

## Einführung in Propensity Scores



**Oliver Kuß**

Institut für Biometrie und Epidemiologie, Deutsches Diabetes-Zentrum (DDZ),  
Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität  
Düsseldorf

und

Centre for Health and Society (chs), Medizinische Fakultät der Heinrich-Heine-  
Universität Düsseldorf

## ÜBERSICHTSARBEIT

# Propensity Score – eine alternative Methode zur Analyse von Therapieeffekten

Teil 23 der Serie zur Bewertung wissenschaftlicher Publikationen

Oliver Kuss, Maria Blettner, Jochen Börgermann

## ZUSAMMENFASSUNG

**Hintergrund:** Nur die Randomisierung garantiert in Therapiestudien eine gleichmäßige Verteilung aller bekannten und unbekanntem Patientenmerkmale auf eine Interventions- und eine Kontrollgruppe und erlaubt dadurch kausale Aussagen über Therapieeffekte. Randomisierte kontrollierte Studien werden jedoch auch für ihre fehlende externe Validität kritisiert. Nichtrandomisierte

**M**an ist sich in der medizinischen Forschung weitgehend einig darüber, dass Therapien primär in randomisierten kontrollierten Studien geprüft werden sollten. Nur die Randomisierung garantiert eine gleichmäßige Verteilung aller bekannten und unbekanntem Patientenmerkmale auf eine Interventions- und

# Einleitung

- Therapien sollen primär in randomisierten Studien geprüft werden.
- Nur die Randomisierung garantiert eine gleichmäßige Verteilung aller bekannten und unbekanntes (!) Störgrößen und Risikofaktoren auf die Therapiegruppen  
→ kausale Aussagen möglich
- **Aber:** Randomisierte Studien sind in manchen Fällen „unnötig, ungeeignet, unmöglich oder ungenügend“ [Black, 1996]
- **Zusätzlich:** Geringe externe Validität [McKee, 1999; Rothwell, 2005]

# Einleitung

- **Alternative:** Nichtrandomisierte Studien
- Bessere *externe* Validität, **aber:** fehlende *interne* Validität:  
Die Therapiezuweisung erfolgt nichtrandomisiert und Interventions- und Kontrollgruppe können sich systematisch bzgl. bekannter und (schlimmer noch) unbekannter Störgrößen, so genannter „Confounder“, unterscheiden
- **Lösung:** Adjustierung für Confounder durch herkömmliche Regressionsmodelle (Standard) oder die Propensity Score-Methode [Rosenbaum/Rubin, 1983]

# Die Propensity Score-Methode

- **Definition:** Der Propensity Score (Abk.: PS) ist die Wahrscheinlichkeit, die zu prüfende Therapie zu erhalten
- Der PS ist i.a. unbekannt und muss in einem **ersten Schritt** geschätzt werden (PS-Modell)
- Schätzung des PS-Modells durch (z.B.) logistische Regression:
  - Zielgröße (abhängige Variable): Therapie
  - Kovariablen (unabhängige Variablen): die zu Therapiebeginn bestehenden Patientenmerkmale (Risikofaktoren, Confounder)
- Berechnung des PS für jeden einzelnen Menschen aus den Modellparametern des PS-Modells

