

PRINZIPIEN UND METHODEN DER KAUSALEN INFERENZ

Prof. Vanessa Didelez

Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS

IQWiG im Dialog – Juni 2019 - Köln

- Motivation: Studien mit GePaRD-Daten
- Kausale Forschungsfragen & kausale Parameter („Estimands“)
- Identifizierbarkeit – strukturelle Annahmen
 - Kein ungemessenes Confounding & Positivität
- Propensity Score: Balanciertheit / Gewichtung
- Methoden zur Confounderadjustierung
 - Outcome-Regression & Standardisierung
 - Stratifikation / Matching
 - Inverse Gewichtung
- Diskussion

BIPS: die GePaRD-Datenbank

- Pharmakoepidemiologische Forschungsdatenbank
 - Abrechnungsdaten gesetzlicher Krankenversicherungen
 - umfasst Informationen von > 20 Millionen Personen
(seit 2004 bei einer der beteiligten Krankenkassen versichert)

Typische Fragestellungen:

- Medikament A vs. B – Langzeit- / seltene Nebenwirkungen?
 - PASS-Studien, comparative effectiveness
- Aktuell: Evaluierung Koloskopie und Brustkrebsscreening
 - RCTs zu alt / lange, Mammographie / Therapien heute anders
- Auch: Off-label-Nutzung, vulnerable Patientengruppen
(Schwangere, Ältere mit Komorbiditäten) etc.

- Effekt einer (konkreten oder hypothetischen) Intervention
 - Screening anbieten oder abschaffen?
 - Medikament A statt B geben?
 - *Nicht*: “Wettersvorhersage” – Prädiktion, Risikogruppen etc.
- bzw. Entscheidungsproblem bzw. **Target Trial**
(Dawid, 2015; Dawid & Didelez, 2010; Hernán et al, 2008 ; Hernán, 2015, 2018)
 - Population
 - was für eine Intervention / Definition Exposition
 - Definition Outcome
 - “Nullzeitpunkt” etc.

⇒ systematische & stimmige Analyse von Beobachtungsdaten

⇒ Biasquellen vermeiden (immortality bias etc.)

(Lévesque et al, 2014; Suissa & Azoulay, 2012)
- Formal: welcher kausale Parameter („Estimand“)?

Notation

- X = Exposition (hier: binär – wenn auch oft unrealistisch!)
- Y = Outcome

- Kovariablen
 - W = alle gemessenen Kovariablen
 - C = Menge Kovariable, die hinreichend zur Konfounderadjustierung (beachte: nicht notw. eindeutig)
 - S = Subgruppen, für die Effekt von Interesse, z.B. Schwangere

Kausale Notation

- Kausale Inferenz – Wunsch:
Effekt von (konkreten / hypothetischen) Interventionen
- Mathematisch ausgedrückt:
mainstream: potential outcomes $Y(x)$ oder
hier: **do-Notation** $P(Y; do(X=x))$ plus **kausale DAGs** (Pearl, 2009)
 - Wichtig: Notation, so dass Zielparameter, strukturelle Annahmen etc. **überhaupt ausgedrückt** werden können!
- **Strukturelles** (kausales) Modell:
wie hängt Y von **Intervention** in X ab
 - also Modell für z.B. $E(Y; do(X=x))$ oder $E(Y|S=s; do(X=x))$

Kausale Parameter (Estimands)

- Entscheidungsträger?
- *Average causal effect* – in der gesamten Population (marginal)

$$ACE = E(Y; do(X=1)) - E(Y; do(X=0))$$

- *Subgruppeneffekt* (bedingter Effekt) – für $S=s$

$$CE(S=s) = E(Y|S=s; do(X=1)) - E(Y|S=s; do(X=0))$$

- Subgruppen: Frauen, Kinder, Schwangere...
 - Spezielle Subgruppen: „treated“ ($X=1$), „principal Stratum“ (latent)
- Oder anderer Kontrast (CRR, COR, abs. Diff. Survivalkurven...)

Kausale Parameter (Estimands)

- Statistische Modelle und ihre Parameter nur „Werkzeuge“
– kausaler Zielparameter separat davon festzulegen

- Festlegung besonders wichtig, wenn **Exposition zeitabhängig**
 - Zu vergleichende Behandlungsstrategien festlegen
„immer“, „niemals“, „Start wenn...“, „Stopp wenn...“, „Switch wenn...“
 - (Optimale) dynamische Strategien
(Robins et al, 2004; Dawid & Didelez, 2010; Chakraborty & Moodie, 2013)

- Exposition meistens zeitabhängig
 - auch in RCTs: non-compliance, switching, etc.

Identifizierung

Wann können (die verfügbaren) Daten überhaupt, im Prinzip (unendl. StiproUmfang) die Fragestellung beantworten?

Dazu: strukturelle Annahmen

- 1. Kein ungemessenes Confounding**
- 2. Positivität**

Auch, aber heute nicht:

- Consistency – Definition Exposition!
- Keine Interferenz
- und: in Praxis, (semi-)parametrische Annahmen

1. Kein ungemessenes Confounding

- Kovariablen - Erinnerung
 - W = alle gemessenen Kovariablen
 - C = Kovariablen, die hinreichend zur Confounderadjustierung (beachte: nicht notw. eindeutig)
 - S = Subgruppen

