

Sensitivitätsanalysen bei der Verwendung von Propensity Score Methoden in nicht-randomisierten Studien und deren Umsetzung mit SAS

Annabel S. Müller-Stierlin
Institut für Epidemiologie und
Medizinische Biometrie,
Universität Ulm
Schwabstraße 13
89075 Ulm
annabel.mueller-stierlin@uni-ulm.de

Benjamin Mayer
Institut für Epidemiologie und
Medizinische Biometrie,
Universität Ulm
Schwabstraße 13
89075 Ulm
benjamin.mayer@uni-ulm.de

Reinhold Kilian
Klinik für Psychiatrie und
Psychotherapie II,
Universität Ulm
Ludwig-Heilmeyer-Straße 2
89312 Günzburg
reinhold.kilian@uni-ulm.de

Rainer Muche
Institut für Epidemiologie und
Medizinische Biometrie,
Universität Ulm
Schwabstraße 13
89075 Ulm
rainer.muche@uni-ulm.de

Zusammenfassung

Propensity Score (PS)-Methoden werden angewandt, um den Selektions-Bias in nicht-randomisierten Studien zu kontrollieren. Der PS beschreibt die bedingte Wahrscheinlichkeit mit der ein Studienteilnehmer unter Betrachtung ausgewählter Kovariablen die Behandlung erhält und somit der Interventionsgruppe angehört. Propensity Score-Analysen umfassen die Variablenauswahl, die PS-Schätzung und die Anwendung verschiedener PS-Verfahren, wie PS-Adjustierung oder PS-Matching. Zur Prüfung der Arbeitsschritte werden Balance Tests und Common Support Checks durchgeführt.

Ziel der IVPOWER-Studie war die Analyse der Wirksamkeit und der Kostenwirksamkeit von Angeboten zur integrierten psychiatrischen Versorgung nach dem Modell des „Netzwerks psychische Gesundheit“ im Vergleich zur Standardversorgung. Es handelt sich um eine prospektive, kontrollierte, nicht-randomisierte Beobachtungsstudie bei 511 Patienten mit psychischen Erkrankungen in verschiedenen Regionen Deutschlands. Zur Kontrolle des Selektions-Bias wurden verschiedene PS-Analysen, unter Verwendung verschiedener Algorithmen zur Variablenauswahl sowohl verschiedener Adjustierungs- bzw. Matchingverfahren, durchgeführt. Im Rahmen dieser Sensitivitätsanalysen wurde untersucht, welche Auswirkung verschiedene PS-Methoden auf die Effektschätzer zeigen.

In diesem Beitrag stellen wir die Vorgehensweise, insbesondere die Umsetzung und Möglichkeiten in SAS, sowie die Ergebnisse der Sensitivitätsanalysen vor. Die Analyse erfolgte unter Verwendung der Prozeduren PROC LOGISTIC zur PS-Schätzung, PROC MIXED zur Effektschätzung und des Matching-Algorithmus von Lanehart et al..

Schlüsselwörter: Propensity Score, Adjustierung, Matching, Variablenauswahl, Sensitivitätsanalysen

1 Einleitung

Die Ausgaben für die psychiatrische Gesundheitsversorgung in Deutschland liegen im internationalen Vergleich in der Spitzengruppe. Bezüglich der Behandlungsqualität bezogen auf psychische Erkrankungen befinden wir uns jedoch nur im Mittelfeld [15]. Der Grund dafür wird in der mangelnden Integration zwischen den Leistungssektoren gesucht. In Deutschland werden stationäre sowie ambulante Leistungen aber auch die berufliche Wiedereingliederung und Wohnunterkünfte von unterschiedlichen Leistungsträgern angeboten. Diese Leistungen werden wiederum über unterschiedliche Kassen abgerechnet und durch verschiedene Sozialgesetzbücher reguliert. Dies erschwert die Koordination und Integration von Leistungsangeboten. Die Integrierte Versorgung (IV) zielt darauf ab verschiedene Sektoren stärker zu vernetzen. Durch eine Änderung im Sozialgesetzbuch (140 a ff SGB V) wurde die rechtliche Grundlage für die Finanzierung einer IV geschaffen. Krankenkassen können nun mit medizinischen und nicht-medizinischen Leistungsanbietern Verträge abschließen, um komplexe psychosoziale Leistungen zu gewährleisten.

Einen solchen Vertrag hat auch die Techniker Krankenkasse mit verschiedenen Trägern gemeindepsychiatrischer Angebote unter der Bezeichnung „NetzWerk für psychische Gesundheit“ (NWpG) abgeschlossen. Vertragsbestandteil ist die Erbringung einer IV durch multiprofessionelle gemeindepsychiatrische Teams. Inzwischen beteiligen sich auch andere Krankenkassen und im Jahr 2015 waren insgesamt 12700 Menschen in die IV eingeschrieben [3]. In der IVPOWER-Studie wurde die Wirksamkeit und Effizienz der IV von Menschen mit psychischen Erkrankungen im NWpG unter Realweltbedingungen bei besonderer Berücksichtigung der Verbesserung von Empowerment und Lebensqualität untersucht.

Eine Randomisierung der Teilnehmer zu Studiengruppen, welche als Goldstandard in klinischen Studien betrachtet wird, war nicht möglich, da es sich bei der IV im NWpG um sogenannte Selektivverträge der Krankenkassen mit Anbietern von IV-Leistungen handelt. Dies führt dazu, dass sowohl Krankenkassen, Leistungsträger aber auch Patienten selbst entscheiden können, inwieweit eine Einschreibung in die IV erfolgt. Die Krankenkasse eröffnet das Angebot der Teilnahme an der IV insbesondere den Patienten, bei denen die Annahme besteht, dass sie von diesem Versorgungskonzept profitieren werden. Diese ausgewählten Patienten haben nun die Möglichkeit den Vorschlag ihrer Krankenkasse anzunehmen oder diesen abzulehnen.

Da bei einer nicht-randomisierter Zuweisung von Studienteilnehmern die Gefahr eines Selektionsbias und damit der Verzerrung der Untersuchungsergebnisse besteht, müssen statistische Verfahren zur Biaskontrolle angewendet werden. In diesem Fall wurden Propensity Score (PS)-Methoden verwendet um einen möglichen Selektionsbias zu korrigieren.

1.1 Fragestellung

Die grundlegende Fragestellung war, inwiefern ein Programm zur IV von Menschen mit psychischen Erkrankungen zur Verbesserung der psychiatrischen Behandlung, insbesondere im Hinblick auf die Hauptzielgrößen Empowerment und Lebensqualität, führt.

Aufgrund der nicht-randomisierten Gruppeneinteilung, sollte an dieser Stelle die Robustheit der Ergebnisse geprüft werden. In umfassenden Sensitivitätsanalysen wurde untersucht, ob die Ergebnisse unter Verwendung von verschiedenen PS-Verfahren repliziert werden können.

2 Methoden

2.1 IVPOWER-Studie

Es handelt sich hierbei um eine prospektive, kontrollierte Beobachtungsstudie mit vier Messzeitpunkten im Abstand von sechs Monaten. Die Studie wurde von der Ethikkommission der Universität Ulm am 3. Mai 2013 (Antragsnummer: 129 / 13) und von der Ethikkommission der Technischen Universität Dresden am 30. September 2013 (Antragsnummer: EK 259072013) genehmigt. Sie ist im deutschen Register klinischer Studien mit der Identifikationsnummer DRKS00005111 eingetragen und wurde nach den Richtlinien der Deklaration von Helsinki 2013 durchgeführt. Das Studienprotokoll zur IVPOWER-Studie ist zweisprachig publiziert [14; 13]. Sämtliche Ergebnisse, einschließlich der hier dargestellten Sensitivitätsanalysen, sind in einer Dissertation ausführlich dargestellt [9].