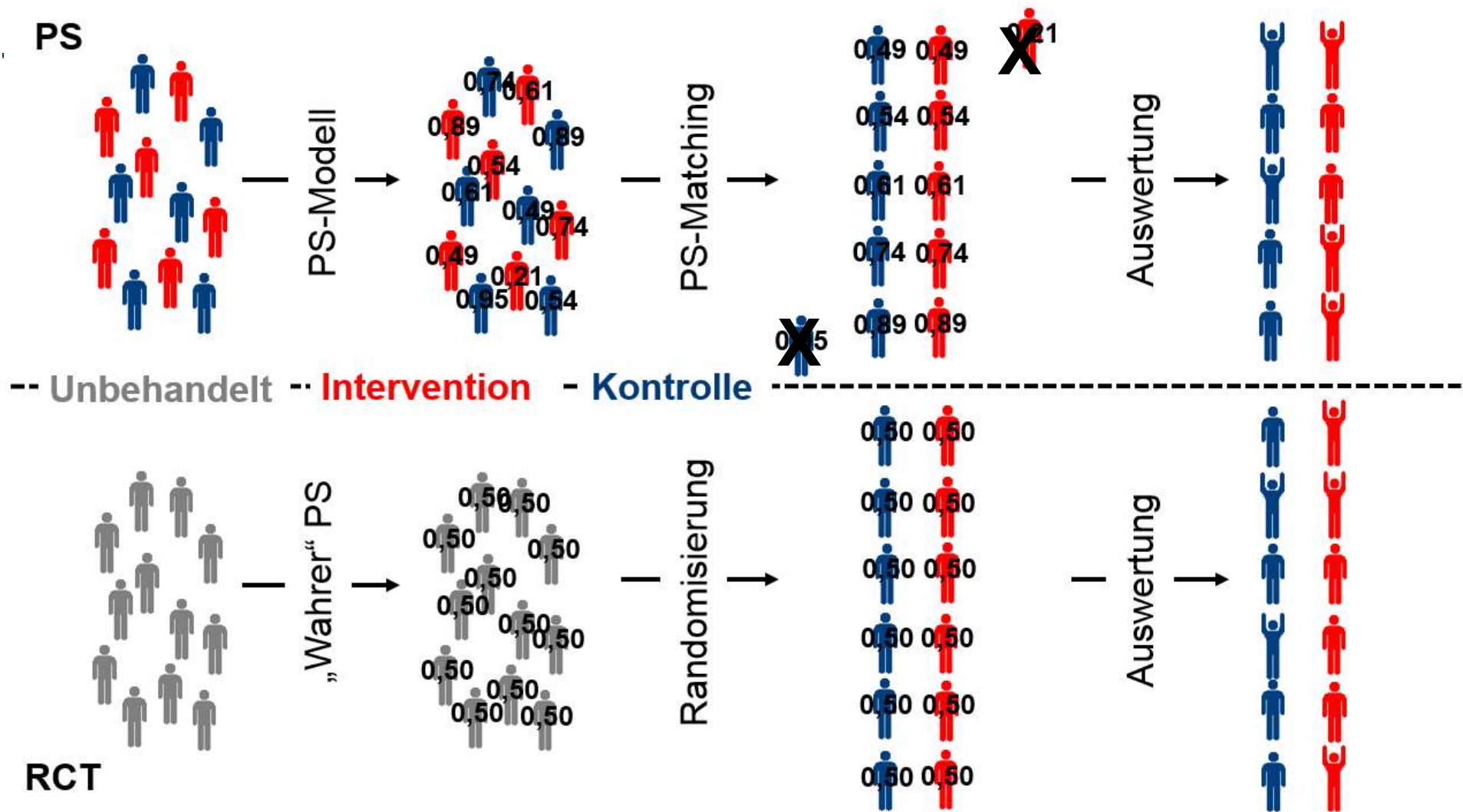
# Die Propensity Score-Methode

- **Frage:** Welche Kovariablen sollen ins PS-Modell aufgenommen werden?
- **Antwort:**
  - Viele! (ruhig auch Interaktionen und nicht-lineare Terme)
  - Vor allem diejenigen, die den späteren Therapieerfolg (und nicht etwa die Therapiezuweisung) vorhersagen, sonst droht Effizienzverlust ohne Biasgewinn [Brookhart, 2006]  
Merkmale, die nur die Therapiezuweisung erklären („Instrumental variables“) sind „Bias amplifier“! [Ding, 2017]