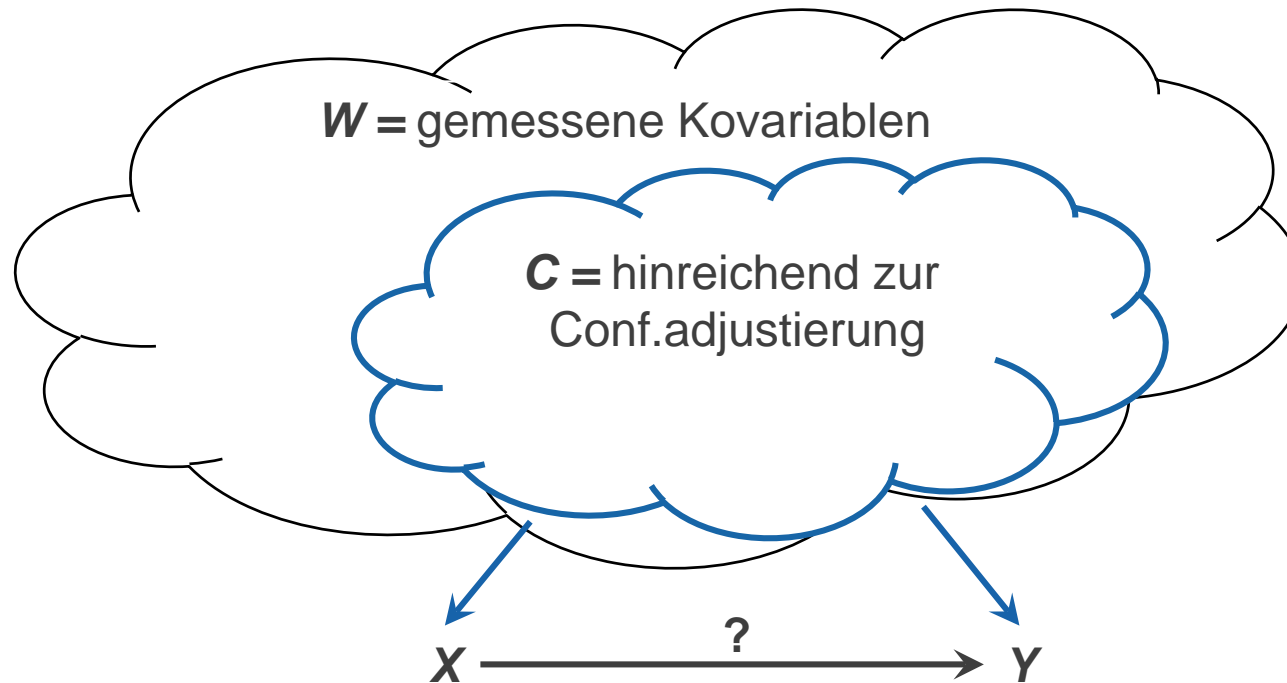
- Hinreichend zur Confounderadjustierung – formal:
$$X \text{ unabh. von } Y(x) \text{ gegeben } C$$

- Ungemessenes Confounding: W enthält **kein** solches C

- Im **zeitabhängigen** Fall: komplizierter!

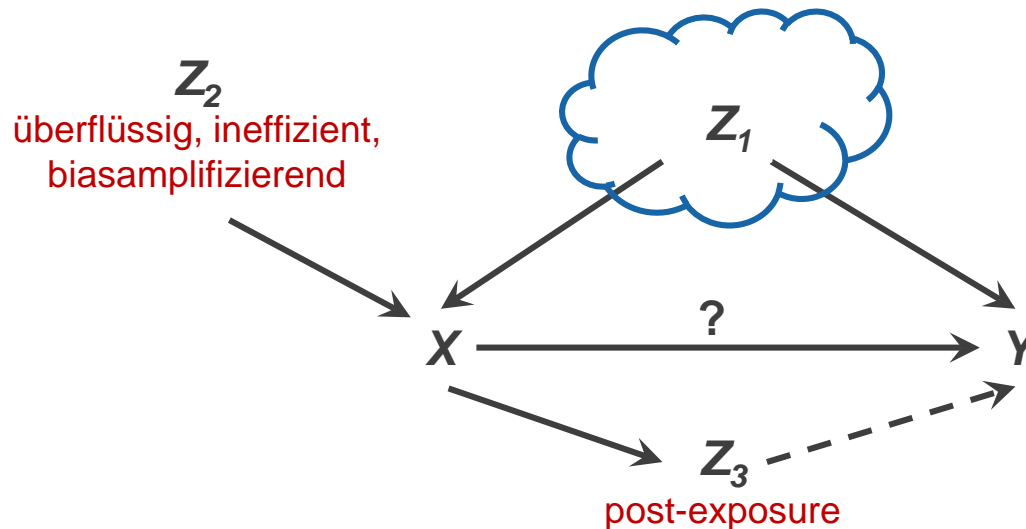
1. Kein ungemessenes Confounding

- Annahme: W enthält ein mögliches C
- Einfacher wäre: W *ist* selbst ein mögliches C



1. Kein ungemessenes Confounding

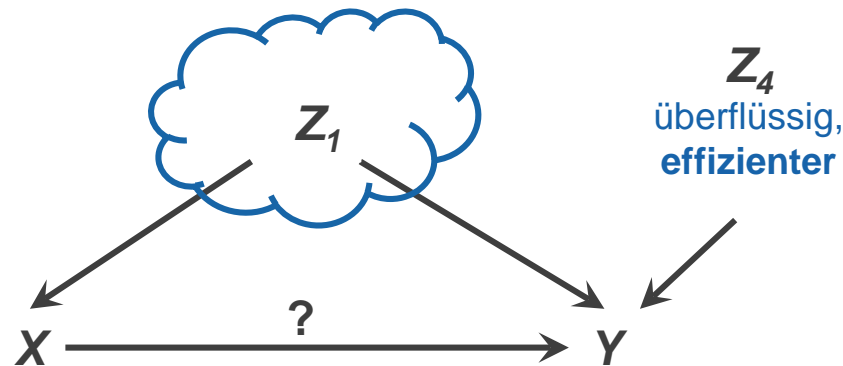
- Was kann schiefgehen? $C = Z_1, Z_2, Z_3 \dots ?$



