Leiden University Medical Center

Adjustment in observational studies: use of propensity scores and instrumental variables

ESCMID Seville

Saskia le Cessie

Department of Medical Statistics and Bio-informatics

Department of Clinical Epidemiology

LEIDEN

Challenge

How to estimate causal effects from observational data?

1. Try to identify all confounders and adjust for them

- Stratification
- Regression methods
- Propensity scores methods
- Marginal structural models (for time dependent confounding)

2. Pseudo-randomization

- Instrumental variable analysis
- Regression discontinuity design

Example: Steroids in cardiac surgery

Effect of prophylactic treatment with steroids in cardiac surgery

- Study population: adult patients, who underwent cardiac surgery
- 115 used steroids as prophylaxis, 361 not
- Outcomes: ventilation time, ICU- and hospital stay, 30 day mortality, infections.

Question: do steroids prevent complications?

Compare patients who did and who did not receive steroids

Outcome	No steroids (n=361)	Steroids (n=115)	Crude odds ratio /ratio means
Mortality (30 days)	10 (2.8%)	4 (3.5%)	1.26 (0.39,4.10)
Ventilation time (hour)	10 (7,19)	11 (6,18)	1.00 (0.76,1.32)
ICU stay (days)	1 (1,3)	2 (1,4)	1.14 (0.94,1.39)
Hospital stay(days)	7 (6,11)	8 (6,13)	1.14 (0.98,1.31)

No beneficial effect of steroids

But data is not balanced: confounding by indication

	No prophylactic corticosteroids (n=361)		Prophylactic corticosteroids (n=115)		Diff
Male	246	(68.1)	69	(60.0)	-8.1
Age (yrs)	64.5	(13.5)	63.9	(12.9)	-0.6
BMI (kg/m^2)	26.6	(4.2)	26.4	(4.2)	-0.2
Diabetes mellitus	54	(15.0)	15	(13.0)	-2.0
EuroSCORE category					
1-2%	115	(32.0)	23	(20.4)	-11.7
3-5%	110	(30.6)	35	(31.0)	0.3
≥6%	134	(37.3)	55	(48.7)	11.3
Type of surgery					
Off-pump CABG	36	(10.0)	6	(5.2)	-4.8
On-pump CABG	100	(27.7)	29	(25.2)	-2.5
Valve	116	(32.1)	39	(33.9)	1.8
Combination/ Other	109	(30.2)	41	(35.7)	5.5

confounding by indication

PD

HD

HR=0.61

Outcome mortality, adjusting for possible confounders in logistic regression model

mort_30	Odds Rati	5			% Conf.	Interval]
steroids					1212478	2.016095
age						.9972507
gender					91002	2.763815
essre						
logess					918805	11.32967
_Iokrecod_2					.0035355	6.615585
_Iokrecod_3					.0041166	2.448379
_Iokreco				0.2	.0055753	3.595269
dm_recod						
_Isurgeon__2		3	e+0	0.01	0.993	0 .
_Isurgeon__4		1	(omitted)			
_Isurgeon__5		718336.7	1.30e+09	0.01	0.994	0 .
_Isurgeon__6		2444266	4.43e+09	0.01	0.994	0 .
_Isurgeon__7		1458558	2.64e+09	0.01	0.994	0 .
_Isurgeon__9		2444387	4.43e+09	0.01	0.994	0 .
_Isurgeon~10		1	(omitted)			
_cons		1.20e-06	.0021775	-0.01	0.994	0 .

Model explodes

Too few events

Different approach: propensity scores

Rosenbaum and Rubin (1983)

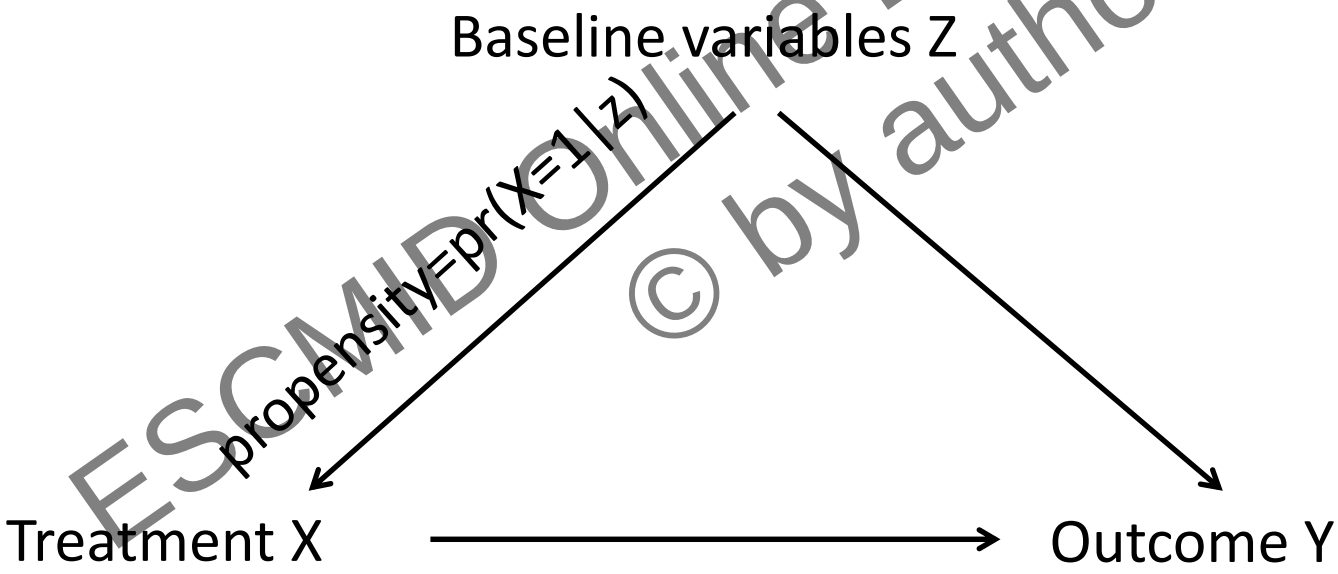
Some patients have a higher probability to receive a certain treatment (a higher “propensity”)

Use this propensity to balance groups

Propensity scores

Propensity score = Probability to receive treatment X=1, given baseline variables Z

