

## Nutzung medizinischer Daten aus der Versorgung für die Forschung

Stausberg J<sup>1</sup>, Drösler S<sup>2</sup>

<sup>1</sup>Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Deutschland

<sup>2</sup>Fachbereich Wirtschaftsingenieurwesen und Gesundheitswesen, Hochschule Niederrhein, Krefeld, Deutschland  
stausberg@uni-essen.de

**Einleitung** Die klassische, papierbasierte Krankenakte stellt weiterhin die zentrale Ablage für Aufzeichnungen über die Versorgung von Patienten dar. Ihre Nutzung für die Forschung wird aber nicht nur durch die notwendige Überführung in eine für die automatisierte Weiterverarbeitung angemessener Form erschwert, sondern auch durch Mängel hinsichtlich Vollständigkeit und Korrektheit [1]. Die Elektronische Patientenakte kündigt sich zwar an, ist aber laut einer aktuellen Umfrage erst in 8,6 % der Krankenhäuser verfügbar [2]. Bereits früh wurde eine Ergänzung der primären Aufzeichnungen zur Versorgung durch eine rechnergestützte Basisdokumentation vorgeschlagen, die für alle Patienten wenige Merkmale umfasst, diese aber vollständig und qualitätsgesichert [3]. Diese Merkmale sollten gezielt für eine mehrfache Verwendung ausgewählt werden, z. B. zur Qualitätssicherung, als Leistungsnachweis, zum Retrieval von Patientengruppen und zur Hypothesengenerierung. Trotz erheblicher Anstrengungen, besonders deutlich sichtbar in der Onkologie [4], hat sich dieser Ansatz einer rechnergestützte Basisdokumentation nicht durchsetzen können. Über die Versorgung verfügbare elektronische Daten sind daher überwiegend auf gesetzliche Pflichten zur Übermittlung an Dritte zurückzuführen. Auf der einen Seite entstehen hierbei große Datenbestände, wie sie in den USA bereits intensiv diskutiert wurden [5], auf der anderen Seite sind die Daten damit auch automatisch bei den Leistungserbringern für weitere Verwendungszwecke vorhanden [6]. Die damit in Deutschland verfügbare rechnergestützte Basisdokumentation basiert somit fast ausschließlich auf rechtlichen Regelungen, und kaum mehr auf einer eigenen Motivation der damit befassten Personen. Ob und in welcher Form diese Daten für die Forschung nutzbar sind, bedarf daher einer besonderen Abwägung.

**Verfügbarkeit von Routinedaten** Relevante Datenbestände entstehen in verschiedensten Versorgungsbereichen durch die Verpflichtung, im Rahmen der Abrechnung medizinische Angaben an die Versicherungen zu übermitteln. Die entsprechenden Vorgaben finden sich im Sozialgesetzbuch (SGB), für Krankenhäuser in § 301 SGB V, für niedergelassene Ärzte in § 295 SGB V, für Leistungserbringer im Bereich Heil- und Hilfsmittel in § 302 SGB V und für ambulante Pflegedienste in § 105 SGB XI. Eine Besonderheit in der stationären Versorgung liegt in einem weiteren, größtenteils identischen Datenbestand, der unabhängig vom Abrechnungsvorgang für die Weiterentwicklung der Diagnosis Related Groups (DRGs) beim Institut für das Entgeltssystem im Krankenhaus (InEK) geführt wird. Für das Jahr 2004 wurden dort 17.730.030 voll- oder teilstationäre Behandlungsfälle mit medizinischen Daten und 3.738.107 mit Kostendaten erfasst [7]. Ebenfalls aus den Krankenhäusern liegt ein umfangreicher Datenbestand bei der Bundesgeschäftsstelle Qualitätssicherung (BQS) vor; aus dem Jahre 2004 waren es 2.388.883 Datensätze [8]. Im niedergelassenen, vertragsärztlichen Bereich liegt die Zahl der Datensätze noch um ein vielfaches höher bei 565.365.400 Abrechnungsfällen (2001) und rund 761 Millionen Arzneimittelverordnungen (2002).

**Qualität von Routinedaten** Aus Sicht der Forschung ist die Vollständigkeit dieser Daten nicht immer gegeben. Die Daten des InEK umfassen im medizinischen Teil nur rund 82 % aller Krankenhäuser, letztendlich zur Kostenkalkulation der DRGs wurden 21 % der rund 17,7 Millionen Behandlungsfälle herangezogen. Hingegen verfügt die externe Qualitätssicherung bei der BQS über 98 % aller Datensätze. Sowohl bei den vertragsärztlichen Leistungen als auch bei der Arzneimittelversorgung fehlt der Bereich der privaten Krankenversicherung. Gleiches gilt für die Medikation stationärer Behandlungsfälle. Alle vorgenannten Datenbestände enthalten personenbezogene Daten. Die betroffene Person ist immer zu identifizieren, wenn auch zum Teil nur über den Melder. Eine Zusammenführung ist derzeit ausschließlich den Versicherungen möglich. Während die Vollständigkeit der großen Datenbestände recht gut abgeschätzt werden kann, ist dies für die Richtigkeit und Nutzbarkeit schon schwieriger. Ein Gutachten zur Verwendung dieser Datenbestände für den Risikostrukturausgleich stellt Angaben zu Diagnosen aus dem stationären Bereich und Daten zur Arzneimittelversorgung gute Noten aus, während Angaben zu Diagnosen aus dem vertragsärztlichen Bereich noch nicht nutzbar waren [9]. In der Kostenkalkulation für die DRGs wird mit 0,7 % eine geringe Fehlerrate berichtet. Andererseits beruhen wichtige Festlegungen auf sogenannten Normalliegern, die nur rund die Hälfte der Datensätze ausmachen [7]. Die Daten der externen Qualitätssicherung können auf Grund des weit ausgebauten Prüfverfahrens zumindest als plausibel gelten [8]. Über die Qualität der Heil- und Hilfsmittelabrechnungsdaten und Angaben der ambulanten Pflegedienste liegen den Autoren keine Informationen vor.

