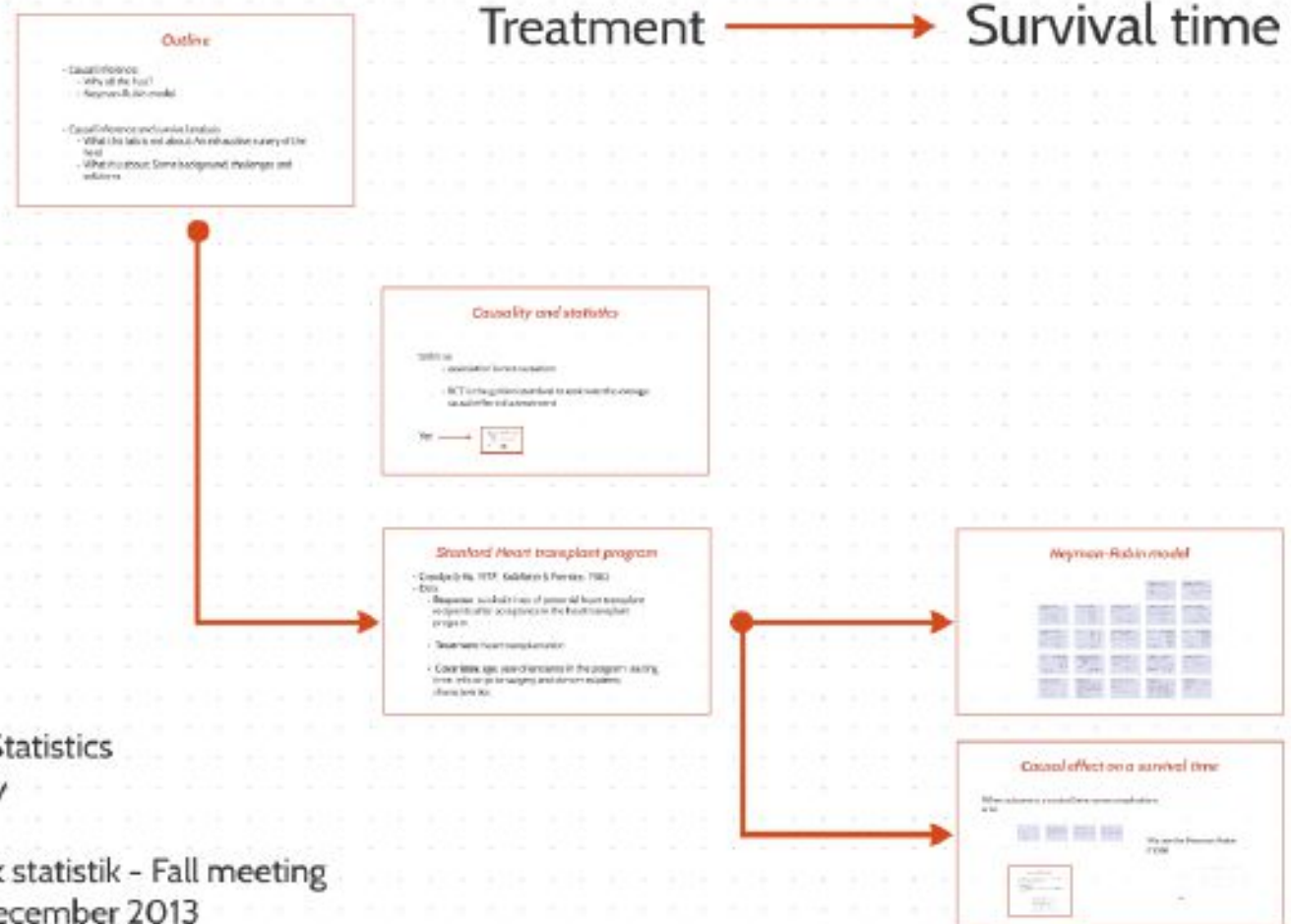


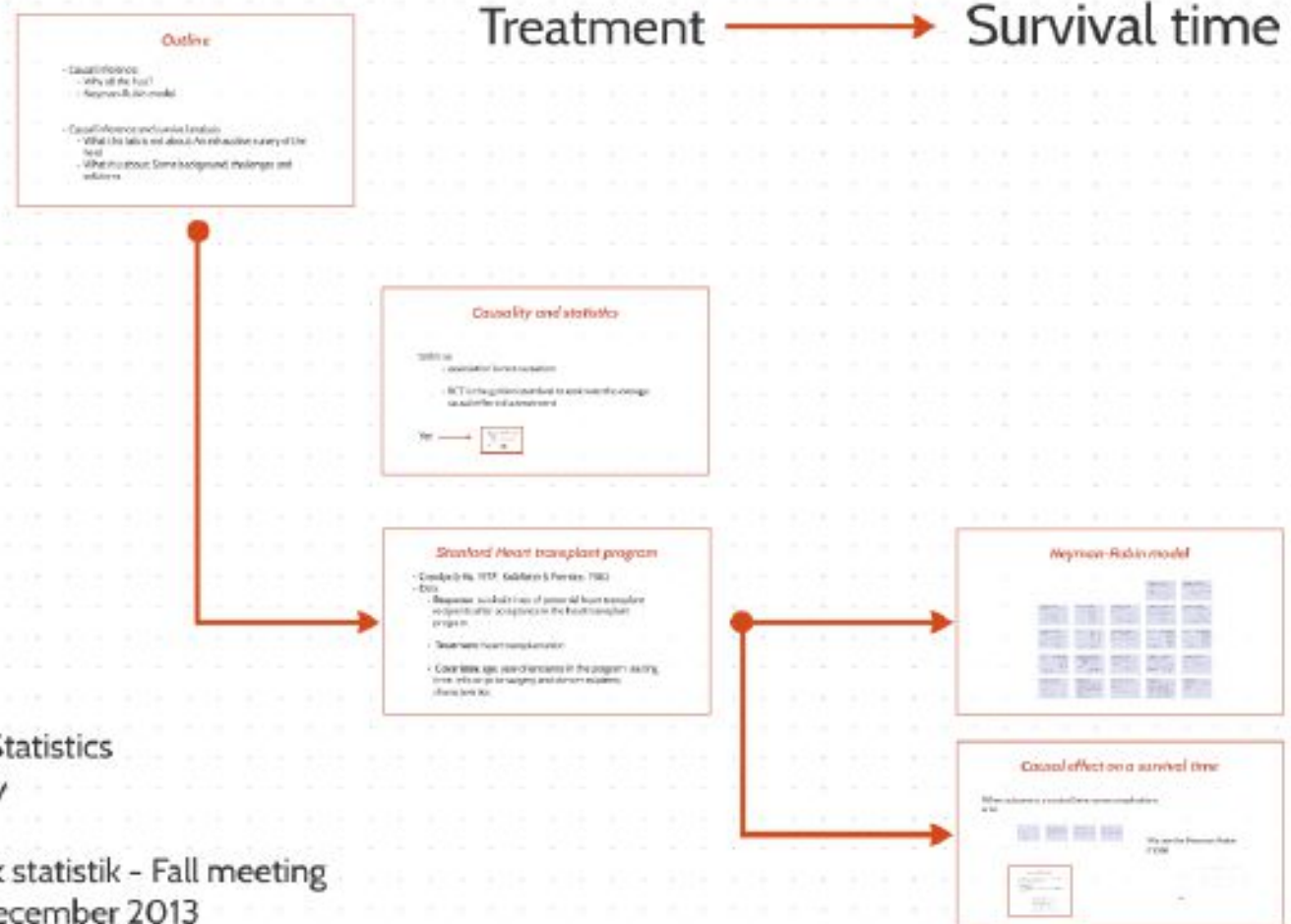
Causal inference and survival analysis



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?



Why?

