



Datenquellen

Bei dem ersten Gedanken an eine gesundheitsökonomische Untersuchung ertönt reflexartig: „In Deutschland gibt es keine Daten!“ Und mehr noch: „Der übertriebene Datenschutz macht jegliche empirische Untersuchung unmöglich.“ Solche Klagen kennen die Kolleginnen und Kollegen in den klinischen Forschungsabteilungen nicht. Sie sind es gewohnt, die benötigten Daten für die jeweilige Fragestellung auf der Basis eines genauen Protokolls selbst zu erheben. In der Gesundheitsökonomie ist das anders, weil sich viele Fragestellungen nicht auf die genau definierten Bedingungen einer Studie beziehen, sondern auf die reale Welt, in der Menschen erkranken, behandelt werden und wo Ausgaben entstehen.

Prof. Dr. med. Bertram Häussler, Dr. rer. pol. Thomas Ecker

Daten! Daten? Daten!

In der realen Welt ist man überwiegend darauf angewiesen, auf Quellen zurückzugreifen, die man nicht selbst erzeugt hat. In den klinischen Forschungsabteilungen, die bisher die alleinige Zuständigkeit für Daten hatten, löst dies oft starke Kopfschmerzen aus, weil derartige Daten als „unrein“ gelten – sind sie von ganz anderer Art, als man gewohnt ist. Oft begegnet man ihnen dort mit Ablehnung, ist allerdings auch ratlos, wie man mit den gewohnten Instrumentarien weiterkommen soll, wenn man an die erforderlichen Budgets denkt.

Dabei ginge es allen Beteiligten besser, wenn sie sich an das einfache Prinzip erinnern, dass die Brauchbarkeit von Daten sich alleine am Zweck misst. Daten „an sich“ sind nicht valide. Sie sind es immer nur vor dem Hintergrund einer konkreten Fragestellung. Wenn man das anerkennt, stellt man nach einer Weile fest, dass es in Deutschland eben doch eine Menge von Daten gibt – man muss sie allerdings kennen. Dann sind auch bezahlbare Studien machbar. Dies soll im Folgenden gezeigt werden.

Welche Daten für welche Fragestellung?

In der ersten Folge dieser Serie (DZKF 7/8, 2004) hatten wir einen Überblick über wichtige Fragestellungen der Gesundheitsökonomie gegeben:

- *Wie groß ist der Bedarf nach einer Behandlung? Ist er gedeckt? Ist die Qualität ausreichend?*
- *Was kostet die Behandlung einer Krankheit oder einer Gesundheitsstörung?*
- *Welchen zusätzlichen Nutzen bringt der Einsatz eines Arzneimittels oder einer medizinischen Technologie in der realen Welt?*

Diese knappe Übersicht macht schon deutlich, dass Daten aus ganz unterschiedlichen Quellen angemessen

sein können: Um die Belastung der Bevölkerung durch eine lebensbedrohliche Erkrankung beschreiben zu können, kann schon die Todesursachenstatistik ausreichend sein. Um die Frage nach der Compliance oder Persistenz in Bezug auf eine bestimmte Therapie beantworten zu können, werden hingegen längsschnittliche und personenbezogene Daten über die Verordnung von Arzneimitteln benötigt.

Übersicht über Typen von Datenquellen

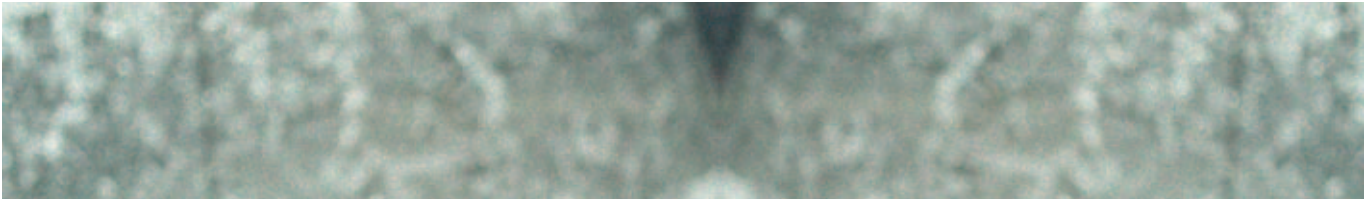
Datenquellen lassen sich nach zahlreichen Kriterien beschreiben und klassifizieren. Nur die wichtigsten können im Folgenden aufgeführt werden (siehe dazu auch Beispiele in Tabelle 1:

