

Propensity Scores



Routinedaten in der Versorgungsforschung

Prof. Dr. med. Benno Neukirch
Health Care Management
Reinartzstr. 49
47805 Krefeld

benno.neukirch@hs-niederrhein.de

Agenda

- **Ausgangssituation: Beobachtungsstudie**
- **internationale Entwicklung**
- **Auswertung von Beobachtungsstudien**
 - **Propensity Score**
- **Beispiel Telemedizin bei Herzinsuffizienz**
- **Fazit**

Beobachtungsstudien

- **Intervention**
 - **DMP**
 - **IV nach § 140 a – d SGB V**
- **Vergleichsgruppe ist die Population, die**
 - **die Regelversorgung erhält**
- **Auf die Gestaltung der Gruppen**
 - **wenig Einfluss**
- **Selektion durch die Ärzte**
- **Selbstselektion durch die Patienten**

Beobachtungsstudien

- **kein Eingriff in die ärztliche Therapieentscheidung**
 - Therapie nicht zum Zweck des Erkenntnisgewinns
 - sondern zur erfolgreichen Behandlung des Patienten
- **keine Randomisierung und Verblindung**
- **Realität:**
 - Therapieerfolg als Zusammenspiel vieler Faktoren
 - Arzt-Patient-Kontakt, Compliance usw.

Beobachtungsstudien

- **nur für bereits zugelassene Therapien**
- **nicht genehmigungspflichtig**
 - **da keinen Eingriff in ärztliche Therapieentscheidung**

- **Verwendung in der Gesundheitsökonomie**
 - **naturalistisches Design**
 - **„True costs in a real world“**

Folge:

- **Vergleichsgruppe?**
 - muss gefunden werden
- **Datenbasis:**
 - Routinedaten

Routinedaten

- **Vorteile sekundärer Datenquellen:**
 - keine teure separate Datenerhebung
 - kurzfristige Datenbereitstellung
 - große Kohorten, langer Zeitraum
 - tägliche Routine wird abgebildet
 - mittlerweile standardisierte Verfahren
- **Nachteile:**
- **bislang noch keine klinischen Daten**
- **low-utiliser sind unterrepräsentiert**

Agenda

- **Ausgangssituation: Beobachtungsstudie**
- **internationale Entwicklung**
- **Auswertung von Beobachtungsstudien**
 - **Propensity Score**
- **Beispiel Telemedizin bei Herzinsuffizienz**
- **Fazit**

Internationale Entwicklung

- **zunehmende Nachfrage nach solchen Studien**
 - **Stakeholder wollen Ergebnisse**
- **Umfrage Academy Health (2009)**
- **85% aller Studien im Bereich**
- **comparative effectiveness research (CER)**
- **sind nicht randomisiert:**
- **observational studies**

Internationale Entwicklung

- **Druck auf gesundheitspolitische Entscheidungen**
 - Anstieg der Gesundheitsausgaben
 - demographischer Wandel
- **Welche positiven Effekte hat**
 - eine neue Technologie
 - gegenüber der bisherigen Versorgung
- **aus der Perspektive der**
 - Patienten
 - Leistungserbringer
 - Kostenträger
- **und welche Kosten sind damit verbunden?**

Guidelines der ISPOR Dezember 2009

- **ISPOR:**
- **International Society for Pharmacoeconomics and Outcomes Research**
- **Good Research Practices for Comparative Effectiveness Research [1 - 3]**

Agenda

- **Ausgangssituation: Beobachtungsstudie**
- **internationale Entwicklung**
- **Auswertung von Beobachtungsstudien**
 - **Propensity Score**
- **Beispiel Telemedizin bei Herzinsuffizienz**
- **Fazit**