- languages for causal reasoning have been developed or associations are hard to get to do things
- PCIT has a long history of success
- kind of observational data not from before

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: 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:

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

★ n_c control individuals for which we observe:

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

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.


$$\cancel{y(1), y(0) \perp z}$$

$\bar{y}^t - \bar{y}^c$ 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.:

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

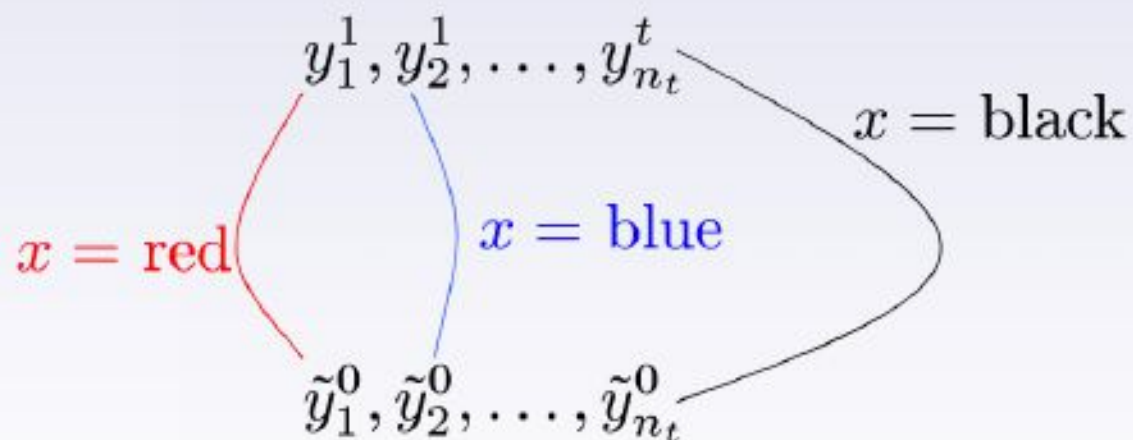
[Unconfoundedness assumption]

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

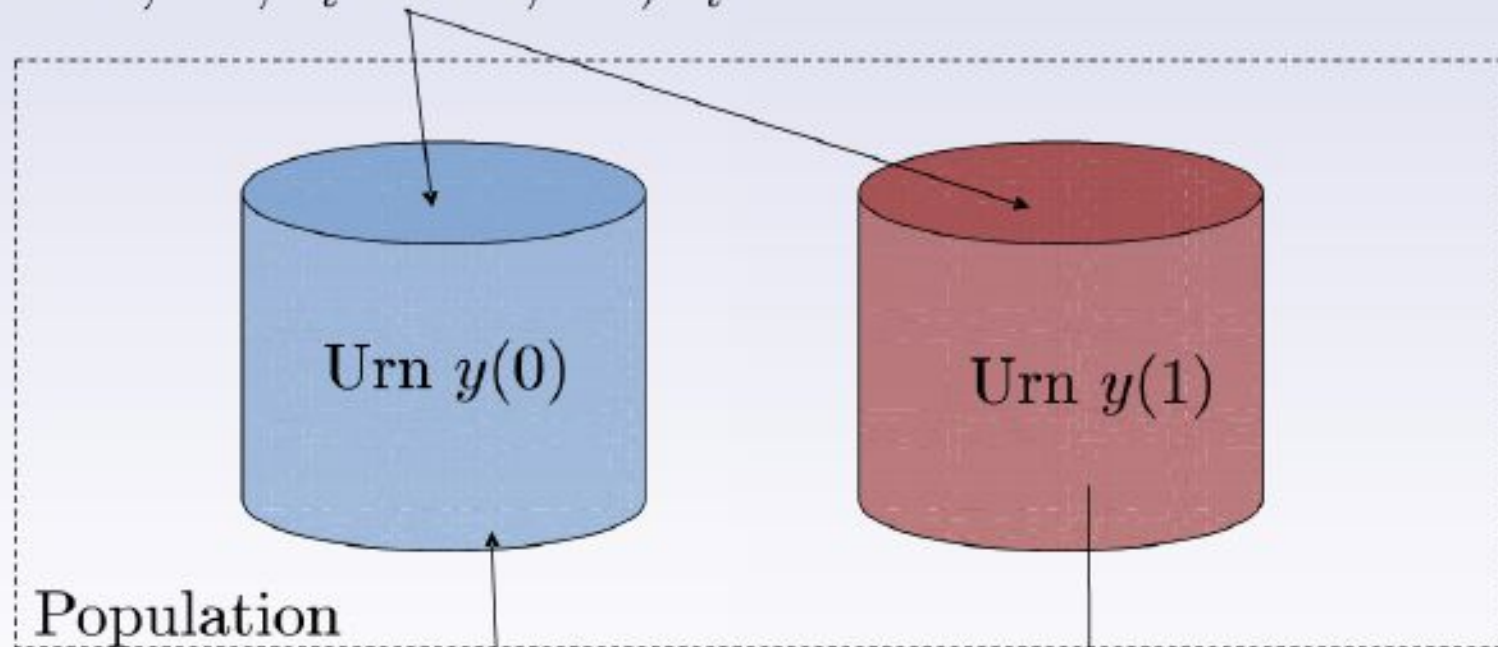
Reassigning treatment randomly and the resulting estimator

