(\widehat{\Delta}_h(t; W) \right) \\ = \frac{\widehat{h}^1(t; W)(1 - \widehat{h}^1(t; W))}{\sum_{i:D=1} I(T_i^1 \geq t) - 1} + \frac{\widehat{h}^0(t; W)(1 - \widehat{h}^0(t; W))}{\sum_{i:D=1} I(T_i^0 \geq t) - 1}. \end{aligned}$$

The estimator of the variance is positively biased (conservative inference) unless $T_i^1 = T_i^0$ for $i = 1, \dots, 2n_1$



Survival functions

- Denote by $T_{(1)}^j(W) \leq T_{(2)}^j(W) \leq \dots \leq T_{(m_j)}^j(W)$ the $m_j \leq 2n_1$ not censored survival times if untreated/treated ($j = 0, 1$), sorted in ascendant order, and define the survival functions:

$$F^j(t; W) = \prod_{i: T_{(i)}^j < t} (1 - h^j(T_{(i)}^j(W); W))$$

- The estimand of interest is the difference in survival functions

$$\Delta_s(t; W) = F^1(t; W) - F^0(t; W).$$



Inference

- An estimator of $\Delta_s(t; W)$ is obtained by replacing the hazards by their estimators, yielding Kaplan and Meier (1958) type estimators
- The asymptotic variance of the Kaplan-Meier estimator is obtained with the Greenwood's (1926) formula
- Inference expected to be conservative when the treatment effect is not zero (unit-treatment additivity sense)
- A simulation study shows that a Wald test based on the Greenwood's variance has fairly good properties (size and power)



Averaging over waiting times

When few observations, you cannot perform inference conditional on W .

Then, we can **average over the observed waiting times**.

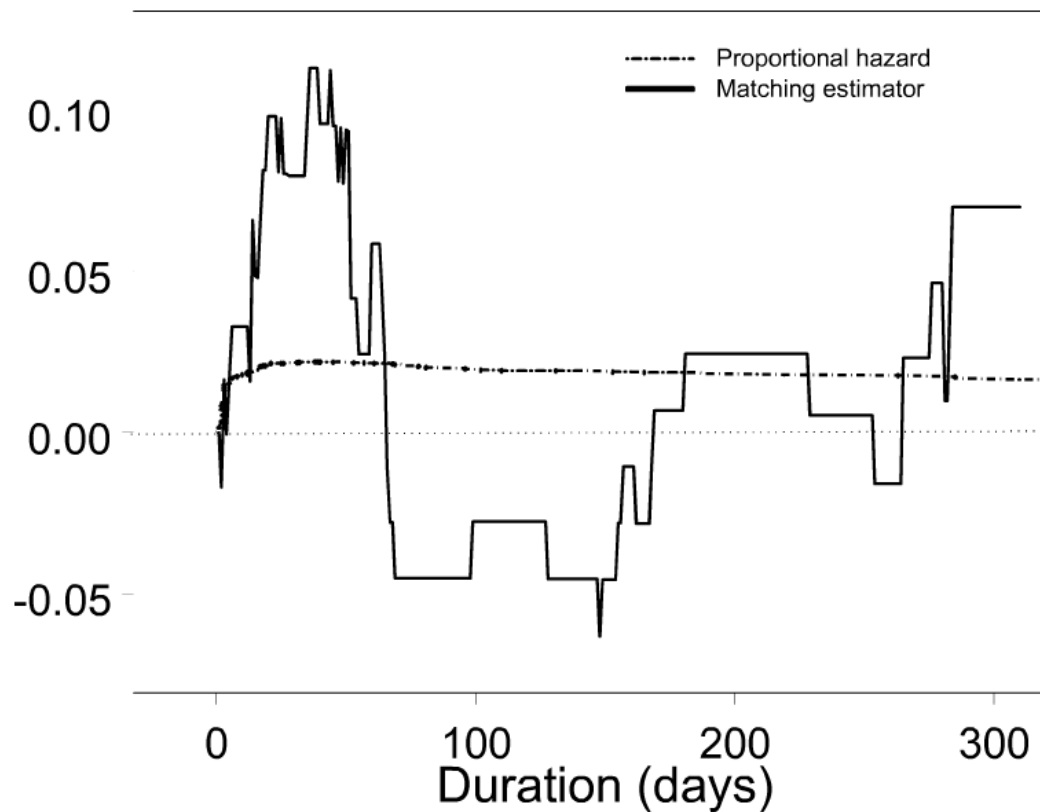
However: Interpretation of survival functions problematic.

Inference problematic unless few treated and many controls.



