

Neue Methoden zum Vergleich von zwei Gruppen in interventionellen Beobachtungsstudien

Swora, M., Pfaff, H., Karbach, U.

Hintergrund

Häufig tritt in der statistischen Praxis das Problem auf, die Wirkung einer Intervention im Rahmen eines Experimental-Kontrollgruppen-Designs statistisch nachzuweisen. Hierbei gilt das randomisierte Experiment als der Goldstandard [1]. Leider sind randomisierte Experimente in der Versorgungsforschung oft nicht realisierbar. Um dennoch eine Vergleichbarkeit beider Gruppen zu gewährleisten und Selektionsbias zu minimieren, möchte man für beobachtete Confounder adjustieren. Neben der gewöhnlichen Regressionsadjustierung wird häufig die Propensity Score Methode [2] angewandt.

Von großer Relevanz ist die Frage, welche statistischen Methoden verfügbar sind, um im einleitend genannten Setting Selektionsbias zu vermeiden. Hierbei ist die besondere Berücksichtigung neu entwickelter Methoden angezeigt. Weiterhin stellt sich die Aufgabe, die Vor- und Nachteile der verfügbaren Verfahren gegenüberzustellen.

Methode

Im Rahmen eines narrativen Reviews wurde eine selektive Literaturrecherche (Cross-Reference-Suche) zu existierenden Methoden zur Korrektur von Selektionsbias durchgeführt. Es wurden Methoden eingeschlossen, die darauf abzielen, die Propensity Score Technik zu verbessern. Bisherige Ergebnisse zur vergleichenden Performance dieser Methoden wurden begutachtet. Weiter wurden verschiedene Eigenschaften der Verfahren verglichen, um eine Einschätzung ihrer Vor- und Nachteile zu ermöglichen. Ergänzend wurden neue Entwicklungen zur Durchführung unterstützender Analysen miteinbezogen.

Ergebnisse

Es wurden vier neue Adjustierungsmethoden zur Korrektur von Selektionsbias sowie mehrere neue Methoden zur Durchführung unterstützender Analyseschritte gefunden. Näher betrachtete Verfahren sind in Abbildung 1 dargestellt (neue Adjustierungsverfahren in rot, neue unterstützende Verfahren in kursiv).

Hinsichtlich der Adjustierungsverfahren wurden die drei Gewichtungsmethoden CBPS weighting [3], Entropy Balancing [4] und Kernel Balancing [5] genauer untersucht. Diese weisen den Beobachtungen Gewichte zu und gleichen so die Verteilungen von Confoundern zwischen den beiden Gruppen an. Der Vorteil gegenüber Matchingverfahren liegt in einer höheren statistischen Power, da alle Beobachtungen genutzt werden. In Simulationsstudien wurde eine Überlegenheit von CBPS und Entropy Balancing im Vergleich zu herkömmlichen Propensity Score Weighting unter verschiedenen Szenarien festgestellt [6,4]. Für Kernel Balancing liegt noch keine ausführliche vergleichende Simulationsstudie vor.