# Die Propensity Score-Methode

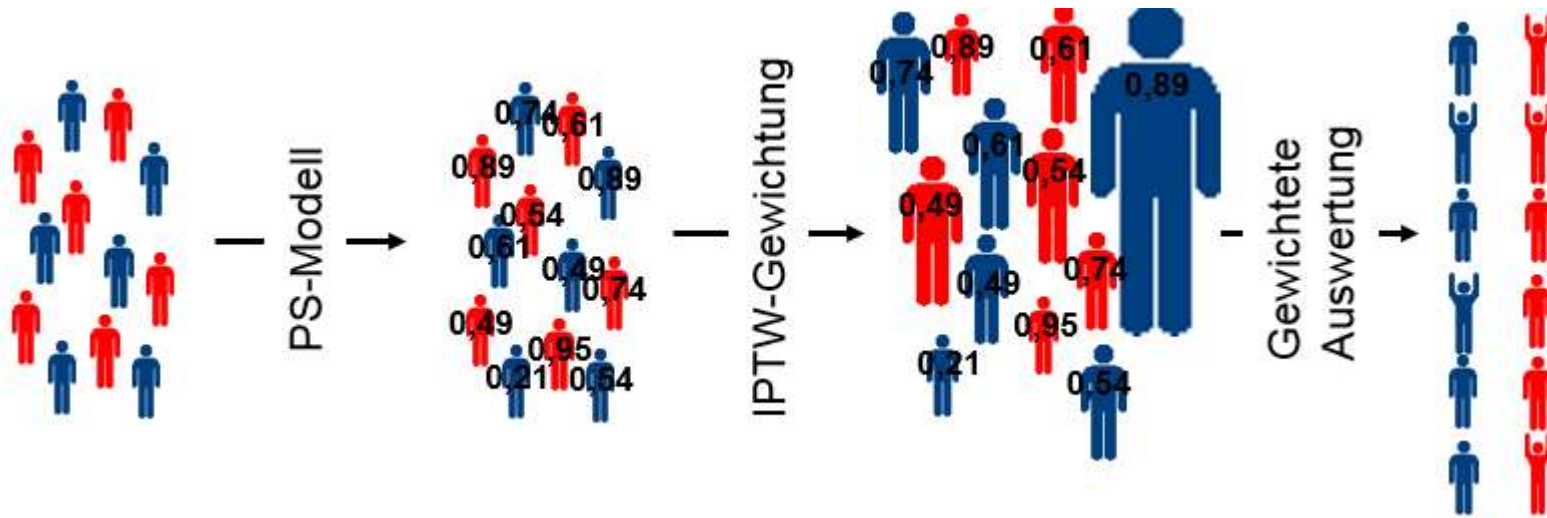
- **Zweiter Schritt:** Schätze den eigentlich interessierenden Therapieeffekt unter Zuhilfenahme des PS
- **Vier Methoden:**
  1. PS-Matching
  2. IPTW(=Inverse probability of treatment weighting)-Methode
  3. Stratifizierung nach dem PS
  4. Regressionsadjustierung für den PS
- **(Nicht nur) mein Favorit:** PS-Matching [Austin, 2007; Morgan, 2006].

