

Einfluss der Wahl der Datenquelle (Primär- oder Routinedaten) auf das Ergebnis eines Qualitätsvergleichs zwischen Kliniken

B. Maier¹, E. Vettorazzi², S. Behrens³, L. Bruch⁴, R. Busse⁵, D. Schmidt⁶, H. Schühlen⁷, R. Thieme⁸, H. Theres⁹, K. Wegscheider¹⁰

¹Berliner Herzinfarktregister, TU Berlin, Berlin; ²Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg; ³Kardiologie, Vivantes Humboldt Klinikum, Berlin; ⁴Kardiologie, Unfallkrankenhaus Berlin, Berlin; ⁵Management im Gesundheitswesen, Technische Universität Berlin, Berlin; ⁶AOK Berlin Brandenburg, Berlin; ⁷Vivantes Auguste-Viktoria-Klinikum, Berlin; ⁸Kardiologie, Jüdisches Krankenhaus, Berlin; ⁹Kardiologie, Charite, Berlin; ¹⁰Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg

Hintergrund: Abrechnungs-Routinedaten werden immer häufiger zur Qualitätssicherung (QS) eingesetzt, ohne dass bisher ihre *Validität* für Zwecke der QS überprüft worden wäre. Dies geschieht im QS-AMI Projekt; hier werden für Patienten mit akutem Myokardinfarkt (AMI) erhobene Qualitätsdaten aus einem regionalen Herzinfarktregister mit Routinedaten verglichen.

Fragestellung: Im Rahmen der QS-AMI Studie wurde u.a. untersucht, inwieweit Prognosemodelle, die auf Primär- oder auf Routinedaten derselben Patienten beruhen, zu vergleichbaren Ergebnissen kommen und welchen Einfluss die Wahl der Datenquelle auf ein Ranking nach der Krankenhausmortalität hat.

Methode: Die Datenbasis bestand aus aufeinander abbildbaren Krankenkassen- und Register-Daten von 2161 Patienten, die 2009-2011 stationär behandelt wurden. Eine eindeutige Verknüpfung der Datensätze gelang über das Alter, Geschlecht und Klinikaufnahmezeitpunkt. Für den Klinikvergleich wurde als Outcomeparameter die Krankenhaussterblichkeit betrachtet. Zunächst wurde für jede der beiden Datenquellen ein Mehrebenenmodell mit der gleichen Auswahl von Patientencharakteristika als Kovariaten berechnet. Nach Imputation fehlender Werte erhielten wir Empirical Bayes Schätzer adjustiert für die Unterschiede zwischen den Patienten. Zur Bewertung der prognostischen Güte verwendeten wir die Fläche unter der ROC-Kurve. In einem letzten Schritt wurden die resultierenden Rankings auf Basis der AOK- und Register-Daten verglichen.

Ergebnisse: Es zeigten sich bei verschiedenen Kovariaten deutliche Unterschiede in Effektstärken und Prognose-Relevanz im Krankenkassen- im Vergleich zum Register-Datensatz. Obwohl sich beide Adjustierungsmodelle in einzelnen Effekten erheblich unterschieden, war die prognostische Güte beider Modelle vergleichbar mit Flächen unter der ROC-Kurve von 0,846, 95% KI 0.817, 0.876 (Krankenkasse) und 0,874, 95% KI 0.848, 0.899 (Register) ($p=0.07$). Beide Krankenhausvergleiche zeigten in der Krankenhaussterblichkeit Unterschiede zwischen den Kliniken auf, allerdings innerhalb der Signifikanzgrenzen. Die Rangreihenfolgen der Kliniken waren dabei stark von der Wahl der Datenquelle abhängig. (Spearman-Korrelationskoeffizient r_s : 0.62). Nur die Extreme wurden vergleichbar abgebildet: Bei Verwendung des Shrinkage-Schätzers wurden die 3 besten und schlechtesten Kliniken ähnlich gerankt.

Diskussion und praktische Implikationen: Bei gleichem Patientenkollektiv (gepoolt über 3 Jahre), gleicher Methodik und vergleichbarer Prognosegüte ergeben sich auf der Basis von Routinedaten grundsätzlich andere Rankings als auf der Basis von Primärdaten. Bei annähernd gleicher Versorgungsqualität von Kliniken sind Mortalitäts-Rankings somit nicht nur wie bereits bekannt starken, häufig unterschätzten Zufallseffekten ausgesetzt, sondern auch wesentlich von der Wahl der Datenquelle/der Datenqualität abhängig. Lediglich in den Extremen (also bei echten Effekten?) stimmen die Rankings überein. Leistungsvergleiche per Ranking sollten deshalb nur verwendet werden, um potentielle schwarze und weiße Schafe zu identifizieren. Das ändert sich auch nicht mit sehr großen Datenbanken (big data), weil die Patientenanzahl pro Krankenhaus nicht mit wächst.