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

| 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_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



We use the Neyman-Rubin model

Some concluding remarks

- Causal inference in observational studies (formalizing key data-generating sequence and control group)
- With population-wide registries
 - Sample is population
 - Large control group as a sort of background is less sensitive to bias for good designs

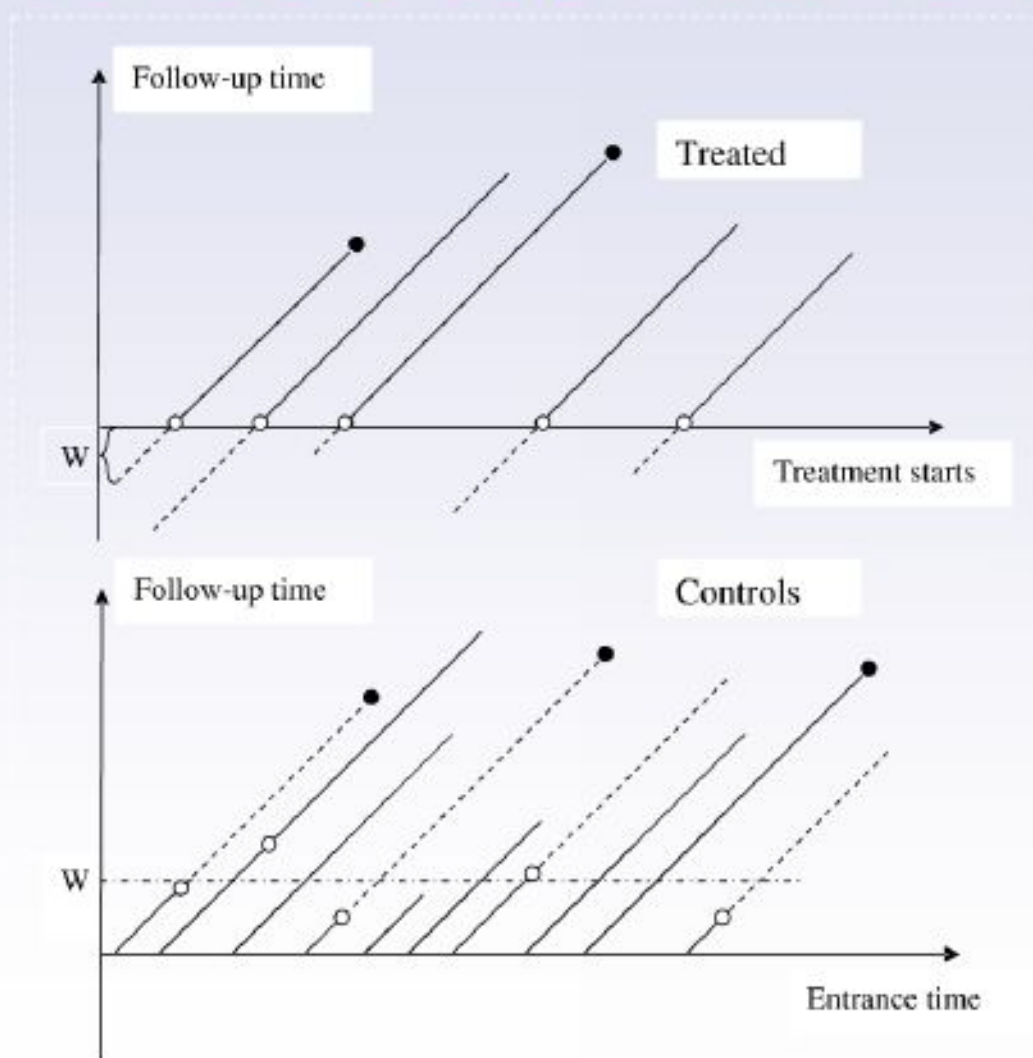
Some references

David A. Glynn, 2019, *Survival Analysis: A Practical Approach*, John Wiley & Sons, Inc.
 David A. Glynn, 2019, *Survival Analysis: A Practical Approach*, John Wiley & Sons, Inc.
 David A. Glynn, 2019, *Survival Analysis: A Practical Approach*, John Wiley & Sons, Inc.