Propensity score = pr(X=1 | Z=z)



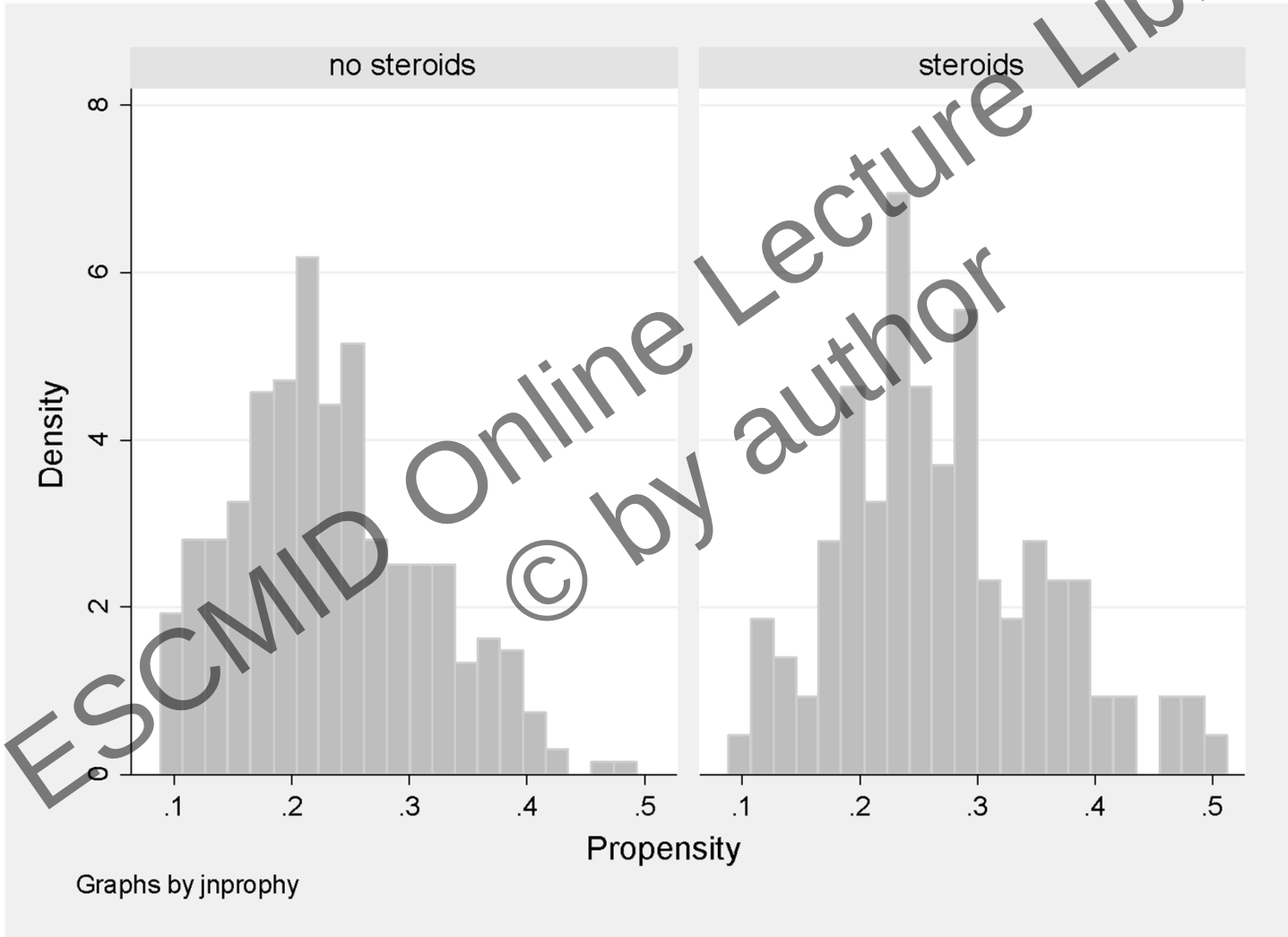
Steroids data

Propensity = probability to receive steroids given baseline characteristics.

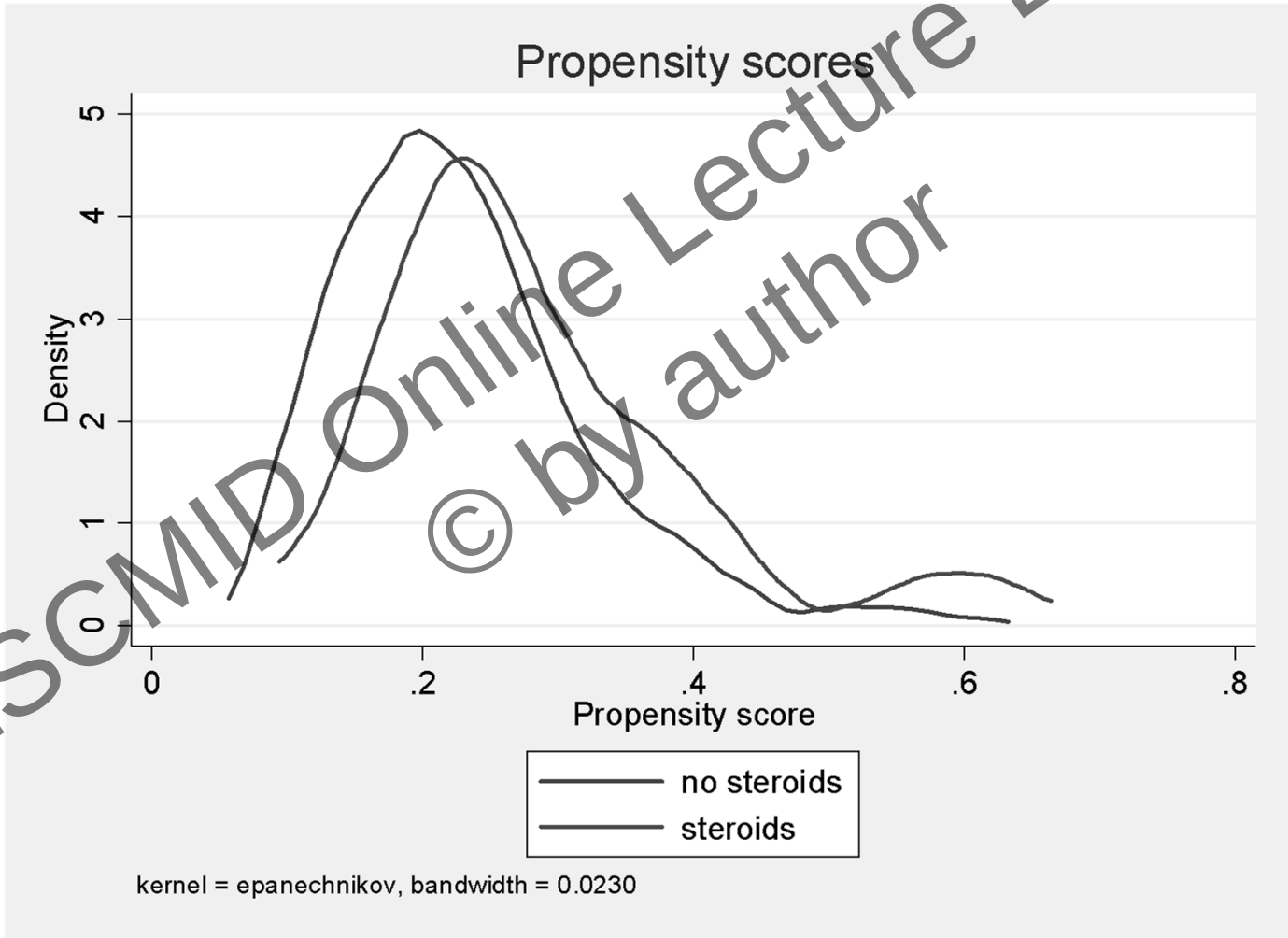
This probability (propensity) can be estimated with for example logistic regression with

