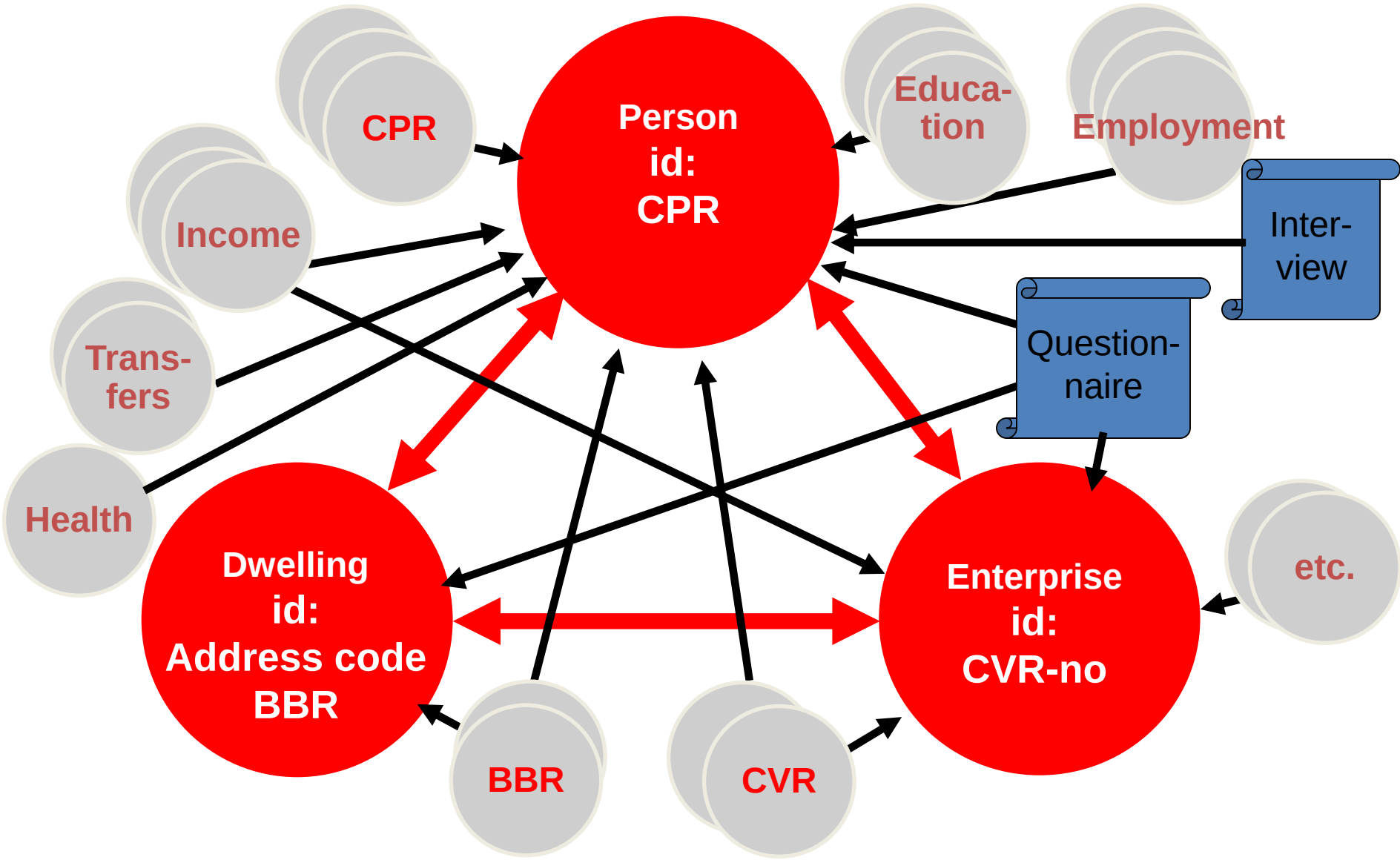
NordForsk PhD course in Register-Based Epidemiology

Bias in register-based studies

Lau Caspar Thygesen

The overall statistical system



Why register-based research

- Easy access to data – utilize existing data
- Large sample size – total population (rare diseases?)
- Population-based studies / real-world data / complete
- Great statistical power
- Follow-up easy
- No need to contact individuals
- No non-response bias (participation, reporting)
- Easy to do due to information technology
- Valuable time has passed – latency analyses
- Administrative data high quality
- Independent data

Selectionbias

- No self selection bias
- No loss to follow-up / attrition bias

AND

- Nordic populations relatively stable and homogeneous demography
- Universal health care system

Minor problem in register-based studies?

Maybe not?

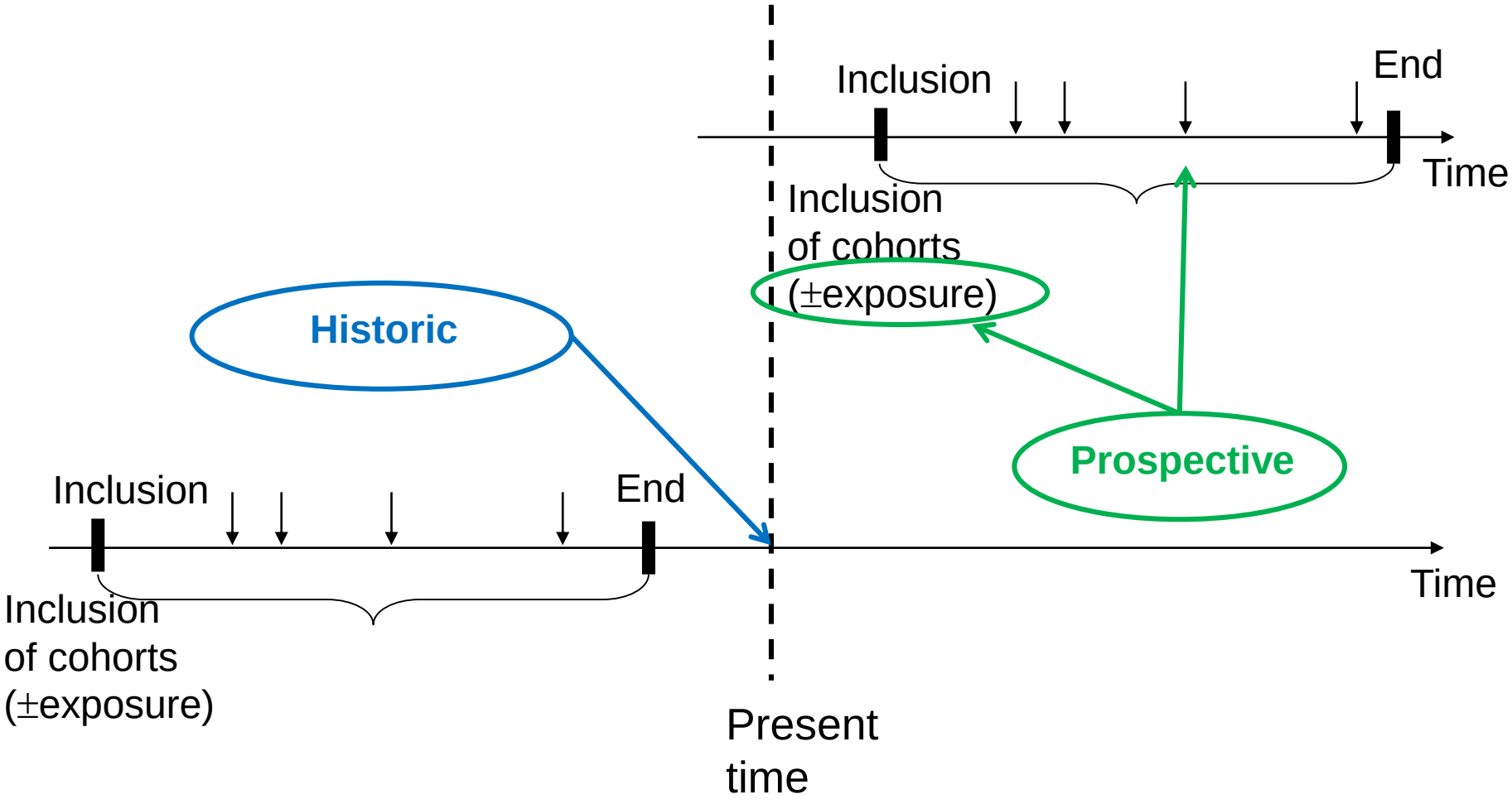
- Salmon effect

- Only unselected for diseases that always require contact to e.g. hospital

Bias in register-based studies

- Same bias as in all observational studies
  - Vulnerable to systematic (and random) errors
- Data is predetermined
- Confounding / non-comparability
- Validity / misclassification
- Truncation bias
- Immortal time bias
- Data dredging
- Statistical tests – are they relevant?

Retrospective?



Truncation

- Truncated by start of registration (left truncation)

- In the start of registration for a register difficult to distinguish between prevalent and incident cases

Consequence

- Overestimate the incidence especially in the first years of registration
- Prevalence underestimated
- Especially for diseases with low morbidity and few contacts to hospitals

What to do?

- Exclude first years from risk time to remove prevalent cases
- Dependent on contacts to hospitals and disease model
  - Few or many contacts
- Modig (2017) on Monday

Incident users

- Also for exposure
- In pharmacoepidemiology only incident users chosen (like randomised studies)
- Prevalent users:
  - Represent both start and late effects
  - Don’t adjust for intermediate variables

As Groucho Marx once said 'Getting older is no problem. You just have to live long enough'.

Queen Elizabeth II, at her 80th birthday celebration in 2006)

Some time ago, while conducting research on U.S. presidents, ***I noticed that those who became president at earlier ages tended to die younger****. This informal observation led me to scattered sources that provided occasional empirical parallels and some possibilities for the theoretical underpinning of what I have come to call the* ***precocity-longevity hypothesis****. Simply stated, the hypothesis is that those who reach career peaks earlier tend to have shorter lives.*

McCann. Personality and Social Psychology Bulletin 2001;27:1429–39

Failed kidney transplants and mortality

- Patients whose kidney transplants (allografts) have failed must return to long-term dialysis
- But should the failed allograft be removed or left in?
- US Renal Data System to study ‘a large, representative cohort of 10 951 patients returning to dialysis after failed kidney transplant’