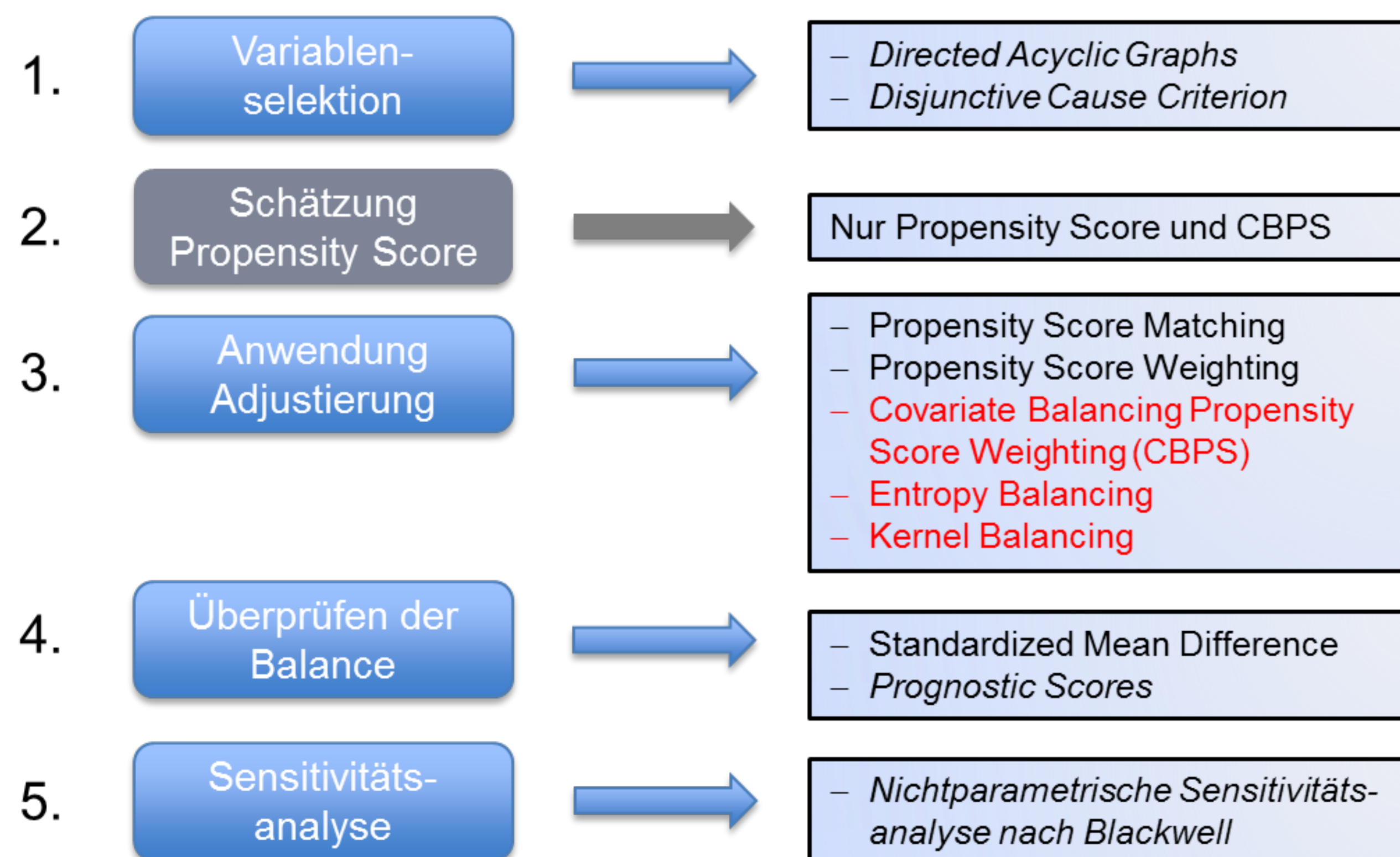


Abbildung 1. Ablauf einer Analyse zur Schätzung des Interventionseffektes mit Korrektur von Selektionsbias.

Der rechte Teil zeigt eine Auswahl zur Verfügung stehender Methoden für jeden Analyseschritt. Neuere Gewichtungsmethoden für die eigentliche Adjustierung sind rot gesetzt, neuere unterstützende Verfahren kursiv.

Die gefundenen unterstützenden Analyseverfahren sind auf alle untersuchten Methoden gleichermaßen anwendbar. Zur Confounderauswahl stellen Directed Acyclic Graphs bzw. das Disjunctive Cause Criterion neue Ansätze dar [7]. Zur Überprüfung der erzielten Balance wurden sogenannte Prognostic Scores vorgeschlagen [8]. Eine von Blackwell entwickelte nichtparametrische Sensitivitätsanalyse ist breit anwendbar sowie leicht zugänglich [9].