- Bias-Amplifikation: *z.B. Pearl, 2011*
- Direkter kausaler Effekt: Thema für sich

1. Kein ungemessenes Confounding

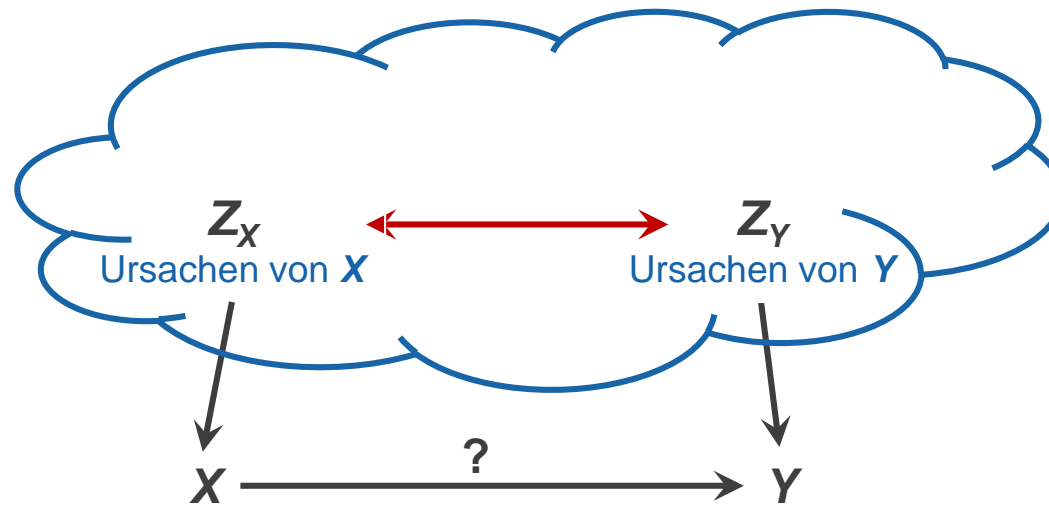
- Nicht so schlimm:



- Z_4 in C führt oft zu effizienteren Schätzern
 - außer bei Matching auf C – dann minimale Menge C wünschenswert
(DeLuna et al, 2011)

1. Kein ungemessenes Confounding

- Ohne Z_2, Z_3 – mögliches C ?



- Alle Ursachen von X oder Y : mehr als nötig, aber sicherer, wenn sonst kein strukturelles Vorwissen (*Vanderweele & Shpitser, 2011, 2013*)
- Beispiel: Z_Y enthält CRC-Risikofaktoren, Z_X familiäre Vorbelastung oder „healthy screenee“ Indikatoren

Beachte:

(Witte & Didelez, 2019)

- Confounderselektion **nicht ohne strukturelles Vorwissen** möglich
- Confounderselektion \neq Variablenselektion
 - Zusammenhang von C mit beidem, X und Y , muss berücksichtigt werden
(Shortread & Ertefai, 2017)
- Automatisierte Verfahren müssen annehmen:
 $W = \mathbf{ist}$ hinreichend zur Conf.adjustierung
 - Nicht: W **enthält** C
- In jedem Fall:
Wahl von C sollte besser begründet werden, als derzeit üblich
 - getrennt von interessierenden Subgruppen ($S=s$)

Propensity Score (PS)

(Rosenbaum & Rubin, 1983)

16

Definition - Propensity Score (PS)

$$PS_i = Pr(X=1 \mid C=c_i)$$

- Wahrscheinlichkeit für Person i (mit $C=c_i$) exponiert zu sein
- PS viele nützliche wichtige Eigenschaften
- Leider: in Praxis muss PS modelliert/geschätzt werden

- Wenn X **stetig, zeitabhängig** etc. PS nicht mehr so schön

2. Positivität

- Können nur Gruppen vergleichen, für die beide (alle) Expositionen *möglich* sind
 - inhaltlich begründen
 - empirisch überprüfen

