

# Gut gelinkt ist halb gewonnen: Es könnte alles so einfach sein, ist es aber nicht



Prof. Dr. Falk Hoffmann, MPH



Dr. Sascha Abbas

Stellen Sie sich vor, Sie möchten die Frage klären, ob nichtsteroidale Antirheumatika (NSAR) mit einem verminderten Risiko für Prostatakrebs einher gehen. Ihre zeitlichen und finanziellen Mittel reichen allerdings nicht für eine Primärdatenerhebung. Sie nutzen daher eine große Gesundheitsdatenbank, um die Population unter Risiko zu selektieren sowie um Informationen zur Arzneimittelexposition zu gewinnen. Die Fälle generieren Sie über ein Krebsregister, weitere klinische Informationen wie den Gleason-Score, den PSA-Wert oder das Tumorstadium zum Zeitpunkt der Diagnose erhalten Sie durch Einsicht in die Krankenakten. Ihre Studienperiode umfasst 16 Jahre, die meisten relevanten Daten stünden Ihnen jedoch auch für einen deutlich längeren Zeitraum zur Verfügung. Alles nur geträumt? Nein, denn eine solche Studie gibt es tatsächlich [1]. Sie stammt allerdings aus der kanadischen Provinz Saskatchewan, die nicht nur geografisch Meilen weit von Deutschland entfernt liegt. Wieso wäre eine solche Studie hierzulande nicht möglich? Einerseits haben wir kein mit Kanada vergleichbares öffentliches Gesundheitssystem, jedoch mit der gesetzlichen Krankenversicherung (GKV) zumindest einen Rahmen, in dem etwa 85% der Bevölkerung versichert sind und deren Gesundheitsdaten strukturiert vorliegen. Zudem sind in Saskatchewan teils elektronische Daten bis in die 1960er Jahre zurück vorhanden. In Deutschland könnten viele gesetzliche Krankenkassen mittlerweile auch für die meisten Leistungsbereiche über Daten verfügen, die mind. 10 Jahre zurückreichen, wenn es denn die Löschfristen zulassen würden. Weiterhin existiert in Saskatchewan eine lebenslange Gesundheitsnummer (Health Services Number), welche eine Verknüpfung zwischen verschiedenen Daten einfach ermöglicht. Seit der Einführung der elektronischen Gesundheitskarte wird hierzulande zumindest in der GKV die Krankenversicherungsnummer einmalig vergeben und bleibt lebenslang, also auch beim Wechsel der Kasse, gleich. Jedoch findet sich diese personenidentifizierende Nummer nicht bei anderen Sozialversicherungsträgern oder in (Krebs-)Registern. Und last but not least ist in Deutschland ein systematischer und einfacher Zugang zu Krankenakten oder sonstigen Befunden für Forscher, die diese kombiniert mit Krankenkassendaten auswerten wollen, bislang nicht möglich. Gerade dieser Punkt wäre in Ergänzung von GKV-Routinedaten wichtig, da in diesen zwar Diagnosen und Leistungen vorliegen, allerdings keine Informationen über Behandlungsergebnisse oder Befunde.

Nichtsdestotrotz begann auch hierzulande in den letzten Jahren – einhergehend mit der zunehmenden Nutzung von GKV-Routinedaten – das zarte Pflänzchen des personenbezogenen Verknüpfens (Linkage) mit anderen Datenkörpern zu wachsen. Eine Reihe von bisher durchgeführten Projekten zum Thema Linkage von Krankenkassendaten ist in diesem Schwerpunktheft zusammengefasst. Der Beitrag von Ohlmeier et al. beschreibt die Verknüpfung ausgehend von Kassendaten mit dem Krankenhausinformationssystem (KIS) einer Klinik am Beispiel der Herzinsuffizienz. Die Autoren verdeutlichen, dass die theoretisch möglichen Zusatzinformationen aus dem KIS in der Realität oft deutlich begrenzter ausfallen als erwartet. Maier et al. diskutieren das Linkage von Daten des Berliner Herzinfarktregisters mit denen einer Krankenkasse und beschreiben, wie diese beiden pseudonymisierten Datensätze miteinander verknüpft werden können. Schmidt et al. weisen in ihrem Beitrag auf die Schwierigkeiten der Zusammenführung von Primärdaten einer bevölkerungsbezogenen Studie und Sekundärdaten hin, die aufgrund der unterschiedlichen Datenstruktur und -reichweite resultieren. Am Beispiel des Schlaganfalls thematisieren die Autoren außerdem die Validierung von Diagnosen zur Endpunktdefinition bei der Nutzung mehrerer Datenquellen. Der Beitrag von Bitzer et al. widmet sich der Befragung von Patientengruppen, die auf Basis ihrer Kassendaten selektiert wurden, und klärt die Frage, was diese Kombination von Informationen für die Qualitätsberichterstattung leisten kann. Hierbei wird vor allem deutlich, wie sich beide Ansätze ergänzen und für welche Fragestellungen dies besonders (oder weniger) gut geeignet ist. Die beiden Arbeiten von March et al. und Stallmann et al. thematisieren den umgekehrten Weg, nämlich ausgehend von Primärerhebungen, die Kassendaten der Teilnehmer zuzuspielen. March et al. beschreiben ihre Erfahrungen, wie problematisch bei solchen Studien die Vielzahl von über 100 verschiedenen gesetzlichen Kassen, mögliche Fusionen und Kassenwechsel sind. Die Autoren diskutieren auch alternative Zugänge zu Daten aller Kassen mit entsprechenden Vor- und Nachteilen und stellen heraus, dass ein Linkage (bisher) ohne Kooperation mit einzelnen Kassen nicht möglich ist. Dieses Problem wird auch im Beitrag von Stallmann et al. deutlich. Am Beispiel der Nationalen Kohorte, dem bisher wohl ambitioniertesten Projekt hiesiger epidemiologischer Forschung, wird der (teils steinige) Weg beschrieben, den Forscher zu gehen haben, um zusätzlich von möglichst vielen Teilnehmern Routinedaten nutzen zu können.

