Die Limitationen des im „Informationssystem Versorgungsdaten“ vorliegenden Datenschatzes

Ein Blick in die „Datentankstelle“ des DIMDI

Auf dem sogenannten Marktplatz des 14. Deutschen Kongresses für Versorgungsforschung (DKVF), auf dem sich Dienstleister als auch Datenanbieter im Bereich Versorgungsforschung und -management präsentieren konnten, standen oft alleine und verlassen zwei Kongressteilnehmer, die zwar mit Daten umgehen können, aber von denen Versorgungsforscher nur zu oft sagen: Von diesen Beamten an Daten zu kommen, ist fast unmöglich. Diesem vielleicht allzu vorschnellen und einfachen Vorurteil stellten sich Dr. Jochen Dreß und Dirk Hellthaler vom Deutschen Institut für Medizinische Dokumentation und Information in Köln im Gespräch mit „Monitor Versorgungsforschung“. Das DIMDI gehört zum Geschäftsbereich des Bundesministeriums für Gesundheit (BMG) und betreibt in dessem Auftrag im Zuge des Datentransparenzgesetzes das „Informationssystem Versorgungsdaten“.

>> Für die „Datenaufbereitungsstelle“, die nach § 303d des SGB V die sogenannten „DaTraV-Daten“ zur Verfügung stellen soll, sind innerhalb des DIMDI fünf Stellen eingerichtet worden. Davon sind derzeit zwei dem IT-Bereich (Infrastruktur, System- & Datenbankadministration, IT-Sicherheit), eine der Verwaltung und zwei dem fachlich-medizinischen Bereich zugeordnet. Die beiden zuletzt genannten Stellen haben Dr. Jochen Dreß und Dirk Hellthaler inne. Ihre Aufgabe ist es, die eingehenden Anträge inhaltlich zu bearbeiten, was auch die Skriptprüfung und -entwicklung mit einschließt. Dreß ist zudem der fachliche Leiter der Datenaufbereitungsstelle, die auch über eine technische Leitung verfügt, die einem der beiden Mitarbeiter aus dem IT-Bereich zugeordnet ist.

Nun könnte man den Kölner Datenbeamten natürlich vorwerfen, was von einigen Versorgungsforschern auch getan wird, dass sie die Anträge zu langsam bearbeiten würden. Dazu muss man allerdings zweierlei wissen: Erstens sind Dreß und Hellthaler die einzigen beiden Mitarbeiter, denen die inhaltliche Überprüfung obliegt; zum zweiten dauert die Bearbeitung eines DaTraV-Datenantrags einfach seine Zeit, was aber wiederum absolut nichts mit der angenommenen durchschnittlichen Arbeitsgeschwindigkeit eines deutschen Beamten zu tun hat, sondern vielmehr mit der Komplexität der Auswertungen, der Menge der Datensätze und der datenschutzrechtlichen Sensibilität der Daten an sich. „Unser Ziel ist es, jeden Antrag durchzubringen“, verdeutlicht Dreß. Und Hellthaler setzt gleich hinzu: „Bisher gab es keinen Antrag, bei dem wir nicht per Mail, Telefon oder auch einem persönlichen Treffen mit den Antragstellern unterstützend eingreifen mussten.“

Eine weitergehende Abstimmung und/oder Beratung ist insbesondere dann erforderlich, wenn im Antrag

- die Fragestellung nur sehr allgemein formuliert wurde,
- die Methodik nur in groben Zügen oder unvollständig dargelegt, also z.B. keine Auswertungspopulation definiert wurde,
- bei einer Beantragung der Datenfernverarbeitung kein oder ein nicht lauffähiges Skript eingereicht wurde.

Des Weiteren bedarf es im Vorfeld der Bereitstellung einer Ergebnismenge auch einer Abstimmung bezüglich der Maßnahmen zur Wahrung der Identität der Versicherten, da diese Maßnahmen zwingend zu Modifikationen einer Ergebnismenge führen, die jedoch oft durch eine Variation der Methodik abgemildert werden kann.

Wichtig ist in diesem Kontext, dass seit Juni 2014 das DIMDI auch mit der Skriptentwicklung beauftragt werden kann. Insgesamt ist dies bisher 15 mal (ohne den zurückgezogenen Antrag 14) geschehen.