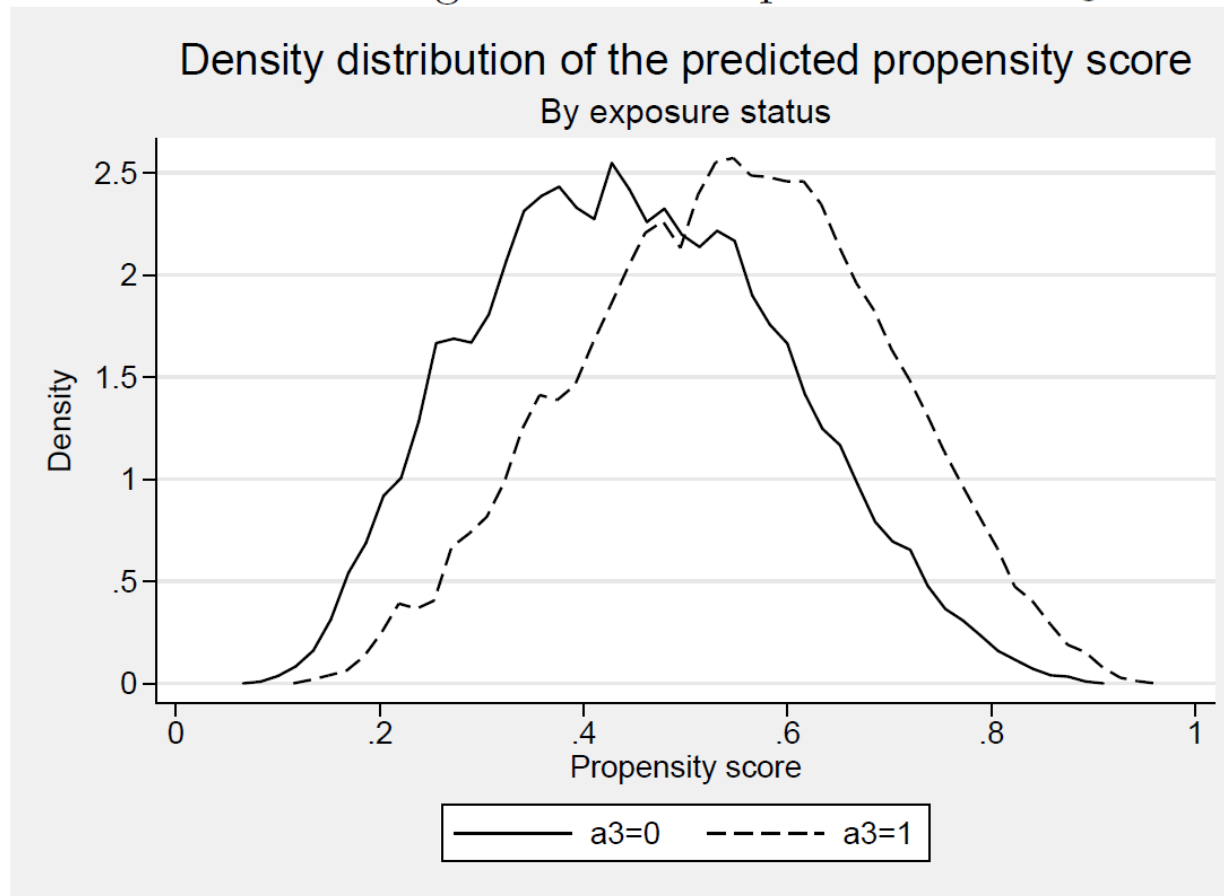
- Mit **Propensity Score**

$$0 < P(X=1 | C=c) < 1 \quad \text{für alle } c$$

- Prüfe: “overlap”
- Beispiel: PatientInnen mit Allergien würden nie Medikament B bekommen wg Kontraindikation \Rightarrow nicht in Vergleich aufnehmen

2. Positivität

- Beispiel – guter Overlap



2. Positivität

Beachte:

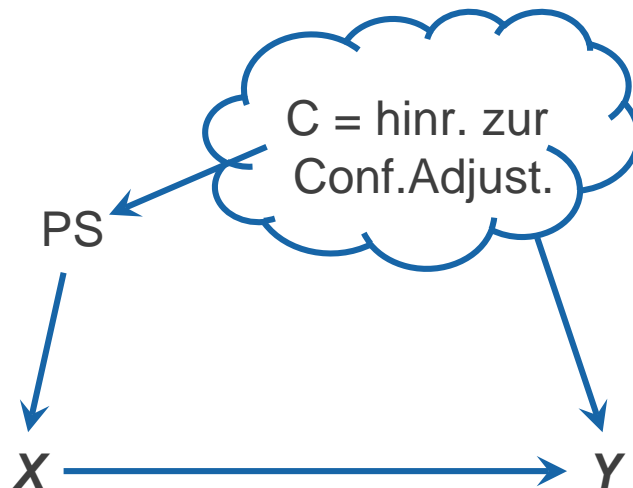
- Egal welche Methode zur Confounderadjustierung nachher verwendet wird, Positivität sollte immer überprüft werden!

- Wenn verletzt (und nicht bloß Zufall): revidiere Zielpopulation / Fragestellung
 - Auch: ungemessene Confounder („confounding by indication“) Grund für fehlende Positivität

 - Maschinelles Lernen: Charakterisierung der Überlappungsmenge wenn \mathcal{C} hoch-dimensional *(Johansson et al, 2019)*

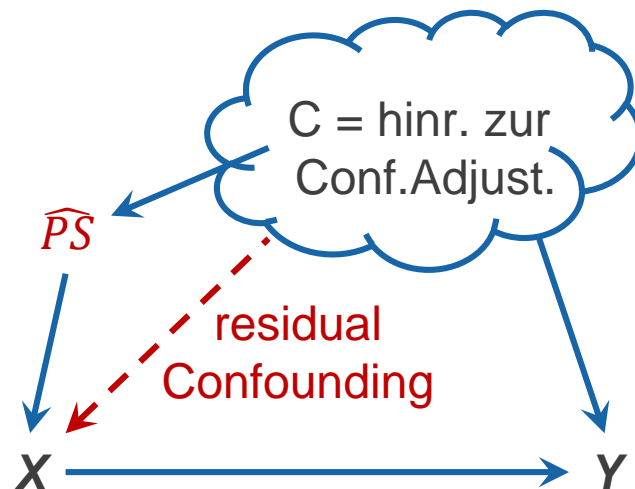
PS – Balancing Score

- Zentral: **Balancing**-Eigenschaft des Propensity Scores:
 - X unabhängig von C gegeben PS
- ⇒ verwende PS statt C ??