## Bibliografie

DOI <http://dx.doi.org/10.1055/s-0034-1398522>  
 Gesundheitswesen 2015;  
 77: 72–73  
 © Georg Thieme Verlag KG  
 Stuttgart · New York  
 ISSN 0941-3790

## Korrespondenzadresse

**Prof. Dr. Falk Hoffmann, MPH**  
 Carl von Ossietzky Universität  
 Oldenburg  
 Fakultät für Medizin und  
 Gesundheitswissenschaft  
 Department für Versorgungsforschung  
 26111 Oldenburg  
[falk.hoffmann@uni-oldenburg.de](mailto:falk.hoffmann@uni-oldenburg.de)

Was wird in der Zusammenschau aus diesen Beiträgen deutlich? In jedem Fall ist eine Verknüpfung von Krankenkassendaten mit weiteren Datenkörpern möglich. Dadurch kann oftmals ein erhebliches Mehr an Informationsdichte resultieren und die Vorteile der verschiedenen Datenquellen können miteinander kombiniert werden. Manchmal stellen sich allerdings im Anschluss an solche Studien auch mehr Fragen, insbesondere dann, wenn es zu klären gilt, welchen Informationen aus welcher Datenquelle man am ehesten „trauen“ soll. Gerade für Validierungsstudien muss man deshalb klären, welche Angaben man tatsächlich als Goldstandard ansieht. Fakt ist jedoch: Validierung von Sekundärdaten kann nicht ohne Linkage funktionieren! Es zeigt sich weiterhin, dass ein Linkage bislang jedenfalls mit einem sehr hohen zeitlichen und personellen Aufwand verbunden ist, der vor allem aus den gesetzlichen und datenschutzrechtlichen Rahmenbedingungen bzw. deren Auslegung resultiert.

Von vielen Autoren in diesem Heft wird auch das für die Zukunft der Routinedatenforschung in Deutschland wichtige Thema der sektorübergreifenden Versorgungsdaten nach den §§ 303 a–f des SGB V (die sog. DaTraV-Daten) diskutiert. Hierfür wurden mittlerweile die bereits von den Kassen an das Bundesversicherungsamt übermittelten Daten des morbiditätsorientierten Risikostrukturausgleichs (Morbi-RSA) aufbereitet. Damit sind erstmalig und kontinuierlich Daten (nahezu) aller gesetzlichen Krankenkassen für die Forschung zugänglich. Wer also heutzutage Routinedaten analysieren möchte, benötigt im Prinzip keine Kooperationen mit einzelnen Kassen mehr. Allerdings sind die Möglichkeiten auch deutlich begrenzter. So sind neben der zeitlichen Verzögerung bei der Zugänglichkeit (zwischen Erhebung und Bereitstellung liegen immerhin 4 Jahre) viele relevante Informationen – zumindest derzeit – nicht in den DaTraV-Daten

enthalten (u. a. Regionalvariablen, Leistungsziffern im ambulanten und stationären Bereich, Daten zur Pflegeversicherung oder Rehabilitation) [2]. Und auch ein personenbezogenes Linkage mit anderen Datenkörpern ist nicht möglich. Trotzdem sind diese Daten nicht per se als unbrauchbar anzusehen und werden hoffentlich auch auf Seiten der Politik als „work in progress“ verstanden. Erst kürzlich wurde per Gesetz der Weg für die Aufnahme eines Regionalkennzeichens geebnet. Viel wird für die zukünftige Entwicklung und die Nutzbarkeit dieser DaTraV-Daten von dem für Ende 2015 geplanten Evaluationsbericht abhängen. An der inhaltlichen Ausgestaltung dieses Berichtes sollen auch Forscher und andere Nutzungsberechtigte beteiligt werden [3]. Die Beiträge dieses Schwerpunktheftes verdeutlichen, dass man die DaTraV-Daten kontinuierlich weiterentwickeln muss und perspektivisch auch über die Möglichkeit nachdenken sollte, diese personenbezogen mit weiteren Datenkörpern verknüpfen zu können.

Wir wünschen allen Lesern eine spannende Lektüre!

**Interessenkonflikt:** Beide Autoren haben in den letzten Jahren Projekte mit verschiedenen Krankenkassen durchgeführt.

#### Literatur

- 1 Mahmud SM, Franco EL, Turner D et al. Use of non-steroidal anti-inflammatory drugs and prostate cancer risk: a population-based nested case-control study. *PLoS One* 2011; 6: e16412
- 2 Mansky T, Robra B, Schubert I. Qualitätssicherung: Vorhandene Daten besser nutzen. *Dtsch Arztebl* 2012; 109: A1082–A1085
- 3 N.N. „Ein Anfang ist gemacht“. Interview mit DIMDI-Direktor Dr. Dietrich Kaiser und Dr. Jochen Dreß, Leiter des Versorgungsdaten-Infosystems. *Monitor-Versorgungsforschung* 2014; 4: 20–22