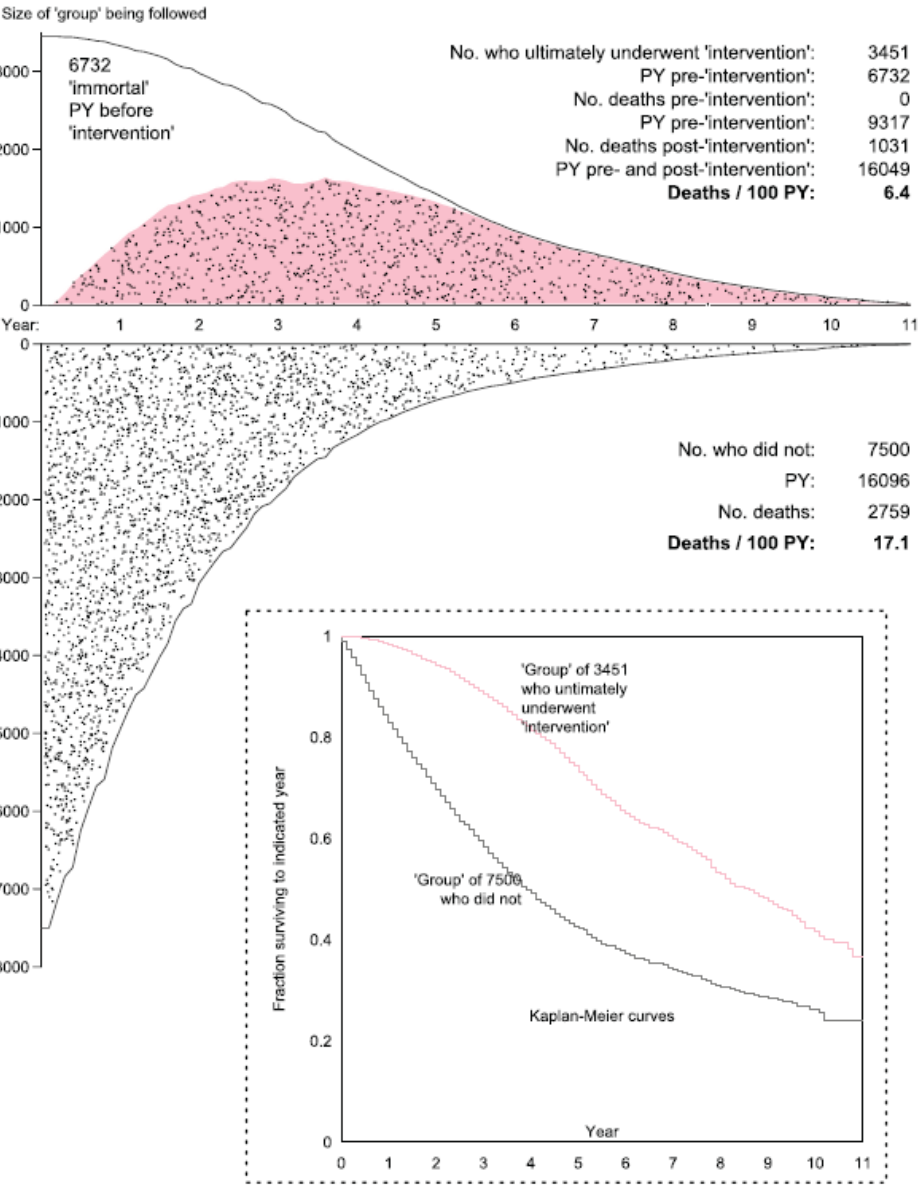
Failed kidney transplants and mortality

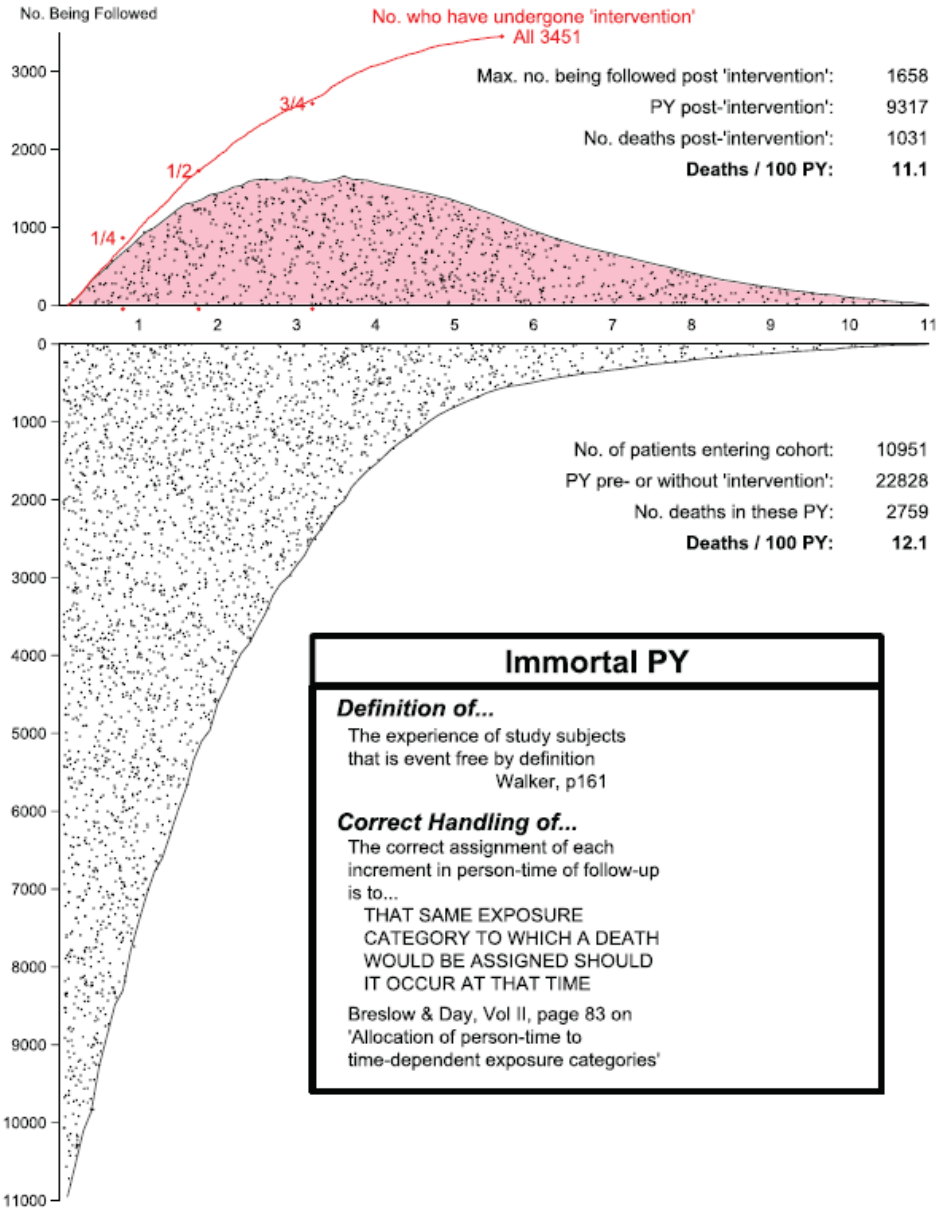
- 32% died of the 3451 in the allograft nephrectomy group
- 36% died of the 7500 in the non-nephrectomy group

Adjusted analysis:

- Receiving an allograft nephrectomy was associated with a 32% lower adjusted relative risk for all-cause death (HR=0.68 (0.63-0.74))

Hanley & Foster. Int J Epidemiolog 2014;43:949-61.





Immortal PY

Definition of...

The experience of study subjects that is event free by definition
Walker, p161

Correct Handling of...

The correct assignment of each increment in person-time of follow-up is to...
THAT SAME EXPOSURE CATEGORY TO WHICH A DEATH WOULD BE ASSIGNED SHOULD IT OCCUR AT THAT TIME

Breslow & Day, Vol II, page 83 on 'Allocation of person-time to time-dependent exposure categories'

Hanley & Foster. Int J Epidemiolog 2014;43:949-61.

Solution

- Do not look at individuals but at risk times
- Time-dependent covariates
- A person can move in and out of exposure
- Correct allocation of risk time

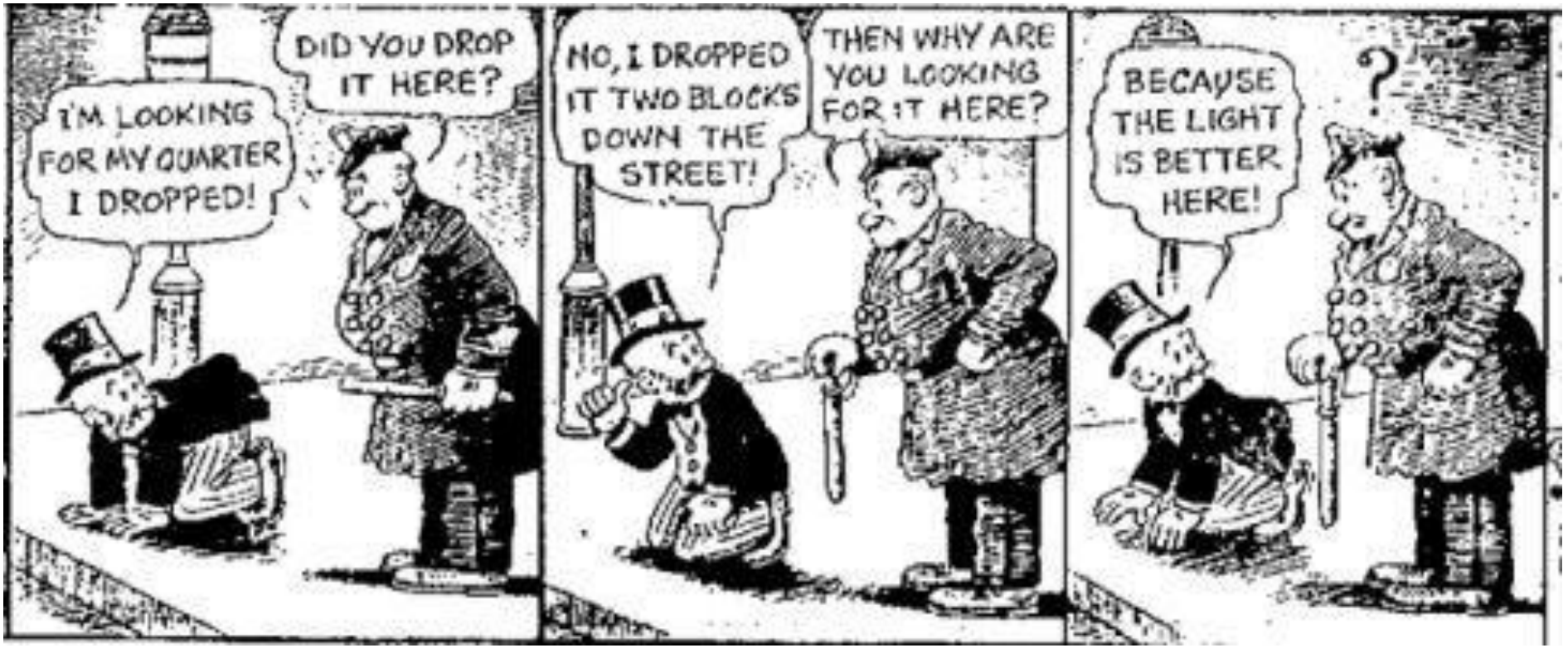
Terminology

- ‘Immortal’ time is not broad enough
- ‘Event-free time, by definition or by construction’
- More general and thus a more appropriate term

Data dredging
Misleading post hoc analysis

- Process of first identifying the data, and then proceed to the question is not good science
- Explorative studies may be appropriate

Looking for answers the wrong places



Dans. Ann Int Med 1993;119:855-7

The threat

- Large datasets
- Relatively cheap and easy to do large scale studies
- Better computer software
- Data explosion! Big data!
- Young investigators: Pressure to publish

Solution: Pre-definition

Pre-definition of

- Tables
- Main analysis
- Exposures
- Outcomes
- SAP (statistical analysis plan)
- Registration
- Other things…

- My experience: Pre-defined tables

Statistical testing

- Unimportant differences become highly significant in large studies
- Both significance level and effect size important

- Clinical or public relevance?

Is significance important?

Pukkala et al (2009) Occupation and cancer – follow-up of 15 million people in five Nordic countries

Observed prostate cancer among wood workers=18,707

SIR=0.97 (0.95-0.98) (table 47)

Observed all malignant cases male electrical workers=34,222

SIR=1.02 (1.01-1.03) (table 80)

Observed all malignant among economical inactive women =654,706

SIR=0.99 (0.98-0.99) (table 81)

JGIM

REVIEWS

Six Persistent Research Misconceptions

Kenneth J. Rothman, DrPH[1,2]

[1]Research Triangle Institute, Research Triangle Park, NC, USA; [2]Boston University School of Public Health, Boston, MA, USA.

Misconception 6. Significance testing is useful and important for the interpretation of data.

BASIC AND APPLIED SOCIAL PSYCHOLOGY, 37:1–2, 2015

ISSN: 0197-3533 print/1532-4834 online
DOI: 10.1080/01973533.2015.1012991

Editorial

David Trafimow and Michael Marks
New Mexico State University

The *Basic and Applied Social Psychology* (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP.

With the banning of the NHSTP from BASP, what

a strong case for rejecting it, con
provide a strong case for conclud
parameter of interest is likely t
interval. Therefore, confidence in
from BASP.

Bayesian procedures are more
problem with Bayesian procedu
on some sort of Laplacian assum
bers where none exist. The Lapla

International Journal of Epidemiology

(IF-5 9.804)

In the IJE we actively discourage the use of the term "statistically significant" or just "significant" and such statements in method sections as "findings at $p<0.05$ were considered significant“

Where used, we ask authors to provide effect estimates with confidence intervals and exact P values, and to refrain from the use of the term "significant" in either the results or discussion section of their papers

Our justification of this position is given in the Sterne J, Davey-Smith G. "Sifting the evidence - What's wrong with significance tests?" BMJ 2001: 322:226-231

See also Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. The American Statistician 2016: DOI:10.1080/00031305.2016.1154108

Epidemiology (IF 5.986)

Significance Testing:

For estimates of causal effects, we strongly discourage the use of categorized P-values and language referring to statistical significance

We prefer instead interval estimation, which conveys the precision of the estimate with respect to sampling variability

We are more open to testing with respect to modeling decisions, such as for tests of interaction (see editorial) and for tests for trend, and with respect to studies using high-dimensional testing, such as genome-wide association or other genomic platforms

JAMA | Original Investigation

Association Between Serotonergic Antidepressant Use During Pregnancy and Autism Spectrum Disorder in Children

Hilary K. Brown, PhD; Joel G. Ray, MD, MSc, FRCPC; Andrew S. Wilton, MSc; Yona Lunsky, PhD, CPsych; Tara Gomes, MHSc; Simone N. Vigod, MD, MSc, FRCPC

IMPORTANCE Previous observations of a higher risk of child autism spectrum disorder with serotonergic antidepressant exposure during pregnancy may have been confounded.

OBJECTIVE To evaluate the association between serotonergic antidepressant exposure during pregnancy and child autism spectrum disorder.

DESIGN, SETTING, AND PARTICIPANTS Retrospective cohort study. Health administrative data sets were used to study children born to mothers who were receiving public prescription drug coverage during pregnancy in Ontario, Canada, from 2002-2010, reflecting 4.2% of births. Children were followed up until March 31, 2014.

Editorial page 1533

Related article page 1553

Supplemental content

MAIN OUTCOMES AND MEASURES Child autism spectrum disorder identified after the age of 2 years. Exposure group differences were addressed by inverse probability of treatment weighting based on derived high-dimensional propensity scores (computerized algorithm used to select a large number of potential confounders) and by comparing exposed children with unexposed siblings.

RESULTS There were 35 906 singleton births at a mean gestational age of 38.7 weeks (50.4% were male, mean maternal age was 26.7 years, and mean duration of follow-up was 4.95 years). In the 2837 pregnancies (7.9%) exposed to antidepressants, 2.0% (95% CI, 1.6%-2.6%) of children were diagnosed with autism spectrum disorder. The incidence of autism spectrum disorder was 4.51 per 1000 person-years among children exposed to antidepressants vs 2.03 per 1000 person-years among unexposed children (between-group difference, 2.48 [95% CI, 2.33-2.62] per 1000 person-years; hazard ratio [HR], 2.16 [95% CI, 1.64-2.86]; adjusted HR, 1.59 [95% CI, 1.17-2.17]). After inverse probability of treatment weighting based on the high-dimensional propensity score, the association was not significant (HR, 1.61 [95% CI, 0.997-2.59]). The association was also not significant when exposed children were compared with unexposed siblings (incidence of autism spectrum disorder was 3.40 per 1000 person-years vs 2.05 per 1000 person-years, respectively; adjusted HR, 1.60 [95% CI, 0.69-3.74]).

CONCLUSIONS AND RELEVANCE In children born to mothers receiving public drug coverage in Ontario, Canada, in utero serotonergic antidepressant exposure compared with no exposure was not associated with autism spectrum disorder in the child. Although a causal relationship cannot be ruled out, the previously observed association may be explained by other factors.