Es gab über Deutschland verteilt fünf Studienregionen (Schleswig-Holstein, Nordrhein-Westfalen, Berlin, Sachsen und Bayern), wodurch eine hohe Rekrutierungsbreite mit unterschiedlichem Urbanisierungsgrad, variierenden lokalen Versorgungssystemen und Leistungsanbietern abgedeckt wurde. In jedem Zentrum sollten 100 Patienten rekrutiert werden. Die Hälfte der Teilnehmer sollte zum Zeitpunkt der Rekrutierung in die IV eingeschrieben sein. Für die Kontrollgruppe wurden Patienten über ambulante und sozialpsychiatrische Einrichtungen rekrutiert, welche ausschließlich Zugang zur Standardversorgung hatten.

Unter Berücksichtigung der Zielsetzung der IV im NWpG, die Fähigkeit der Patienten zu einem selbstverantwortlichen Umgang mit ihrer Erkrankung zu stärken und ihre Möglichkeiten zu einer selbstbestimmten Lebensweise zu erweitern, wurde die Verbesserung des Empowerments im Verlauf von 18 Monaten als primäres Zielkriterium gewählt. Neben dem Empowerment-Gesamtwert (EPAS), bildet das Selbstbeurteilungsinstrument die fünf Empowerment-Dimensionen Alltagsleben (EPAS-1), soziale Beziehungen (EPAS-2), Behandlungspartizipation (EPAS-3), Selbstwirksamkeit (EPAS-4) und Selbstwertgefühl (EPAS-5) ab.

Zur Auswertung wurden gemischte Regressionsmodelle mit einem variablen Zeiteffekt, einem fixen Gruppeneffekt und einem fixen Interaktionseffekt zwischen der Zeit und der Untersuchungsgruppe durchgeführt. Fehlende Messwerte wurden dabei nicht imputiert, sondern im gemischten linearen Modell durch die Gewichtung der Parameterschätzung berücksichtigt. Messwiederholungen und Studienzentren wurden in der Kovarianzstruktur des variablen Zeiteffektes berücksichtigt. Die Datenanalysen wurden nach dem intention-to-treat Prinzip durchgeführt: Der Gruppeneffekt bezieht sich auf die Gruppenzugehörigkeit zu Baseline und Versorgungsgruppenwechsel wurden in der

Hauptanalyse nicht berücksichtigt. Der Interventionseffekt wird im Interaktionseffekt zwischen der Zeit und der Behandlungsgruppenzugehörigkeit abgebildet. Der Interaktionseffekt zeigt, ob sich die zeitliche Veränderung der Zielgrößen zwischen Patienten der IV im NWpG und Patienten der Standardversorgung unterscheidet.

Aufgrund der Verträge der Krankenkassen und der Träger war eine Randomisierung nicht möglich. Um den Selektionsbias zu kontrollieren, wurden PS-Methoden angewandt.

2.2 Propensity Score-Methoden

Der PS beschreibt die bedingte Wahrscheinlichkeit mit der ein Studienteilnehmer unter Betrachtung ausgewählter Kovariablen die Behandlung erhält, beziehungsweise in diesem Fall in die IV eingeschrieben ist. Rosenbaum und Rubin haben gezeigt, dass es sich hierbei um einen Balancing Score handelt, was so viel bedeutet, wie dass alle Patienten mit gleichem PS die gleiche Verteilung der zugrundeliegenden Kovariablen haben – durch Kontrolle des PS werden also auch die zugrundeliegenden Kovariablen kontrolliert [4; 11].

Eine PS-Analyse umfasst die Auswahl an Variablen, welche dem PS zugrunde liegen sollen, die Schätzung des PS und die Anwendung des PS in Verbindung mit der Effektschätzung. Die PS-Analyse sollte stets um einen Common Support Check und Balance-Tests vor und nach Anwendung des PS ergänzt werden. Die Common Support Area ist der Bereich zwischen dem minimalen PS der Interventionsgruppe und dem maximalen PS der Kontrollgruppe. In diesem Bereich sollte sich der Großteil der Studienteilnehmer befinden, denn dies ist insbesondere für das PS-Matching eine wichtige Voraussetzung. Die Balance-Tests dienen der Prüfung der erfolgreichen Bias-Kontrolle. Der Schematische Ablauf der Propensity Score-Analyse wird in Abbildung 1 dargestellt.

2.2.1 Variablenauswahl

Eine Vorauswahl an möglichen Kovariablen zur Schätzung des PS umfasst Variablen, die standardmäßig in statistischen Analysen kontrolliert werden (wie Alter und Geschlecht), aber auch Variablen, die den Krankheitszustand oder das allgemeine und soziale Umfeld der Person näher definieren und somit die Zielgröße, die Veränderung des Empowerments im Studienverlauf, beeinflussen können. Im Rahmen dieser Analyse wurden 39 Variablen vorausgewählt. Diese Variablen wurden in der Makrovariable `&set0` hinterlegt.

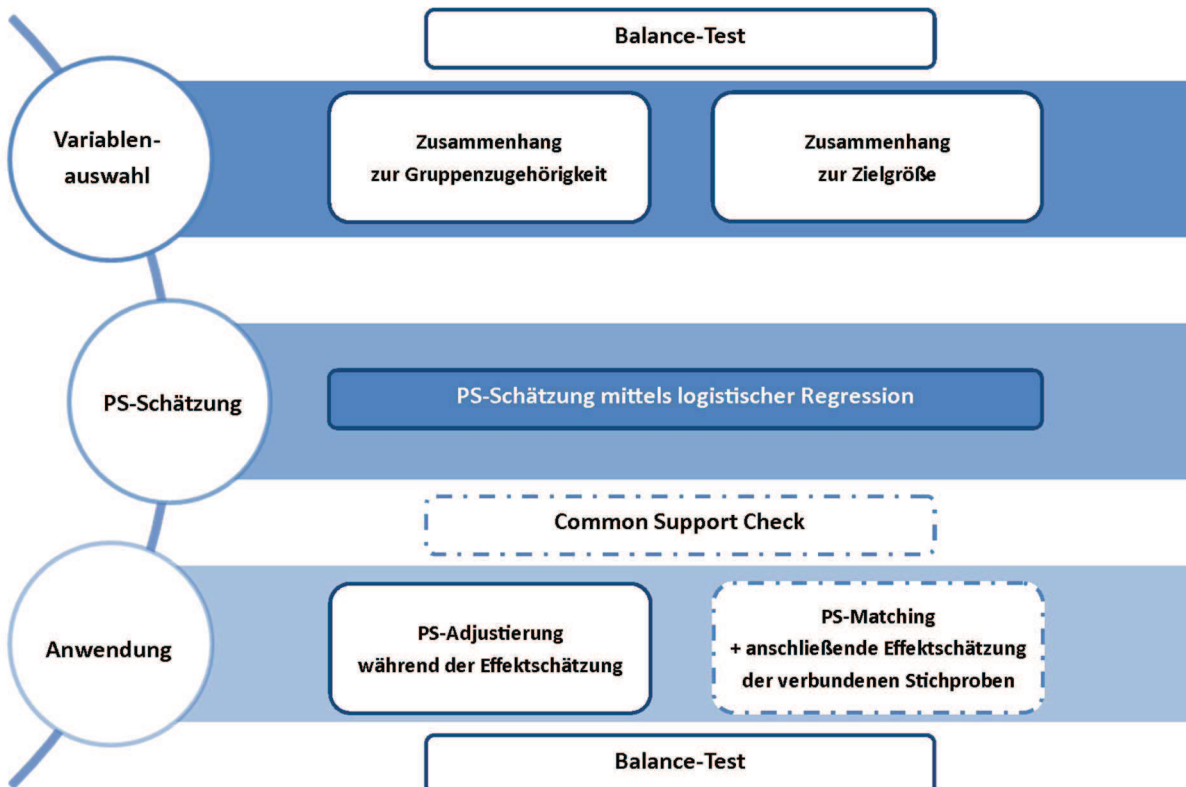


Abbildung 1: Schematischer Ablauf der Propensity Score-Analyse (eigene Darstellung)

Ein Ansatz zur Schätzung des PS ist es, diejenigen Variablen heranzuziehen, welche mit der Zielgröße oder mit der Gruppenzugehörigkeit im Zusammenhang stehen. Die Vorauswahl an Variablen, welche an Visite 1 erhoben wurden, wurde bezüglich ihres Zusammenhanges zur Gruppenzugehörigkeit beziehungsweise mit dem Outcome (Veränderung des EPAS-Gesamtwertes von Visite 1 zu Visite 4; `epastotalch`) untersucht.