Note 1: Control group must include treated

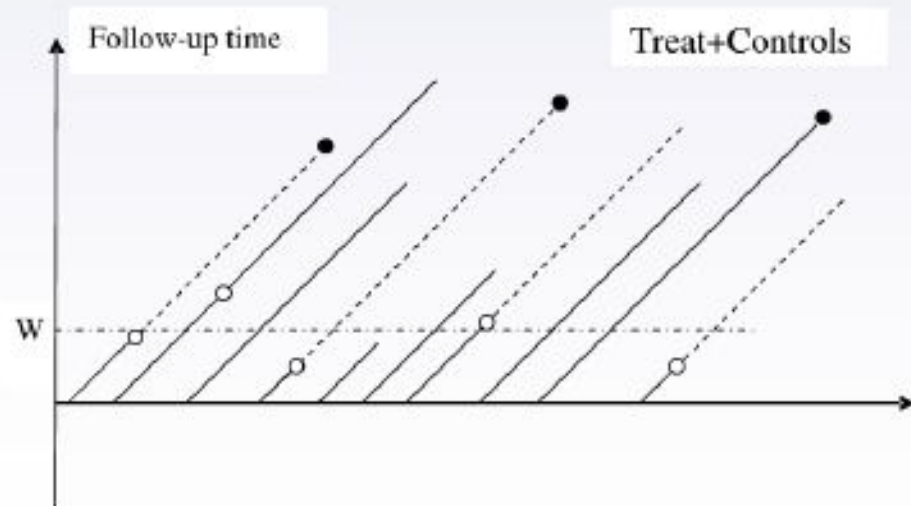
Lexis diagrams



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

Randomized controlled trial (RCT)
 Experimental design: Randomized allocation of subjects to two groups: treatment and control.
 Outcome: A continuous variable (e.g., blood pressure).
 Hypothesis: The treatment group will have a lower mean blood pressure than the control group.
 Null hypothesis: There is no difference in mean blood pressure between the two groups.
 Alternative hypothesis: The treatment group has a lower mean blood pressure than the control group.

Observed status of the variables

At baseline, the variables are:

| Variable | Control | Treatment |
|----------------|----------|-----------|
| Age | 65 | 65 |
| Sex | Male | Male |
| Weight | 70 kg | 70 kg |
| Blood pressure | 160 mmHg | 160 mmHg |

At follow-up, the variables are:

| Variable | Control | Treatment |
|----------------|----------|-----------|
| Age | 66 | 66 |
| Sex | Male | Male |
| Weight | 71 kg | 71 kg |
| Blood pressure | 150 mmHg | 140 mmHg |

Assumptions

1. **Randomization**: The assignment of subjects to treatment and control groups is random.
 2. **Blinding**: The subjects and the researchers are blinded to the treatment assignment.
 3. **Intention-to-treat**: The analysis is based on the original assignment, regardless of whether the subjects completed the study or not.
 4. **Missing data**: The missing data are missing at random (MAR).

Estimands

The estimand is the difference in the mean blood pressure between the treatment and control groups at follow-up.

Results

The difference in the mean blood pressure between the treatment and control groups at follow-up is -10 mmHg (95% CI: -15 to -5 mmHg). This difference is statistically significant (p = 0.001).

Inference (on missing...)

The missing data are missing at random (MAR). The missing data are imputed using multiple imputation by chained equations (MICE).

Sampling distribution

The sampling distribution of the difference in the mean blood pressure between the treatment and control groups at follow-up is normal with a mean of -10 mmHg and a standard deviation of 5 mmHg.

Concomitant to treatment

The concomitant variables are age, sex, and weight. These variables are measured at baseline and follow-up.

End of study, drop out, etc.

The subjects who dropped out of the study are excluded from the analysis. The subjects who completed the study are included in the analysis.