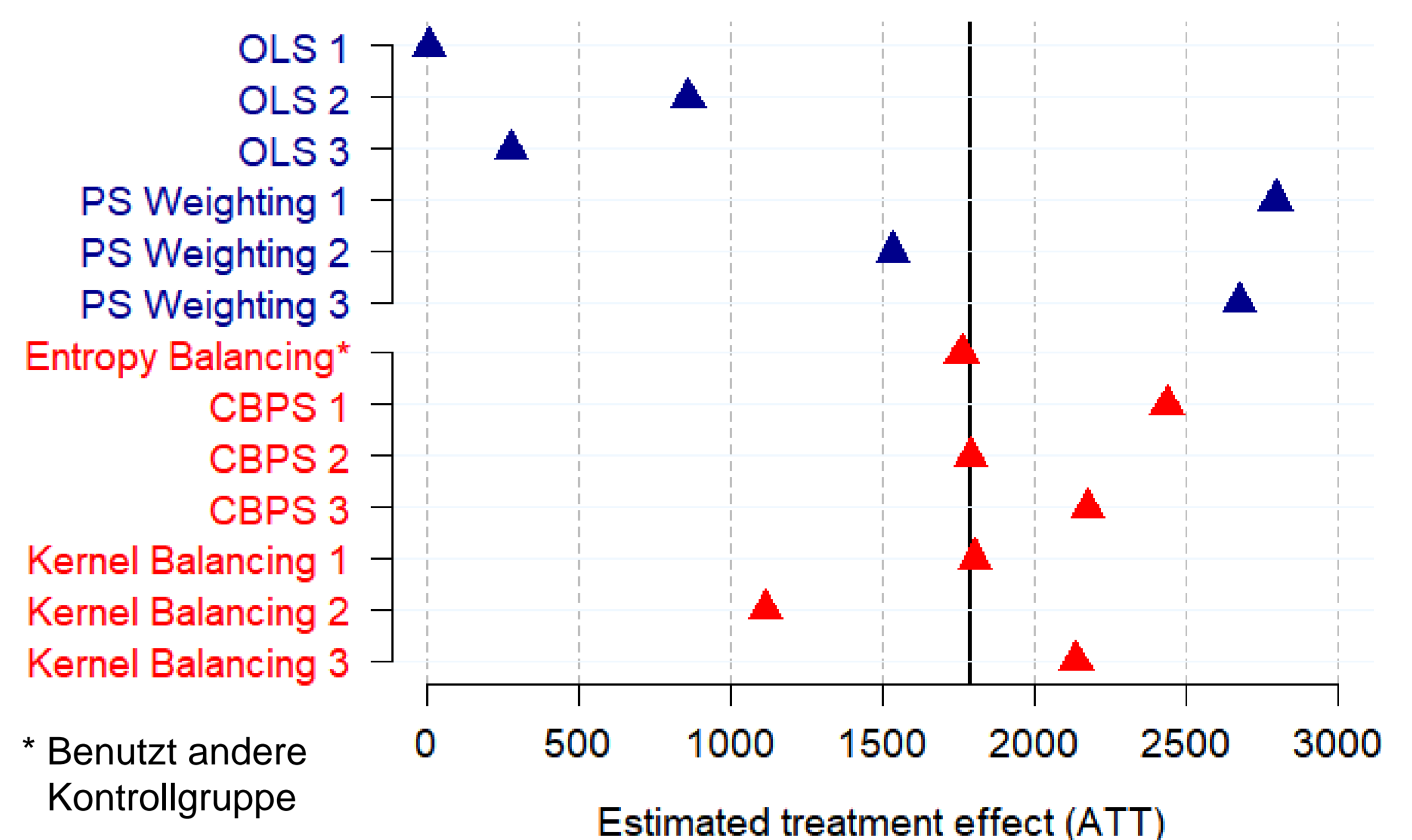
Kontakt

Michael Swora
+49 (0)221 478 97151
michael.swora@uk-koeln.de



Deutscher Kongress
für Versorgungsforschung

16. Kongress I 4. - 6.10.2017 | Urania - Berlin



* Benutzt andere Kontrollgruppe

Abbildung 2. Reanalyse von Dehejia and Wahba: Schätzung des Effektes eines Jobtrainings-Programmes auf das Einkommen (in US-Dollar) mit verschiedenen Methoden und Spezifikationen.