- Outcome: steroids yes/no
- Independent variables: baseline variables (Euroscore, type of surgery, surgeon, age, gender etc.)

Steroids data: distribution of calculated propensity score



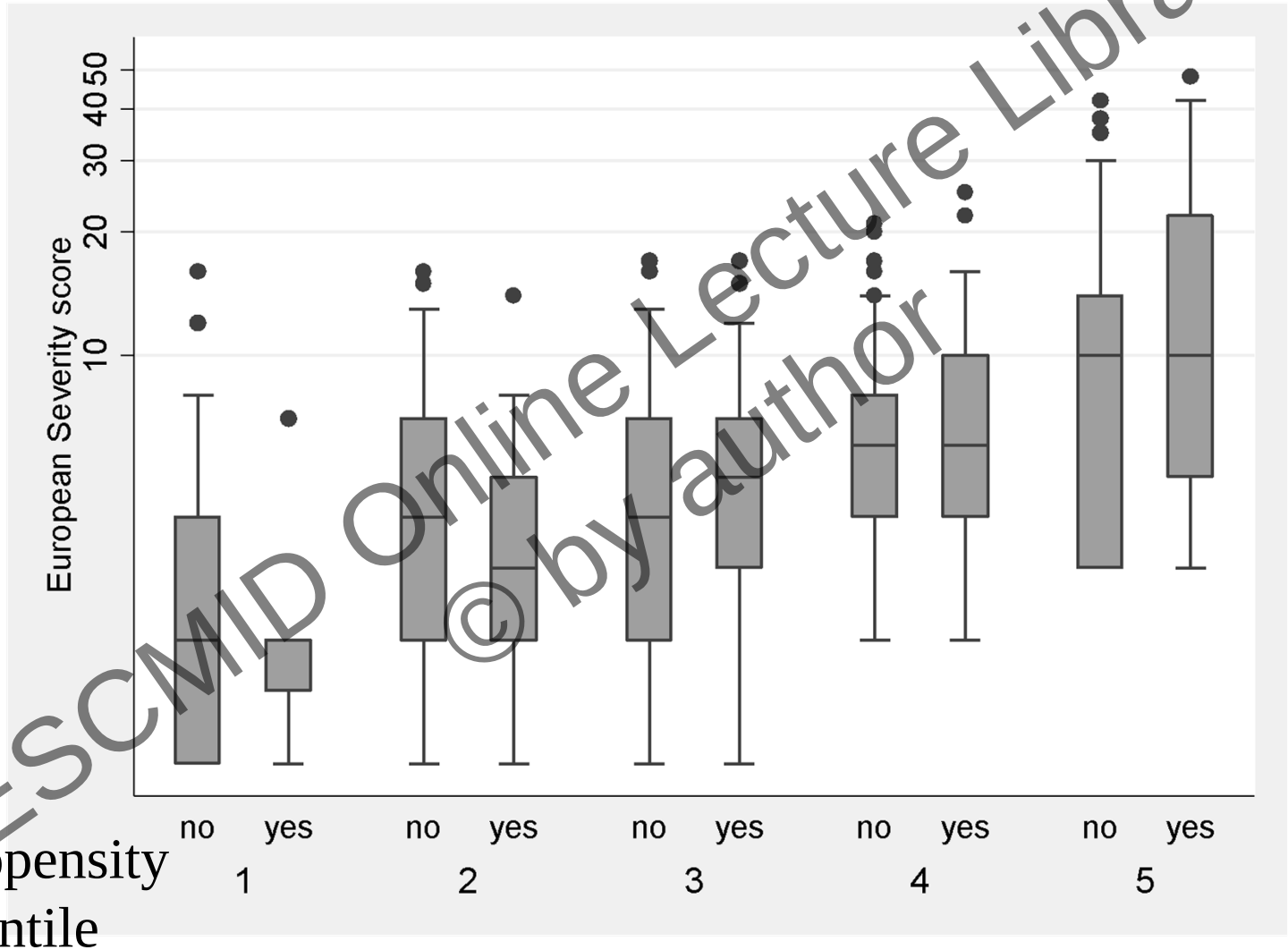
Distribution propensity score



Properties of propensity score

- Within a group with same propensity, treated and untreated persons have similar distribution of the observed covariates
- Treatment assignment is independent of confounders (ignorable) for any given value of the propensity score
- Propensity score is a balancing score
- Note: holds for measured confounders

For subjects with same propensity, observed covariates are similar distributed in treated and untreated



How to use a propensity score?