Results

The difference in the mean blood pressure between the treatment and control groups at follow-up is -10 mmHg (95% CI: -15 to -5 mmHg). This difference is statistically significant (p = 0.001).

Results

The difference in the mean blood pressure between the treatment and control groups at follow-up is -10 mmHg (95% CI: -15 to -5 mmHg). This difference is statistically significant (p = 0.001).

Survival functions

The survival function is the probability of surviving beyond a certain time point. The survival function is estimated using the Kaplan-Meier method.

Inference

The inference is based on the difference in the survival function between the treatment and control groups. The difference is statistically significant (p = 0.001).

Averaging over waiting times

The waiting times are averaged over the subjects in the treatment and control groups. The average waiting time is 10 minutes.

Heart transplant program

The heart transplant program is a randomized controlled trial. The subjects are assigned to the treatment group (heart transplant) or the control group (medical treatment).

Employment, activity program

The employment, activity program is a randomized controlled trial. The subjects are assigned to the treatment group (employment, activity program) or the control group (usual care).

Results

The difference in the mean blood pressure between the treatment and control groups at follow-up is -10 mmHg (95% CI: -15 to -5 mmHg). This difference is statistically significant (p = 0.001).

Exact test to one matching

The exact test to one matching is used to test the null hypothesis of no difference in the mean blood pressure between the treatment and control groups. The test is statistically significant (p = 0.001).

Concluding data

The concluding data are the results of the trial. The results show that the treatment group has a lower mean blood pressure than the control group.



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

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

for $j = 0, 1$.



Results

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

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

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**.

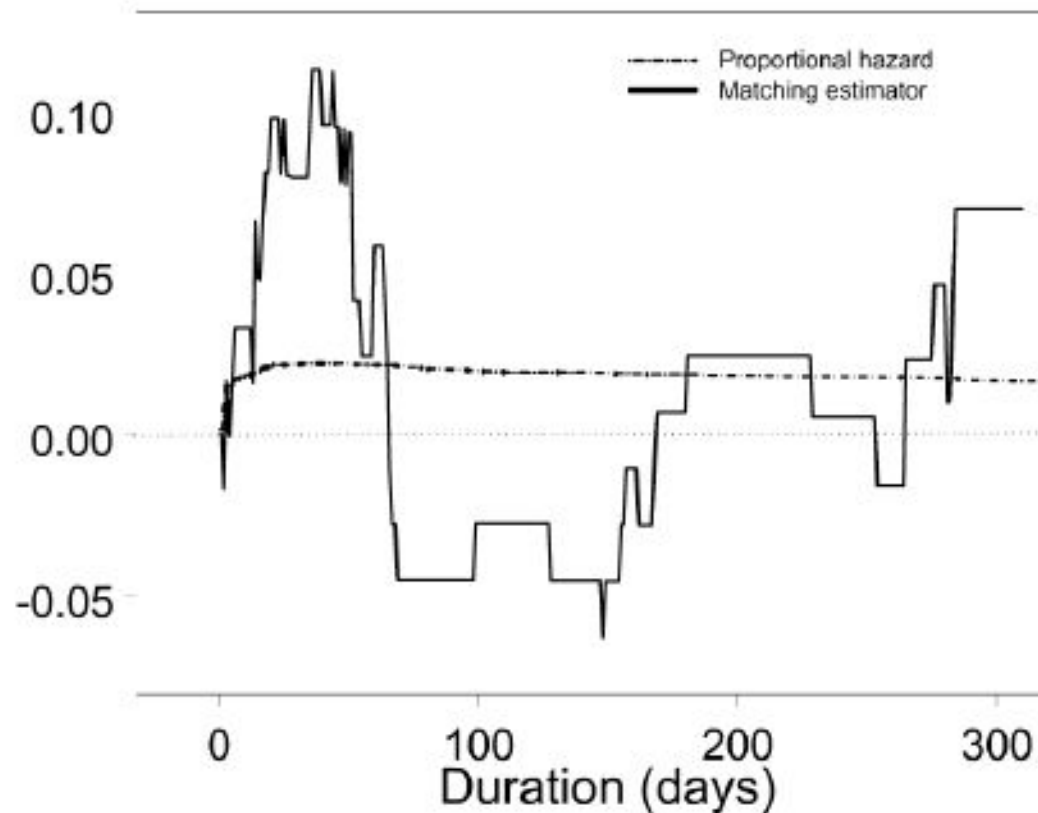
However: Interpretation of survival functions problematic.