Um sicherzustellen, dass für jeden Teilnehmer ein PS geschätzt werden kann, wurden die fehlenden Werte in den Kovariablen einfach imputiert mittels FCS (fully conditional specification) unter Verwendung der logistischen Methode für binäre (`&set0c`) und der Regressionsmethode für kontinuierliche (`&set0n`) Variablen.

```
proc mi data=work.Fragebogen seed=54321
  nimpute=1 out=work.Fragebogen_impcov
  minimum=0 0 ...
  maximum=1 1 ...;
  class &set0c;
  fcs nbiter=10 logistic(&set0c) regression(&set0n);
  var &set0;
run;
```

Zur Bewertung des Zusammenhanges der Variablen mit der Gruppenzugehörigkeit, wurde die standardisierte Differenz mit Hilfe des SAS-Makros von Yang & Dalton bestimmt. Die absolute standardisierte Differenz bietet den Vorteil unabhängig von der Stichprobengröße verwendet werden zu können [16]. Eine standardisierte Differenz größer 0,10 bzw. 0,20 weist auf einen moderaten bzw. starken Zusammenhang zur Gruppenzugehörigkeit hin.

```
%stddiff( inds =ps_p,  
groupvar = versorgung,  
numvars = &set0n,  
charvars = &set0c,  
outds = stddiff_result_pre );
```

Inwieweit die Variablen einen Einfluss auf die Veränderung des EPAS-Gesamtwertes (epastotalch) zeigten, wurde mittels der Prozedur GLM (siehe am Beispiel des Berufsstatus; employment_stat) geprüft. Bei einem p-Wert kleiner 0,10 wurde von einem Zusammenhang zum Outcome ausgegangen.

```
proc glm data=work.Fragebogen_impcov;  
class employment_stat;  
model epastotalch=employment_stat;  
run; quit;
```

Für die PS-Schätzung im Rahmen der Hauptauswertung wurden anschließend diejenigen Variablen ausgewählt, welche entweder mit dem Outcome assoziiert waren ($p < 0,10$) oder für welche Gruppenunterschiede aufgezeigt wurden (absolute standardisierte Differenz $d > 0,10$). Diese wurden in der Makrovariable &set1 hinterlegt. Die PS-Schätzung erfolgte anschließend anhand logistischer Regressionsmodelle.

```
ods select 'Odds-ratios';  
ods output 'Odds-ratios'=work.ps_pot;  
PROC LOGISTIC data = work.Fragebogen_impcov;  
class &set1c;  
model group (event='1.00')=&set1;  
output out= work.ps PREDICTED = pst;  
RUN;
```

2.2.2 Propensity Score-Adjustierung

Es stehen verschiedene PS-Verfahren zur Verfügung, wie die PS-Adjustierung und das PS-Matching. Im Rahmen einer Bachelorarbeit konnte gezeigt werden, dass beide Verfahren auf den Datensatz der IVPOWER-Studie angewendet werden können [12]. Bei Sekundärdatenanalysen mit vergleichsweise hohen Fallzahlen wird meist das PS-Matching der PS-Adjustierung vorgezogen. Anhand simulierter Daten wurde jedoch demonstriert, dass bei einer Fallzahl gleich 500 die PS-Adjustierung dem PS-Matching hinsichtlich der Effektschätzung und der Bias-Reduktion überlegen ist [5; 6].

Um einen möglichen Selektions-Bias zu kontrollieren, erfolgte die Effektschätzung im Rahmen dieser Studie aus diesen Gründen zunächst mit PS-Adjustierung durch Ergänzung des gemischten Regressionsmodells um den Haupteffekt des PS.

```
proc mixed data=data1 covtest noclprint order=internal;  
class region nr group;  
model epastotal = group visit group*visit ps/solution;  
random int visit / subject=nr(region) type=un;  
run;
```

Im Anschluss wurden Balance-Tests durchgeführt, um festzustellen, wie stark die Verteilungen der Kovariablen zwischen beiden Versorgungsgruppen nach der Adjustierung

tatsächlich übereinstimmen. Hierfür wurde die gewichtete, bedingte standardisierte Differenz nach Austin verwendet (Syntax nicht dargestellt) [1].

2.2.3 Propensity Score-Matching

Alternativ erfolgte die Kontrolle des Selektions-Bias durch Verwendung eines PS-Matching-Verfahrens. Es handelte sich hierbei um ein 1:1-Greedy Matching anhand der Nearest-Neighbor-Methode mit Caliper. Hierzu wurde der Matching-Algorithmus von Lanehart et al. genutzt (Syntax nicht dargestellt) [7].

Durch die Verwendung eines Caliper wird ein Toleranzbereich für die maximale Distanz zwischen zwei Matching-Partnern definiert. Für die Wahl des Caliper werden verschiedene Werte empfohlen und es gibt keine allgemeingültigen Standards. Häufig wird die 0,2-fache Standardabweichung des Logits des PS gewählt [2]. In diesem Datensatz entsprach dies einem Caliper gleich 0,252252.

```
data work.ps;
set work.ps;
logit = ln (pst/(1-pst));
run;

proc means data=work.ps Mean Std;
var logit;
run;
```

Das SAS-Makros von Yang & Dalton wurde erneut verwendet um anschließend die Balance des gepaarten Datensatzes anhand der absoluten standardisierten Differenz zu bestimmen [16].

Die Effektschätzung erfolgte ausschließlich für die paarweise zugeordneten Patienten mittels gemischter Regressionsmodelle unter Berücksichtigung der Abhängigkeit der Matching-Partner. Die Abhängigkeit zwischen den Studienzentren wurde in dieser Analyse vernachlässigt.

```
proc mixed data=data2 covtest noclprint order=internal;
class match nr group;
model epastotal = group visit group*visit /solution;
random int visit / subject=nr(match) type=un;
run;
```

2.2.3 Übersicht der Sensitivitätsanalysen

Um auszuschließen, dass die Auswahl der Methodik einen Einfluss auf das Ergebnis hat, wurde untersucht inwieweit unter Anwendung verschiedener PS-Verfahren, die Ergebnisse in Sensitivitätsanalysen repliziert werden konnten. Zum einen wurden verschiedene Kriterien zur Variablenauswahl für die PS-Schätzung herangezogen und zum anderen wurde der PS bei der Effektschätzung auf unterschiedliche Weise berücksichtigt.