- *Sind die Daten routinemäßig vorhanden oder werden sie speziell erzeugt?*

Bereits vorhandene Daten erlauben natürlich einen zügigen Projektstart und sind in der Regel mit geringeren Kosten verbunden. Ihr Nachteil ist allerdings, dass sie in Bezug auf Präzision häufig nicht solche Ansprüche erfüllen, wie sie Mitarbeiter aus der klinischen Forschung gewohnt sind. Meist lässt sich aber beschreiben, in welchen Grenzen die erzeugten Ergebnisse Gültigkeit haben.

- *Sind die Datenhalter staatliche Stellen, Körperschaften (wie Krankenkassen oder Kassenärztliche Vereinigungen) oder private Anbieter?*

Von der Antwort auf diese Frage ist meist die Verfügbarkeit von Daten abhängig. In der Regel sind Daten staatlicher Stellen frei verfügbar, eingeschränkt allerdings durch Datenschutzbestimmungen. Diese erlauben es z. B. nicht, dass Daten bis auf die Ebene einzelner Einrichtungen desagregiert werden. Ein einzelnes Krankenhaus oder gar eine Abteilung kann somit nicht Gegenstand der Betrachtung sein.



■ *Haben die Daten einen direkten Populationsbezug?*

Der Populationsbezug entscheidet darüber, ob Ergebnisse auf die gesamte Bevölkerung hochgerechnet werden können. Bei Daten wie der Sterblichkeitsstatistik ist dies ohne Weiteres möglich. Bei regionalisierten Daten von Kassenärztlichen Vereinigungen ist oftmals die Bezugsbevölkerung nur mit zusätzlichen Modellannahmen herzustellen.

Beispiele von Datenquellen

Im Folgenden werden drei Typen von Datenquellen exemplarisch beschrieben.

■ *Öffentliche Statistiken*

Die Todesursachenstatistik ist eine herausragende Datenquelle in der Gesundheitsökonomie, weil sie einen entscheidenden klinischen „Endpunkt“ abbildet, nämlich den Tod und seine Ursachen. In Bezug auf definierbare Bevölkerungen gibt sie Auskunft über die Zahl von Todesfällen nach demographischen Kriterien. Wenn zusätzlich Angaben über die Inzidenz einzelner Erkrankungen bereit stehen (z. B. in Krebsregistern), können Neuerkrankungen mit Behandlungen und Sterbefällen in Verbindung gebracht werden. Mit gewissen Einschränkungen lassen sich Aussagen über die Effektivität von Interventionen unter realen Bedingungen treffen.

■ *Routinedaten der Krankenversicherungen*

Eine weltweit stark zunehmende Bedeutung kommt Untersuchungen zu, die sich auf Routinedaten von Krankenversicherungen („claims data“) stützen (siehe Abbildung 1). Routinedaten von deutschen Krankenversicherungen beziehen sich zum einen auf die versicherten Personen bzw. Patienten. Zum anderen geben sie Auskunft über die in Anspruch genommenen Leistungen (Details siehe Tabelle 2).