Compare treated and untreated subjects with the same propensity score

This can be done in different ways

1. Stratification
   - Define strata based on propensity score values
   - Calculate effects within each stratum and pool effects
2. Matching
   - Match each treated subject to an untreated subject with the same propensity value
3. Propensity score as covariate in regression model
4. Weighting

Inverse probability weighting

Use weights based on propensity score

w_i=1/propensity if treatment X=1

w_i=1/(1-propensity) if X=0

Treated: low propensity → high weight

high propensity → low weight

Untreated: high propensity → high weight

low propensity → low weight

Propensity score

- Balances observed confounders
- In general does not balance hidden confounders (randomization does)
- Similar to multiple regression methods which also only accounts for observed confounders

Which variables to include in propensity model?

- Variables which affect both treatment and outcome (confounders) should be included in propensity score
- Using variables only related to treatment, gives less precise results
- Using variables only related to outcome can increase precision of results

What is a good propensity score?

The propensity score should balance the data

- Check balance of covariates between treatment groups
  - Look at (standardized) mean differences, not only at p-values

This does **not** imply that the AUC (area under the curve) of the propensity model should be high

- There should be overlap between treatment groups
- In a randomized trial, the AUC of propensity model will be close to 0.50

When is a propensity score useful ?

In follow-up study with:

- Rare outcome
- Common treatment
- Multiple confounders

In this situation: limited power to adjust for confounders in outcome model

For matching in follow-up study when only few individuals can be followed and with one treatment/exposure of interest

- Matching on propensity scores gives better balance that matching on a few individual covariates

Advantage of propensity score

- The construction of the score is done without looking at the outcome (analogous to RCT where randomization is done, before outcome is known)
  - No danger of overfitting of relation between treatment and outcome
- Balance is easier to check than the fit of an outcome regression model.
- Overlap between distribution of confounders can be studied (check for positivity)

Advantage standard methods (regression, Mantel Haenszel)

More efficient: smaller confidence intervals

In practice regression using propensity score and regression methods often give comparable results

The steroids data

Compare outcomes

	Steroids (n=171)	No steroids (n=309)	Crude odds ratio (95% CI)	Odds ratio adjusted for Euroscore (95% CI)	Odds ratio Propensity score in regression model
Mortality (30d)	4 (3.5%)	10 (2.8%)	1.3 (0.39, 4.1)	0.84 (0.23, 3.0)	0.67 (0.19, 2.4)
Infection	15 (13%)	52 (15%)	0.88 (0.47, 1.6)	0.72 (0.37, 1.4)	0.66 (0.34, 1.3)

But

- Did we account for all confounding?
- Reviewers were not convinced.

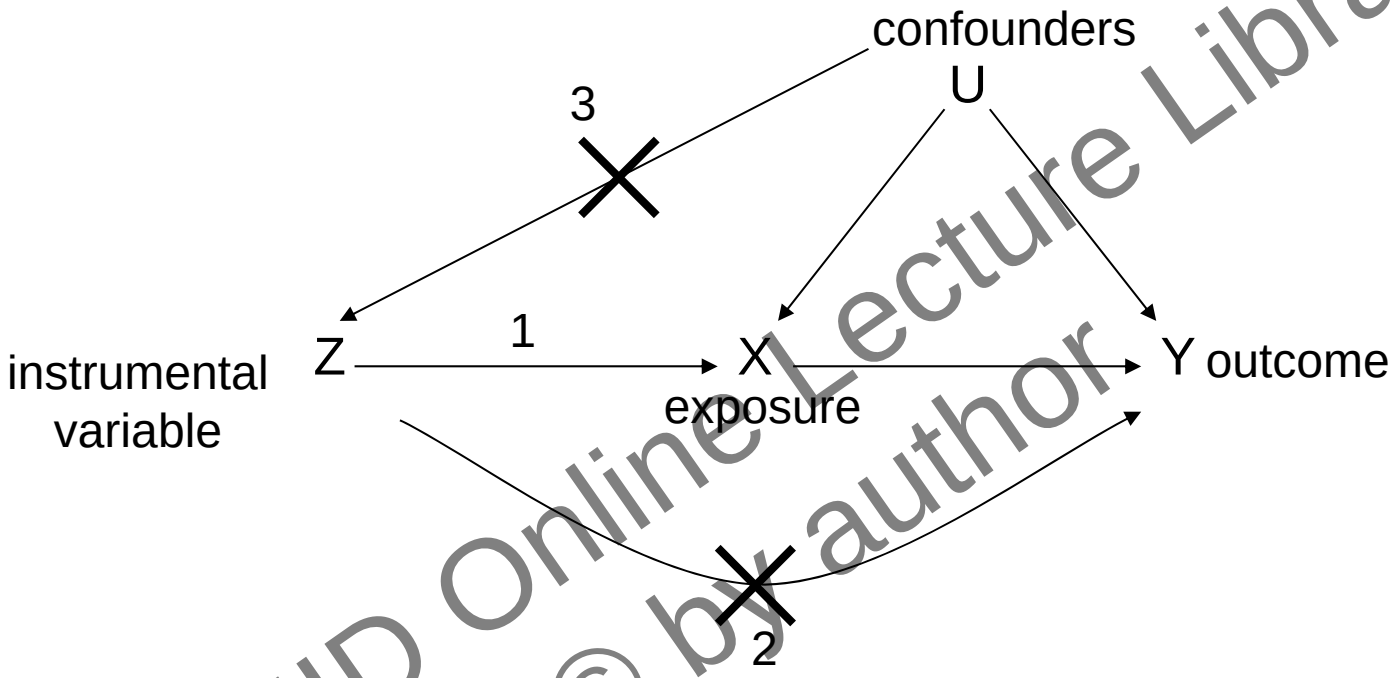
→ Alternative approach. Instrumental variables

Instrumental variable analysis

Use a variable that patients more or less randomly distributes over treatments/ exposure groups

- Taxes on cigarettes
- Genetic variants
- Residence: a new treatment is introduced in phases
- Residence: determines (in part) in which center a patient is treated, which can influence treatment
- Preference of physician: physicians prescribe different medications for similar patients