Da in der Literatur verschiedene Entscheidungskriterien (Operationalisierungen und Grenzwerte) diskutiert werden, wurden verschiedene Sets an Variablen für die PS-Schätzung verwendet:

- Set 1: Zusammenhang mit Gruppe ($d > 0,10$) **oder** Zielgröße ($p < 0,10$)
- Set 2: starker Zusammenhang mit Gruppe ($d > 0,20$) **oder** Zielgröße ($p < 0,10$)
- Set 3: Zusammenhang mit Gruppe ($d > 0,10$) **und** Zielgröße ($p < 0,10$)
- Set 4: Zusammenhang mit Gruppe ($d > 0,10$)
- Set 5: starker Zusammenhang mit Gruppe ($d > 0,20$)
- Set 6: Zusammenhang mit Zielgröße ($p < 0,10$)
- Set 7: Zusammenhang mit Gruppe ($d > 0,10$) **oder** Zielgröße ($p < 0,10$)
inkl. Krankenkassenzugehörigkeit

Da die IV im NWpG primär durch die Techniker Krankenkasse angeboten wurde, steht die Krankenkassenzugehörigkeit in einem sehr starken Zusammenhang zur Versorgungsgruppe (Techniker Krankenkasse: TAU: 12,7 %, IV: 70,0 %). Ein PS, welcher diese Variable miteinbezieht, würde zwischen den beiden Gruppen zu stark diskriminieren. Da angenommen wurde, dass die Krankenkassenzugehörigkeit nicht die Zielgröße beeinflusst, wurde die Variable Krankenkassenzugehörigkeit im Rahmen der Hauptauswertung nicht in die PS-Schätzung einbezogen. Im Rahmen einer Sensitivitätsanalyse sollte jedoch geprüft werden, welchen Einfluss diese Entscheidung auf die Ergebnisse zeigte (siehe Set 7).

Da sich die Rekrutierungsstrategien in den verschiedenen Studienregionen deutlich voneinander unterscheiden, ist es denkbar, dass die Ausprägung des Selektionsbias zwischen den Studienregionen variierte. Aus diesem Grund erfolgt in dem sogenannten Set 8 die PS-Schätzung getrennt für die fünf Versorgungsregionen basierend auf der ursprünglichen Variablenauswahl (Set 1). Anschließend erfolgte unter Verwendung der verschiedenen PS der acht Sets die PS-Adjustierung.

In einer weiteren Analyse wurden Variablen, welche nach PS-Adjustierung mittels Set 1 noch stark unbalanziert waren ($d > 0,20$) zur multivariaten Adjustierung neben dem PS hinzugezogen.

```
proc mixed data=datal covtest noclprint order=internal;  
class region nr group;  
model epastotal = group visit group*visit pst  
var1 var2 var3/solution;  
random int visit / subject=nr(region) type=un;  
run;
```

Des Weiteren wurde das PS-Matching mit zehn verschiedenen Caliper durchgeführt. Neun äquidistante Caliper (0,1 bis 0,9) wurden neben dem Caliper entsprechend dem 0,2-fachen der Standardabweichung des Logits des PS verwendet.

3 Ergebnisse

3.1 IVPOWER-Studie

Die Baseline-Erhebung hat mit 251 Teilnehmern in der Kontrollgruppe und 260 Teilnehmern in der Interventionsgruppe stattgefunden. Insgesamt haben 83 Patienten die Studie vorzeitig abgebrochen. Somit liegt die Dropout-Rate bei 16,2 %. Die Baseline-Daten zeigen, dass es bereits zu Beginn der nicht-randomisierten Studie wesentliche Unterschiede in den Charakteristika der beiden Versorgungsgruppen gibt. In der Kontrollgruppe hatten mehr Patienten eine schizophrene Diagnose als in der IV-Gruppe. Patienten in der Kontrollgruppe waren bereits länger erkrankt, als Patienten in der IV-Gruppe und hatten dementsprechend auch bereits mehrere stationäre Aufenthalte aufgrund der psychischen Erkrankung. Verschiedene Kennwerte deuteten darauf hin, dass die Patienten der IV-Gruppe einen höheren Sozialstatus aufwiesen und besser in soziale Strukturen integriert waren.

Tabelle 1: Deskriptive Darstellung der Studienpopulation zur Baseline-Erhebung

	Kontrolle	Intervention	p
Berufstätigkeit, Anteil	23,5%	46,4%	<,0001
Krankengeldbezug, Anteil	17,2%	44,6%	<,0001
Bezug von Hartz-IV-Leistungen, Anteil	16,7%	10,4%	0,0360
Bezug von Leistungen der Sozialhilfe, Anteil	8,0%	2,3%	0,0040
alleinlebend, Anteil	50,6%	41,5%	0,0400
Kontakt zur Familie, > 1x pro Monat, Anteil	80,4%	87,9%	0,0200
Versicherte der Techniker Krankenkasse, Anteil	12,7%	70,0%	0,0001
Diagnose nach ICD-10, F20-F29, Anteil	19,1%	10,0%	0,0030
Diagnose nach ICD-10, F30-F39, Anteil	76,1%	81,2%	0,1630
Erkrankungsdauer in Jahren, Mittelwert	14,24	10,77	0,0010
schwere psychische Erkrankung, Anteil	26,4%	33,5%	0,0820
Anzahl an Klinikaufenthalten, Mittelwert	3,83	1,77	<,0001
Einnahme von psych. Medikamenten, Anteil	79,3%	69,2%	0,0090
Betreuer, Anteil	9,2%	3,1%	0,0040
betreute Wohnform, Anteil	12,4%	1,9%	<,0001
Motivation zur ambulanten Versorgung, Anteil	71,3%	79,2%	0,0380

Für die PS-Schätzung in der Hauptauswertung wurden 31 Variablen ausgewählt (Set 1), für welche entweder ein Zusammenhang zum Zielkriterium (Differenz des EPAS-Gesamtwertes von Visite 1 und Visite 4) oder ein Gruppenunterschied besteht. Es zeigt sich eine große Bandbreite an berücksichtigten Variablen, welche sowohl soziodemographische Variablen, aber auch Informationen zur medizinischen Vorgeschichte und zu den Zielgrößen umfasst.

- soziodemographische Variablen: Alter, Berufstätigkeit, Kinder, alleinlebend, Geschlecht, Schulabschluss, in Partnerschaft, Bezug von Hartz IV, Bezug von Sozialhilfe, Kontakt zur Familie
- medizinische Vorgeschichte: Erkrankungsdauer, Anzahl an Klinikaufenthalten, Krankengeldbezug, schizophrene Erkrankung, affektive Erkrankung, Anzahl psychische Diagnosen, Einnahme psych. Medikamente, Motivation ambulant versorgt zu werden, gesetzlicher Betreuer, betreute Wohnform, Body Mass Index, Raucher
- Zielgrößen: psychische Beeinträchtigung (HoNOS), Versorgungsbedarf & Bedarfsdeckung (CAN), Lebensqualität (WHOQoL-bref), Behandlungszufriedenheit (ZUF-8), direkte Krankheitskosten

Trotz der Baseline-Unterschiede, zeigt der Common Support Check eine große Überlappung der Histogramme der PS beider Gruppen. Tatsächlich befinden sich fast 500 Patienten in der Common Support Area ($n = 497$, 97,3%), weshalb davon auszugehen ist, dass sich die Gruppenunterschiede zu Baseline durch PS-Methoden kontrollieren lassen (siehe Abbildung 2).