Für einige Leistungsbereiche sind die zugrunde liegenden Diagnosen verfügbar. Klinische Daten sind in Deutschland nur über sehr spezielle Konstruktionen verknüpfbar. Der besondere Vorteil dieser Daten liegt in der teilweise langen Historie, die zu den Versicherten gespeichert sind, und der Tatsache, dass die Leistungen aller Leistungserbringer zu den pseudonymisierten Personen verbunden sind. Ein großer Nachteil besteht darin, dass solche Daten von den Krankenkassen nur in ganz speziellen Konstellationen zur Verfügung gestellt werden

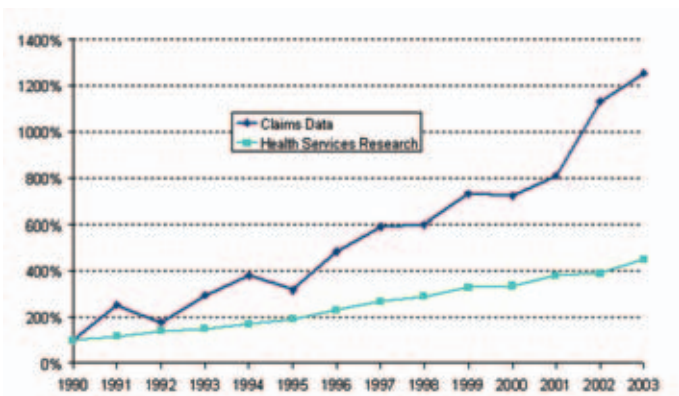


Abbildung 1: Entwicklung von Publikationen aus dem Bereich „Versorgungsforschung“ im Vergleich zu Untersuchungen, die sich auf Routinedaten (claims data) stützen. Index 1990=100 Prozent - Quelle: Recherche in Medline

■ *Haben die Daten einen Personenbezug und sind damit verknüpfbar über Leistungserbringer und die Zeit?*

Immer wenn Analysen gefordert sind, die sich auf Abfolge von Behandlungsmaßnahmen und daraus resultierenden Ergebnissen beziehen, werden Daten benötigt, die eine Verknüpfung von einzelnen Interventionen über die Zeit hinweg und über verschiedene Leistungserbringer hinweg ermöglichen. Dies ist häufig nicht der Fall bei reinen Fallstatistiken. Wenn diese eine Zahl von Fällen ausweisen, kann nur geschätzt werden, von wie vielen Patienten diese Fälle verursacht worden sind und welche anderen Fälle diesen Patienten möglicherweise noch zuzurechnen sind.



dürfen. Eine freie Verfügbarkeit – z. B. in Form eines „Public Use File“ – wird derzeit im Rahmen der Vorarbeiten zu einem Datentransparenzgesetz diskutiert.

In eigenen Untersuchungen haben wir z. B. die Effektivität der Eradikationstherapie bei peptischen Ulzera unter realen Bedingungen untersucht. Dabei zeigte sich, dass die Rehospitalisierung nach einer solchen Therapie im Vergleich zu einer konventionellen säurehemmenden Behandlung um etwa 90 Prozent niedriger liegt.

■ *Für bestimmte Fragestellungen eigens erzeugte Daten*

In prospektiven Untersuchungen können sowohl in experimentellen als auch in nicht-experimentellen Ansätzen gesundheitsökonomisch relevante Variablen einbezogen werden. Diese beziehen sich –

neben den klassischen klinischen Daten – zum einen auf die Bereiche „Lebensqualität“ bzw. krankheitsspezifische Instrumente zur Messung der Beeinträchtigung. Zum anderen werden Daten zur Inanspruchnahme von Leistungen bzw. zum Verbrauch von Ressourcen des Gesundheitssystems benötigt. In Bezug auf letztere liegt die Herausforderung vor allem darin, dass die Erhebung von Daten an einzelnen Patienten möglichst das gesamte Spektrum der in Anspruch genommenen Leistungen abbilden soll. Da Patienten oftmals nicht genau angeben können, welche diagnostischen oder therapeutischen Verfahren bei ihnen zum Einsatz kamen, kann der Rückgriff auf die behandelnden Einrichtungen notwendig werden. Patiententagebücher oder Online-Dokumentationen mit Hilfe von elektronischen Erfassungssystemen (PDA usw.) können hilfreich sein. Der zeitliche und finanzielle Aufwand eigener Erhebungen kann beträchtlich sein.