Einen empirischen Vergleich der Adjustierungsverfahren bietet Abbildung 2. Hierbei wurden Reanalysen eines Datensatz von Dehejia and Wahba [10], der ursprünglich auf LaLonde [11] zurückgeht, teilweise repliziert und teilweise selbst berechnet. Bei den Daten handelt es sich um die Untersuchung der Wirksamkeit eines Jobtrainings-Programmes in einem randomisierten Experimental-Kontrollgruppen-Design. LaLonde ersetzte die experimentelle Kontrollgruppe durch eine externe Kontrollgruppe, die sich in wichtigen Confoundern von der Experimentalgruppe signifikant unterscheidet. Dies ermöglicht es, die Performance von Methoden zur Korrektur von Selektionsbias am experimentellen Benchmark (1794 Dollar) zu messen. Die in Abbildung 2 dargestellten Berechnungen nutzen die PSID1 Kontrollgruppe.

Diskussion

In Beobachtungsstudien im Experimental-Kontrollgruppen-Design sollten adäquate Verfahren zur Kontrolle von Selektionsbias verwendet werden. Hierzu existieren eine Reihe neuer Methoden. Diese bedürfen jedoch noch weiterer Untersuchungen, um ihre Performance besser beurteilen zu können. Dies trifft vor allem auf das Verfahren Kernel Balancing zu. Um weitere Erkenntnisse zu diesem Thema zu erlangen, werden im Rahmen eines Doktorandenprojekts diese neuen Methoden durch eine Monte Carlo Studie auf ihre Performance überprüft. Eine Limitation des vorliegenden Reviews ist die Selektivität der Literaturrecherche, sodass womöglich noch weitere relevante neue Methoden zur Korrektur von Selektionsbias existieren könnten.

Literatur

1. Kabisch M., Ruckes C., Seibert-Grafe M., Blettner M. Randomized Controlled Trials: Part 17 of a Series on Evaluation of Scientific Publications. *Deutsches Ärzteblatt International*. 2011; 108(39): 663–668.
2. Li M. Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. *Organizational Research Methods*. 2012; 16(2): 188-226.
3. Imai K, Ratkovic M. Covariate Balancing Propensity Score. *J. R. Statist. Soc. B*. 2014; 76(1): 243–263.
4. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. 2012; 20(1): 25-46.
5. Hazlett C. *Inference in Tough Places: Essays on Modeling and Matching with Applications to Civil Conflict* [Dissertation]. Cambridge: Massachusetts Institute of Technology; 2014.
6. Wyss R., Ellis AR., Brookhart MA., Girman CJ., Jonsson Funk M., LoCasale R., Stürmer T. The Role of Prediction Modeling in Propensity Score Estimation: An Evaluation of Logistic Regression, bCART, and the Covariate-Balancing Propensity Score. *Am J Epidemiol*. 2014; 180(6): 645-655. doi: 10.1093/aje/kwu181.
7. VanderWeele TJ., Shpitser I. A new Criterion for Confounder Selection. *Biometrics*. 2011; 67(4):1406-1413. doi: 10.1111/j.1541-0420.2011.01619.x.
8. Start EA., Lee BK., Leacy FP. Prognostic Score-based Balance Measures for Propensity Score Methods in Comparative Effectiveness Research. *J Clin Epidemiol*. 2013; 66(80): 84-90. doi: 10.1016/j.jclinepi.2013.01.013.
9. Blackwell M. A Selection Bias Approach to Sensitivity Analysis for Causal Effects. *Political Analysis*. 2014; 22(2): 169-182.
10. Dehejia RH., Wahba S. Propensity Score Matching Methods for nonexperimental Causal Studies. *The Review of Economics and Statistics*. 2002; 84(1): 151-161.
11. LaLonde R. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review*. 76(4): 604-620.