Data collection is predetermined

- Not controlled by the researcher
- Research topic needs to suit the database
- Hard to know exactly how data were generated
- Very difficult to validate

Data collection is predetermined

- Limit the usefulness of coded diagnoses
  - Variation in coding
    - Between persons?
    - Between departments?
    - Institutions?
    - Over time: New coding
- Errors in coding
- Limitation in specificity in the available codes
- Bound to used definitions and administrative practices
  - ‘Administrators view of the world!’
  - Registers contain information on the citizens in relation to public administrators
  - Researchers distant from the actual data collection

Data quality

- Big issue – hard to evaluate
- Like religion?

Data quality

1. Completeness of registration of individuals

2. Validity of the information
   - Accuracy and degree of completeness of the registered data

Goldberg et al. Epidemiologic Reviews 1980;2:210-20.

Methods to evaluate completeness

- Compare sources
- Comprehensive records review
- Aggregated methods
- Capture – recapture

Example: Compare sources

- Danish Registry on Regular Dialysis and Transplantation (NRDT)
- Data on all Danish patients being actively treated for end-stage renal disease
- All patients in NRDT or NPR were included
- Incident patients in the period 2001-4 were identified in NPR
- NRDT was compared with NPR

Hommel et al. Nephrol Dial Transplant 2010;25:947-51

Table 1. Completeness of The Danish National Registry on Regular Dialysis and Transplantation (NRDT) compared to The Danish National Patient Registry (NPR)

	Number
Dialysis and renal transplantation procedures in NPR 2001–2004	
Number of patients receiving a renal transplant	653
Number of patients receiving dialysis at least 12 times and for at least 90 days with a ratio of days per dialysis of ≤7 or at least one code of chronic dialysis	3479
Incident RRT patients registered in NPR 2001–2004	
With dialysis at least 12 times for at least and 90 days and ratio of days per dialysis ≤7 days or at least one code of chronic dialysis or renal transplantation	3020
Receiving a renal transplant	185 (6.1%)
Receiving peritoneal dialysis	767 (25.4%)
Receiving haemodialysis	2068 (68.5%)
Incident chronic RRT patients in NPR 2001–2004	3020
Registered in NRDT with start of dialysis 2001–2004	2255 (74.7%)
Or registered in NRDT with start of dialysis 2000 or 2005	2421 (80.2%)
Or registered in NRDT with start of dialysis 1990–2006	2934 (97.2%)

Hommel et al. Nephrol Dial Transplant 2010;25:947-51

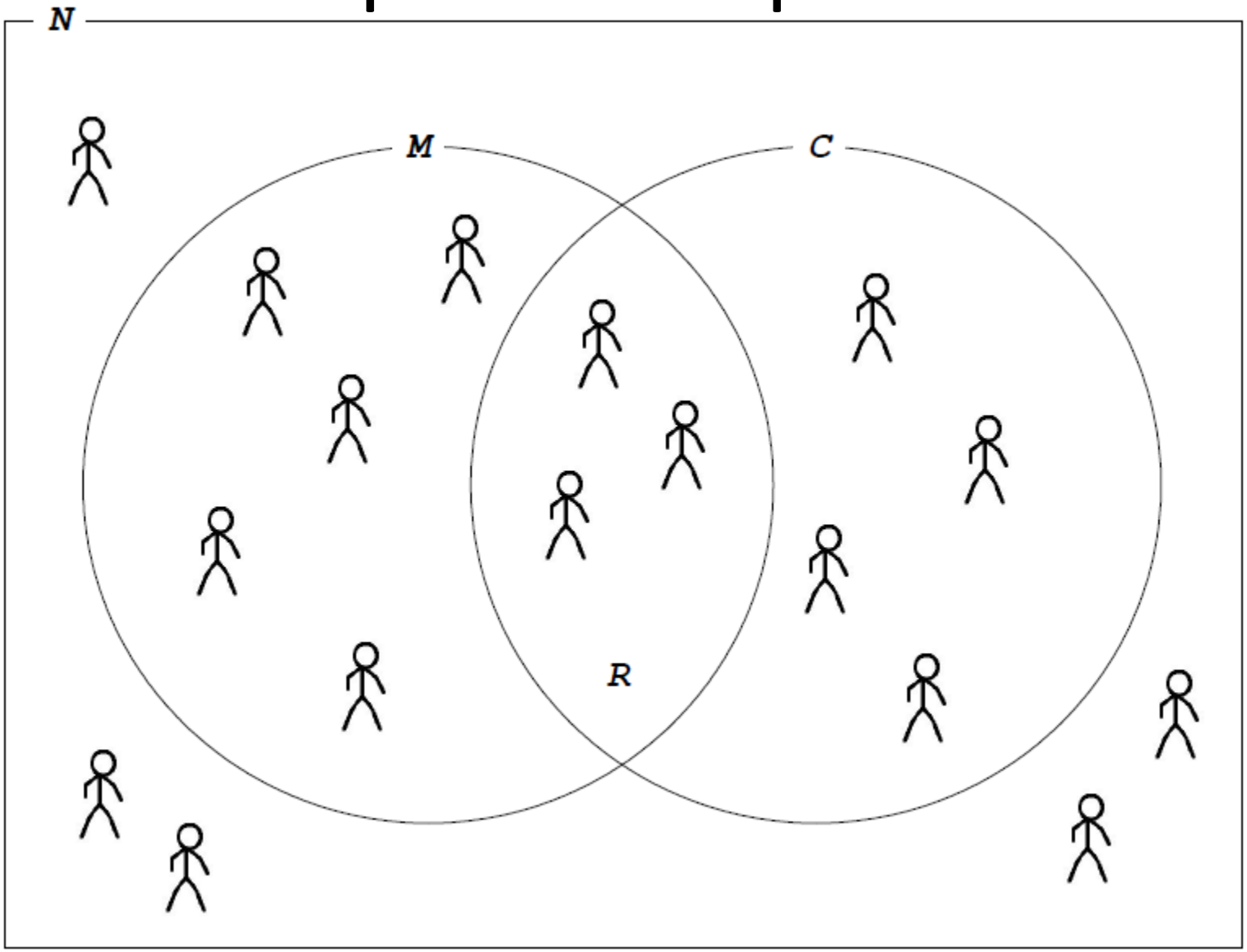
Methods to evaluate completeness

- Compare sources
- Comprehensive records review
- Aggregated methods
- Capture – recapture

Capture-recapture

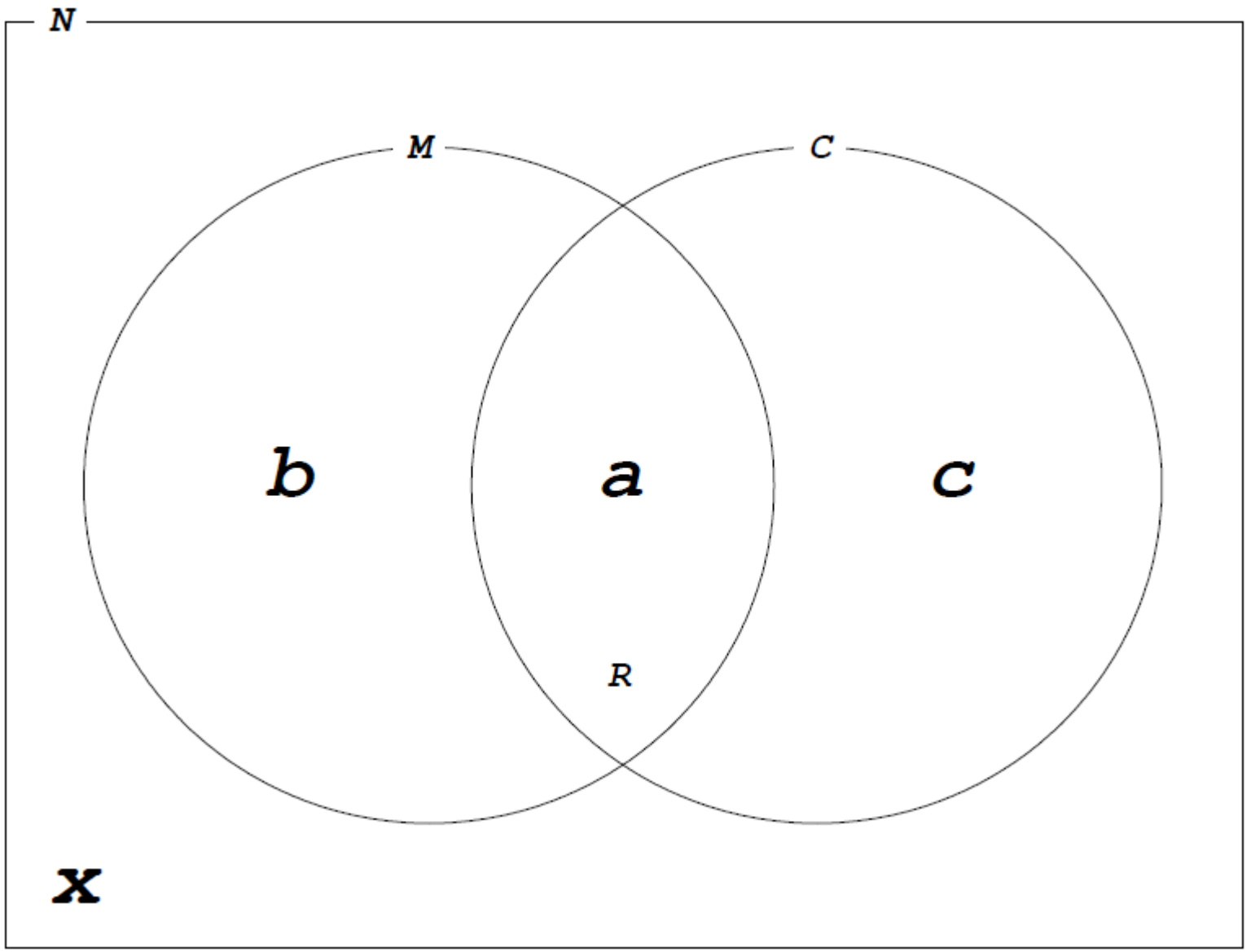
- Used to estimate the sensitivity of two case-finding methods
- The sensitivity of a case-finding method is a measure of how well the method performs at finding cases

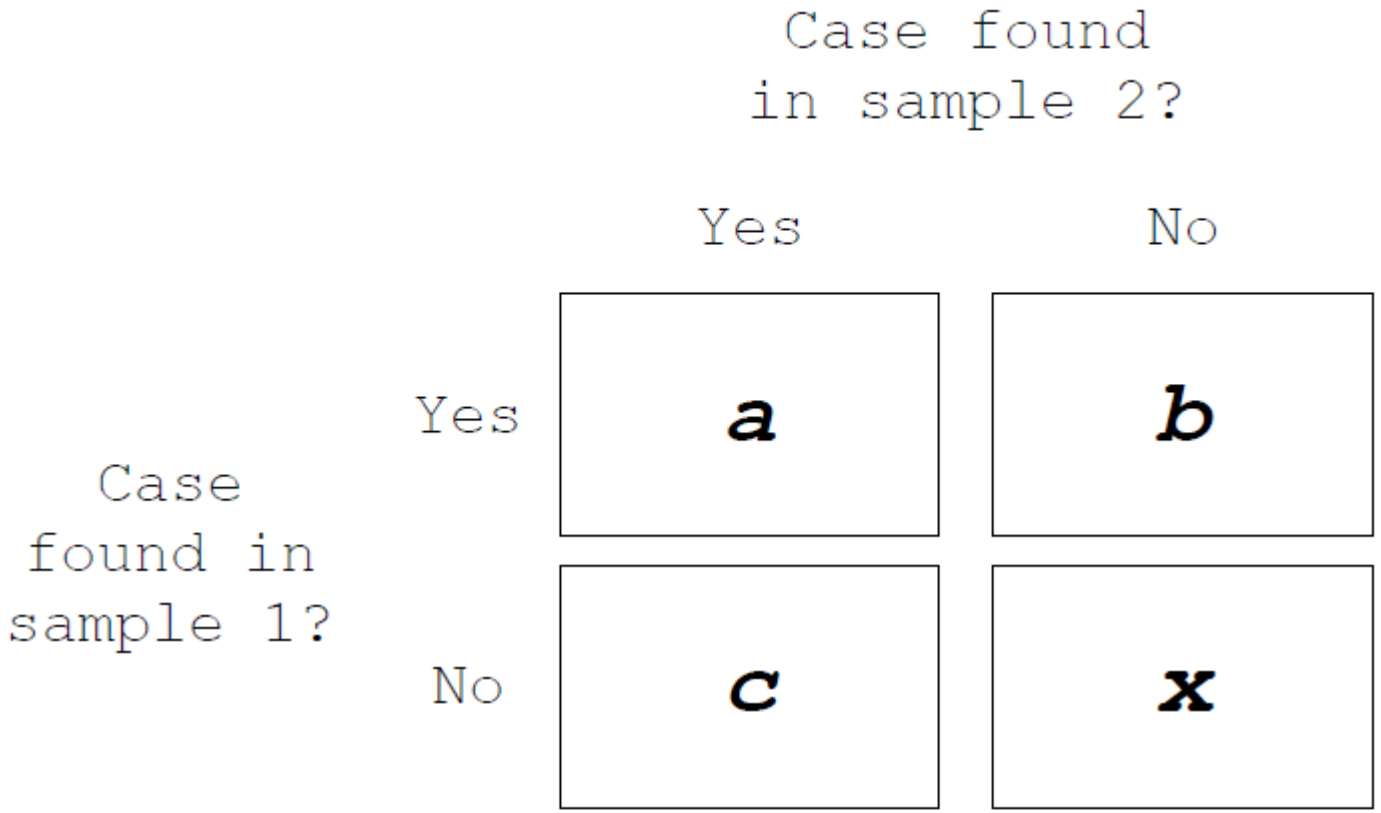
Capture-recapture



Capture-recapture

- One sample will usually be collected using a census (or census-like) sampling method
- The other sample will be collected using a rapid case-finding method


N
M
C
b
a
c
R
x



From *Figure 2:*

$$M = a + b$$
$$C = a + c$$
$$R = a$$
$$N = a + b + c + x$$

Capture-recapture

- If we assume:
  - The population is closed (i.e. there is no change in the population during the investigation)
  - The presence of a case in the second sample is not influenced by the presence of the same case in the first sample
  - Cases sampled on both occasions can be identified and matched
  - Each case has an equal chance of being included in each sample
- then we can calculate a value for the unknown x cell in the table

Capture-recapture

- Under these assumptions (especially the assumption of *independence*)
- the probability of a case being present in the second sample if it is present in the first sample:

  P(in sample 2 | in sample 1)= a / (a + b)

- and the probability of a case being present in the second sample if it is not present in the first sample:

  P(in sample 2 | **not** in sample 1)= c / (c+x)

- are the same

$$\frac{a}{a+b}=\frac{c}{c+x}$$

$$c\times(a+b)=a\times(c+x)$$

$$b\times c=a\times x$$

$$x=\frac{b\times c}{a}$$

Capture-recapture

- Knowing x allows us to estimate the total number of cases in the study population:

$$N = a + b + c + x$$

$$N = a + b + c + \frac{b \times c}{a}$$

$$N = \frac{(a+b) \times (a+c)}{a}$$

$$N = \frac{M \times C}{R}$$

Example

- A capture-recapture study found:
  - Cases found by central screening (***M***) : 30
  - Cases found by active case-finding (***C***) : 43
  - Cases found in both methods (***R***) : 22
- The total number of cases in the study population is estimated to be:
- N = M * C / R = 30 * 43 / 22 = 58.6 = 59

- Sensitivity of case-finding (%) = 43 / 59 = 73%

Remember period

- Registers cover events only during a defined time interval
- Events before not included (truncation)
- Right censoring
  - Events happening after end of registration
  - Handled by survival analysis techniques
- Interval censoring
  - E.g. many pharmacoepidemiological registers do not cover drug exposure during hospital admissions

Is completeness necessary?