		routinemäßig vorhanden			erzeugt für allgemeine Zwecke	erzeugt für spezielle Fragestellung
		gesetzliche Grundlage	Versicherungen	Leistungserbringer		
mit direktem Populationsbezug	ohne Personenbezug	Bevölkerung	Diagnose-Statistik	Leistungsstatistik Ärzte	Survey (z. B. BGS)	repräsentative Befragung
		Todesursachen	Arzneimittel-Statistik (AVR, Gamsi etc.)	z. B. DPM von IMS		
		meldepflichtige Erkrankungen		z. B. IFAP-Datenbank		
		Krankenhaus-Fälle Bund		z. B. DRG-Datenbanken		
	mit Personenbezug		Inanspruchnahme von Leistungen		Panel (z. B. SOEP)	
					Register (z.B. MONICA)	
ohne direkten Populationsbezug	ohne Personenbezug			Krankenhaus-Fälle einzelne KHS		retrospektive Aktenanalyse
	mit Personenbezug				klinisches Register z. B. Mediplus von IMS	klinisch-epidemiologische Studien

Tabelle 1: Systematische Übersicht über Datenquellen für gesundheitsökonomische Untersuchungen mit einzelnen Beispielen.



available for standard sickness fund		available only for few sickness funds			available in few projects involving outpatient physicians	
person insured	services used					
year of birth / sex		Inpatient services	outpatient prescriptions	nursing and physiotherapy services	medical aids	physician outpatient services
region by postal code	date	admission / discharge	day of pick-up	day of provision	day of pick-up	day of provision
occupation	diagnoses	ICD10 ≤ 9 per discharge	---	optional	optional	ICD10 ≤ 20 per quarter
start / end insurance	type of service	not specified	package ID (link to ATC etc.)	specified	specified	procedure codes (> 1,000)
reason for termination (e.g. death)	provider	type of hospital region by postal code	physician specialty region by postal code	specified	specified	physician specialty region by postal code
	cost	indirect (cost per day)	pharmacy prices of statutory health insurance	prices of statutory health insurance	prices of statutory health insurance	value points multiplied with actual prizes per value point

Tabelle 2: Übersicht über den Inhalt von Routinedaten der gesetzlichen Krankenkassen in Deutschland.

Schlussbemerkung

Eine Übersicht über mögliche Datenquellen muss notwendiger Weise – wie hier in dem Beitrag – exemplarisch bleiben, weil die Gesamtheit verfügbarer und eigens erzeugter Datenquellen mittlerweile sehr groß ist. Jede Quelle hat Vor- und Nachteile. Eine Entscheidung über den Einsatz konkreter Datenquellen kann und muss daher immer vor dem Hintergrund der Fragestellung erfolgen, die sehr genau zu beschreiben ist. Eine Auswahl muss dabei immer abwägen zwischen der Aussagekraft und den Kosten, die mit einer konkreten Vorgehensweise einhergehen. Bei Bedarf können auch mehrere Datenquellen miteinander verknüpft werden. |

Kontakt

Prof. Dr. med. Bertram Häussler

Clinische Studien Gesellschaft mbh, Berlin
www.csg-germany.com
 E-Mail: csg@csg-germany.com

Dr. rer. pol. Thomas Ecker

IGES Institut für Gesundheits- und Sozialforschung GmbH
 Wichmannstr. 5
 D-10787 Berlin
 Tel.: +49 30 23 08 09 0
 Fax: +49 30 23 08 09 11
 E-Mail: iges@iges.de
www.iges.de

¹ Das IGES-Institut hat im September 2004 im Rahmen der „5th European Conference on Health Economics“ an der London School of Economics eine Satellitenveranstaltung organisiert, die sich mit der Reichweite derartiger Analysen befasste. Eine Buchpublikation hierzu ist in Vorbereitung. Die Präsentationen finden sich unter www.iges.de.