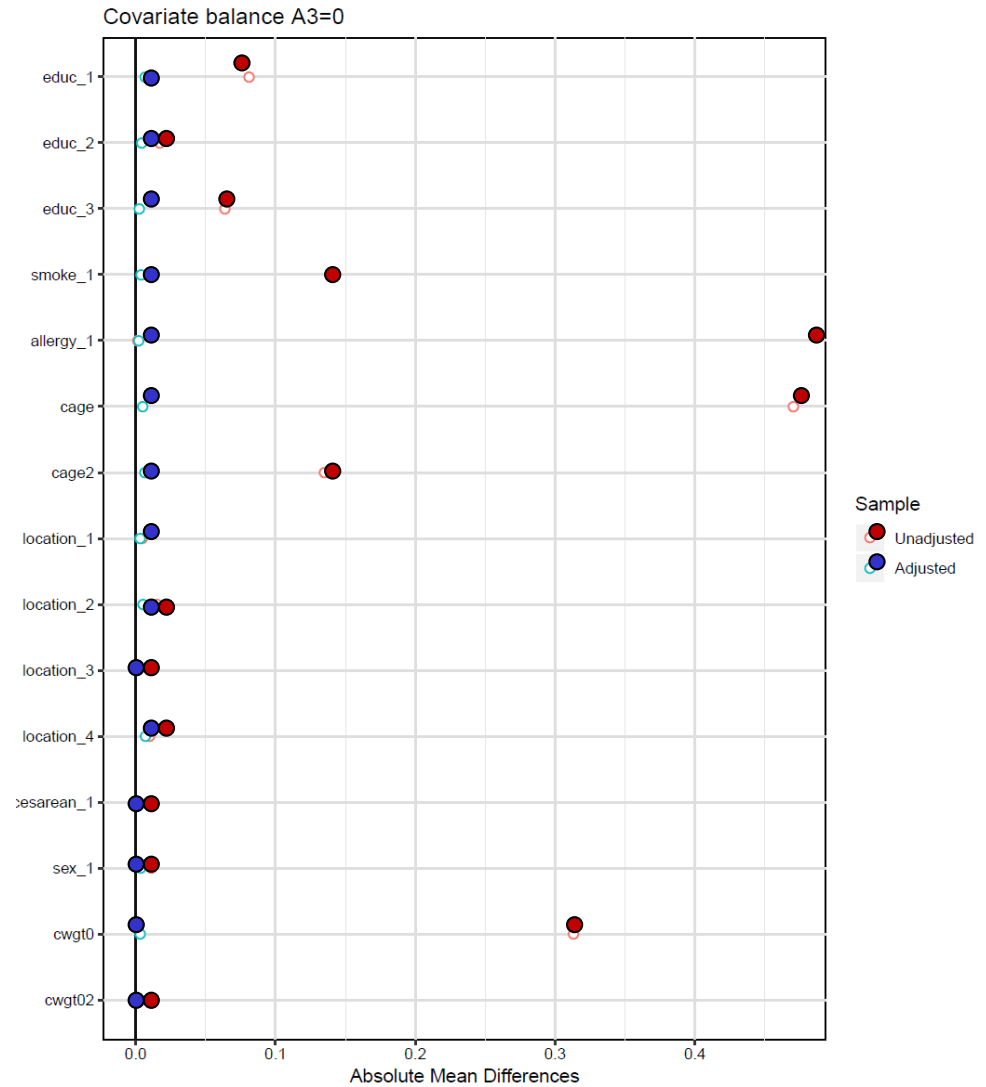
PS – Balancing Score

- Wenn PS geschätzt, prüfe Balancierung
 - bei gegebenem \widehat{PS} muss C in $X=1$ und in $X=0$ gleich verteilt sein
 - um Residualconfounding zu vermeiden
- ⇒ Balance-Diagnostiken (*Austin, 2009, 2019*)



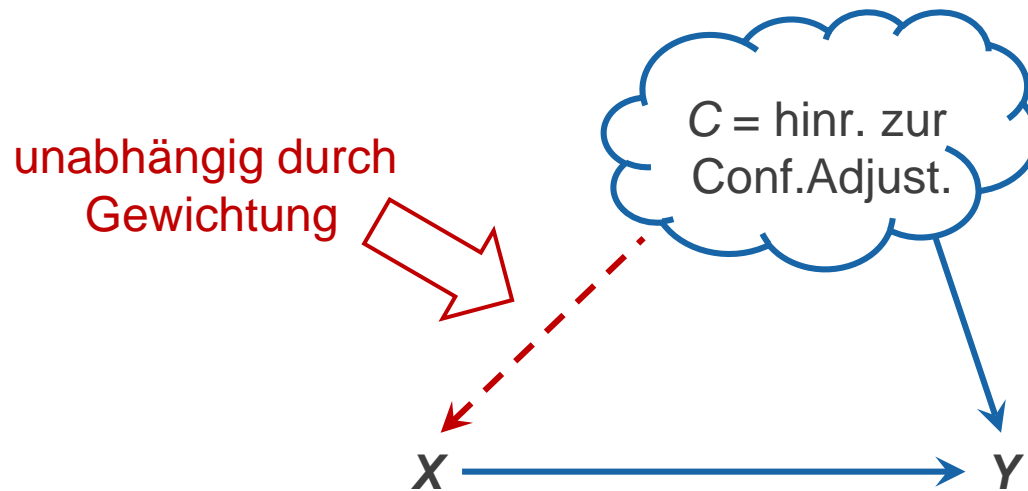
PS – Balancing Score

- Beispiel
abs. mean difference



PS – Inverse Probability Weighting (IPW)