**Diskussion** Umfangreiche Routinedaten liegen in Deutschland bereits an vielen Stellen verteilt vor. Es kann daher nicht von einem Mangel an Informationen über die Gesundheitsversorgung gesprochen werden. Derzeit sind allerdings nur die Versicherungen in der Lage, umfassende Analysen zu einzelnen Versicherten oder Versichertengruppen vorzunehmen. Diese Daten sind daher für die Versorgungsforschung im Sinne einer Sekundärdatenanalyse weiterhin hochinteressant [10]. Hierbei kann von hoher Vollständigkeit und mäßiger Vollständigkeit, aber schlecht fassbarer Richtigkeit ausgegangen werden. Für die klinische Forschung liegt in denen zur Übermittlung an Dritte bei den Leistungserbringern angelegten Datenbeständen ein großer, bislang noch nicht gehobener Schatz zur Rekrutierung geeigneter Studienteilnehmer, zur (automatischen) Plausibilisierung von Erhebungen und ggf. auch zur Teildatenübernahme vor. Dies beschränkt sich allerdings momentan weitgehend auf die Verwendung von Daten eigener Patienten, da die individuelle Zusammenführung von Informationen verschiedener Leistungserbringer durch das Fehlen eindeutiger Identifikationsmerkmale sowie die datenschutzrechtliche Situation erschwert wird. Für die epidemiologische Forschung ergeben sich neue Optionen aus dem Titel Datentransparenz nach § 303a-f SGB V. Für den hier eröffneten nationalen Datenpool ist explizit die Nutzung für Längsschnittanalysen vorgesehen, wenn auch mit Blick auf Fragen der angemessenen Versorgung. Neue Chancen ergeben sich auch durch die Einführung der elektronischen Gesundheitskarte als Zugang zu medizinischen Daten in der Verantwortung des Versicherten. So wird zukünftig die automatische Nutzung von medizinischen Daten für die Forschung möglicherweise ohne Einbindung der Leistungserbringer möglich sein. Die Folgen sind derzeit nur schwer abzuschätzen. Allerdings sind Forschungsaspekte bei Architekturvorschlägen und -entscheidungen zur Telematikinfrastruktur noch nicht berücksichtigt. Der Datensatz auf der Gesundheitskarte zur Notfallversorgung wird aber bereits einen Umfang haben, der den jetzt verfügbaren großen Datenbeständen in seinem Merkmalsumfang nicht nachsteht [11]. Auf Grund der bislang nur rudimentären Auseinandersetzung mit der Nutzung medizinischer Daten aus der Versorgung für die Forschung besteht allerdings noch großer Nachholbedarf, das hierfür angemessene Methodeninventar aufzubauen.

## Literatur

- [1] Püschmann H, Haferkamp G, Scheppkoat K-D, Vinz H, Wegner M. Vollständigkeit und Qualität der ärztlichen Dokumentation in Krankenakten. Dtsch Arztebl 2006; 103: A 121-6.
- [2] Fachhochschule Osnabrück. Pressemitteilung IT-Report GW. 20.03.2006. <http://www.wiso.fh-osnabrueck.de/14095.html> [Zugriff am 27.3.2006].
- [3] Thurmayr R. Elektronische Auswertungen von Krankengeschichten mit Datenverarbeitungsanlagen. Der Chirurg 1969, 40: 251-8.
- [4] Dudeck J, Wagner G, Grundmann E, Hermanek P. Basisdokumentation für Tumorkranke. München: Zuckschwerdt, 1999.
- [5] Measuring quality, outcomes, and cost of care using large databases. The Sixth Regenstrief Conference. Annals of Internal Medicine 1997; 127: number 8 (part 2).
- [6] Drösler S, Lange P, Stausberg J. Aus Routinedaten kann man mehr machen. Administrative Daten als Basis zur Beurteilung der Ergebnisqualität. Krankenhaus Umschau 2005; 74: 43-7.
- [7] InEK. Abschlussbericht. Weiterentwicklung des G-DRG-Systems für das Jahr 2006. Klassifikation, Katalog und Bewertungsrelationen. Teil I: Projektbericht. Siegburg, den 20. Dezember 2005.
- [8] Mohr VD, Bauer J, Döbler K, Eckert O, Fischer B, Woldenga C, Hrsg. Qualität sichtbar machen. BQS-Qualitätsreport 2004. Düsseldorf, 2005
- [9] Rechke P, Sehlen S, Schiffhorst G, Schröder WF, Lauterbach KW, Wasem J. Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Endbericht. November 2004.
- [10] Swart E, Ihle P, Hrsg. Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Huber, 2005.
- [11] Bundesministerium für Gesundheit und Soziale Sicherung. Medizinische Datensätze der elektronischen Gesundheitskarte – Notfallversorgung und Prüfung der Arzneimitteltherapiesicherheit. Version 2.0, 10. Juni 2005.