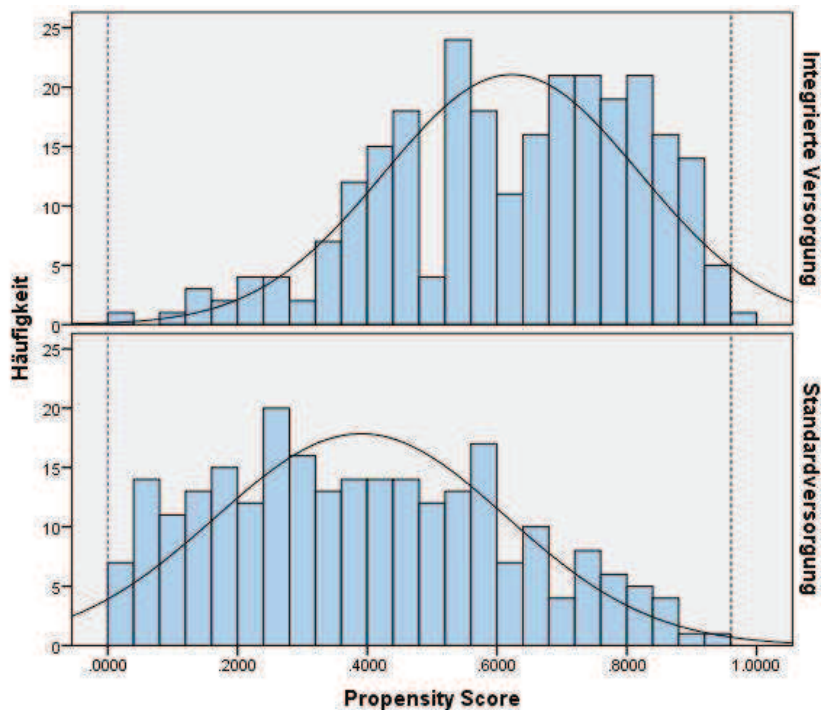


Abbildung 2: Graphische Darstellung der Propensity Scores an der Baseline-Messung

In Tabelle 2 sind die Ergebnisse bezogen auf den Empowerment-Gesamtwert und die EPAS-Teilskala zur Behandlungspartizipation (EPAS-3) dargestellt. Beim Gesamtwert handelt es sich um die primäre Zielgröße der IVPOWER-Studie. Die Kenngröße EPAS-3 wird ebenfalls dargestellt, weil für diesen Parameter ein Interventionseffekt aufgezeigt werden konnte.

Die Parametrisierung der Regressionsmodelle erfolgte ohne Mittelwertzentrierung. Folglich beziehen sich alle Effekte auf die Referenzkategorien der restlichen Modellvariablen. Das heißt der Schätzer für IV zeigt inwieweit sich die Kennwerte zu Baseline

zwischen den Gruppen unterschieden. Der Schätzer für die Zeit zeigt inwieweit sich die Kennwerte über die Zeit in der Kontrollgruppe verändert haben.

Relevant ist insbesondere der Interaktionseffekt. Der Schätzer für Interaktion Zeit * IV zeigt inwieweit sich die zeitliche Veränderung zwischen der IV-Gruppe und der TAU-Gruppe unterscheidet.

Die Ergebnisse der gemischten Regression für den EPAS-Gesamtwert belegen keinen signifikanten Interaktionseffekt zwischen der Zeit und der Behandlungsgruppenzugehörigkeit ($b = 0,02$; $p = 0,2579$), jedoch einen signifikanten positiven Zeiteffekt ($b = 0,07$; $p < 0,0001$). Während sich der Empowerment-Gesamtwert also für alle Studienteilnehmer signifikant verbesserte, erzielten Patienten der IV keine größere Verbesserung als Kontrollpatienten.

Für die EPAS-Teilskala zu ‚Behandlungspartizipation und Medikamente‘ (EPAS-3) zeigte sich sowohl ein signifikanter positiver Zeiteffekt ($b = 0,04$; $p = 0,0051$) als auch ein signifikanter positiver Interaktionseffekt zwischen der Zeit und der Behandlungsgruppe ($b = 0,06$; $p = 0,0014$). Während sich also die Behandlungspartizipation in beiden Behandlungsgruppen im Studienverlauf signifikant verbesserte, erzielten Patienten der IV eine stärkere Verbesserung als TAU-Patienten.

Tabelle 2: Ergebnisse der gemischten Regressionsmodelle für den EPAS-Gesamtwert und für die EPAS-Teilskala zur Behandlungspartizipation (Dimension 3)

EPAS-Gesamtwert	b	se	p	95 %-CI	Modellinformation
Konstante	3,27	0,063	<,0001	3,15 - 3,39	n = 511
IV	-0,05	0,065	0,4056	-0,18 - 0,07	BIC = 1635,4
Zeit	0,07	0,009	<,0001	0,05 - 0,09	
Zeit * IV	0,02	0,013	0,2579	-0,01 - 0,04	
PS	0,23	0,120	0,0536	0,00 - 0,47	
EPAS-Dimension 3	b	se	p	95 %-CI	Modellinformation
Konstante	3,71	0,067	<,0001	3,58 - 3,85	n = 511
IV	-0,16	0,072	0,0259	-0,30 - -0,02	BIC = 2740,7
Zeit	0,04	0,013	0,0051	0,01 - 0,06	
Zeit * IV	0,06	0,018	0,0014	0,02 - 0,09	
PS	0,01	0,125	0,9284	-0,23 - 0,26	

Die Balance-Tests zeigten, dass nach der PS-Adjustierung nur noch drei von ursprünglich 16°Variablen stark unbalanziert (absolute standardisierte Differenz > 0,2) waren. Unbalanziert blieben die Krankenkassenzugehörigkeit, der prozentuale Anteil an ungedeckten Bedarfen und die Anzahl an bisherigen Krankenhausaufenthalten.

3.2 Sensitivitätsanalysen

Im Folgenden werden die Ergebnisse für die Sensitivitätsanalysen unter Verwendung unterschiedlicher Sets an Variablen zur PS-Schätzung und bei Anwendung verschiedener PS-Methoden dargestellt.

3.2.1 Variablenauswahl

Um abzuklären, inwieweit die Auswahl der Variablen, die in den PS einbezogen wurden, einen Einfluss auf die Effektschätzung hatte, wurden verschiedene Selektionsmethoden durchgeführt. Die Auswertung wurde mit diesen PS-Schätzwerten wiederholt.

Unabhängig vom Selektionsverfahren für die PS-Schätzung befand sich der Großteil der Teilnehmer in der Common Support Area (97,5 % bis 98,4 % für Set 1 bis Set 6). Dies spricht für eine große Überlappung der PS-Verteilungen und deutet auf eine hohe Vergleichbarkeit der Studiengruppen und einen geringen Selektions-Bias hin.

Für Set 7, in welchem die Krankenkassenzugehörigkeit in die PS-Schätzung einbezogen wurde, zeigte sich für die IV-Gruppe bzw. die Kontrollgruppe eine rechts- bzw. links-gipflige PS-Verteilung. In diesem Fall umfasst die Common Support Area nur 90,6 % (463) der Patienten. Die geringste Überlappung der PS-Verteilungen der beiden Gruppen zeigte sich, wenn die PS-Schätzung für jede Versorgungsregion separat erfolgte (Set 8). In diesem Fall befinden sich nur 79,6 % der Patienten in der Common Support Area.

Tabelle 3: Anzahl (N) bzw. Anteil (%) an Teilnehmern in der Common Support Area während der Baseline-Messung bei Verwendung unterschiedlicher Propensity Score-Schätzungen

	Common Support Area	
	N	%
Set 1	498	97,5%
Set 2	498	97,5%
Set 3	502	98,2%
Set 4	500	97,8%
Set 5	500	97,8%
Set 6	503	98,4%
Set 7	463	90,6%
Set 8	407	79,6%