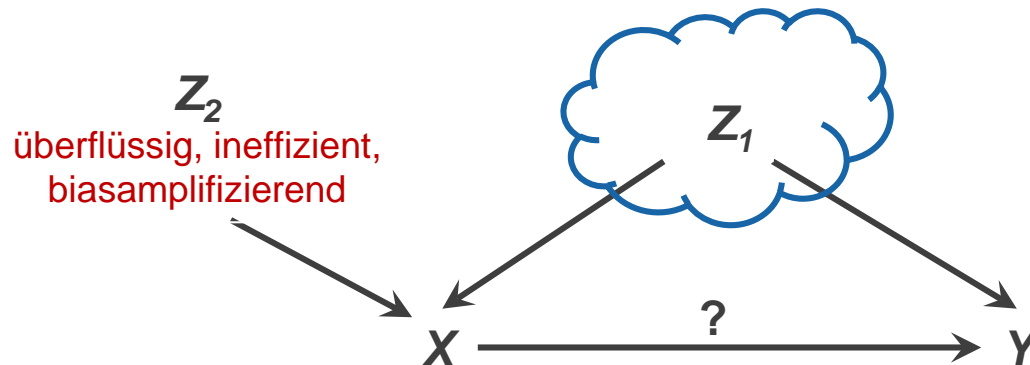
- Zentral – Gewichtung
 - PS^{-1} wenn exponiert,
 - $(1-PS)^{-1}$ wenn nicht exponiert
- in gewichteter Population: X unabhängig von C
 - kein Confounding in gewichteter Population



Erinnerung

PS-Modellierung

- Verführerisch, Variablen wie Z_2 zu benutzen...
 - sollte aber vermieden werden
 - nur C im Modell $P(X/C)$ nötig



- Beachte auch: Vorhersagegüte von PS-Modell irrelevant
 - „Machine learning“ für PS nur bedingt sinnvoll
 - nur **Balanciertheit** oder **korrektes Modell** X gegeben C wichtig.

- Outcome-Regression *mit Standardisierung*
 - mit Kovariablen \mathcal{C}
 - mit Propensity Score statt \mathcal{C}

- Stratifikation / Matching
 - auf Kovariablen \mathcal{C}
 - auf Propensity Score statt \mathcal{C}

- Inverse Probability Weighting (auch Bayesianisch)
 - durch PS bzw (1-PS)

- Doppelt robuste Verfahren
 - Kombination: Outcome-Modell & Expositionsmodell

Alle obigen Methoden benötigen zur konsistenten Schätzung:

- Kein ungemessenes Confounding (gegeben \mathcal{C})
 - alle benutzen selbes \mathcal{C}
- Positivität

- **Alle Methoden *äquivalent***, wenn saturierte (nicht-parametrische) Modelle und gleicher kausaler Parameter

- **Empfehlung: in Sensitivitätsanalyse zumindest mit einer anderen Methode vergleichen** *(Bsp: Abrahams et al, 2018)*

- Regression Y auf X und C
 - d.h. Modell z.B. für $E(Y|X,C)$
 - Modell muss korrekt spezifiziert sein
 - Standardisierung über empirische Verteilung von C für marginale Effekte (o. Subgruppen)

- Regression von Y auf X und PS (und evtl. S)
 - d.h. Modell für $E(Y|X,PS)$
 - Outcome-Modell und **PS-Modell** müssen korrekt spezifiziert sein
 - Standardisierung über PS-Vtlg.

Diskussion

- Warum Standardisierung?

$$E(Y; do(X = x)) = \sum_c E(Y|C = c, X = x) Pr(C = c)$$

- Standardisierung auch: bei Nicht-Kollabierbarkeit (z.B. odds-ratios)
- Erweiterung auf *zeitabhängige* Exposition: **g-Formel** (Robins, 1986)
 - zeitabhängige Confounder rausintegrieren
 - kein einzelnes Outcome-Modell mehr

Diskussion

- Zusätzl. Stratifizieren nach X (wenn binär) oft sinnvoll und flexible Modelle für $E(Y|X,C)$ oder $E(Y|X,PS)$ wünschenswert
- ...da korrektes Outcome-Modell zur konsistenten Schätzung des kausalen Parameters nötig
 - obwohl $Y-C$ - oder $Y-PS$ -Zusammenhang dabei nicht interessiert
- $E(Y|X,PS)$ selten verwendet; kann aber nützlich sein, wenn Y seltenes Outcome
 - im linearen Fall Robustheit: Outcome-Modell darf hinsichtlich PS fehlspezifiziert sein *(Vansteelandt & Daniel, 2014)*
- Gefahr der **Extrapolation** (separat Positivität prüfen!)

- Gewichtetes Mittel über Strata
 - Populär: Quintile des PS, besser mehr wenn N groß
(Lunceford Davidian, 2004)
 - Im Prinzip: Strata über \mathcal{C} , aber unpraktisch, wenn zu viele
 - Hier: PS muss \mathcal{C} innerhalb Strata balancieren!

- Matching auf individueller Ebene *(review: Stuart, 2010)*
 - Kriterium (z.B. nearest neighbour Mahalanobis) – angewendet auf \mathcal{C} oder PS
 - Prüfe **Balanciertheit**, sonst wiederhole
 - Stutzen: wenn kein Match zu finden ist – Positivität?

Diskussion

- Im Prinzip: nicht-parametrisch
 - Korrektes PS-Modell nicht nötig, **nur Balanciertheit** von C
(Waernbaum, 2012)
 - Balancierungs-Diagnostiken im Hoch-Dimensionalen?
- Positivität wird automatisch gesichert (sonst kein Match)
- Outcome wird zum Design nicht benötigt – Vor/Nachteil?
- Guidelines zu Matching beachten *(review: Stuart, 2010)*
 - PS-Matching kann verbessert werden durch vorgeschaltetes exaktes Matching auf wenige wichtigsten Confounder
 - PS-Matching-Paradox *(King & Nielsen, 2016)*
- Verallgemeinerungen auf stetige / zeitabhängige Expositionen nicht überzeugend