Beobachtungsstudie

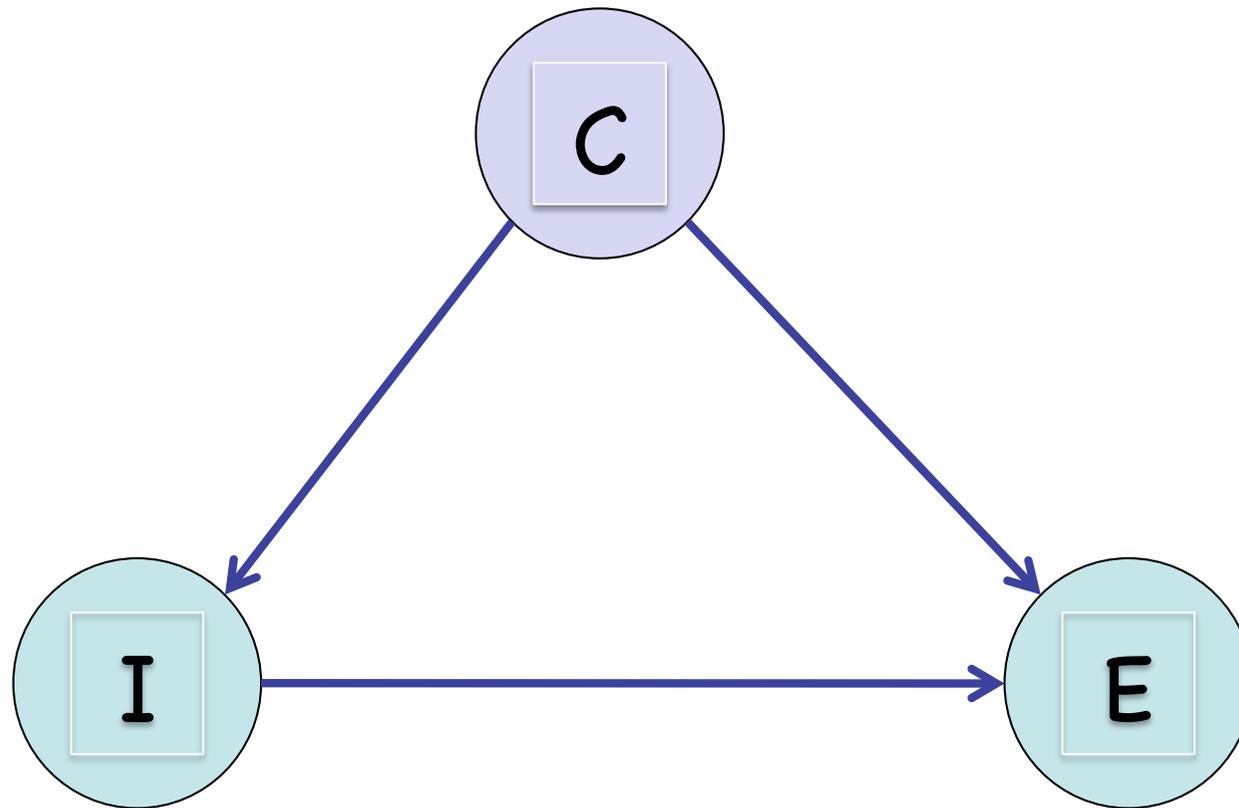
- **Erhebliche Unterschiede zwischen**
 - **Interventions- und**
 - **Kontrollgruppe**
- **Confounding**

- **Ein- und Ausschlusskriterien:**
 - **zu eng:**
 - **einige werden ausgeschlossen**
 - **zu weit:**
 - **„Gieskanne“**

Confounding

- **Betrachtung möglicher Verzerrungen**
 - z.B. Confounding by indication
- **Measured – unmeasured confounding**
 - unmeasured, but measurable
 - not measurable
- **Causal Graph**
 - Diagramm zur Darstellung der Beziehungen

Causal Graph



Confounding und Auswertung

- **mögliche Strategien**
- **Stratifizierung**
 - klassisches ‚matched pair‘
 - viele Strata mit kleinen Fallzahlen
 - klinisch sehr wichtige Informationsquelle
- **Regression (Standard!)**
 - lineare Regression (OLS) bei kontinuierlichen Variablen (als abhängige Variable)
 - logistische Regression bei binären Variablen
 - **Selektion der Einfluss-Variablen**
 - **unabhängige Variablen**

Confounding und Auswertung

- **weitere mögliche Strategien:**
- **Propensity Scores**
- **difference-in-difference Analyse**
- **Instrumentelle Variablen**

Propensity Score

- **Ursprung des Propensity Scores**
 - Paul Rosenbaum, Donald Rubin
- **Verwendung steigt sprunghaft an**
 - 1997 in PubMed: 7 Publikationen
 - 2007 in PubMed: 189 Publikationen
- **Propensity Score (PS):**
 - bedingte Wahrscheinlichkeit, mit der – unter Betrachtung einer Anzahl von Variablen – ein Kandidat die Intervention erhalten würde.