- The demand depends on the research question
- In some analytical studies completeness may be less important than whether the misclassification is random or differential
  - Incomplete case ascertainment in follow-up studies could be critical
  - But may be less critical in case-control studies if case identification is unrelated with exposure of interest

Data quality

Two fundamental concerns:

1. Completeness of registration of individuals

2. **Validity of the information**
   - **Accuracy and degree of completeness of the registered data**

Goldberg et al. Epidemiologic Reviews 1980;2:210-20.

Validity

- Often the question: How high is the validity of register data
- Validity is the extent to which a variable measures what it is intended to measure
- Important measures
  - Sensitivity / specificity
  - Positive and negative predictive value
- Record review is often used for the validation

Validity of National Patient Register

- Validity epilepsy diagnosis in Danish National Patient Register (LPR)
- Randomly selected 200 patients with epilepsy diagnosis in LPR
  - 50 born before 1977
  - 50 born after 1977
  - 50 with a first diagnosis of epilepsy with complex focal seizures
  - 50 patients with a first diagnosis of primary generalized epilepsy
- Extracted information from medical records
  - Age
  - Gender
  - Date of first seizure
  - Date of first registration in LPR
  - Seizure type
  - EEG findings
  - CT/MRI findings
- One author classified the patients according to criteria

Christensen et al. Epilepsy Research 2007;75:162-70

Validity of National Patient Register

- Records from 57 departments at 41 hospitals
- Missing for 12 patients
- Epilepsy diagnosis confirmed for 153 patients (PPV = 81%)
- ICD-8: PPV = 84%
- ICD-10: PPV = 79%
- Specialized department: PPV = 83%

Christensen et al. Epilepsy Research 2007;75:162-70

Validity of National Patient Register

- Among 35 who did not meet criteria, 14 had one unprovoced seizure:
  - PPV (seizure disorder) = 89%
- Among the rest:
  - Observation for epilepsy (n=6)
  - Syncope (n=3)
  - Headache (n=2)
  - Myoclonia (n=2)
  - Bradycardia / respiratory problems (n=2)
  - Behavioral problems (n=2)
  - Ferebrile seizure (n=1)
  - Psychogenic seizures (n=1)
  - Dyskinesias (n=1)
  - Mental retardation (n=1)
- Complex focal epilepsy (28 out of 47 fulfilled criteria): PPV = 60%
- Primary generalize epilepsy (17 our of 48 fulfilled criteria): PPV = 35%

Christensen et al. Epilepsy Research 2007;75:162-70

Example: Ruptura uteri

- National Patients Register showed that 956 patients with ruptura uteri in the period 1980-1987
- Each gynecological department were contacted for medical records on these patients
- Careful eximantion showed that only 129 (14.1%) were correctly registered with ruptura uteri
- Instead registered with
  - Observatio pro/ruptura uteri imminens
  - Ruptura colli uteri
  - Ruptura vaginae
  - Ruptura perinei
  - Episiotomi

Devantier & Kjer. Ugeskr Laeger 1991;153:516-7

Ludvigsson "External review and validation of the Swedish national inpatient register."

- Launched in 1964 (psychiatric diagnoses from 1973)
- Complete coverage from 1987
- >99% of all somatic (including surgery) and psychiatric hospital discharges are registered
- January 2010 searched the medical databases, Medline and HighWire, using the search algorithm "validat* (inpatient or hospital discharge) Sweden“
- Contacted 218 members of the Swedish Society of Epidemiology and an additional 201 medical researchers to identify papers that had validated the register
- 132 papers were reviewed
- PPV was found to differ between diagnoses, but is generally 85-95%
- In conclusion, the validity of the Swedish IPR is high for many but not all diagnoses

BMC public health 2011;11: 450

Sund "Quality of the Finnish Hospital Discharge Register: a systematic review."

- The Finnish Hospital Discharge Register is one of the oldest individual level hospital discharge registers
- Since 1969
- Several reference databases were searched for validity articles published until January 2012
- Focus of validation, register years examined, number of compared observations, external source(s) of data, summary of validation results, and conclusions concerning the validity of FHDR were extracted
- 32 different studies comparing patient data to external information identified
- Most of the studies examined validity in the case of vascular disease, mental disorders or injuries
- PPV for common diagnoses between 75 and 99%
- Completeness and accuracy in the register seem to vary from satisfactory to very good in the register as long as the recognised limitations are taking into account

Scand J Public Health 2012;40(6): 505-515

Schmidt: "The Danish National Patient Registry: a review of content, data quality, and research potential."

- 114 papers, validating 1–40 codes/algorithms each and 253 in total
- PPVs ranged from below 15% to 100%.
- May result from different reference standards used

- Majority: Cross-sectional studies with medical record review as reference standard
- Other reference standards used:
  - Patient interviews
  - Danish Cancer Registry
  - Research database
  - Clinical registries
  - A military conscription system database
  - Danish prescription registries
  - Radiology reports
  - Clinical Laboratory Information
  - Danish National Pathology
  - Hospital pharmacy systems
  - GP verification
  - Autopsy reports

Clin Epidemiol 2015;7:449-90

Setting and calendar year

- PPV depends on the prevalence of disease
- Higher PPV in specialized departments
- Calendar year seems to increase quality, given the continuous improvement in diagnostic criteria and procedures used

PPV is dependent on prevalence

PREV=0.75						
	Syg	Rask				
Test=+	231	32	263		PPV	0.88
Test=-	27	54	81		NPV	0.59
	Se=0.90	Sp=0.63	344			
PREV=0.25						
	Syg	Rask				
Test=+	77	95	172		PPV	0.45
Test=-	9	163	172		NPV	0.95
	Se=0.90	Sp=0.63	344			

"Predictive values observed in one study do not apply universally"

Altman & Bland. BMJ 1994;309:102

Checklist of validation studies

- Title, keywords, abstract
- Introduction
- Methods
- Results
- Discussion

Benchimol. J Clin Epidemiol 2011;64:821e829

EPIDEMIOLOGY Announces the "Validation Study" Submission Category

Timothy L. Lash[a,b] *and Andrew F. Olshan*[c,d]

The editors of EPIDEMIOLOGY are pleased to announce a new manuscript submission category for validation studies, which we broadly define as studies that have the objective of improving the quality of the evidence obtained from other epidemiologic research. Although EPIDEMIOLOGY has published such studies[1] and previously encouraged such studies,[2] they have appeared rarely. We decided to make this category available to authors for two reasons. First, we hope that papers published in this category will enable access to the information required to support quantitative bias analyses by other authors or to otherwise increase confidence in other authors' research results. Second, we hope that a defined sub-

Epidemiology 2016;27:613-4

Clinical Epidemiology

Dovepress

open access to scientific and medical research

Open Access Full Text Article

EDITORIAL

Helping everyone do better: a call for validation studies of routinely recorded health data

This article was published in the following Dove Press journal:
Clinical Epidemiology
12 April 2016
Number of times this article has been viewed

Vera Ehrenstein[1]
Irene Petersen[1,2]
Liam Smeeth[3]
Susan S Jick[4]
Eric I Benchimol[5,6]

There has been a surge of availability and use for research of routinely collected electronic health data, such as electronic health records, health administrative data, and disease registries. Symptomatic of this surge, in 2012, *Pharmacoepidemiology and Drug Safety* (PDS) published a supplemental issue containing several reviews of validated methods for identifying health outcomes using routine health data,[1]

What to do next?

American Journal of Epidemiology
Vol. 138, No. 11
Printed in U.S.A.

Use of the Positive Predictive Value to Correct for Disease Misclassification in Epidemiologic Studies

Hermann Brenner[1] and Olaf Gefeller[2]

Misclassification problems of the disease status often arise in large epidemiologic cohort studies in which the outcome is classified on the basis of record linkage with routinely collected error-prone data sources, such as cancer registries or mortality statistics. If the misclassification is nondifferential, i.e., independent of the exposure status, this leads to bias toward the null in estimates of relative risk. A variety of methods have

Null-findings

Table 1. Influence of nondifferential misclassification of exposure on observed relative risk in a cohort study of 200,000 persons with 1,500 outcomes among exposed persons (incidence proportion = 1,500/100,000 = 1.5%) and 500 among unexposed persons (0.5%)

	Specificity					
Sensitivity	**0.60**	**0.70**	**0.80**	**0.90**	**0.95**	**1.00**
0.60	1.22 (82)	1.35 (73)	1.50 (63)	1.68 (53)	1.79 (47)	1.91 (41)
0.70	1.36 (72)	1.50 (63)	1.65 (54)	1.83 (45)	1.94 (40)	2.05 (35)
0.80	1.56 (60)	1.70 (52)	1.86 (44)	2.04 (35)	2.14 (31)	2.25 (26)
0.90	1.85 (44)	2.00 (37)	2.16 (30)	2.33 (23)	2.43 (19)	2.54 (15)
0.95	2.09 (33)	2.22 (27)	2.37 (21)	2.54 (15)	2.64 (12)	2.74 (8)
1.00	2.43 (19)	2.54 (15)	2.67 (11)	2.82 (6)	2.90 (3)	3.00 (0)

Thygesen. J Clin Epidemiol 2017;85: 17-20

What to do next?

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2005;**34**:1370–1376

doi:10.1093/ije/dyi184

A method to automate probabilistic sensitivity analyses of misclassified binary variables

Matthew P Fox,[1,2]* Timothy L Lash[2,3] and Sander Greenland[4]

Accepted 9 August 2005

Background Misclassification bias is present in most studies, yet uncertainty about its magnitude or direction is rarely quantified.

Methods The authors present a method for probabilistic sensitivity analysis to quantify likely effects of misclassification of a dichotomous outcome, exposure or covariate. This method involves reconstructing the data that would have been observed had the misclassified variable been correctly classified, given the sensitivity and specificity of classification. The accompanying SAS macro implements the method and allows users to specify ranges of sensitivity and specificity of misclassification parameters to yield simulation intervals that incorporate both systematic and random error.