IP(T)W heißt

- Beobachtungen werden gewichtet durch PS^{-1} wenn exponiert, $(1-PS)^{-1}$ wenn nicht exponiert
 - in gewichteter StiPro ist X unabhängig von C – also kein Confounding
- **Marginales strukturelles Modell (MSM)** für gewünschten Parameter

z.B.
$$E(Y; do(X=x)) = a+bx$$

- Gewichte (also PS) schätzen
- Prüfe: **Balanciertheit in gewichteter Stichprobe**
- evtl. extreme Gewichte trunkieren / schrumpfen
- MSM mit gewichteter StiPro anpassen

Diskussion

- Positivität / „overlap“ prüfen!
 - extreme Gewichte trunkieren / schrumpfen – **Varianz/Bias-Trade-Off**
- PS-Modell gegeben C muss **korrekt** sein (nicht nur Balanciertheit)
- MSMs semi-parametrisch (Schätzgleichungen)
 - oft: große Standardfehler im Vgl. zu Outcome-Regression
 - **aber vielleicht auch ehrlicher?**
 - effiziente Wahl der Gewichte wünschenswert, Theorie schwierig

Diskussion

- IPW eigentlich motiviert durch zeitabhängige Expositionen
(Robins et al, 2000; Hernán et al, 2001)
 - lässt sich darauf erweitern
 - dann: *stabilisierte* Gewichte effizienter
 - Gewichte dann selbst auch zeitabhängig
 - dynamische Strategien
(Orellana et al, 2010; Gran et al, 2010; Hernán, 2018)
 - stetiges X schwierig...

Doppelt robuste Schätzung

(Robins & Rotnitzky, 2001)

35

- Erweiterung der IPW Schätzung – „augmented IPW“ = AIP(T)W
- Schätzgleichung für kausalen Parameter verwendet beides:
 - PS-Modell für Gewichte *und*
 - Outcome-Modell für $E(Y|X,C)$
- Robustheit: konsistente Schätzung wenn
 - entweder PS-Modell korrekt
 - oder Outcome-Modell korrekt spezifiziert
 - bzw. für hoch-dimensionales C : gewisse Spärlichkeitsannahmen
(*Belloni et al, 2017; Chernozhukov et al, 2017; Dukes et al, 2018; Athey et al, 2018; Smucler et al, 2019; Rotnitzky et al, 2019*)
 - Theorie nicht einfach
- Software? `iWeigReg` package (*Tan & Shu, 2013*)

Bemerkungen

- Standardfehler mit Bootstrap
 - außer bei Matching (*Abadie & Imbens, 2008*)
- Wahl zwischen (Varianten von) Methoden: **Varianz vs. Bias**
- Kritiken der Methoden beziehen sich oft auf unüberlegte / naive / falsche Verwendung
 - Wünschenswert: **Sensitivitätsanalyse mit mehreren Methoden**
- Balanciertheit (PS oder IPW) – kann nur an *gemessenen* Kovariablen überprüft werden
 - ⇒ hilft nicht bei **ungemessenem Confounding**
- Confounding aber nicht das einzige Problem
Ungeeigneter Analyseplan / Methoden liefern verzerrt Ergebnisse, selbst wenn kein ungemessenes Confounding (*Hernán et al, 2008*)

Ungemessenes Confounding?

- Beispiel:
comparative effectiveness studies – confounding by indication
 - sehr schwer in GePaRD abzubilden
 - Positivität oft stark verletzt

- Zusatzinformationen? 2-Phasen-Designs (*GePaRD: Enders et al, 2016*)

- Bias-Modellierung
(*Lash et al, 2011; Dorie et al, 2016; Vanderweele & Ding, 2017*)

- Instrumentelle Variablen? (*Greenland, 2000; Hernán and Robins, 2006*)
 - Präferenz des Arztes für ein Medikament
(*Brookhart et al, 2006; Kollhorst et al, 2016*)
 - Mendelsche Randomisierung (*Didelez & Sheehan, 2006*)

Fazit

- Es gibt keine Garantie für kausal belastbare Ergebnisse....
 - Aber wenn wir nicht...
 - zunächst eine kausale Forschungsfrage formulieren, Exposition (als hyp. Intervention) und Outcome entsprechend definieren – motiviert durch **Entscheidungsproblem / target trial**
 - strukturelle Annahmen diskutieren und rechtfertigen (ungemessenes Confounding, Positivität – plausibel?)
 - eine geeignete (robuste) Methode verwenden
- ⇒ **vermeidbaren Bias vermeiden**
- Sensitivitäts- / Biasanalysen durchführen**
- ⇒ dann sind die Ergebnisse wohl eher nicht kausal belastbar und nicht mit RCTs vergleichbar
- **No causes in – no causes out!** *(Cartwright, 1994)*

Danke für die Aufmerksamkeit!

Übersicht: Goetghebeur et al - für STRATOS (2019, eingereicht)
website: www.ofcaus.org

www.leibniz-bips.de

Kontakt

Vanessa Didelez

Leibniz-Institut für Präventionsforschung
und Epidemiologie – BIPS
Achterstraße 30
28359 Bremen

didelez@leibniz-bips.de



References

- Abadie A, Imbens GW. On the Failure of the Bootstrap for Matching Estimators. *Econometrica*. 2008;76(6):1537-1557.
- Abrahami Devin, Douros Antonios, Yin Hui, Yu Oriana HY, Faillie Jean-Luc, Montastruc François et al. Incretin based drugs and risk of cholangiocarcinoma among patients with type 2 diabetes: population based cohort study BMJ 2018; 363 :k4880
- Athey, Imbens, Wager (2018). Approximate Residual Balancing: De-Biased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society JRSS-B*
- Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity score matched samples. *Statistics In Medicine*. 2009;28:3083-3107.
- Austin, P. C. (2019). Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Statistical Methods in Medical Research*, 28(5), 1365–1377.
- Austin, P. C., and Stuart, E. A. (2015) Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statist. Med.*, 34: 3661– 3679. doi: 10.1002/sim.6607.
- Bang, H. and Robins, J.M., 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4), pp.962-973.
- Belloni, Chernozhukov, Hansen (2013). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*
- Belloni, Chernozhukov, Fernández-Val, C. Hansen (2013). Program evaluation and causal inference with high-dimensional data. *Econometrica*
- Brookhart, Wang, Solomon Schneeweiss (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology*

