
56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS) und
6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie e.V. (DGEpi)
Mainz, 26. bis 29. September 2011

Wewering P, Linder R

WINEG | Wissenschaftliches Institut der TK für Nutzen und Effizienz im Gesundheitswesen, Hamburg

Benchmarking von Propensity Score basierten Matchingalgorithmen

Abstract

Einleitung

In der Versorgungsforschung ist es häufig nicht möglich, den Erfolg einer Intervention mit Hilfe von randomisierten, kontrollierten Studien (RCT) zu messen. Ersatzweise wird häufig mit einer Stratifizierung gearbeitet, meist nach Alter und Geschlecht, um Selektionseffekten zu begegnen. Möchte man zusätzlich Komorbiditäten und weitere wichtige Einflussgrößen mit berücksichtigen, ergibt sich sehr schnell ein Dimensionsproblem und dieser Ansatz ist nicht mehr praktikabel. Eine Alternative ist die Methodik des Propensity Score Matching [1]. Angewendet auf GKV-Routinedaten erlaubt das Matching die retrospektive Generierung einer geeigneten Kontrollgruppe.

In der Literatur sind unterschiedliche Matchingalgorithmen beschrieben, bislang jedoch nicht systematisch verglichen worden.

Methoden

Um zwei Gruppen hinsichtlich beobachteter Kovariablen anzugleichen, wird für die Gesamtheit aller Personen eine logistische Regression berechnet mit den Kovariablen als unabhängige Variablen sowie der tatsächlichen Gruppenzuordnung (Teilnehmergruppe • Nichtteilnehmergruppe) als abhängige Variable. Mit den sich hieraus ergebenden Beta-Koeffizienten wird anschließend der individuelle Propensity Score bestimmt.

Ausgehend vom Propensity Score kann anschließend ein Matching durchgeführt werden, mit dessen Hilfe versucht wird, die beiden Gruppen hinsichtlich der beobachteten Kovariablen anzugleichen. Hierbei kommen verschiedene Verfahren zum Einsatz, zum Beispiel k-Nearest Neighbor Matching oder Greedy Matching. Viele dieser Verfahren lassen dem Anwender anhand von Einstellungen die Möglichkeit, auf den Erfolg des Matching Einfluss zu nehmen. Es wird gezeigt, inwieweit sich hierdurch die Resultate des Matching verändern und ob beziehungsweise wie prinzipielle Aussagen über die Güte eines Matching getroffen werden können.

Ergebnisse

Auf Basis von drei Datensätzen (DMP-Datensatz der Techniker Krankenkasse [2] sowie zwei Datensätzen zum National Supported Work-Experiment [3]) wird ein Vergleich verschiedener Matchingmethoden hinsichtlich deren Güte vorgenommen. Hierbei kommen insgesamt acht verschiedene Matchingverfahren zum Einsatz, die sich in vier grundlegende Typen einteilen lassen (k-Nearest Neighbor Matching, Caliper Matching, Interval Matching, Greedy Matching) und sich hinsicht-

lich der eingestellten Parameter unterscheiden. Vergleicht man diese verschiedenen Verfahren anhand von in der Literatur vorgestellten Gütemaßen (Sample Percent Reduction in Bias, Standardized Bias, Vergleich der p-Werte [4]), zeigt sich, dass Caliper Matching und Greedy Matching in Bezug auf die unterschiedlichen Datensätze und Gütemaße robust gute Ergebnisse liefern. In speziellen Konstellationen sind andere Verfahren im Vorteil, beispielsweise das Interval Matching beim DMP-Datensatz hinsichtlich des p-Wert-Vergleichs.

Diskussion

In Ermangelung von RCTs ermöglicht das Propensity Score Matching die Generierung einer Kontrollgruppe unter Einbeziehung einer großen Anzahl an Kovariablen zur Minimierung von Selektionsfehlern. Beim Vergleich der Matchingalgorithmen lassen sich Caliper Matching und Greedy Matching empfehlen, eine Erweiterung des Benchmarking auf zusätzliche Datensätze ist geplant.

Kontakt

Philipp Wewering | philipp.wewering@wineg.de
WINEG | Wissenschaftliches Institut
der TK für Nutzen und Effizienz
im Gesundheitswesen

Bramfelder Straße 140
22305 Hamburg

Literatur

- [1] Rosenbaum P, Rubin D: The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55
- [2] Linder R, Ahrens S, Köppel D, Heilmann T, Verheyen F: The benefit and efficiency of the disease management program for type 2 diabetes. *Dtsch Arztebl Int* 2011; 108(10): 155–62.
- [3] Dehejia R, Wahaba S: Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 2002; 84(1): 151-161.
- [4] Baser O: To much ado about propensity score models? Comparing methods of propensity score matching. *Value In Health* 2006;6:377-985.