Propensity Scores

Regressionsrechnung:

$$\ln \frac{p(\text{outcome})}{1 - p(\text{outcome})} = \beta_0 + \beta_1 E_1 + \beta_2 C_2 + \dots + \beta_k C_k$$

Propensity Score:

$$\ln \frac{p(\text{exposure})}{1 - p(\text{exposure})} = \beta_0 + \beta_2 C_2 + \dots + \beta_k C_k$$

Propensity Score

- **Was bewirkt der Propensity Score?**
 - balanciert die Confounder
 - ähnlich RCT, aber anders
- **notwendig zur Berechnung:**
- **geeignete Kovariablen (engl. covariates)**
- **verschiedene Methoden zur Auswahl**
 - Korrelation zur Exposition und
 - Korrelation zum Outcome

Schritt 1: Auswahl der Confounder

- **Welche Kovariablen gibt es?**
- **verschiedene Methoden:**
 - **Verwendung von gebräuchlichen Indizes**
 - Charleson Comorbidity Index
 - Elixhauser Index
 - **Benutzung von naheliegenden Störgrößen**
 - meist zwischen 20 und 40 Variablen
 - manchmal unsystematisch
 - **Empirische Findung der maximalen Confounder**
 - nach Schneeweiß et al. [4]
 - wenn viele Routinedaten vorhanden

Schritt 2: Berechnung des Propensity Scores

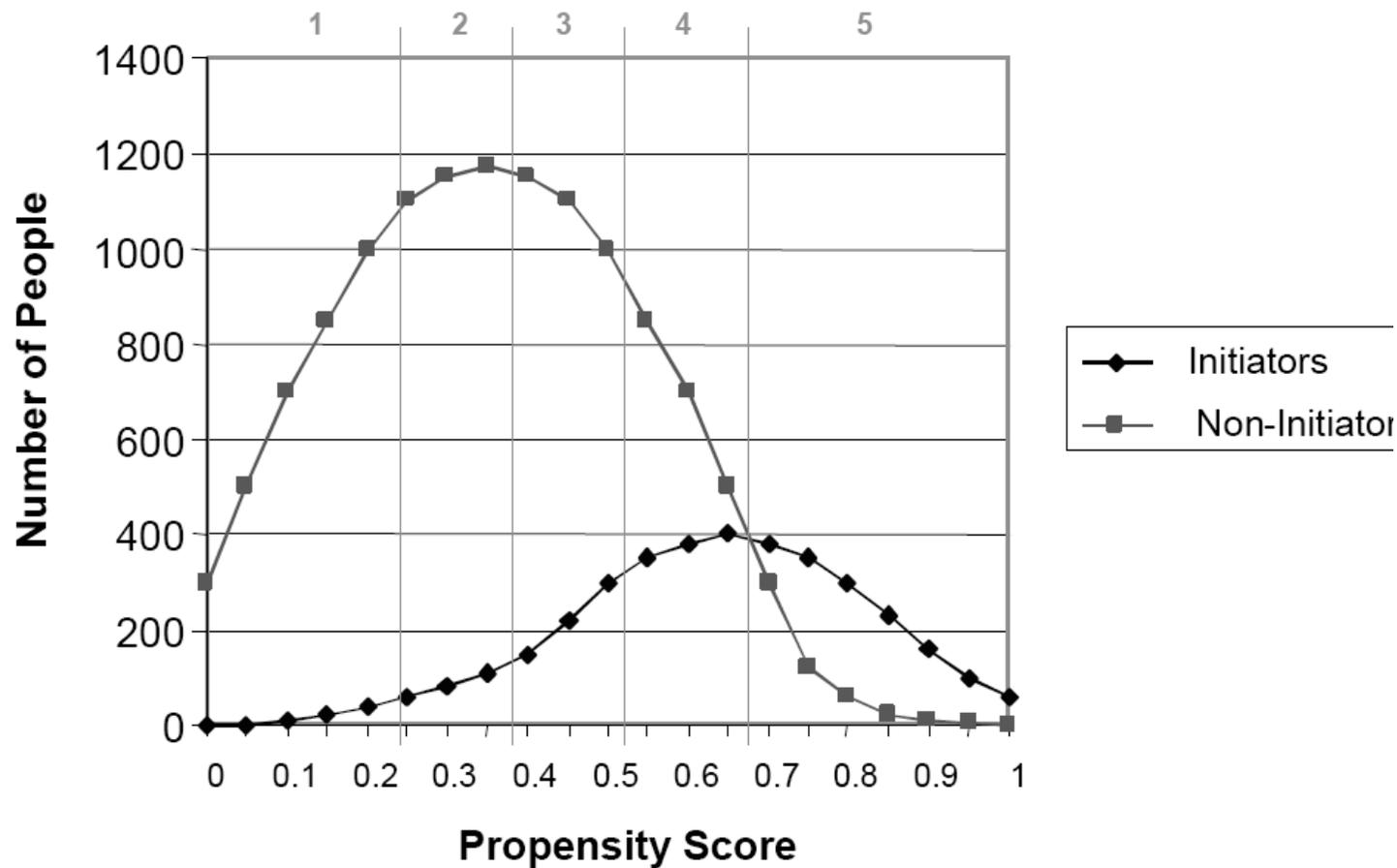
- **Kennzeichnung mit binärer Variable**
 - Intervention
 - Kontrolle
- **Einstellen des Sets der Kovariablen**
- **Berechnung**
- **Speichern der Wahrscheinlichkeit**
 - = propensity score