References

- Cartwright (2003). Nature's Capacities and Their Measurement. *Oxford Scholarship Online*
- Chakraborty, Moodie (2013). Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference and Personalized Medicine. *Book*
- Chernozhukov, Chetverikov, Demirer, Duo, Hansen, Newey, Robins (2017). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*
- Dawid (2015). Statistical Causality from a Decision-Theoretic perspective. *Annual Review of Statistics and Its Application*
- Dawid, Didelez (2010). Identifying the consequences of dynamic treatment strategies: A decision theoretic overview. *Statistics Surveys*
- De Luna, Waernbaum, Richardson, Covariate selection for the nonparametric estimation of an average treatment effect, *Biometrika*, Volume 98, Issue 4, December 2011, Pages 861–875
- Dehejia RH, Wahba S. Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics* 2002;84:151–161.
- Didelez, Sheehan (2006). Mendelian randomisation as an instrumental variable approach to causal inference, *Statistical Methods in Medical Research*
- Ding, VanderWeele (2016). Sensitivity Analysis Without Assumptions. *Epidemiology*
- Dorie, Harada, Carnegie, Hill (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine*
- Dukes, Avagyan, Vansteelandt (2018). High-dimensional doubly robust tests for regression parameters. *arXiv*

References

- Enders, Kollhorst, Engel, Linder, Pigeot (2018). Comparison of multiple imputation and two-phase logistic regression to analyze two-phase case-control studies with rich phase 1: A simulation study. *Journal of Statistical Computation and Simulation*
- Gran, Røysland, Wolbers, Didelez, Sterne, Ledergerber, Furrer, von Wyl, and Aalen, (2010). A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Statistics in Medicine*.
- Greenland (2000). An introduction to instrumental variables for epidemiologist. *International Journal of Epidemiology*
- Greenland (2005). Multiple-bias modelling for analysis of observational data. *JRSSA*
- Hernán (2010). The Hazards of Hazard Ratios. *Epidemiology*
- Hernán (2015). Counterpoint: epidemiology to guide decision-making: moving away from practice-free research. *American Journal of Epidemiology*
- Hernán (2018). The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *American Journal of Public Health*
- Hernán (2018). How to estimate the effect of treatment duration on survival outcomes using observational data. *British Medical Journal*
- Hernán, Brumback, Robins (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association – Application and Case Studies*
- Hernán, Alonso, Logan (2008). Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*
- Hernán, Hernández-Díaz, Robins (2004). A structural approach to selection bias. *Epidemiology*

References

- Hernán, Robins (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*
- Hernán, Robins (2018). Causal Inference. *Book in preparation*
- Kang, Schafer (2007). Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science*
- King, Nielsen (2016). Why Propensity Score Should Not Be Used for Matching. Working Paper
- Kollhorst B, Abrahamowicz M, Pigeot I. The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs. *Journal of Clinical Epidemiology*
- Lash, Fox , Fink (2011). Applying quantitative bias analysis to epidemiologic data. *Springer Science & Business Media*
- Lévesque, Hanley, Kezouh, Suissa (2014). Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *British Medical Journal*
- Li YP, Propert KJ, Rosenbaum PR. Balanced risk set matching. *Journal of the American Statistical Association* 2001;96(455):870–882.
- Lunceford, Davidian (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects. *Statistic in Medicine*
- McCandless, Gustafson (2017). A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding. *Stats in Med*
- Orellana, Rotnitzky, Robins (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes. *The International Journal of Biostatistics*
- Pearl (2000, 2009). Causality – models, reasoning and inference. *Cambridge University Press*
- Pearl (2011). Invited Commentary: Understanding Bias Amplification. *American Journal of Epidemiology*

References

- Robins (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker effect. *Mathematical Modelling*
- Robins (2001). Data, design and background knowledge in etiologic inference. *Epidemiology*
- Robins, Hernán, Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*
- Robins, Hernán, Siebert (2004). Effects of multiple interventions. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*
- Rotnitzky, Smucler, Robins (2019). Characterization of parameters with a mixed bias property. arXiv
- Rosenbaum, Rubin (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, Volume 70, Issue 1, Pages 41–55,
- Rubin (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*
- Rubin (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*
- Schneeweiss (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies. *Pharmacoepidemiology and Drug Safety*
- Shortreed, S. M., & Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, 73, 1111–1122.
- Smucler, Rotnitzky, Robins (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. arXiv
- Stuart EA. Matching methods for causal inference: A review and look forward. *Statistical Science*. 2010;25:1-21.

References

- Suissa, Azoulay (2012). Metformin and the Risk of Cancer: Time-related biases in observational studies. *Diabetes Care*
- Tan Z. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*. 2006;101:1619-1637.
- VanderWeele, Ding (2017). Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine*
- VanderWeele, T. J. and Shpitser, I. (2011), A New Criterion for Confounder Selection. *Biometrics*, 67: 1406-1413.
- Vansteelandt S, Daniel R M. On regression adjustment for the propensity score. *Statistics in Medicine*. 2014;33(23):4053-4072.
- Waernbaum, I. (2012), Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. *Statist. Med.*, 31: 1572-1581.
- Witte, Didelez (2018). Covariate selection strategies for causal inference: classification and comparison. *Biometrical Journal*