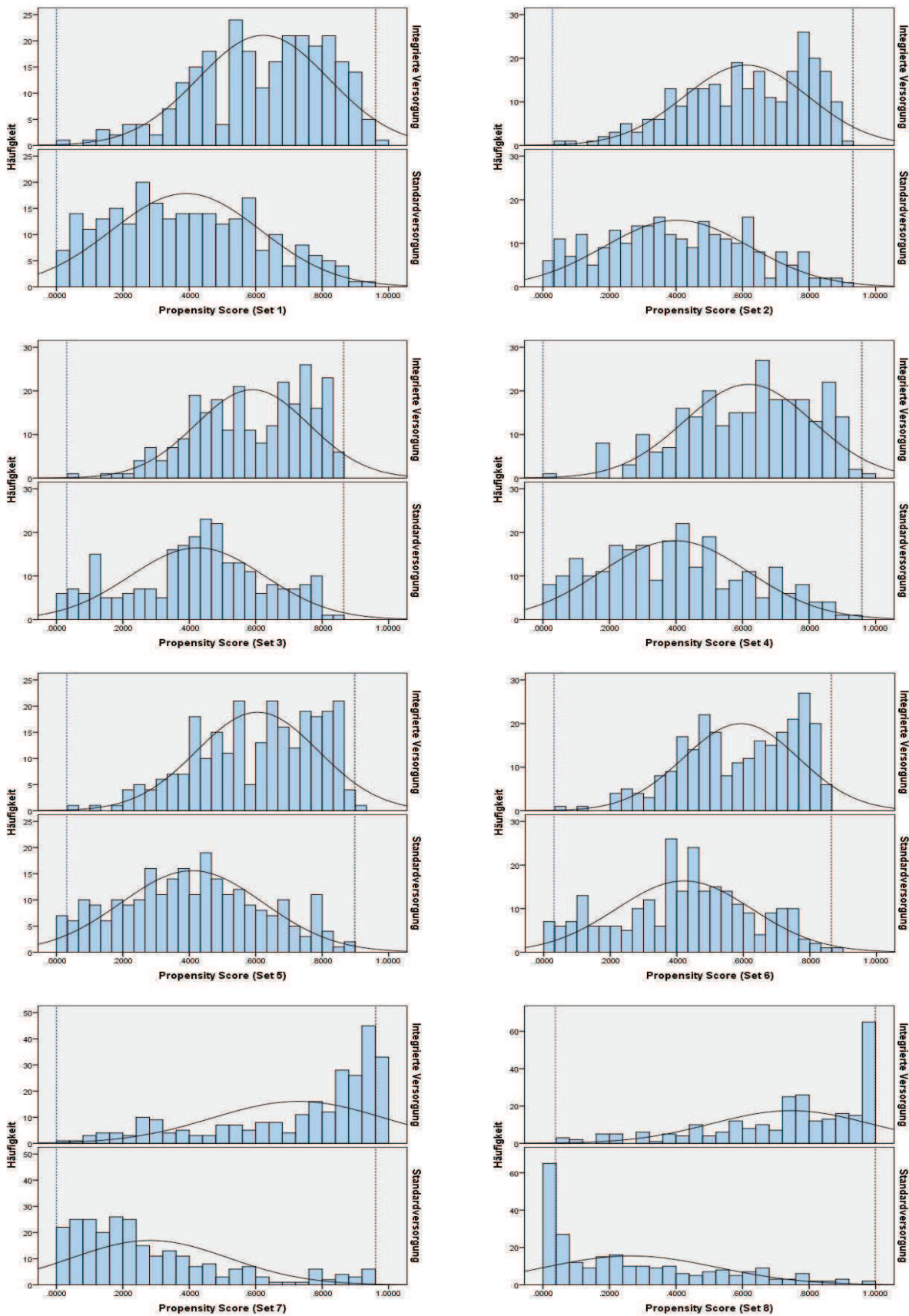


Abbildung 3: Graphische Darstellung der Propensity Scores bei Verwendung unterschiedlicher Propensity Score-Schätzungen an der Baseline-Messung

Die Balance-Tests nach Anwendung der PS-Adjustierung zeigten ein homogenes Bild unter Verwendung der verschiedenen PS-Schätzungen: etwa ein Drittel bzw. 10 % der beobachteten Variablen waren weiterhin unbalanziert ($d > 0,10$) bzw. stark unbalanziert ($d > 0,20$). Die minimale Anzahl unbalanzierter Variablen erzielte die Auswahl von Variablen ausschließlich angesichts dem Zusammenhang zur Gruppenzugehörigkeit (Set 4, $n = 11$ (28,2 %)). Der Anteil stark unbalanzierter Variablen stieg bei der regionalen PS-Schätzung (Set 8) auf 23,1 % an.

Tabelle 4: Balance-Tests nach PS-Adjustierung bei Verwendung unterschiedlicher Propensity Score-Schätzungen

	Balance Tests	
	$d > 0,1$	$d > 0,2$
Set 1	12 (30,8%)	3 (7,7%)
Set 2	13 (33,3%)	3 (7,7%)
Set 3	14 (35,9%)	5 (12,8%)
Set 4	11 (28,2%)	3 (7,7%)
Set 5	14 (35,9%)	3 (7,7%)
Set 6	13 (33,3%)	5 (12,8%)
Set 7	13 (33,3%)	1 (2,6%)
Set 8	15 (38,5%)	9 (23,1%)

Die Ergebnisse in Tabelle 5 und Tabelle 6 zeigen, dass das Verfahren der PS-Schätzung jedoch keinen Einfluss auf die Studienergebnisse bezüglich des EPAS-Gesamtwertes und der EPAS-Teilskala zur Behandlungspartizipation (Dimension 3) hat. Für den EPAS-Gesamtwert gibt es in allen Modellen einen signifikanten positiven Zeiteffekt, jedoch keinen signifikanten Interaktionseffekt. Hinsichtlich der EPAS-Teilskala zur Behandlungspartizipation sind sowohl der Zeit- als auch der Interventionseffekt robust bezüglich der Variablenauswahl. Nur bei Verwendung von Set 7 und Set 8 erreicht der Gruppeneffekt (zu Baseline) keine Signifikanz mehr.

Tabelle 5: Ergebnisse der gemischten Regressionsmodelle für den EPAS-Gesamtwert bei Verwendung unterschiedlicher Propensity Score-Schätzungen

	Set 1		Set 2		Set 3		Set 4	
	b	p	b	p	b	p	b	p
Konstante	3,27	<,0001	3,26	<,0001	3,30	<,0001	3,19	<,0001
IV	-0,05	0,4056	-0,05	0,4031	-0,02	0,7314	-0,09	0,1441
Zeit	0,07	<,0001	0,07	<,0001	0,07	<,0001	0,07	<,0001
Zeit * IV	0,02	0,2579	0,02	0,2578	0,01	0,2644	0,02	0,2556
PS	0,23	0,0536	0,26	0,0391	0,13	0,3294	0,42	0,0005
	Set 5		Set 6		Set 7		Set 8	
	b	p	b	p	b	p	b	p
Konstante	3,15	<,0001	3,28	<,0001	3,32	<,0001	3,34	<,0001
IV	-0,10	0,1134	-0,03	0,5991	-0,07	0,3589	-0,04	0,6023
Zeit	0,07	<,0001	0,07	<,0001	0,07	<,0001	0,07	<,0001
Zeit * IV	0,02	0,2544	0,01	0,2634	0,01	0,2643	0,01	0,2637
PS	0,51	<,0001	0,19	0,1482	0,15	0,1333	0,08	0,4168

Tabelle 6: Ergebnisse der gemischten Regressionsmodelle für die EPAS-Teilskala zur Behandlungspartizipation (Dimension 3) bei Verwendung unterschiedlicher Propensity Score-Schätzungen

	Set 1		Set 2		Set 3		Set 4	
	b	p	b	p	b	p	b	p
Konstante	3,71	<,0001	3,69	<,0001	3,71	<,0001	3,67	<,0001
IV	-0,16	0,0259	-0,17	0,0142	-0,16	0,0227	-0,18	0,0101
Zeit	0,04	0,0051	0,04	0,0051	0,04	0,0051	0,04	0,0051
Zeit * IV	0,06	0,0014	0,06	0,0014	0,06	0,0014	0,06	0,0014
PS	0,01	0,9284	0,08	0,5329	0,01	0,9491	0,12	0,3473
	Set 5		Set 6		Set 7		Set 8	
	b	p	b	p	b	p	b	p
Konstante	3,62	<,0001	3,69	<,0001	3,74	<,0001	3,73	<,0001
IV	-0,20	0,0038	-0,17	0,0165	-0,13	0,1182	-0,13	0,1076
Zeit	0,04	0,0051	0,04	0,0051	0,04	0,0051	0,04	0,0051
Zeit * IV	0,06	0,0013	0,06	0,0014	0,06	0,0014	0,06	0,0014
PS	0,24	0,0705	0,06	0,6720	-0,07	0,5123	-0,05	0,6636

3.2.2 Adjustierung für unbalanzierte Variablen

Da nach der PS-Adjustierung noch drei Variablen stark unbalanziert (absolute standardisierte Differenz > 0,2) waren, wurde in weiteren gemischten Regressionsanalysen für diese Variablen zusätzlich adjustiert. Die Ergebnisse der gemischten Regressionsmodelle für den EPAS-Gesamtwert und die EPAS-Teilskala zur Behandlungspartizipation

(Dimension 3) wurden bestätigt bei weiterer Adjustierung für die unbalanzierten Variablen.