Schritt 3: Analyse der Vergleichsgruppen

- **Bildung der Vergleichsgruppen mit dem PS**
 - **Intervention**
 - **Kontrolle**
- **Dazu gibt es mehrere Techniken**
 - **overlap- Bestimmung**
 - **matching 1:1 oder 2:1 (nearest neighbour)**
 - **stratification**
 - **Quintile, Dezile**
 - **regression**
 - **weighting**

Verteilung des Propensity Scores

Hypothetical Distribution of Propensity Scores



Propensity Score

- **Kritische Bewertung:**
- **wenig empirische Differenzen zwischen**
 - **klassischer Regressionsrechnung und**
 - **propensity score**
- **Vorteile:**
 - **overlap wird bestimmt**
 - **Subgruppenanalyse möglich** → **Kliniker**
 - **Gesamtbilanz möglich** → **Kasse**
- **Nachteile:**
 - **unmeasured confounding**

Agenda

- **Ausgangssituation: Beobachtungsstudie**
- **internationale Entwicklung**
- **Auswertung von Beobachtungsstudien**
 - **Propensity Score**
- **Beispiel Telemedizin bei Herzinsuffizienz**
- **Fazit**

Pilotprojekt Knappschafft Telemedizin bei Herzinsuffizienz

- **alle Teilnehmer hatten vorher eingewilligt**
 - **beim Telemedizin Programm mitzumachen**
- **nach randomisierter Zuweisung**
 - **relevante Non-Compliance in der Interventionsgruppe**
- **Wie kann man den Effekt des Programms abschätzen?**
 - **‚Intention to treat‘ oder ‚As treated‘? [5]**
 - **Berücksichtigung der Selbstselektion: Propensity Score**

Hintergrund und Datengrundlage

- **randomisierte Studie mit Non-Compliance:**
- **Kontrollgruppe: 589 Teilnehmer**
- **Interventionsgruppe: 592 Teilnehmer**
 - **Complier: 367 Zertiva® SHL Telemedizin**
 - **Non-Complier: 225**
- **pseudonymisierte Routinedaten**
 - **stat. Fall- und Kostendaten (ICD und OPS)**
 - **amb. Fall- und Kostendaten (ICD)**
 - **Arzneiverordnungen (ATC und Kostendaten)**
- **1 Jahr vor Beobachtungsstart**

Ergebnisse: Intention – to – Treat

- unadjustiert -	Intervention	Kontrolle
	(n= 592) Complier: 62%	(n=589)
Anzahl Krankenhausaufenthalte	1,2 ($\pm 0,06$)	1,2 ($\pm 0,07$)
Tage im Krankenhaus	10,5 ($\pm 0,8$)	11,0 ($\pm 1,0$)
Krankenhauskosten	5.088€ (± 532)	4.761€ (± 467)
Ambulante Arztkosten	341€ (± 11)	343€ (± 13)
Arzneimittelkosten	1.579€ (± 133)	1.528€ (± 99)