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

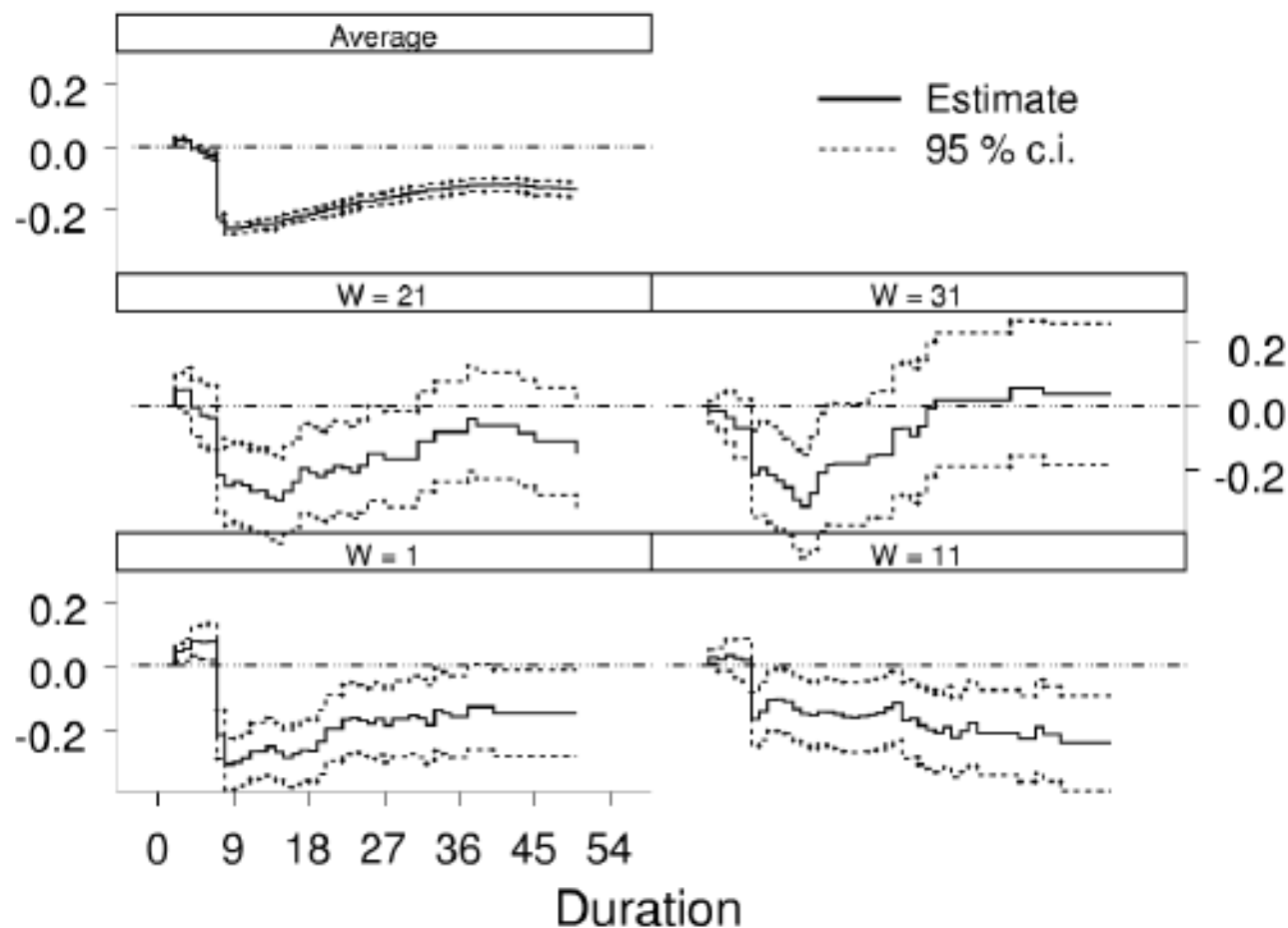


- Eligible: at least 25, registered unemployed at least 12 months in a row
- Register data: 98-02; 631,358 eligible, 3% ended into program; 40% ended in employment
- 630,000 eligible; after matching: 7,651 individuals left



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.