Documentation / metadata

- Statistical metadata is descriptive information or documentation about statistical data
- Statistical metadata facilitates the sharing, querying, and understanding of statistical data over the lifetime of the data
- Increasing demand
  - The need for metadata in the statistical production has been increasingly evident
  - Most statistical offices are striving to introduce metadata systems, or improve existing ones

Documentation / metadata

- Great difference between surveys and registers
  - Surveys have their own data collection
  - Registers often collect information from administrators or from other administrative registers

- Register metadata
  - Industrial classification, product category, education, occupation and regional codes are examples of important classifications
  - These classifications change at regular intervals
  - The volume of the metadata can be very high
  - Every change must be documented

Data quality

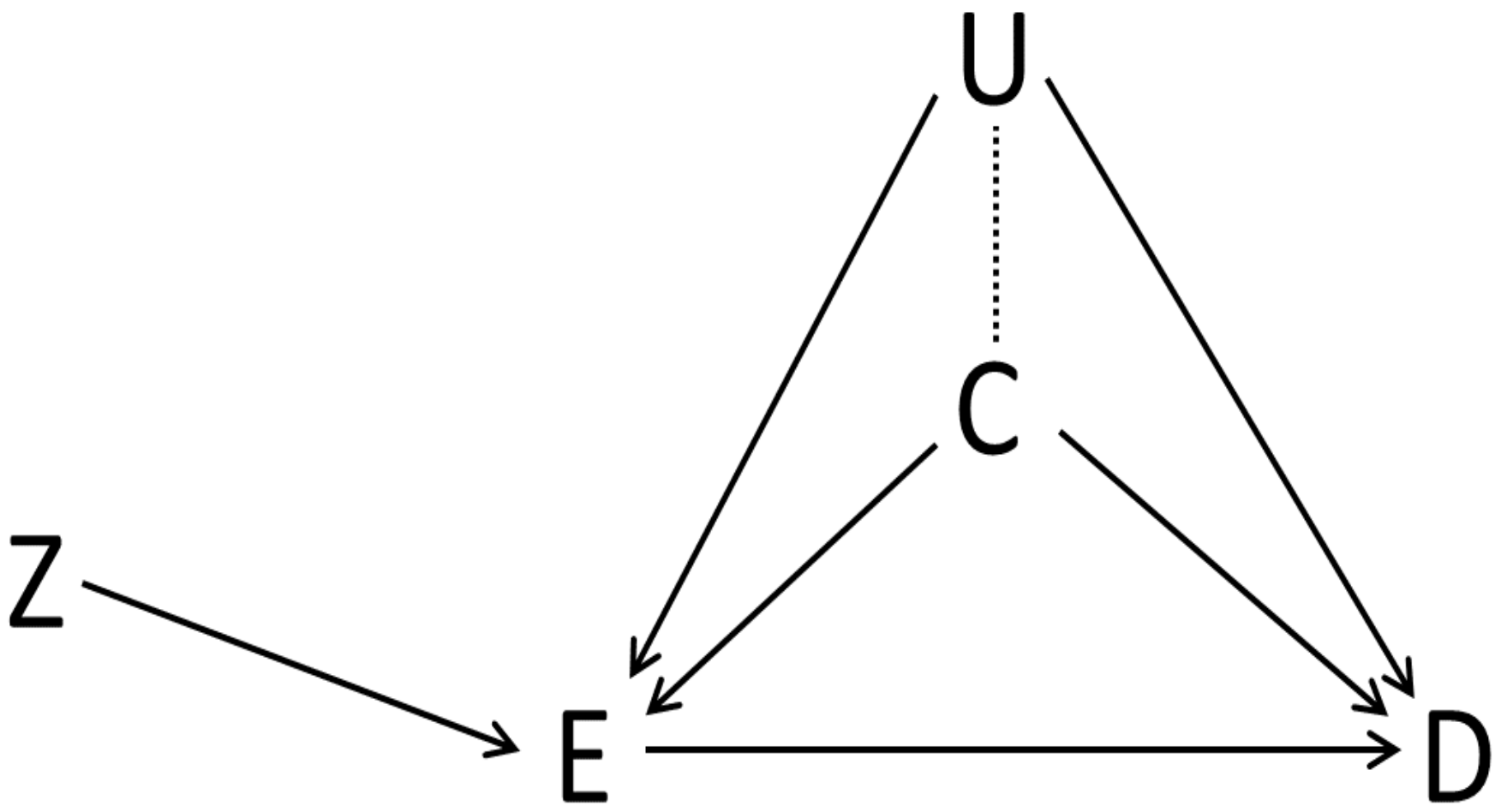
- Dependens on the research question!
  - Occurrence of disease at one specific time
    - Data should be complete
    - Further testing of positive recordings will reduce prevalence
    - Further testing of negative recordings will increase prevalence
  - Absolute occurrence of disease over time
    - Data should be complete
    - Otherwise, specificity and sensitivity should be the same over time

Sørensen et al. Int J Epidemiol 1996;25:435-42.

Data quality

- Relative rates in different populations
  - Non-differential misclassification may bias towards the null
  - Particularly for diseases with low specificity
    - Higher specificity → lower power
- Prognosis of disease
  - Diagnostic process must be unrelated with prognosis
  - The most severe cases are usually under closer scrunity

Sørensen et al. Int J Epidemiol 1996;25:435-42.


U
C
Z
E
D

’Adjustment’

- Want to make exposed and unexposed comparable
- Confounders that require detailed information on
  - clinical parameters
  - lifestyle
  - over-the-counter medications
- are often not measured in registers

- Causing confounding bias

Register-based studies

- Often few and unspecific confounders
- Combined with great statistical strength finding small effects
- Large risk of confounding bias

EUROPEAN UROLOGY 69 (2016) 883–884

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Platinum Priority – Editorial

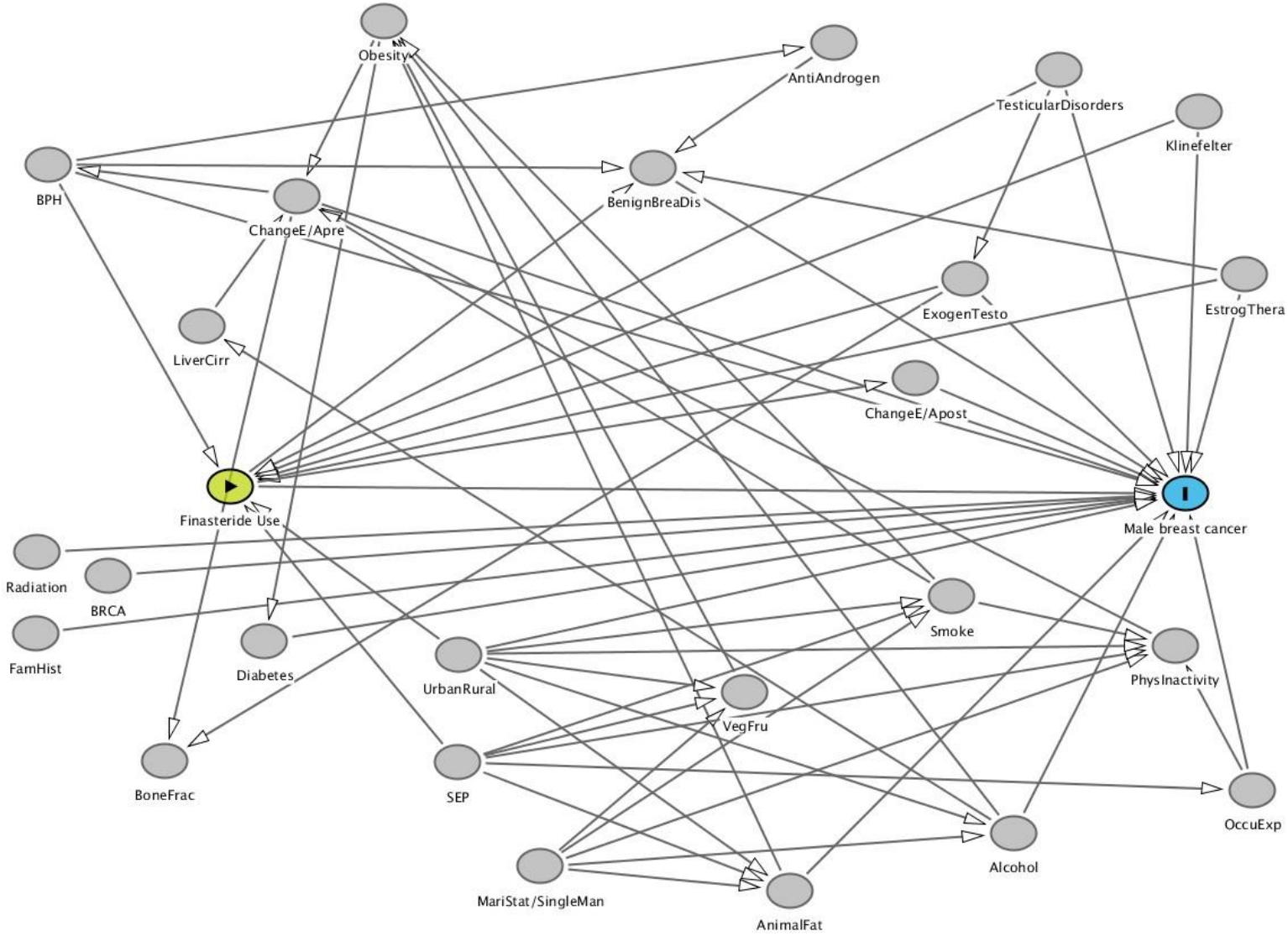
Referring to the article published on pp. 877–882 of this issue

Trouble in Paradise: Unmeasured Confounding in Registry-based Studies of Etiologic Factors

Jennifer R. Rider[a,b,*]

In practice

- How do you choose variables for the DAG?
  - Literature search
  - Experts
  - Look at associations in data
    - Change in estimate (10%)
    - Assocations with exposure and outcome
- Put arrows between vairables
- Becomes complex…


Obesity
AntiAndrogen
TesticularDisorders
Klinefelter
BPH
BenignBreaDis
ChangeE/Apre
ExogenTesto
EstrogThera
LiverCirr
ChangeE/Apost
Finasteride Use
Male breast cancer
Radiation
BRCA
FamHist
Smoke
Diabetes
UrbanRural
PhysInactivity
VegFru
BoneFrac
SEP
OccuExp
Alcohol
MariStat/SingleMan
AnimalFat

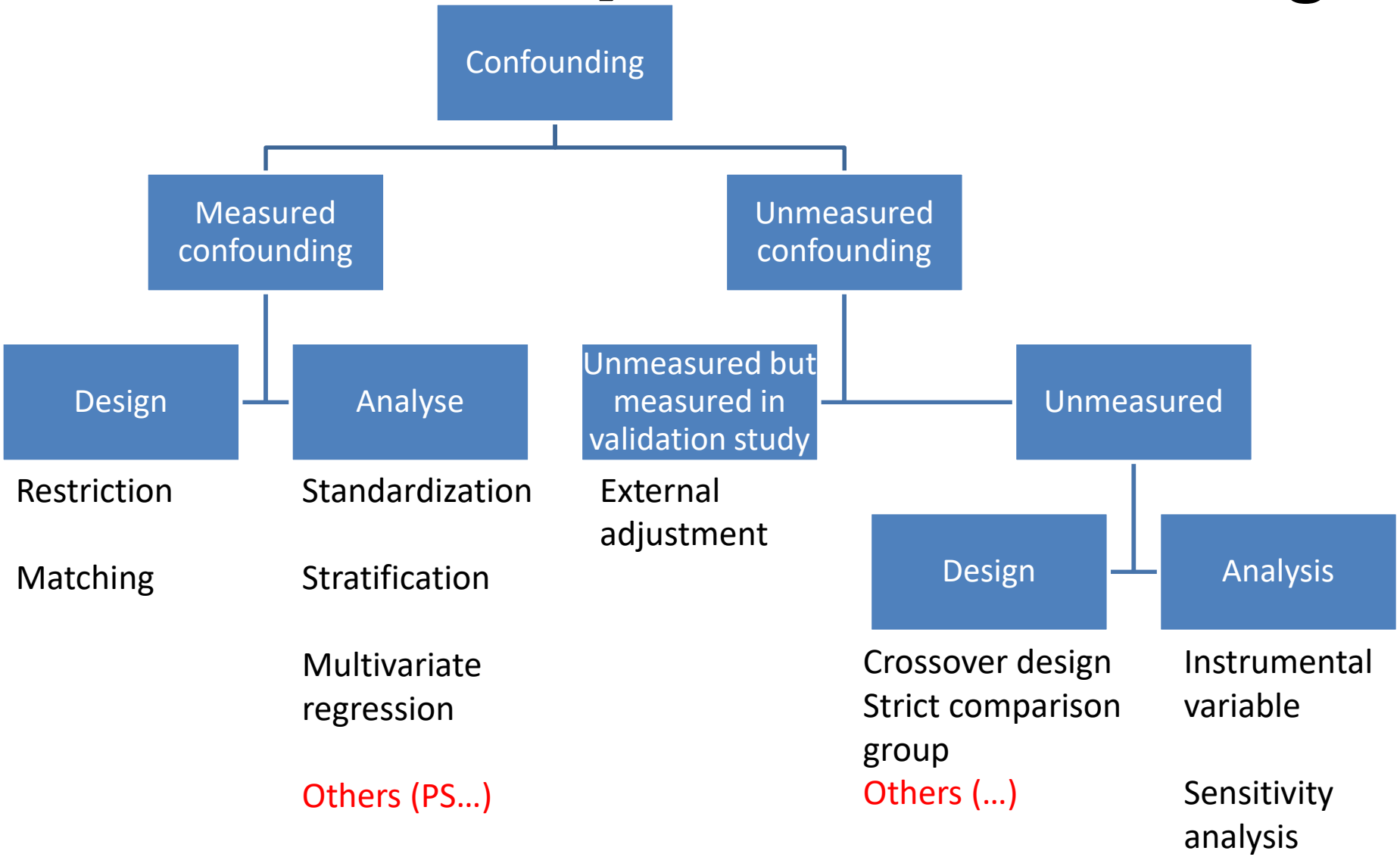
Directed acyclic graph (DAG)

- Often presented relatively simple in articles
- DAGitty is one program to handle complex situations
- Sufficient adjustment sets

Register-based studies

- Often few and unspecific confounders
- Combined with large statistical powoer to identify small effects
- Large risk for confounding bias
- Mimic randomized trial

Methods to adjust for confounding



Schneeweiss et al. Pharmacoepidemiology and drug safety 2006;15:291–303

Restriction

- Make patients more homogeneous
- Inklusion og eksklusionskriterier

Restriction

- Incident users
  - Like randomized studies
  - Do not adjust for intermediate variables
  - No general rules on operanilization
    - Stricht / loose definition
    - Newly marketed drugs
    - Reduced sample size