Mittelwert und Standardfehler des Mittelwerts

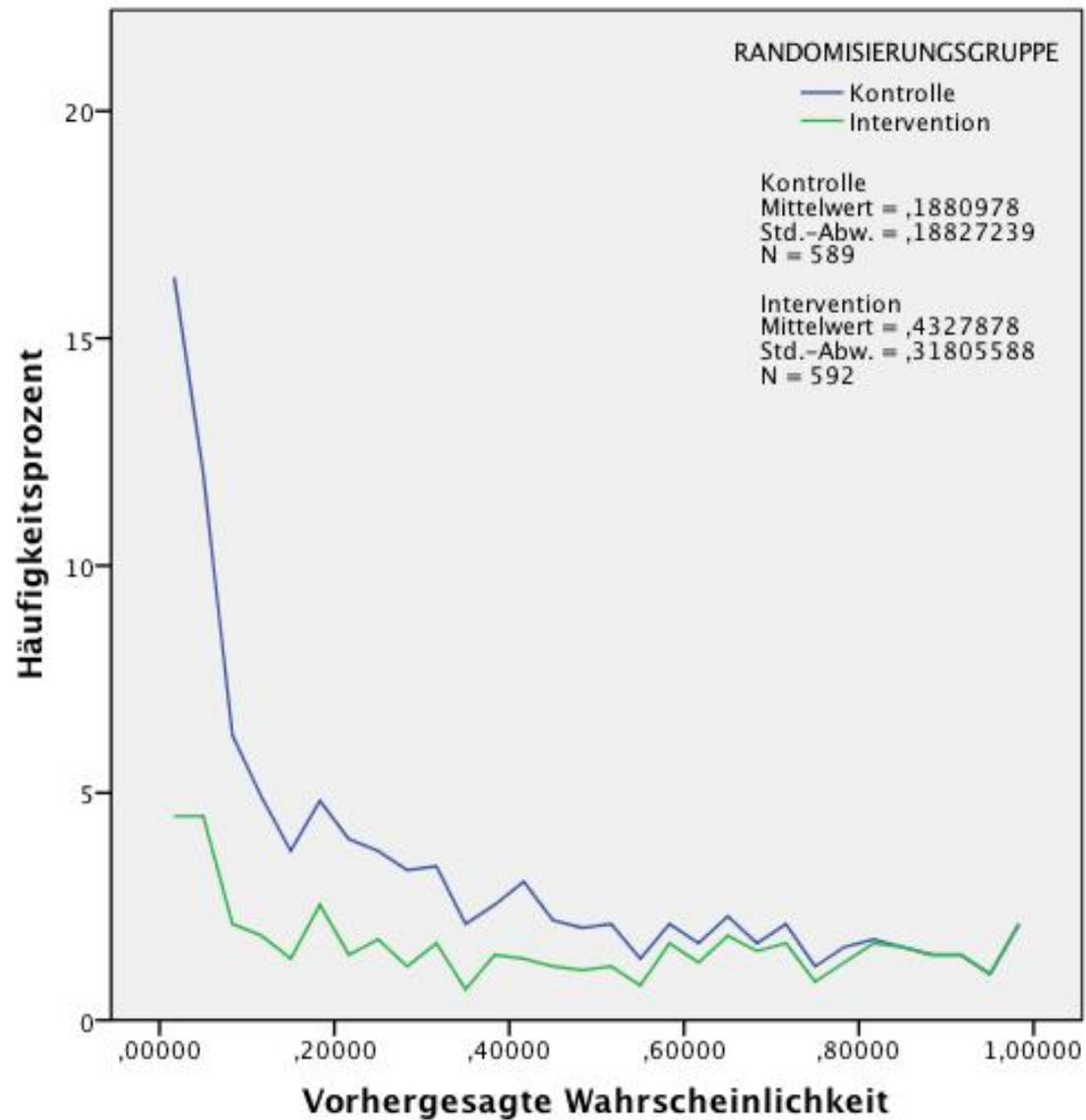
Auswahl der Variablen

- **Suche der 500 Codes mit der größten Dysbalance**
 - maximal 0,5
- **Bestimmung des wiederholten Auftretens eines Codes**
 - once – sporadic – frequent
 - 1.500 binäre Variablen
- **Berechnung eines multiplikativen Biasterms [6]**
 - Berücksichtigung des relativen Risikos
 - bei der Intervention
 - beim Outcome (Krankenhaus ja – nein)
- **Auswahl der TOP 500 Variablen**