Tabelle 7: Ergebnisse der gemischten Regressionsmodelle für den EPAS-Gesamtwert und für die EPAS-Teilskala zur Behandlungspartizipation (Dimension 3) bei Adjustierung für unbalanzierte Variablen

	Gesamtwert		Dimension 3	
	b	p	b	p
Konstante	3,42	<,0001	3,76	<,0001
IV	-0,10	0,1647	-0,16	0,0417
Zeit	0,07	<,0001	0,04	0,0072
Zeit * IV	0,02	0,2355	0,06	0,0009
PS	0,35	0,0142	0,07	0,6448
Anzahl Klinikaufenthalte	0,00	0,7038	0,00	0,9115
Anteil ungedeckter Bedarf nicht bei der Techniker Krankenkasse	-0,41	<,0001	-0,23	0,0091
	-0,08	0,204	0,01	0,8526

3.2.3 Propensity Score-Matching

Im Folgenden wurde die Robustheit der Ergebnisse bei Verwendung eines PS-Matching-Verfahrens mit unterschiedlichen Caliper untersucht.

Wie zu erwarten war, steigt mit wachsender Toleranz, i. S. eines höheren Caliper, die Anzahl an paarweise zugeordneten Patienten, aber auch die Anzahl an unbalanzierten Variablen. Ab einem Caliper gleich 0,8 wurde allen 251 Kontrollpatienten ein IV-Patient zugeordnet und der Selektions-Bias (gemessen an der Anzahl unbalanzierter Variablen) wurde kaum reduziert.

Tabelle 8: Ergebnisse der Balance-Tests bei Verwendung unterschiedlicher Caliper zum Propensity Score-Matching

	N	Power	d > 0,1 n (%)	d > 0,2 n (%)
Caliper = 0,1	306	84,6%	7 (17,9%)	1 (2,6%)
Caliper = 0,2	332	87,7%	8 (20,5%)	1 (2,6%)
Caliper = 0,252252	348	89,3%	7 (17,9%)	1 (2,6%)
Caliper = 0,3	374	91,5%	16 (41,0%)	2 (5,1%)
Caliper = 0,4	412	94,0%	21 (53,8%)	4 (10,3%)
Caliper = 0,5	442	95,5%	24 (61,5%)	5 (12,8%)
Caliper = 0,6	472	96,7%	25 (64,1%)	12 (30,8%)
Caliper = 0,7	488	97,2%	26 (66,7%)	12 (30,8%)
Caliper = 0,8	502	97,5%	28 (71,8%)	15 (38,5%)
Caliper = 0,9	502	97,5%	28 (71,8%)	16 (41,0%)
Adjustierung	511	97,7%	12 (30,8%)	3 (7,7%)

Die Ergebnisse bei Anwendung des PS-Matchings mit unterschiedlichen Caliper entsprachen den Ergebnissen der Hauptauswertung unter Anwendung der PS-Adjustierung.

Tabelle 9: Ergebnisse der gemischten Regressionsmodelle für den EPAS-Gesamtwert bei Anwendung eines Propensity Score-Matching-Verfahrens mit unterschiedlichen Caliper

	0,1		0,2		0,252252		0,3		0,4	
	b	p	b	p	b	p	b	p	b	p
Konstante	3,41	<,0001	3,38	<,0001	3,38	<,0001	3,38	<,0001	3,39	<,0001
IV	-0,05	0,5016	-0,02	0,7404	-0,02	0,8143	-0,02	0,741	-0,05	0,4597
Zeit	0,07	<,0001	0,08	<,0001	0,08	<,0001	0,07	<,0001	0,07	<,0001
Zeit * IV	0,01	0,4485	0,01	0,6044	0,01	0,4433	0,02	0,3174	0,02	0,1415
	0,5		0,6		0,7		0,8		0,9	
	b	p	b	p	b	p	b	p	b	p
Konstante	3,38	<,0001	3,37	<,0001	3,36	<,0001	3,36	<,0001	3,36	<,0001
IV	-0,04	0,4773	-0,02	0,7264	-0,02	0,798	0,00	0,9531	0,00	0,9759
Zeit	0,07	<,0001	0,07	<,0001	0,07	<,0001	0,07	<,0001	0,07	<,0001
Zeit * IV	0,02	0,1871	0,01	0,2835	0,01	0,3052	0,02	0,2126	0,01	0,2715

Tabelle 2: Ergebnisse der gemischten Regressionsmodelle für die EPAS-Teilskala zur Behandlungspartizipation (Dimension 3) bei Anwendung eines Propensity Score-Matching-Verfahrens mit unterschiedlichen Caliper

	0,1		0,2		0,252252		0,3		0,4	
	b	p	b	p	b	p	b	p	b	p
Konstante	3,74	<,0001	3,70	<,0001	3,70	<,0001	3,71	<,0001	3,75	<,0001
IV	-0,20	0,0168	-0,15	0,0747	-0,15	0,0592	-0,18	0,0185	-0,23	0,0020
Zeit	0,06	0,0006	0,06	0,0002	0,06	0,0002	0,05	0,0011	0,04	0,0062
Zeit * IV	0,06	0,0092	0,05	0,0280	0,05	0,0141	0,06	0,0027	0,07	0,0002
	0,5		0,6		0,7		0,8		0,9	
	b	p	b	p	b	p	b	p	b	p
Konstante	3,73	<,0001	3,73	<,0001	3,72	<,0001	3,72	<,0001	3,72	<,0001
IV	-0,20	0,0056	-0,18	0,0073	-0,17	0,0112	-0,16	0,0148	-0,16	0,0148
Zeit	0,04	0,0081	0,04	0,0048	0,04	0,0041	0,04	0,0051	0,04	0,0051
Zeit * IV	0,06	0,0010	0,06	0,0016	0,06	0,0012	0,06	0,0011	0,06	0,0012

4 Diskussion

Es handelte sich hierbei um eine prospektive Studie mit 511 Teilnehmern mit psychischen Erkrankungen aus fünf Regionen in Deutschland, die dem Vergleich der IV im NWpG und der Standardversorgung diente. Aufgrund der Rahmenbedingungen war eine Randomisierung der Studienteilnehmer in die beiden Versorgungsgruppen nicht möglich. Der fortwährenden Kritik bezüglich des potentiellen Selektionsbias der Studie sollte durch PS-Verfahren Rechnung getragen werden. Die Hauptauswertung erfolgte

mittels Adjustierung für einen PS, welcher auf Variablen beruhte, die einen Zusammenhang zur Gruppenzugehörigkeit oder zum Outcome aufwiesen.

Obwohl keine signifikante Überlegenheit der IV für die Wirksamkeit hinsichtlich dem EPAS-Gesamtwert nachgewiesen konnte, wurde eine signifikante Überlegenheit hinsichtlich der Empowerment-Dimension ‚Behandlungspartizipation und Medikamente‘ aufgezeigt. Patienten, die eine IV erhalten haben, zeigen gegenüber Patienten in der Standardversorgung eine stärkere Verbesserung der Einschätzung ihrer Möglichkeiten zur aktiven Einflussnahme auf Entscheidungen im Rahmen ihrer psychiatrischen Behandlung, einschließlich der Verordnung psychiatrischer Medikamente.