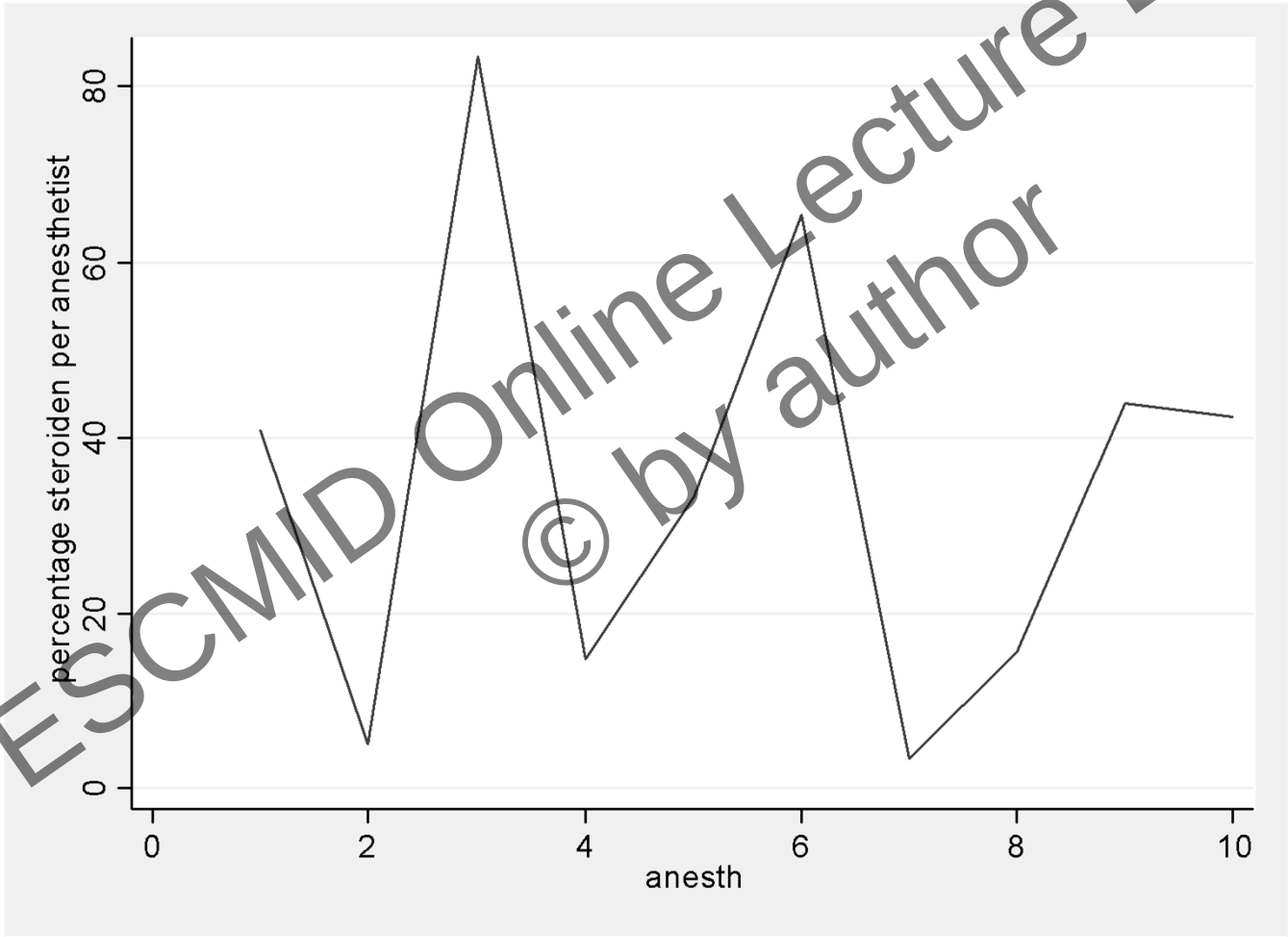
What is an instrumental variable?



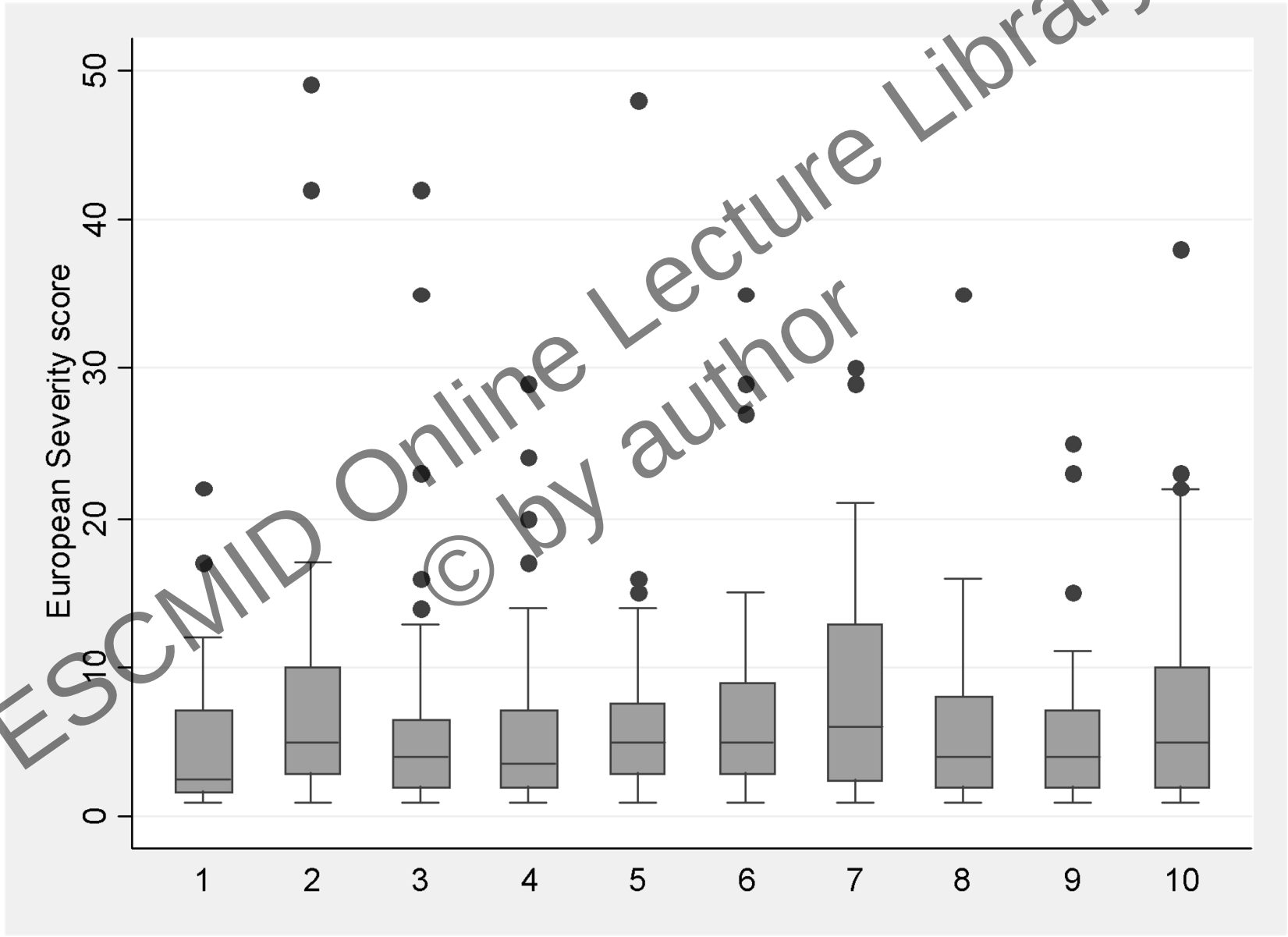
Assumptions:

1. The instrument predicts exposure
2. The instrument only affects the outcome through the exposure
3. The instrument does not share causes with the outcome

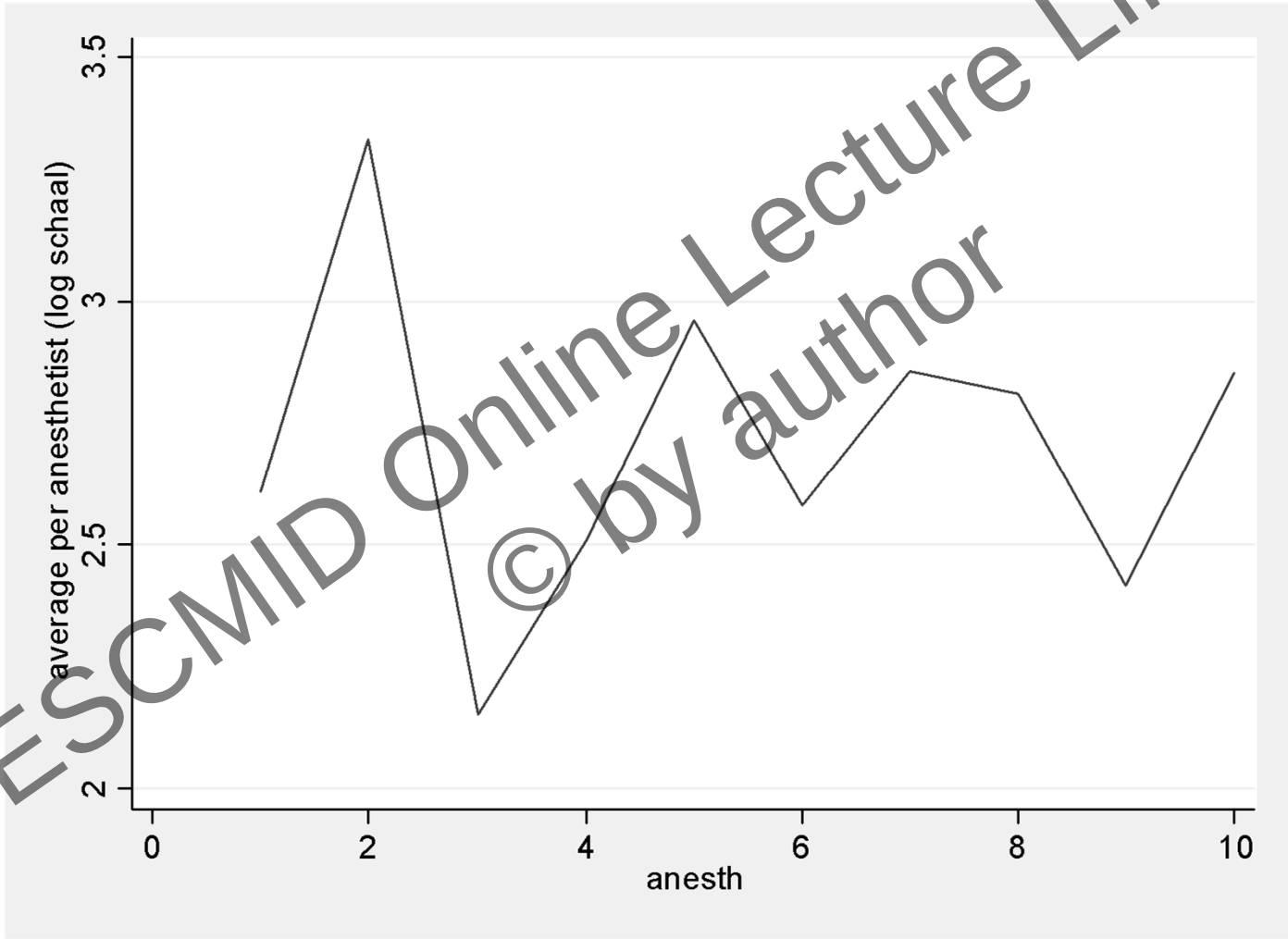
Our example: there is a large difference in use of steroids among anesthesiologists