Active comparator

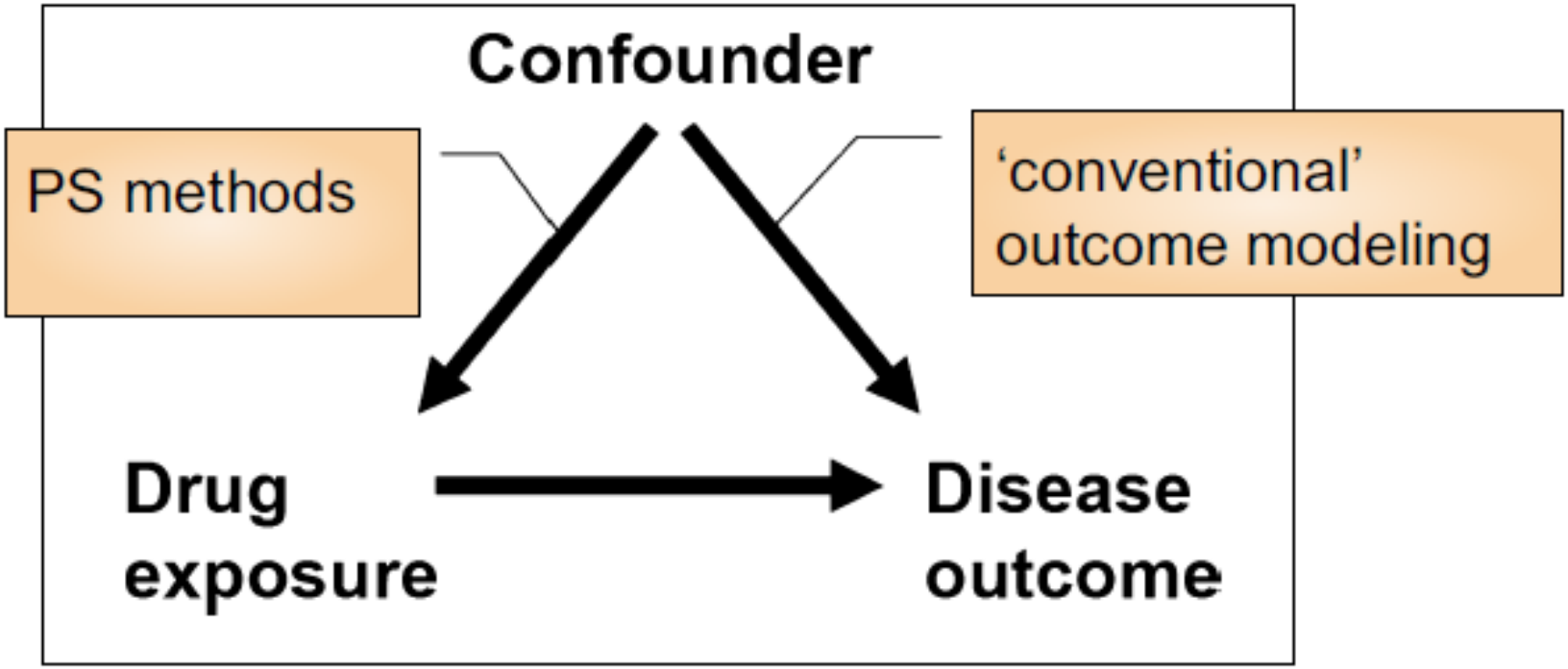
- Compare new users with new users of other drugs with same indication
- Same comorbidity / phase of disease
- Non-users are often different (non-comparable) and may have less contact with health-care system

Analysis phase

- Standardization
- Stratification
- Multivariate regression
- Others (PS…)

Propensity score

- Combining many covariates into a single variable (Miettinen 1976)
- Rosenbaum / Rubin 1983
- Popular in studies of drugs and medical procedures
- Estimate predicted probability (propensity) of drug use, based on characteristics
- Treatment effect measured among patients with same propensity


Confounder
PS methods
'conventional' outcome modeling
Drug exposure
Disease outcome

Propensity score

- Appealing
  - subjects with same PS have same chance of receiving treatment
  - *(assuming all relevant predictors of treatment are included)*
- Like a randomized trials
- Simultaneous control for many variables when small number of outcomes
  - But often used when outcome is common!
- Distribution of covariates similar btw treated and untreated subjects

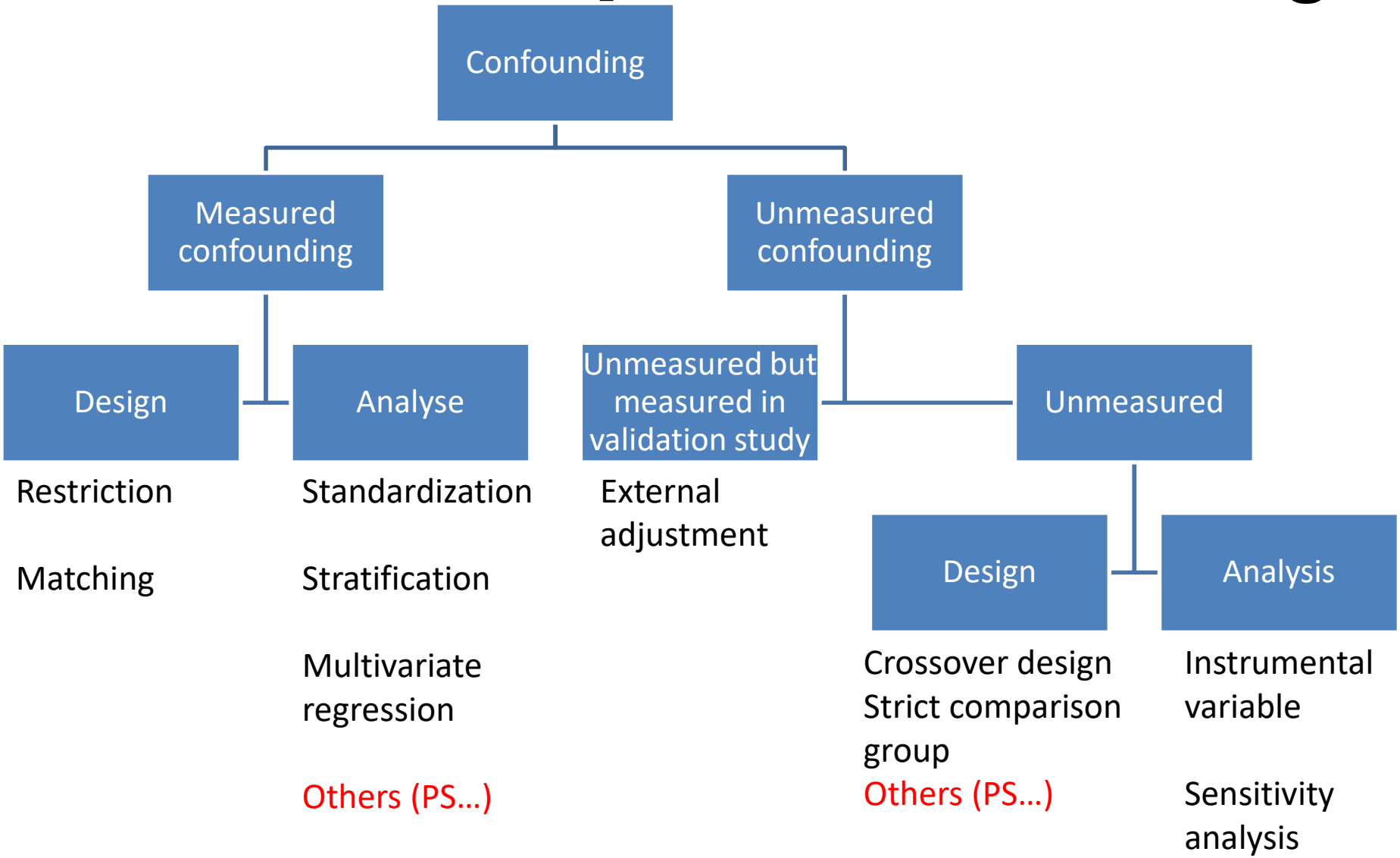
Brug af PS

- Matching
- Stratification
- Regression
- Weighting

PS matching

- Matche on only one variable
- Two groups similar on PS
- Should be checke for every covariates

Methods to adjust for confounding



Schneeweiss et al. Pharmacoepidemiology and drug safety 2006;15:291–303

External adjustment
Propensity score calibration

- Two propensity scores
  - Error-prone PS: confounders in the main study
  - Correct PS: confounders in both main and survey study

- Correct the error-prone PS in the main study

BZRD and cancer risk

Benzodiazepines widely used to treat anxiety and insomnia

Unclear association with cancer

Factors associated with use:

- Lifestyle factors
- Over-the-counter medications
- Clinical characteristics
- Comorbidity
- Treatments
- Self-rated health

Thygesen et al 2017 Br J Clin Pharmacol

Original study

Nationwide case-control study using risk-set sampling

Danish residents 18-85 years alive 2002 and followed until 2009

No cancer prior to index date

Data sources

Danish Cancer Registry

Danish National Prescription Registry

Danish Civil Registration System

Danish National Patient Register

Survey data

National representative Danish Health Interview Surveys (2000, 2005, 2010)

Participants aged 18-85 years (n=35,291)

Frequency matched (age+sex) (n=6,804)

BZRD use and self-reported information on potential confounders

Confounders

Register-based infor:

- **Prescriptions of drugs**
- **Diagnosis of diseases**
- **Charlson Comorbidity Index**

Survey information:

- **Education**
- **Self-rated health**
- **Self-reported comorbidities**
- **Self-reported drug use**
- **Smoking habits**
- **Alcohol intake**
- **Physical activity**
- **Body mass index**

Statistical analyses

Logistic regression models

Propensity score calibration:

Error-prone PS (X_{EP})

Correct PS (X_{GS})

Correct error-prone PS in the main study

Results

94,923 cancer cases

759,334 controls

681 long-term users

4,950 non-users

Survey – self-reported

		Non-user	User of BZRD		
			1 prescrip-tion	2+ prescriptions	
				<499 DDD	≥500 DDD
Self-rated health	Excel/very good	3,425 (69)	256 (57)	395 (55)	245 (36)
Drug use	Heart disease	572 (12)	83 (18)	161 (22)	191 (28)
	Pain	287 (6)	50 (11)	83 (12)	139 (20)
Smoking	Never	1,760 (36)	150 (33)	214 (30)	176 (26)
	15+ cig	710 (14)	62 (14)	98 (14)	141 (21)
Sedentary		663 (13)	86 (19)	158 (22)	227 (33)

Results

	All cancers	Smoking-related cancers	Alcohol-related cancers	Lung cancer
Age- and sex adjusted	1.22 (1.16-1.28)	1.27 (1.20-1.35)	1.08 (1.01-1.16)	1.70 (1.54-1.88)
Error-prone PS adj	1.16 (1.11-1.23)	1.20 (1.13-1.27)	1.06 (0.99-1.14)	1.48 (1.33-1.64)
Propensity score calibrated	1.09 (1.00-1.19)	1.10 (1.00-1.21)	1.03 (0.91-1.17)	1.23 (1.03-1.46)

Comparison group

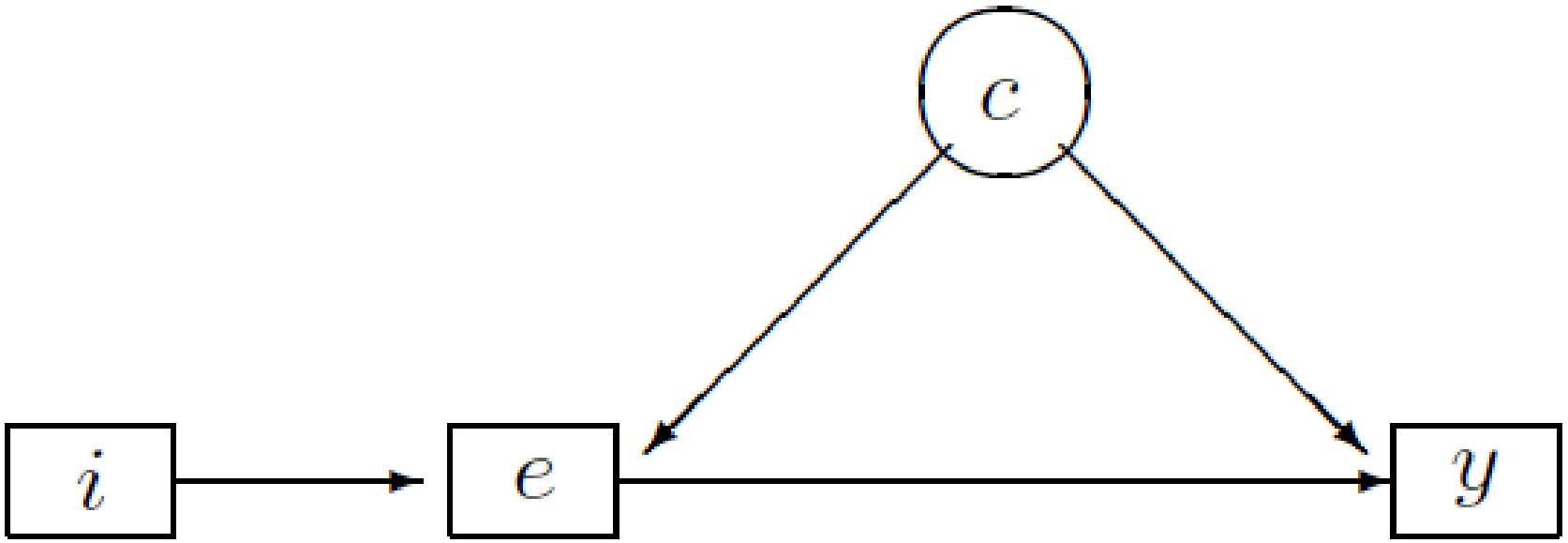
- Comparison of exposed group with a comparable but non-exposed group
- Thereby adjusting for unmeasured confounders
  - Age, sex
  - Education, income, labour-market affiliation
  - Health- and lifestyle behaviours

Comparison group

E.g.

- One occupational group with a comparable but non-exposed occupational group
- One patient group with another comparable non-exposed patient group

Instrumental variable



Example

- Register-based study of chemotherapy for advanced lung cancer (stage IV NSCLC) in the SEER tumor registry
- Aged 65+ and older
- Instrument
  - Unexplained geographic variation
  - Divided health care service areas into quintiles of chemotherapy utilization

Earle et al. J Clin Oncol 2001;19:1064–70

Assessed strength of instrument

- Instrument of geographic location predicted chemotherapy ($r^2 = 0.71$)
- Not independently associated with survival

Table 4. Instrumental Variable Analysis: Characteristics of Patients in the Lowest P_{chemo} Versus the Highest P_{chemo} Health Care Service Area Quintiles

Variable	Patients	
	Low P_{chemo} HCSA	High P_{chemo} HCSA
Patients*		
No.	1109	2059
%	21	39
Chemotherapy,† %	21	39
Mean age, years	72.5	72.7
Female, %	41	40
Non-white,† %	15	19
Without comorbidity,† %	78	73
SES quintiles, %		
1	20	20
2	19	20
3	20	20
4	20	20
5	21	20

Abbreviations: P_{chemo}, the probability of receiving chemotherapy based on geographic area of residence; SES, socioeconomic status.