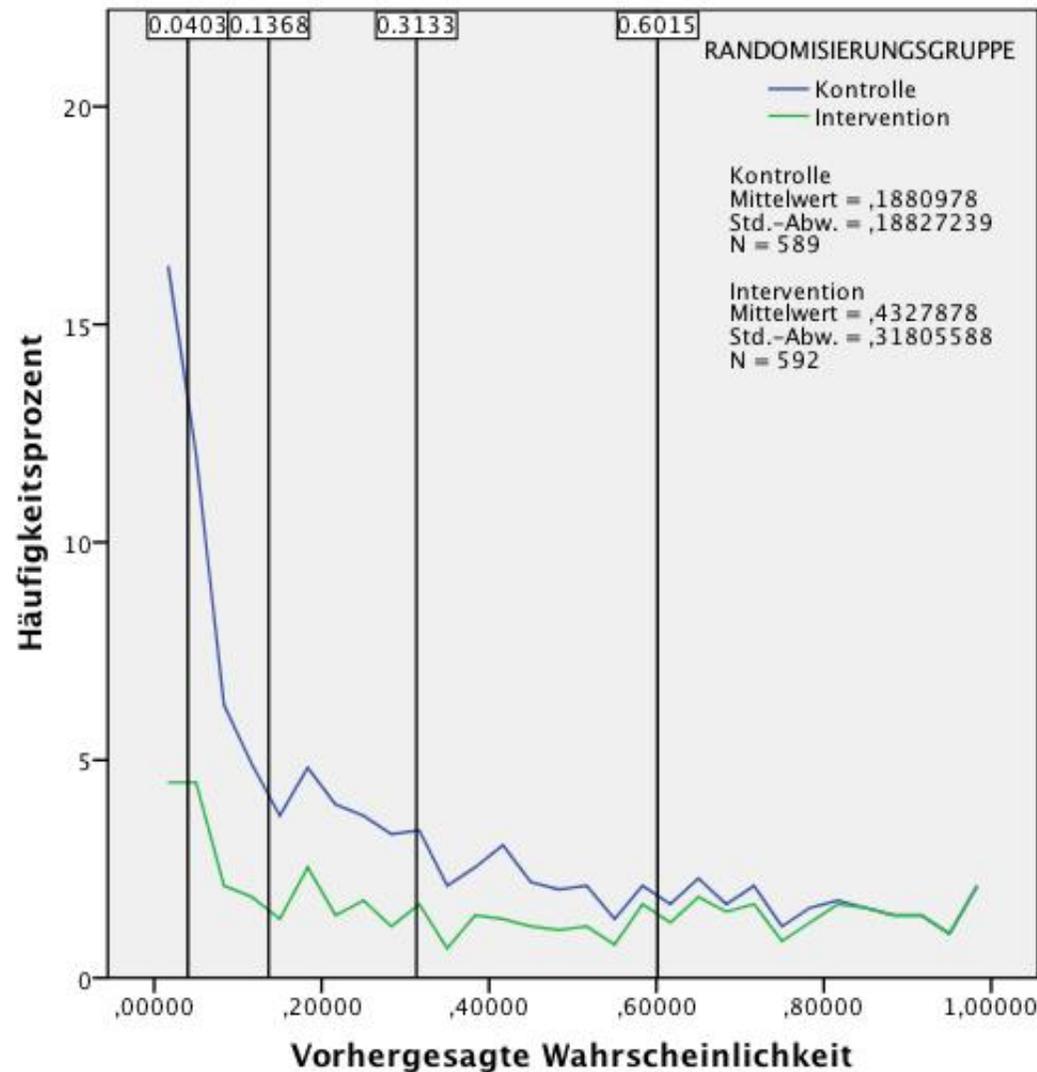
TOP 10 Diagnosen stationär

Kovariable	Bias _{mult}	ICD Text
H36_once_sum	8,353	Affektionen der Netzhaut bei anderenorts klassifizierten Krankheiten
I63_once_sum	4,657	Hirninfarkt
B37_once_sum	4,187	Kandidose
M53_once_sum	4,186	Sonstige Krankheiten der Wirbelsäule und des Rückens, anderenorts nicht klassifiziert
M79_once_sum	4,186	Sonstige Krankheiten des Weichteilgewebes, anderenorts nicht klassifiziert
G25_sporadic_sum	4,176	Sonstige extrapyramidale Krankheiten und Bewegungsstörungen
K74_once_sum	4,176	Fibrose und Zirrhose der Leber
L02_once_sum	4,176	Hautabszess, Furunkel und Karbunkel
Z09_once_sum	4,176	Nachuntersuchung nach Behandlung wegen anderer Krankheitszustände außer bösartigen Neubildungen
B95_sporadic_sum	4,163	Streptokokken und Staphylokokken als Ursache von Krankheiten, die in anderen Kapiteln klassifiziert sind

Verteilung des Propensity Scores



Propensity Quintile



Quintil	Intervention	Kontrolle
< 0,0403	70	166
bis 0,1368	86	150
bis 0,3133	101	137
bis 0,6015	126	110
> 0,6015	209	26

Ergebnisse - gewichtet

- gewichtet -	Intervention	Kontrolle	
	(n= 592)	(n=589)	
Anzahl Krankenhausaufenthalte	0,6 ($\pm 0,04$)	0,9 ($\pm 0,05$)	p < 0,001
Tage im Krankenhaus	5,0 ($\pm 0,5$)	8,8 ($\pm 0,7$)	p < 0,001
Krankenhauskosten	2.558€ (± 384)	3.745€ (± 339)	p < 0,005
Ambulante Arztkosten	165€ (± 8)	270€ (± 10)	p < 0,001
Arzneimittelkosten	748€ (± 63)	1.210€ (± 89)	p < 0,001

Mittelwert und Standardfehler des Mittelwerts sowie die Signifikanz als p-Wert

Agenda

- **Ausgangssituation: Beobachtungsstudie**
- **internationale Entwicklung**
- **Auswertung von Beobachtungsstudien**
 - **Propensity Score**
- **Beispiel Telemedizin bei Herzinsuffizienz**
- **Fazit**

Comparative Effectiveness Research

- **Auswertung:**
 - **Kombination mehrere Ansätze**