Heart transplant program

Estimating $F^1 - F^0$:





Employment subsidy program

Forslund, Johansson and Lindqvist (2004)

- Treatment: employment subsidy for the long-term unemployed – 50% of total wage costs is paid for 6 months
- Response: Unemployment duration (time to employment)
- covariates:
age, sex, "disability", citizenship, education, unemployment history

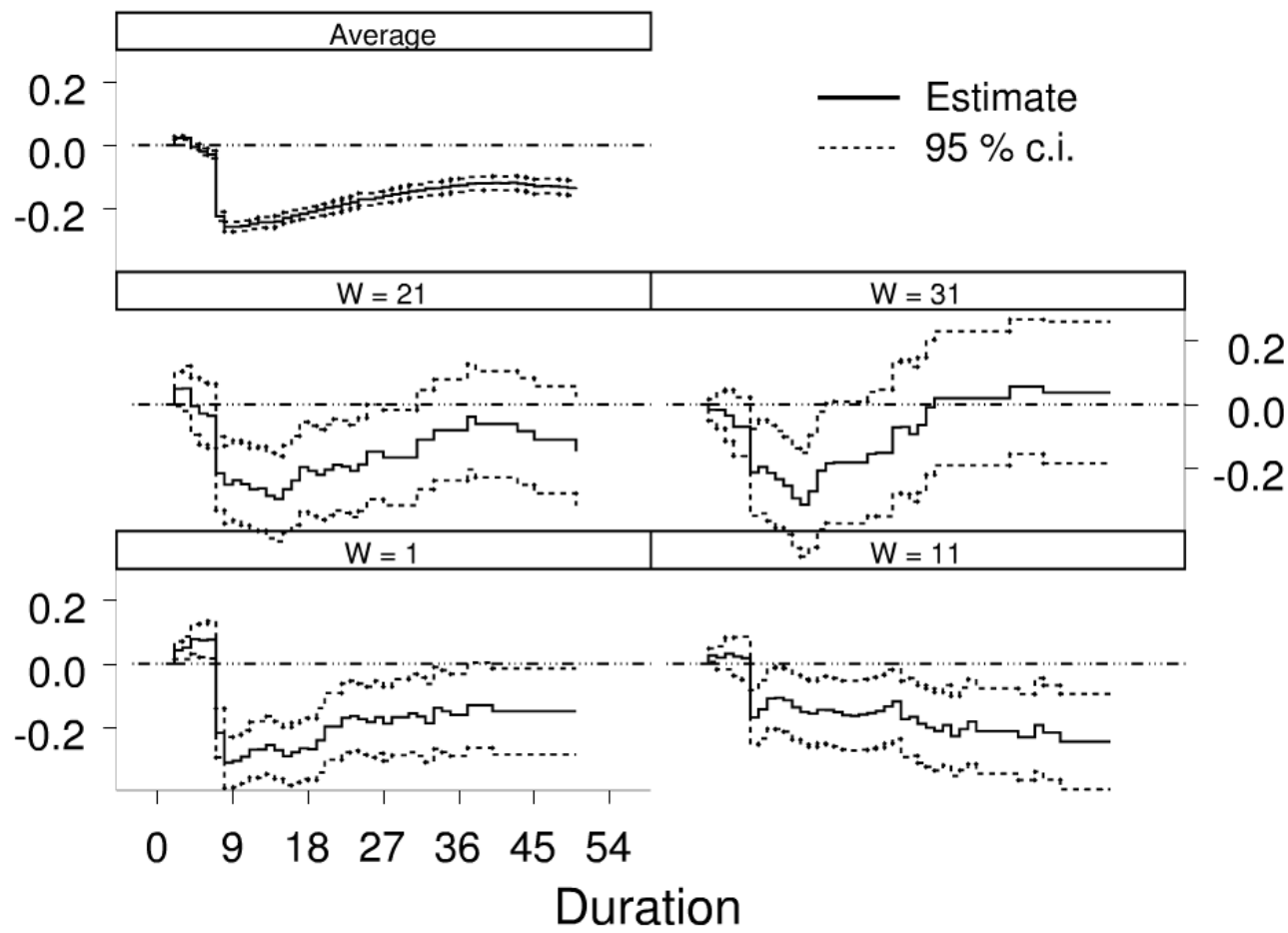


- Eligible: at least 25, registered unemployed at least 12 months in a row
- Register data: 98-02; 631,358 eligible, 3% ended into program; 40% ended in employment
- 630,000 eligible; after matching: 7,651 individuals left



Exact one-to-one matching

Estimating $F^1 - F^0$:



Some concluding remarks

- Causal inference in observational studies: Protocols defining population, treatment assignment and control group
- With population wide registers:
 - Sample is population
 - Large control groups and rich set of background characteristics allow for good designs

Some references

de Luna, X & Johansson P (2010) Non-parametric inference for the effect of a treatment on survival times with application in the health and social sciences, *Journal of Statistical Planning and Inference* 140, 2122-2137. Erratum in same journal 2012.

Neyman J (1990) On the application of probability theory to agricultural experiments. Essay on principles Translated by D.M. Dabrowska and edited by T.P. Speed *Statistical Science*, 5, pp. 465-472. Original text from 1923.

Rubin, DB (1991) Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47 , 1213-1234.

Heitjan, DF & Rubin DB (1991) Ignorability and coarse data. *The Annals of Statistics*, 19, 2244-2253.