*Total no. of patients = 3,168.

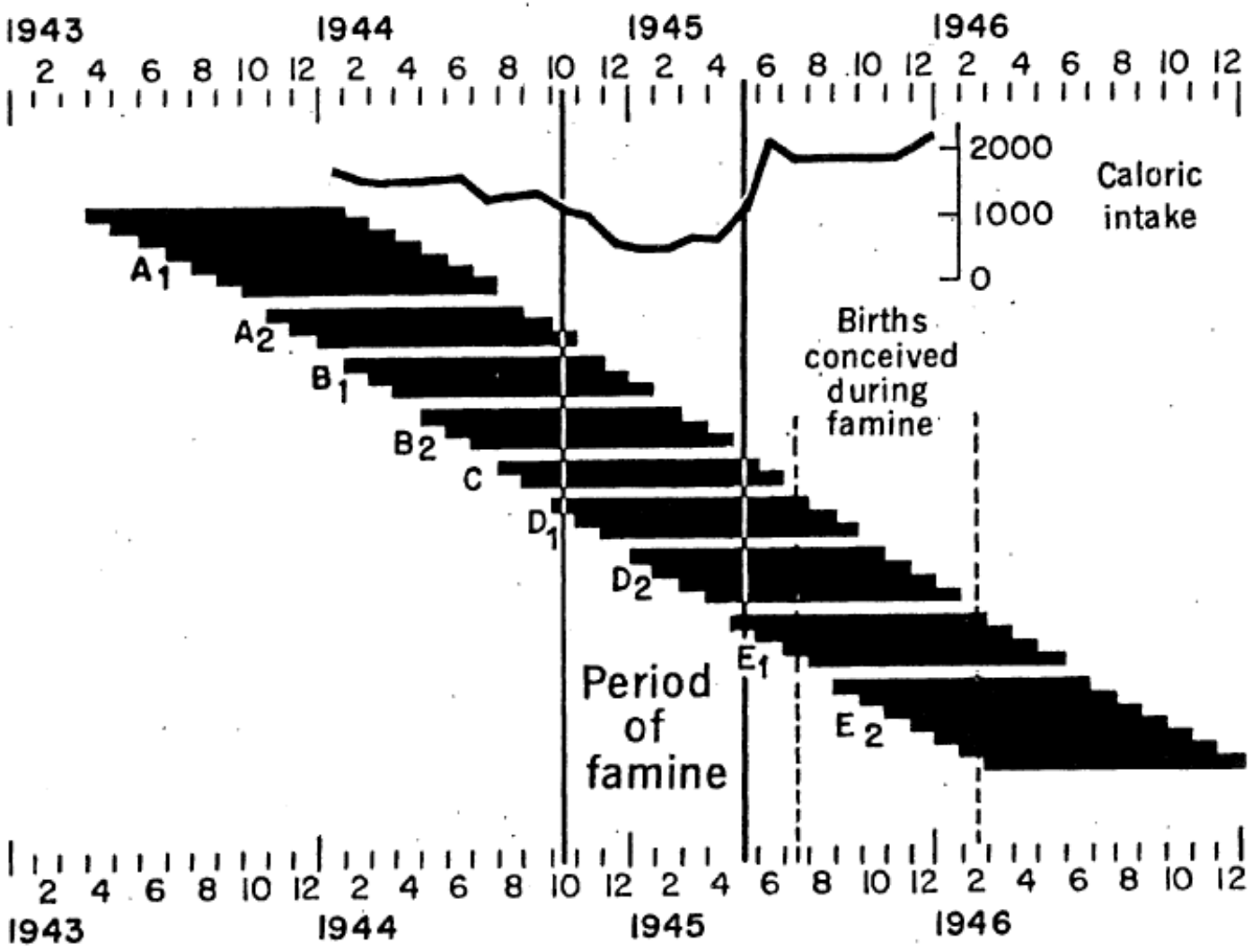
†$P < .05$.

Results

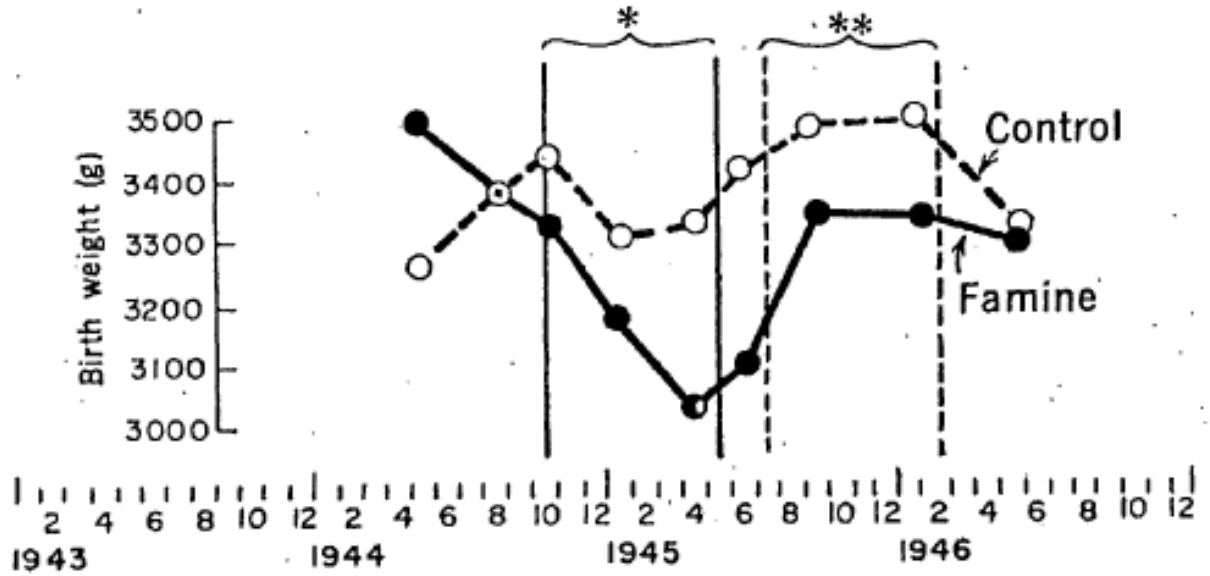
- Increase in median survival of 33 days (14-105) for patients treated with chemotherapy in the high utilization regions

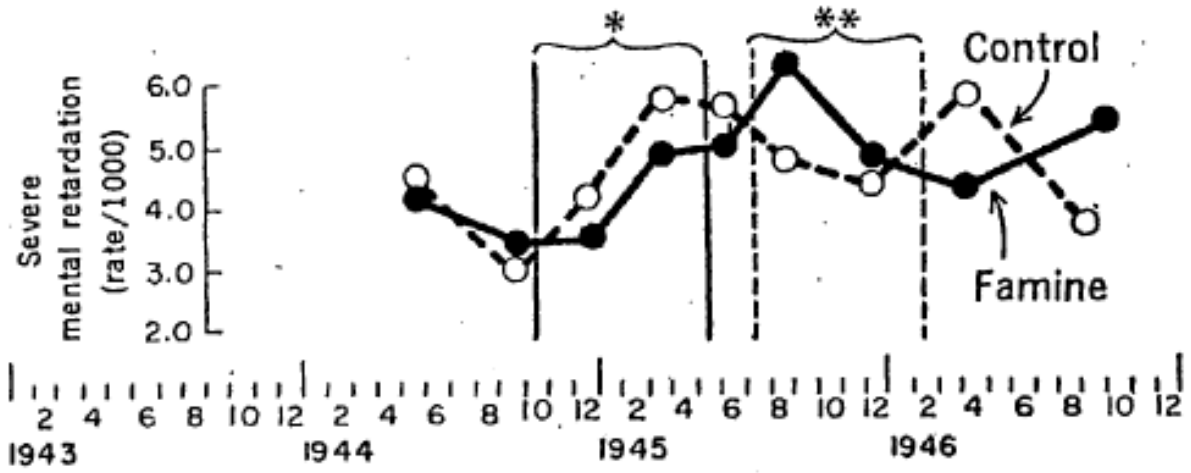
- 1-year survival was increased by 9% (4%-23%)

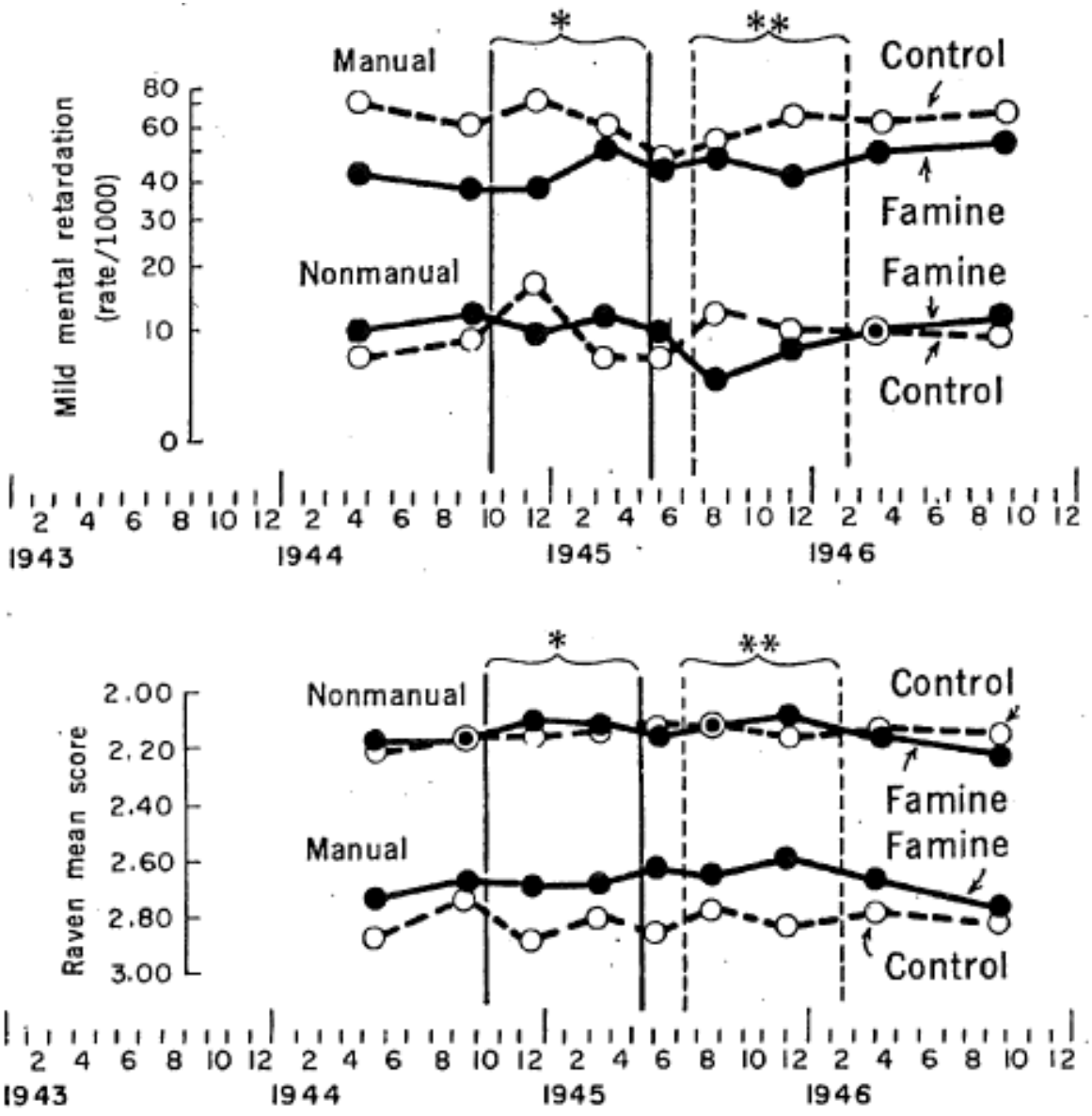
Nutrition and mental performance



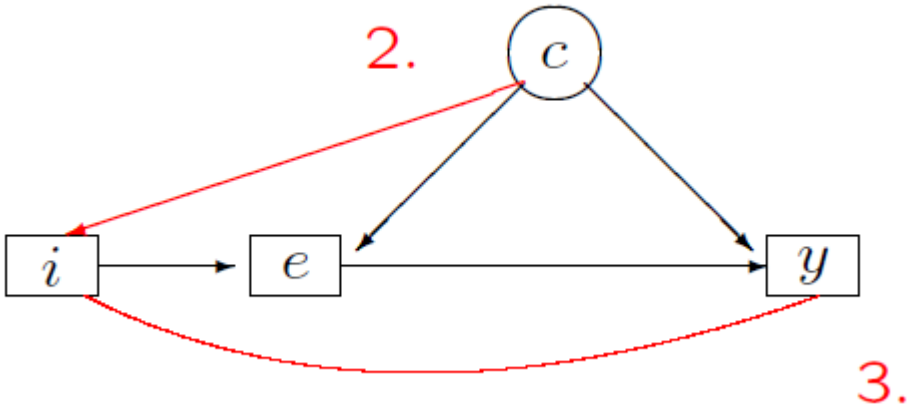
Stein et al. Science 1972;178:708-713


*
**
Birth weight (g)
3500
3400
3300
3200
3100
3000
Control
Famine
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12
1943
1944
1945
1946


*
**
Severe mental retardation (rate/1000)
6.0
5.0
4.0
3.0
2.0
Control
Famine
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12
1943
1944
1945
1946