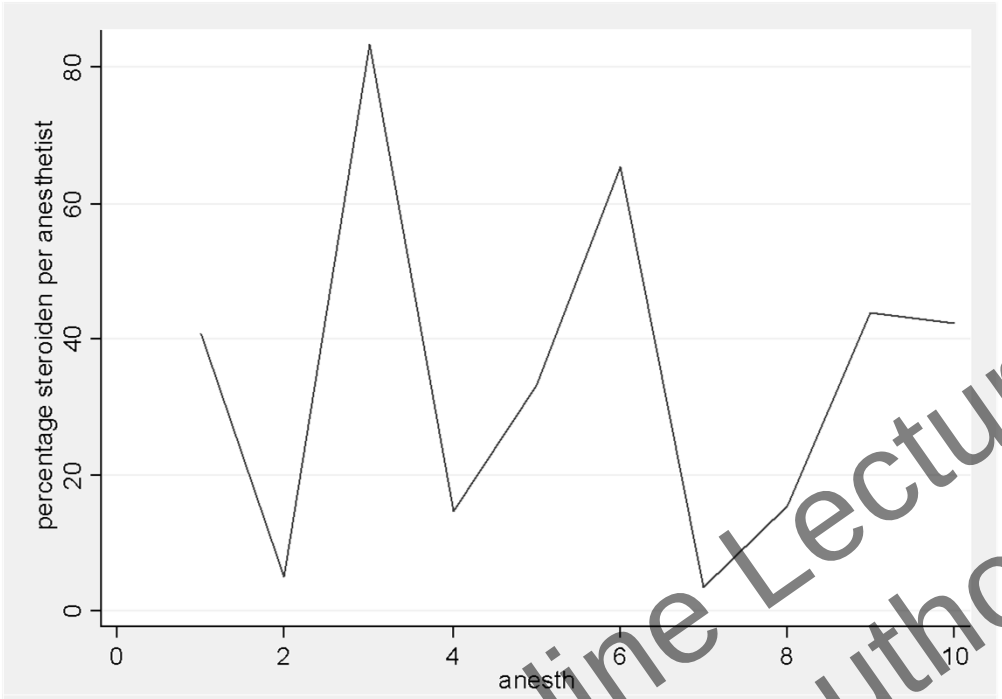


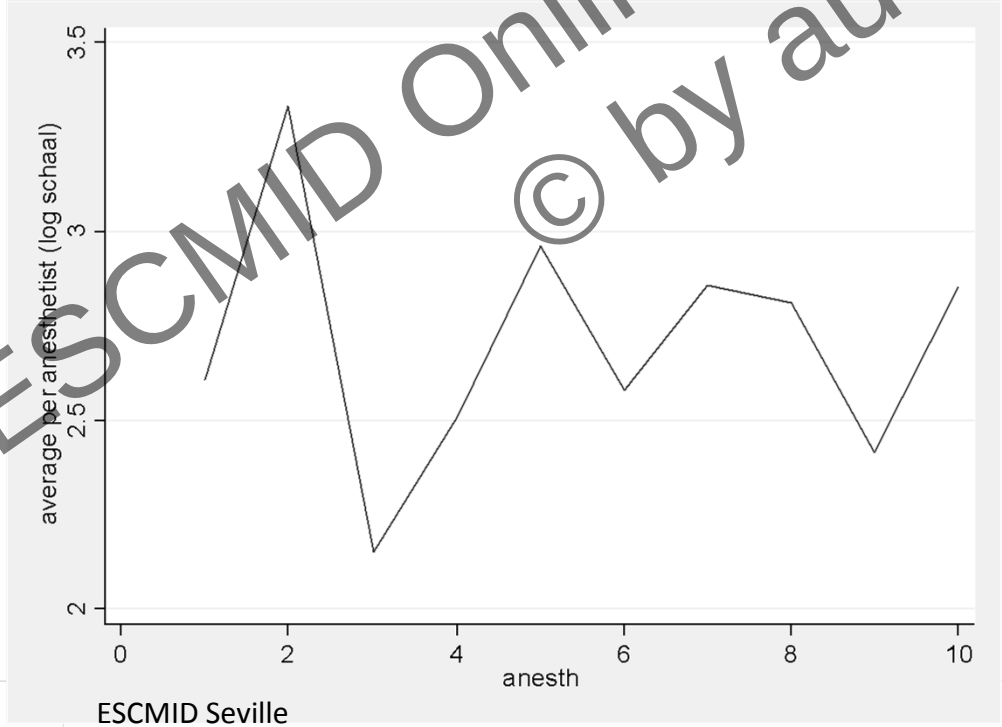
They have quite comparable patients



There seems to be a difference in median ventilation time







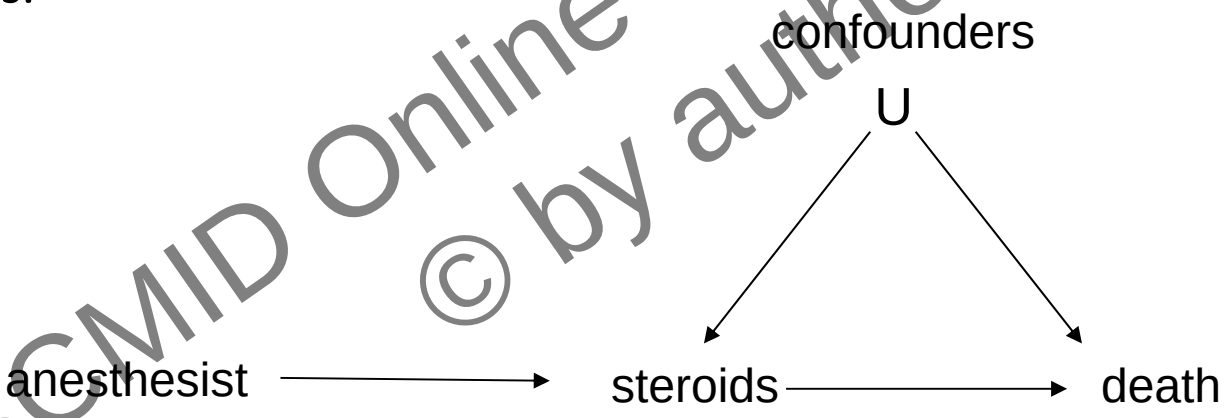
Inverse relation between use of steroids and ventilation time?

Use preference of anesthetists as instrument

How to define preference?

We used:

- The proportion of all earlier patients of the same anesthetists who received steroids.