 - **Klassische Regressionsrechnung**
 - **propensity score, ggf. instrumentelle Variable**
 - **Sensitivitätsanalysen**
- **Diese Forschungsansatz arbeitet mit**
 - **nichtrandomisiertem Design**
 - **Routinedaten und klin. Daten**
 - **betrachtet das Confounding**
- **ist daher geeignet für Beobachtungsstudien**

Durchführung von Beobachtungsstudien

- **guter Ansatz zur Evaluation transsektoraler Versorgungsmodelle**
 - Nutzen; Kosten
 - Qualität
- **Nähe zur Alltagsversorgung**
- **Problem der Selbstselektion**
- **ermöglicht eine Stratifizierung**
- **damit eine Verbesserung der Ressourcenallokation**

Vielen Dank



Literatur

1. Berger, Marc L.; Mamdani, Muhammad; Atkins, David; Johnson, Michael L.: Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I; Value in Health (2009) 12, 8: 1044–1052
2. Cox, Emily; Martin, Bradley C.; Van Staa, Tjeerd ; Garbe, Edeltraut; Siebert, Uwe; Johnson, Michael L.: Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II; Value in Health (2009) 12, 8: 1053-1061
3. Johnson, Michael L.; Crown, William; Martin, Bradley C.; Dormuth, Colin R.; Siebert, Uwe: Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III; Value in Health (2009) 12, 8: 1062-1073

Literatur

4. Schneeweiss, Sebastian; Rassen, Jeremy A.; Glynn Robert J.; Avorn, Jerry; Mogun, Helen; Brookhart, Alan M.: High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data. *Epidemiology* (2009) 20: 512–522
5. Marcus, Sue M; Gibbons, Robert D.: Estimating the Efficacy of Receiving Treatment in Randomized Clinical Trials with Noncompliance. *Health Services & Outcomes Research Methodology* (2001) 2:247–258
6. Hirano, Keisuke; Imbens, Guido W.: Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization. *Health Services & Outcomes Research Methodology* (2001) 2:259–278