Rotnitzky, A & Robins, JM (2005) Inverse Probability Weighting in Survival Analysis. *Encyclopedia of Biostatistics*.

Some references

de Luna, X & Johansson P (2010) Non-parametric inference for the effect of a treatment on survival times with application in the health and social sciences, *Journal of Statistical Planning and Inference* 140, 2122-2137. Erratum in same journal 2012.

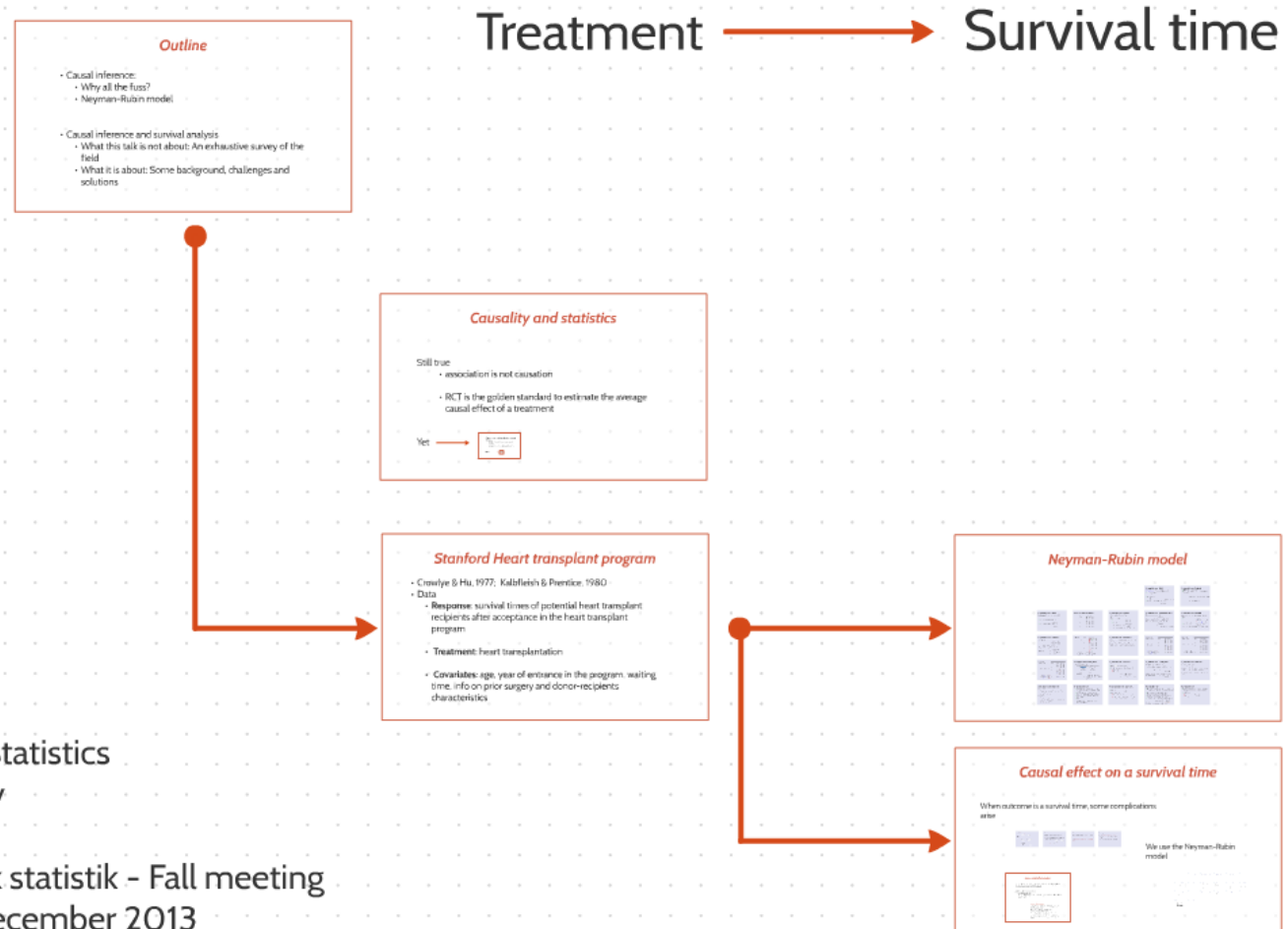
Neyman J (1990) On the application of probability theory to agricultural experiments. Essay on principles Translated by D.M. Dabrowska and edited by T.P. Speed
Statistical Science, 5, pp. 465–472. Original text from 1923.

Rubin, DB (1991) Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47 , 1213–1234.

Heitjan, DF & Rubin DB (1991) Ignorability and coarse data. *The Annals of Statistics*, 19, 2244–2253.

Rotnitzky, A & Robins, JM (2005) Inverse Probability Weighting in Survival Analysis. *Encyclopedia of Biostatistics*.

Causal inference and survival analysis



Xavier de Luna, Dept of Statistics
USBE @ Umeå University

Föreningen för medicinsk statistik - Fall meeting
AstraZeneca, Mölndal, December 2013