# PS-Matching und RCT – Ein Vergleich

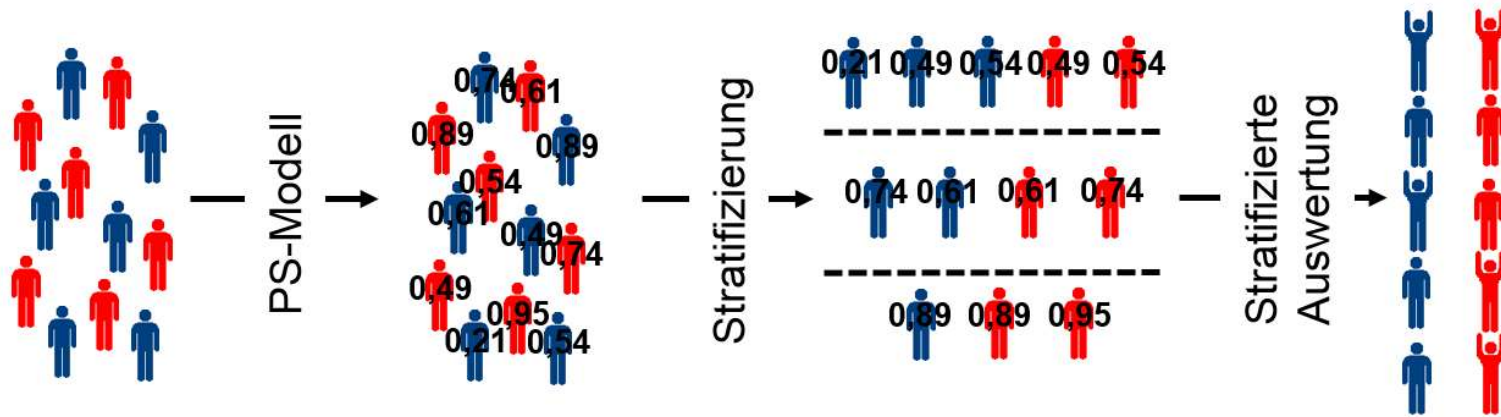




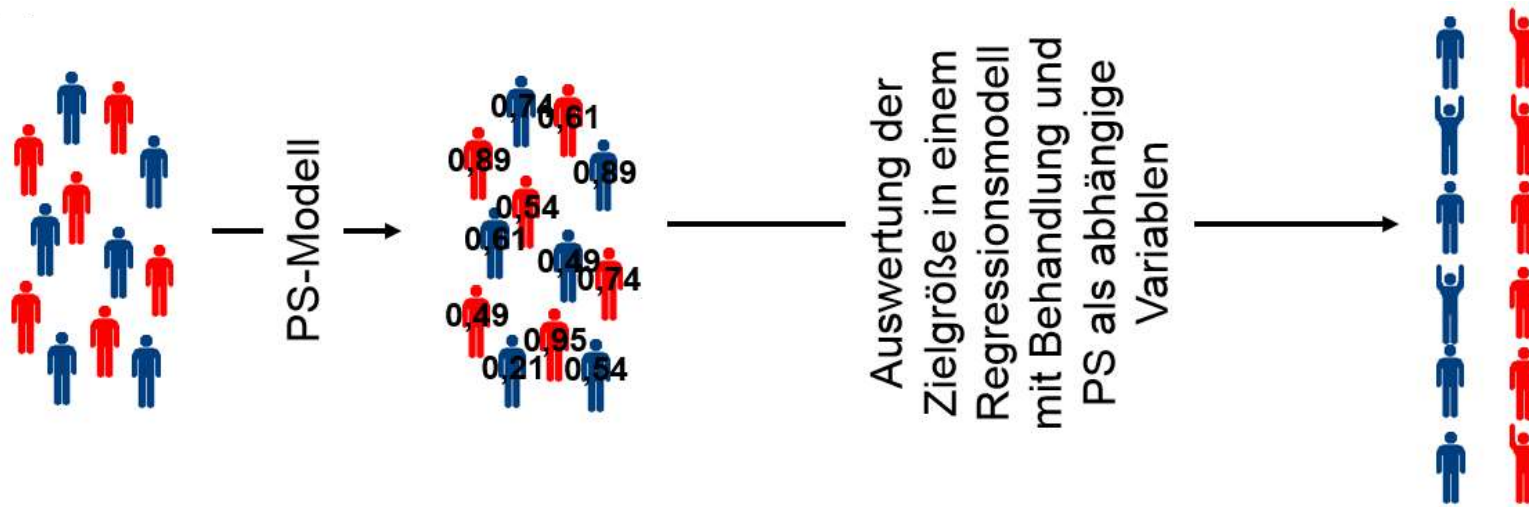
# IPTW-Schätzung



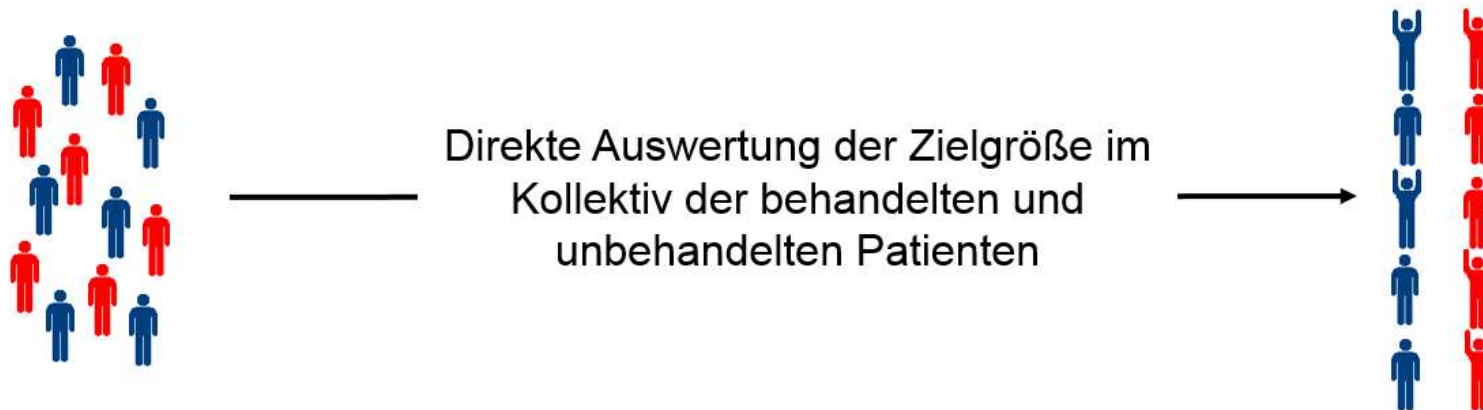
# Stratifizierung nach dem PS



# Regressionsadjustierung für den PS



# Herkömmliche Regressionsadjustierung (ohne PS)



# Die Propensity Score-Methode

## Vorteile PS-Matching:

1. Explizite Darstellung sowohl der Eigenschaften von behandelten und unbehandelten Patienten („Table 1“ in einer randomisierten Studie) als auch der Balanciertheit der Confounder

	PS-gematchte Patienten (n = 788)		
	Less-OPCAB (n = 394)	cCABG (n = 394)	z-Differenz
Alter [Jahre]	69,3 ± 9,1	69,0 ± 8,9	0,46
Männlich [%]	78,2	77,9	0,09
Präoperativer Myokardinfarkt [%]	27,2	26,7	0,16