Da kein Konsens über die optimale PS-Methodik besteht, wurden in umfassenden Sensitivitätsanalysen verschiedene PS-Verfahren angewandt. Die Ergebnisse bezogen auf den EPAS-Gesamtwert und die EPAS-Dimension Behandlungspartizipation konnten jeweils repliziert werden. Es bleibt jedoch zu bedenken, dass möglicherweise in der Vorauswahl der 39 Variablen nicht alle bias-relevanten Merkmale der Studienteilnehmer berücksichtigt wurden. Falls dies der Fall ist, können sogenannte unbekannte Confounder die Analyseergebnisse deutlich verzerren [5].

Für den stetig wachsenden Bereich der Versorgungsforschung, in welchem häufig nicht-randomisierte Studien eingesetzt werden müssen, leiten wir dennoch basierend auf unseren Erfahrungen die Empfehlung ab, Propensity Score Methoden anzuwenden und die Robustheit der Ergebnisse in umfassenden Sensitivitätsanalysen zu prüfen. Da es derzeit jedoch an Literatur zur PS-Methodik bei Primärdaten mangelt, existiert derzeit kaum eine Entscheidungsgrundlage für die Wahl der PS-Methodik für die Primäranalyse. Dieses Defizit soll in weiteren Simulationsstudien behoben werden. Wie in einer solchen Simulationsstudie am Beispiel der PS-Adjustierung und dem PS-Matching gezeigt werden konnte, sind die Empfehlungen bezüglich der PS-Verfahren, welche anhand von großen Datensätzen abgeleitet wurden, nicht ohne weiteres auf die Analyse von Primärdaten übertragbar [5]. Aufbauend auf eine Bachelorarbeit [12], soll in einer solchen Simulationsstudie z. B. auch das Problem der Variablenauswahl adressiert werden. Zur Standardisierung und Vereinfachung weiterer Forschungsbestrebungen soll deshalb ein SAS-Makro programmiert werden. Als Grundlage wird das SAS-Makro von Nill et al. [8; 10] dienen.

Weiterer Forschungsbedarf wurde auch für die Idee der regionalen PS-Schätzung (vgl. Set 8) identifiziert. Der Idee lag die Tatsache zu Grunde, dass sich der Rekrutierungsprozess zwischen den Studienzentren unterscheiden kann und, dass somit regionenspezifische Selektionsprozesse aufgetreten sein können, welchen man durch unabhängige PS-Schätzungen der verschiedenen Regionen besser modellieren könnte. Dieses Problem ist für alle multizentrischen Studien denkbar. Unklar ist jedoch, inwieweit identische PS verschiedener Regionen miteinander vergleichbar sind und somit simultan zur PS-Adjustierung verwendet werden können. Hinzu kommt das Problem der reduzierten Fallzahlen bei der regionalen PS-Schätzung. Die PS-Methodik in multizentrischen Studien soll deshalb zukünftig weiterentwickelt und untersucht werden.

5 Schlussfolgerung

Was die IVPOWER-Studie betrifft, weisen die breite Common Support Area und die Balance-Tests darauf hin, dass die Studiengruppen recht vergleichbar waren und vermutlich kein bedeutender Selektions-Bias vorlag. In umfassenden Sensitivitätsanalysen wurden die Ergebnisse bezüglich dem EPAS-Gesamtwert und bezüglich der EPAS-Teilskala Behandlungspartizipation bestätigt. Es ist deshalb davon auszugehen, dass die IV tatsächlich einen positiven Effekt auf die Behandlungspartizipation hat und, dass es sich dabei um keinen Zufallseffekt aufgrund der multiplen Testprozedur handelt.

Im Allgemeinen gilt, dass die Anwendung von PS-Verfahren mit einer großen Anzahl an methodischen Entscheidungen verbunden ist (z. B. Umgang mit fehlenden Werten bei der PS-Schätzung, Variablenauswahl zur PS-Schätzung, Aufnahme von Interaktionseffekten zur PS-Schätzung, Parametrisierung der PS-Modelle, Auswahl eines Caliper für PS-Matching, u. v. m.), welche das Ergebnis des PS-Verfahrens und somit auch die Effektschätzung beeinflussen können. Bislang existiert nur eine schwache Entscheidungsgrundlage für die Auswahl von PS-Verfahren, insbesondere bei vergleichsweise kleinen Fallzahlen von nicht-randomisierten Studien im Bereich der Versorgungsforschung. Solange es keine klaren Empfehlungen gibt, ist es deshalb ratsam die Robustheit der Ergebnisse von PS-Analysen in umfassenden Sensitivitätsanalysen unter Verwendung verschiedener PS-Verfahren zu prüfen.

Literatur

- [1] P. C. Austin, *Pharmacoepidemiol Drug Saf* 2008, 17, 1202–1217, DOI: 10.1002/pds.1673.
- [2] W. G. Cochran und D. B. Rubin, *Sankhya Ser A* 1973, 35, 417–446.
- [3] Dachverband Gemeindepsychiatrie e.V., *Grundlagen der Integrierten Versorgung*, <http://www.dvgp.org/themen-engagement/integrierte-versorgung/grundlagen-der-integrierten-versorgung.html>, 03/03/2017.
- [4] S. Guo und M. W. Fraser, *Propensity score analysis*, Sage Publications, Thousand Oaks, 2010.
- [5] N. Lamp, *Effektschätzung mit Hilfe des Propensity Scores*, Ruprecht-Karls-Universität Heidelberg, Heidelberg, 2017.
- [6] N. Lamp, A. S. Mueller-Stierlin, V. M. Schoening und R. Kilian, in: 21. Konferenz der SAS-Anwender in Forschung und Entwicklung, 09.-10.03.2017.
- [7] R. E. Lanehart, P. Rodriguez de Gil, E. S. Kim, A. P. Bellara, J. D. Kromrey und R. S. Lee, in: SAS Global Forum, 2012.
- [8] B. Mayer und S. Nill, in: 22. Konferenz der SAS-Anwender in Forschung und Entwicklung, 01.-02.03.2018.

- [9] A. S. Mueller-Stierlin, *Wirksamkeit und Effizienz der Integrierten Versorgung von Menschen mit psychischen Erkrankungen unter Realweltbedingungen bei besonderer Berücksichtigung der Verbesserung von Empowerment und Lebensqualität – die IVPOWER-Studie –*, Universität Ulm, Ulm, 2018.
- [10] S. Nill, *Entwicklung eines zweistufigen Matching- Algorithmus für Beobachtungsdaten aus verschiedenen Registern*, Universität Ulm, Ulm, 2017.
- [11] P. R. Rosenbaum und D. B. Rubin, *Biometrika* 1983, 70, 41–55, DOI: 10.2307/2335942.
- [12] V. M. Schoening, *Propensity Score-Methoden zur Kontrolle des Selektionsbias bei nicht-randomisierten Studien*, Hochschule Ulm, Ulm, 2015.
- [13] A. S. Stierlin, T. Becker, M. Schuetzwohl und R. Kilian, in: *Forschungskongress der Fachkliniken der bayerischen Bezirke* (Hrsg.: M. Wolfersdorf and G. Laux), Roderer, Regensburg, 2014, 154–170.
- [14] A. S. Stierlin, K. Herder, M. J. Helmbrecht, S. Prinz, J. Walendzik, M. Holzmann, T. Becker, M. Schuetzwohl und R. Kilian, *BMC Psychiatry* 2014, 14, 163, DOI: 10.1186/1471-244X-14-163.
- [15] S. Weinmann, *Erfolgsmythos Psychopharmaka*, 1. Aufl., Psychiatrie Verlag, Bonn, 2008.
- [16] D. Yang und J. E. Dalton, in: *SAS Global Forum*, 2012.