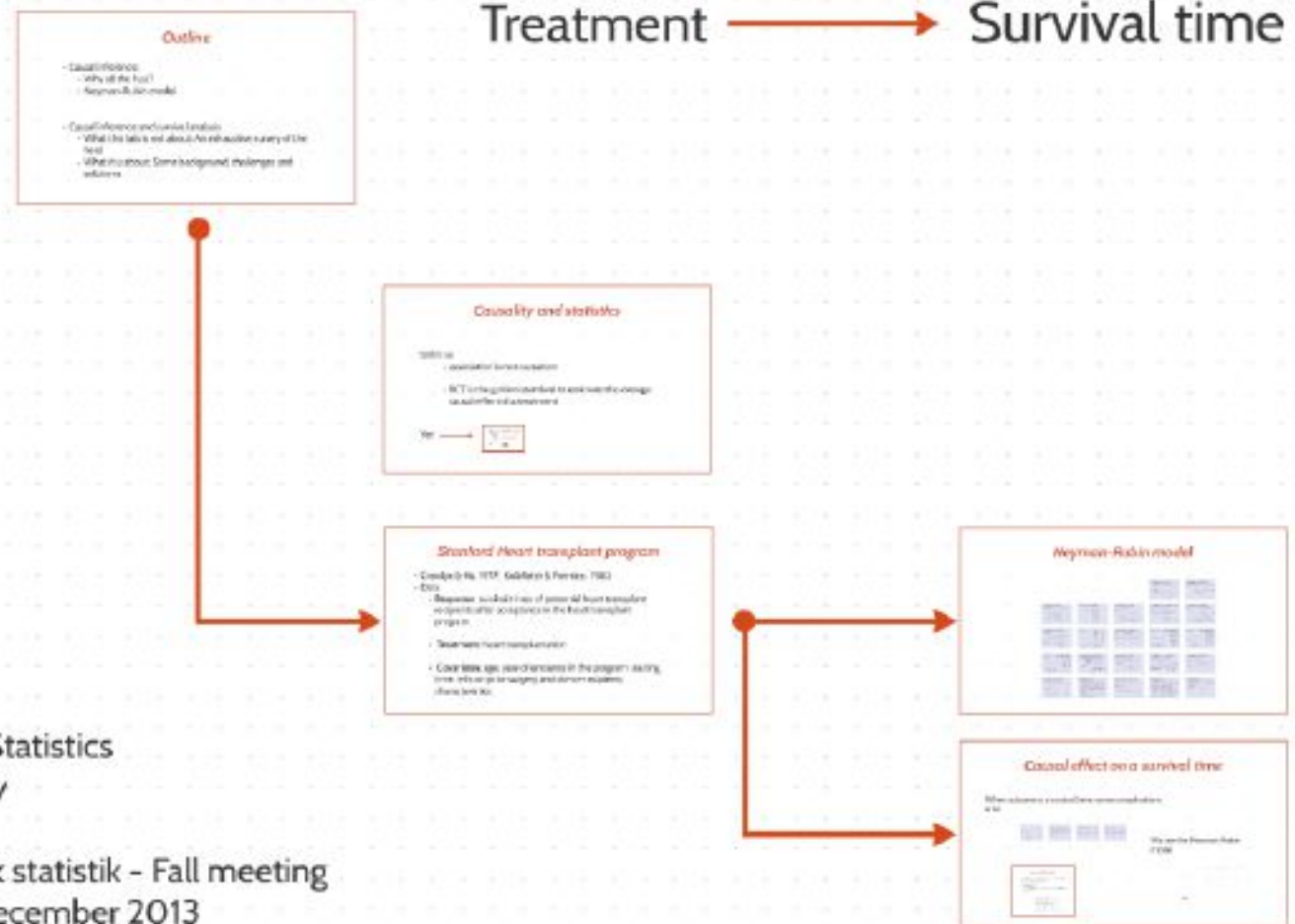
Neyman J (1990) On the application of probability theory to agricultural experiments. Essay on principles Translated by D.M. Dabrowska and edited by T.P. Speed
Statistical Science, 5, pp. 465-472. Original text from 1923.

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