# Die Propensity Score-Methode

## Vorteile PS-Matching:

1. Explizite Darstellung sowohl der Eigenschaften von behandelten und unbehandelten Patienten („Table 1“ in einer randomisierten Studie) als auch der Balanciertheit der Confounder, als auch des Erfolges des Matchings

	Alle Patienten (n = 1.282)			PS-gematchte Patienten (n = 788)		
	Less-OPCAB (n = 395)	cCABG (n = 887)	z-Differenz	Less-OPCAB (n = 394)	cCABG (n = 394)	z-Differenz
Alter [Jahre]	69,3 ± 9,1	67,5 ± 9,4	3,24	69,3 ± 9,1	69,0 ± 8,9	0,46
Männlich [%]	78,2	77,9	0,13	78,2	77,9	0,09
Präoperativer Myokardinfarkt [%]	27,1	35,7	-3,14	27,2	26,7	0,16

# Die Propensity Score-Methode

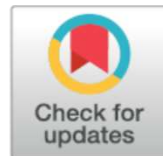
**DDZ**

German Diabetes Center

# ACHTUNG!



ELSEVIER



Journal of Clinical Epidemiology 126 (2020) 9–16

**Journal of  
Clinical  
Epidemiology**

## ORIGINAL ARTICLE

Unknown confounders did not bias the treatment effect when improving balance of known confounders in randomized trials

Oliver Kuss<sup>a,b,\*</sup>, Matthaeus Miller<sup>a</sup>

<sup>a</sup>German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Institute for Biometrics and Epidemiology, Düsseldorf, Germany

<sup>b</sup>Medical Faculty, Heinrich Heine University Düsseldorf, Institute of Medical Statistics, Düsseldorf, Germany

# Die Propensity Score-Methode

## **Nachteil (?) PS-Matching:**

Patienten, für die kein Matching-Partner gefunden wurde, werden ausgeschlossen

→ Reduktion der Fallzahl, Verlust an statistischer Power

## **Aber:**

Es wird klar, über welche Subgruppe überhaupt Aussagen gemacht werden dürfen

# Die Propensity Score-Methode

## **Konsequenz:**

PS-Matching vs. die anderen PS-Methoden ist immer ein Trade-Off [Stuart, 2009] zwischen

- **Bias** (verzerrter Schätzung des Therapieeffektes) und
- **Varianz** (zu ungenaue Schätzung des Therapieeffektes)

# Die Propensity Score-Methode

**TABELLE 1**

**Eigenschaften der vier verschiedenen Methoden zur Berücksichtigung des Propensity Scores (PS) und der herkömmlichen Regressionsanalyse bei der Analyse von nichtrandomisierten Therapiestudien**

	Methode				
	PS-Methode				herkömmliche Regressionsadjustierung
	PS-Matching	IPTW-Schätzung	Stratifizierung	Regressionsadjustierung für den PS	
ermöglicht eine leichte Beurteilung der Vergleichbarkeit von behandelten und unbehandelten Patienten	+	(+)	(+)	-	-
ermöglicht eine Beurteilung der Balanciertheit der Merkmale im Auswertungsdatensatz	+	+	(+)	-	-
nutzt den vollständigen Datensatz (kleinere Varianz des Therapieeffekts bei größerer Gefahr für Bias)	-	+	+	+	+
ähnelt von der Vorgehensweise einem RCT (generiere vergleichbare Gruppen und ignoriere dabei die Zielgrößen)	+	(+)	(+)	-	-
ist robust gegenüber Patienten mit extremem PS	+	-	+	+	+
kommt insgesamt mit weniger statistischen Modellannahmen aus	+	+	(+)	-	-

IPTW, „inverse probability of treatment weighting“; PS, Propensity Score; RCT, randomisierte kontrollierte Studie; „+“ bedeutet „ja“ oder „ist gegeben“, „-“ bedeutet „nein“ oder „ist nicht gegeben“, „(+)“ bedeutet „teilweise“ oder „ist zum Teil gegeben“



# Fazit

- Die PS-Methode ist eine zwar nicht neue, aber doch innovative Methode zur Auswertung von nichtrandomisierten Therapiestudien, die sowohl statistische als auch erkenntnistheoretische Vorteile im Vergleich zur herkömmlichen Regression hat
- Unter den 4 Möglichkeiten zur Berücksichtigung des PS hat das PS-Matching eine Reihe von Vorteilen
- **Aber:** Der PS kann nur für die bekannten und tatsächlich gemessenen Confounder adjustieren!