Mild mental retardation (rate/1000)
80
60
40
30
20
10
0
*
**
Manual
Control
Famine
Nonmanual
Famine
Control
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12
1943
1944
1945
1946
Raven mean score
2.00
2.20
2.40
2.60
2.80
3.00
*
**
Nonmanual
Control
Famine
Manual
Famine
Control
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12
1943
1944
1945
1946

Instrumental variables - assumptions



- Fundamental weakness of method: Not possible to test in data

Case-only designs

- Use only cases and use the cases as their own controls
- Case-crossover design
- Case-time-control design
- Self-controlled case-series design

Designs on Tuesday

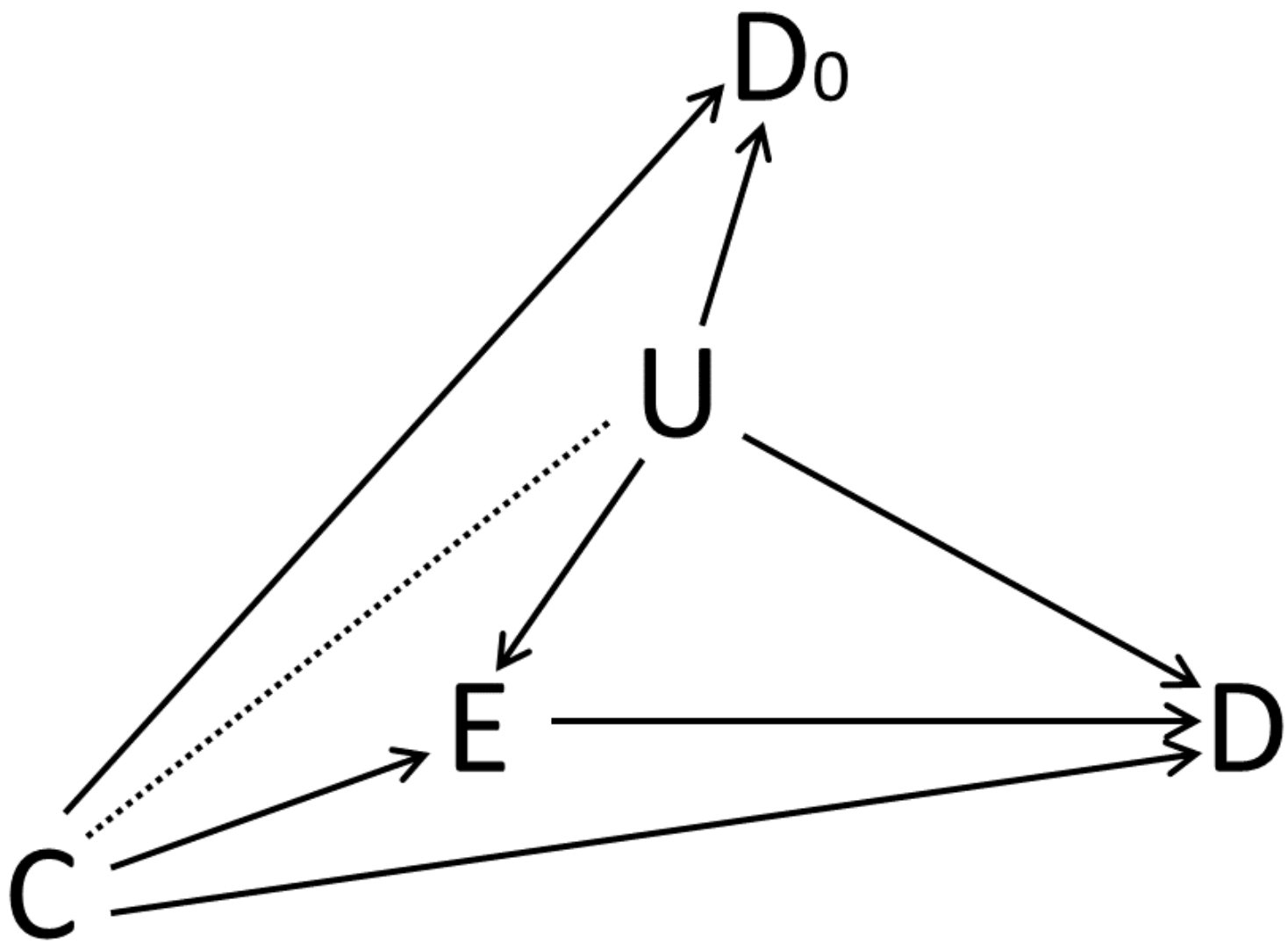
- Øystein on Tuesday

- **Natural experiments**
- Instrumental variable
- **Family**
- **Regression discontinuity**
- **Ignorance**
- **Interrupted time series**

- Closely related to the question of unmeasured confounding
- Aim to mimic randomization

Negative controls exposure or outcome

- In biologic laboratory experiments the use of ‘negative controls’ is a standard method
- Epidemiology: Suggested as method to evaluate unmeasured confounding
- Adherence users more healthy?


D0
U
E
D
C

Decreased all-cause and prostate cancer specific mortality among metformin users

- Negative outcome: Association between metformin use and cataract surgery
- Cataract surgery common elective surgery and is probably associated with health-seeking characteristics
- No association between metformin use and cataract surgery indicating that main analysis not influenced by unmeasured health-seeking characteristics
- Margel. J Clin Oncol 2013;31:3069-75

Another example

- The effect of flu shot receipt in patients before the flu season began found an association between flu vaccines and mortality
- Jackson. Int J Epidemiol 2006;35:337-44

A special case of negative exposures are using fathers as negative controls when studying the influence of maternal exposures

Sensitivity analysis

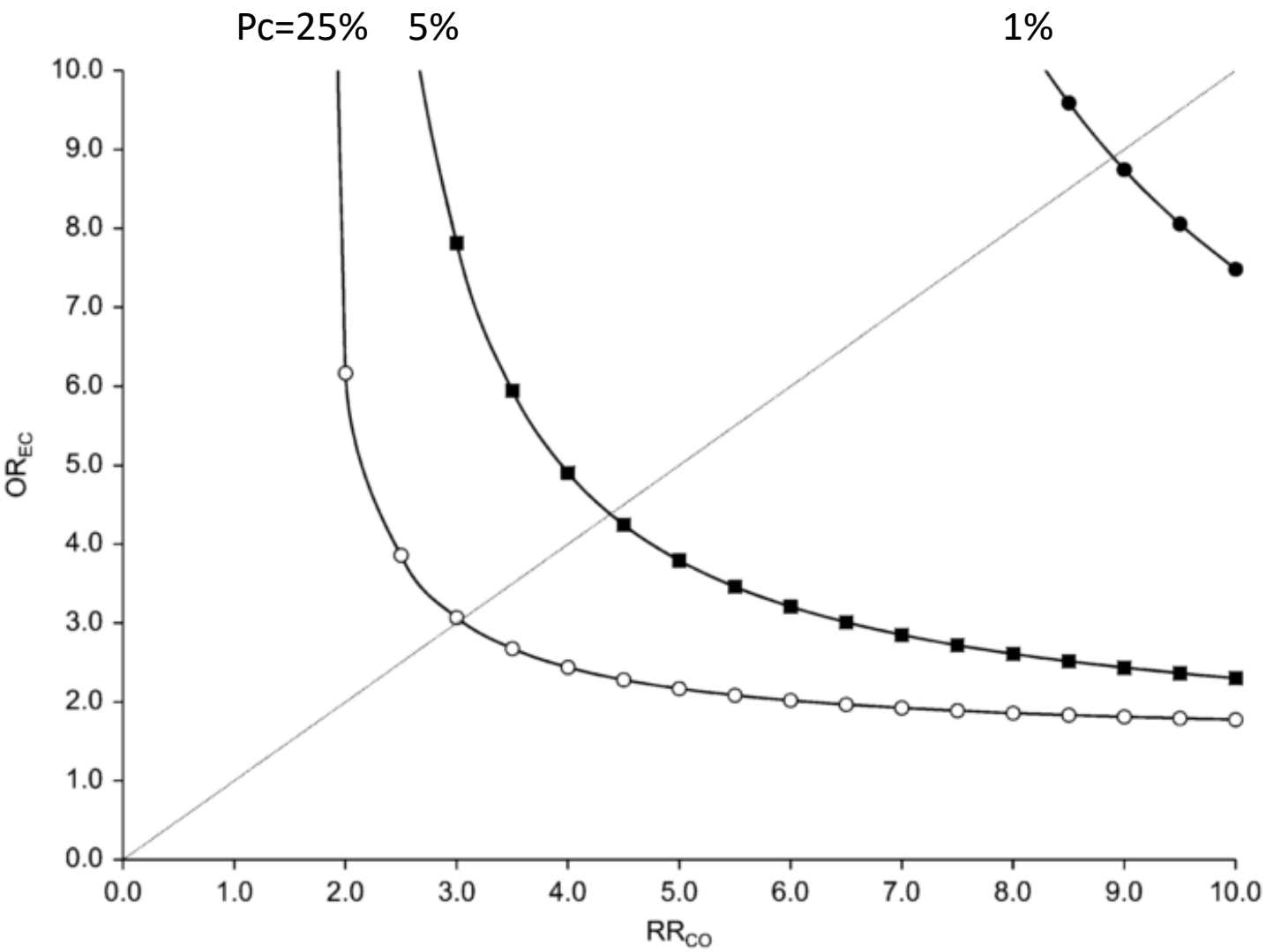
- Sensitivity analyses based on an array of informed assumptions

- Analyses to identify the strength of residual confounding that would be necessary to explain an observed exposure-outcome association

Parameters necessary

1. The strength of association between unmeasured confounder(s) and the event of interest in the unexposed group
2. The strength of the association between unmeasured confounder(s) and the exposure in the underlying study population
3. The prevalence of unmeasured confounder(s) in the underlying study population

Rule-out method



Estimated RR=1.34

Different confounder prevalences

Two areas

Limitation

- Susceptible to misuse
- If many scenarios considered, potential for conflicting results between sensitivity analysis
- Confounders do not operate in isolation
- Joint distribution can increase confounding
- In the real world, exposures are not dichotomized
  - And when does it act – timing and dose
  - Measurement error
- Sensitivity analysis is messy business

E-value

- Alternative approach to sensitivity analyses
- How strong would the unmeasured confounding have to be to negate the observed results?
- Requires no assumptions from investigators
- Intuitive
- Readily applied to the bounds of a 95% CI

$$\text{E-value} = RR + \text{sqrt}\{RR \times (RR - 1)\}$$

Limitations and Misinterpretations of E-Values for Sensitivity Analyses

Have limitations and are prone to misinterpretation

- No general rule can exist about what is a “small enough” E-value
- When there is several confounders
- Readers not familiar with how to interpret a range of E-values
- The automation of E-values may give an excuse not to think seriously about confounding
- Moreover, biases other than confounding may still undermine results
- Exposures often not dichotomized

Ioannidis, Tan, Blum. Ann Intern Med 2019;170:108-111

Use of comorbidity scores

- Health status may be an important confounder in many epidemiological studies
- Comorbidity adjustment has been proposed

- The simplest: Age
  - Unspecific but precise

- **Charlson Index**
  - Deyo CI (ICD-9-CM)
  - Dartmouth-Manitoba CI (ICD-9-CM)
  - Ghali CI (ICD-9-CM)
  - D'Hoore CI (ICD-9 three digits)

- **Chronic Disease Score**
  - Chronic Disease Score
  - Extended Disease Score

Schneeweiss et al. Int J Epidemiol 2000;29:891-8.

Table 3. Weighted index of comorbidity

Assigned weights for diseases	Conditions
1	Myocardial infarct
	Congestive heart failure
	Peripheral vascular disease
	Cerebrovascular disease
	Dementia
	Chronic pulmonary disease
	Connective tissue disease
	Ulcer disease
	Mild liver disease
	Diabetes
2	Hemiplegia
	Moderate or severe renal disease
	Diabetes with end organ damage
	Any tumor
	Leukemia
	Lymphoma
3	Moderate or severe liver disease
6	Metastatic solid tumor
	AIDS

Assigned weights for each condition that a patient has. The total equals the score. Example: chronic pulmonary (1) and lymphoma (2) = total score (3).

Charlson et al. J Chron Dis 1987;40:373-83.

Charlson comorbidity index

- Developed in 1987
- Chart review to predict 1-year mortality
- 604 patients admitted to medical service at New York Hospital during 1 month in 1984
- Validated using 685 breast cancer patients admitted to a Connecticut teaching hospital 1962-1969
- The final index: list of 19 conditions assigned weights
- Based on adjusted HR from Cox model
- Widely used
- Adapted to ICD-9 or ICD-10 codes in administrative databases

Conclusion Schneeweiss (2000)

- Scores provide only modest improvement on age adjustment
- Perform poorly because summarizes a complex construct
- May perform well in one setting and poor in another

Schneeweiss et al. Int J Epidemiol 2000;29:891-8.

Elixhauser

- Comorbidity index of 30 comorbidities defined using ICD-9-CM codes from administrative data
- Predictors of LOS and hospital charges
- No weighting and no index – 30 binary variables

Medication-Based Disease Burden Index

Chronic Disease Score (CDS)

- Medications instead of diagnostic codes
- Original CDS included 17 diseases and was validated against chart review and physician rating of physical disease severity
- CDS-2 updated medications (28 diseases) - weighting based on regression models

RxRisk

- RxRisk was developed as an all-age risk assessment instrument using outpatient pharmacy data to identify chronic diseases and predict future health care costs
- The RxRisk-V was a subsequent modification adapted to the Veterans Health Administration population

The end…

- Creative use of registers gives great possibilities
- Linkage with e.g. surveys and clinical data
- Always remember the limitations:
  - Predetermined data collection
  - Confounding
  - Validity / completeness
  - Truncation
  - Do not condition on the future
  - Data dredging