How to estimate ? Technical details

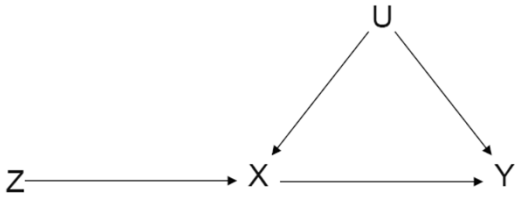
IV analysis often done using a two-stage regression analysis

Stage 1 = regression with X dependent and Z independent variable

Save predicted values of X given Z. E[X|Z]

Stage 2 = regression: Y dependent and E[X|Z] independent

Frequently linear models are used for both stages

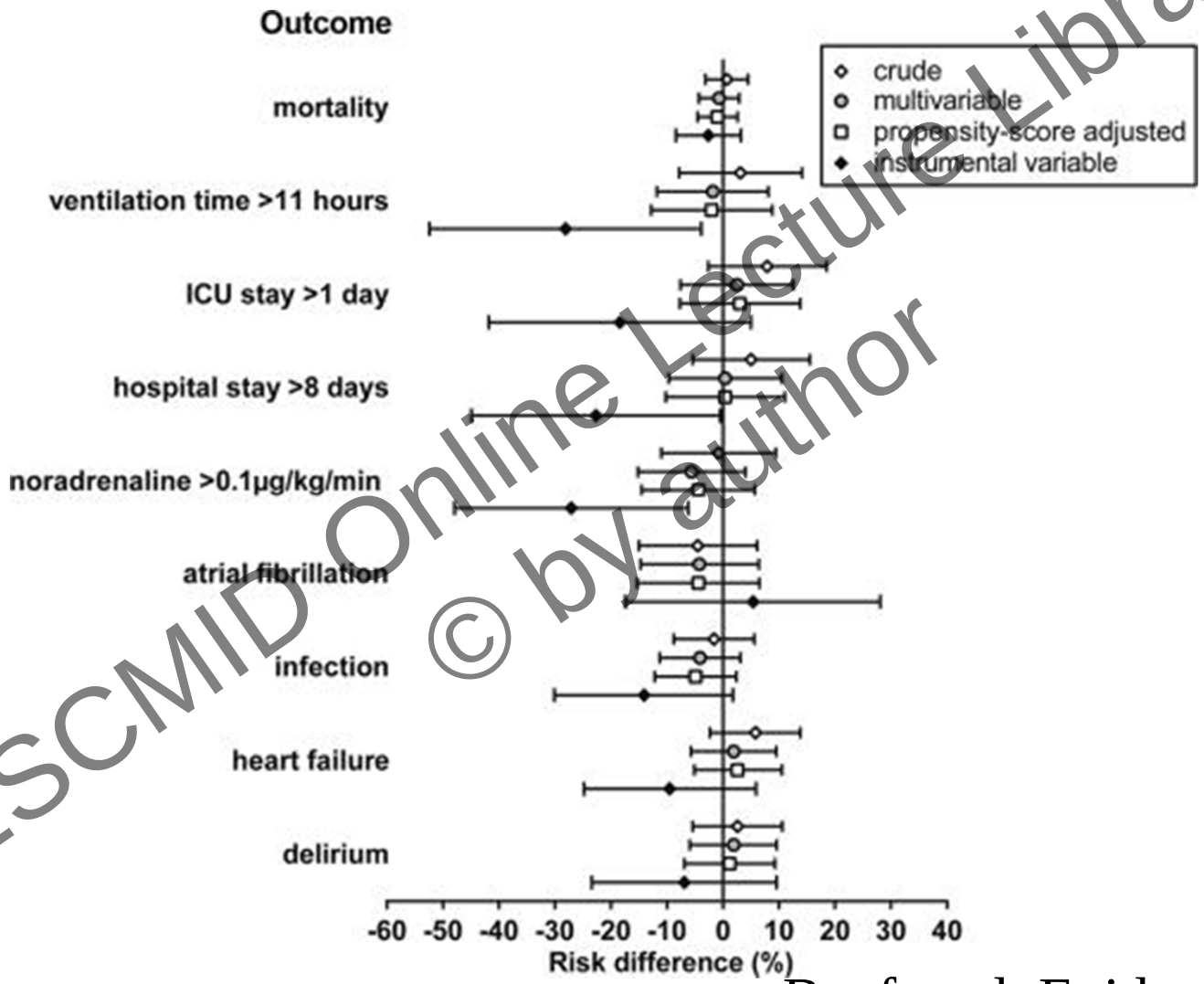


Compare results traditional analysis with results IV analysis

Outcome (used on log scale)	Ratio of means (95% CI)	Regression corrected for euroscore	IV analysis (mean preference)
Ventilation time (h)	1.11 (0.86-1.4)	0.93 (0.74-1.2)	0.50 (0.27, 0.92)
IC stay(days)	1.05(0.92-1.2)	0.99 (0.83-1.2)	0.63 (0.41; 0.98)
Hospital stay(days)	1.09(0.97-1.2)	1.05 (0.93-1.2)	0.79 (0.58, 0.92)

IV analysis: mean ventilation time after using steroids is 0.5 times the mean time without steroids

Binary outcomes



Boef et al, Epidemiology, 2014

What do we observe?

- Significant positive effect of steroids on ventilation time, IC-stay and stay in hospital

- Very different effects (rather extreme).
- Especially with binary IV and binary outcome
- Similar in direction to large randomized trial (Dieleman 2012)
- Very large confidence intervals

Summary

Two fundamentally different ways to deal with confounding

1. Try to identify all confounders and adjust for them
   - Multivariate regression
   - Propensity score
2. Pseudo randomization
   - Instrumental variables

Each approach has his own (untestable) assumptions

To summarize (2)

- propensity score method:
  - assumes no unmeasured confounding
  - useful when the outcome is rare
  - Useful to match in follow-up study when limited numbers can be followed and there is only one exposure.

- Instrumental variable analysis
  - Uses different set of assumptions
  - Assumptions are generally not testable
  - Large numbers are needed (trade-off between bias and variance)
  - Can account for unmeasured confounding
  - Estimates a different type of treatment effect (unless homogeneity of treatment effects is assumed)

References

Austin PC . An introduction to propensity-score methods for reducing the effects of confounding in observational studies. Multivariate Behavio Res 2011;46: 399–424.

Boef AGC et al Physician's preference-based instrumental variable analysis: is it valid and useful in a moderate-sized study? Epidemiology. 2014; 25(6):923-7.

Hernán et al Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2014 Jan;25(1):164.