Hohe Komplexität

Bereits an dieser Stelle wird die Komplexität der Datenabfrage deutlich, die wiederum nicht dem DIMDI oder dessen Mitarbeitern zuzuschreiben ist, sondern die alleine dem im DaTraV-Gesetz festgelegten Modus geschuldet ist, mit dem die Daten weitergegeben werden dürfen. Selbst über die eingegangenen Anträge und die bereitgestellten Ergebnismengen darf nur sehr eingeschränkt berichtet werden. Der Grund dafür ist, dass es einfach keine gesetzliche Grundlage dafür gibt, die Anträge oder auch nur den Namen des Antragstellers zu benennen oder gar zu veröffentlichen. Obendrein sei laut Dreß im Rahmen des Gesetzgebungsverfahrens sehr eindringlich von verschiedener Seite darauf gedrungen worden, dass das Kölner Institut die Anträge (und damit natürlich auch die Datenweitergabe) vertraulich behandeln müsse.

Damit ist es aber auch nicht möglich, dass alle Anträge in einem Register veröffentlicht werden, damit beispielsweise Forscher sehen können, welche Anträge bereits gestellt worden sind, wo bestimmte Fragestellungen bereits bearbeitet werden, auch um Doppelarbeit zu vermeiden und wo man sich vielleicht auch einbringen könnte.

Auf dem Antrag kann jedoch jeder Antragsteller ankreuzen, ob er damit einverstanden ist, dass bestimmte Angaben zu seinem Antrag nach einem Jahr in einem noch zu schaffenden Antragsregister veröffentlicht werden dürfen. Immerhin: Bei 21 von 33 Anträgen wurde dieses Einverständnis auch gegeben. Freiwillig könnten Antragsteller zudem ihren Antrag oder ihre Studie jederzeit ergänzend ins Versorgungsforschungsregister des DNVF (http://www.versorgungsforschung-deutschland.de) eintragen – was aber bisher noch nie geschehen ist.

Andererseits gibt es auch noch keine einzige Rückkoppelung von jenen Forschern, die bereits Datensätze erhalten haben. Nach Wissen des DMIDI gibt es noch keine Publikationen, die auf Basis der mehr als ein Dutzend bereits ausgelieferter DaTraV-Ergebnismengen entstanden wären. Was aber auch daran liegen mag, dass Antragsteller nicht

Zeitlinie / Fakten

- Im **Februar 2014** wurde der Pilotbetrieb mit der Möglichkeit, die Daten per Datenfernverarbeitung auszuwerten, aufgenommen. Im Rahmen der Datenfernverarbeitung ist vom Antragsteller ein SQL-Skript einzureichen.
- Ab **Juni 2014** konnte mit Erreichen der ersten Ausbaustufe auch das DIMDI mit der Skriptentwicklung beauftragt werden.
- Ab **April 2015** ist auch die Verwendung von SAS zur Programmierung der Auswertungsskripte möglich (Ausbaustufe 2).
- **Ende 2015** soll die 3. Ausbaustufe erreicht werden. Ab dann kann die Nutzung eines Arbeitsplatzes für Gastwissenschaftler beantragt werden.
- Von **Februar 2014 bis Dezember 2014** sind 17 Anträge eingegangen, im **Jahr 2015** sind es bisher 16. Damit haben das DIMDI insgesamt 33 Anträge erreicht. Zum Stand 17.11.2015 sind 17 Anträge, inklusive einer Ablehnung aus formalen Gründen, beschieden worden. Ein Antrag wurde zurückgezogen.

nur Universitäten und Hochschulen sind. So wurden bereits für diverse andere Zwecke Ergebnismengen zur Verfügung gestellt, wobei eine Veröffentlichung der Ergebnisse aber oft gar nicht vorgesehen ist.

Zeitverzug von vier Jahren

Ein Blick auf den durch das DaTraV-Gesetz vorliegenden Datenstamm zeigt zudem: Da fehlt etwas! Das Hauptproblem der vorliegenden Datenbasis sind die fehlenden Daten insbesondere zu Diagnosen sowie ambulant verordneten und abgerechneten Arzneimitteln derjenigen Versicherten, die das GKV-System verlassen haben, da sie beispielsweise verstorben sind.

Dem nachgeordnet ist das Problem des Zeitverzugs der Datenlieferung vom BVA an das DIMDI – der beträgt immerhin vier Jahre. Dieser Verzug rührt aber wiederum daher, dass den über das DIMDI verfügbaren Datenmengen nur jene Morbi-RSA-Daten zu Grunde liegen, die je nach Berichtsjahr und erst nach einer Durchführung von Plausibilitätsprüfungen und Korrekturmeldungen durch das Bundesversicherungsamt (BVA) an das DIMDI übermittelt werden. Durch diese Prüfungen und Korrekturen beträgt der zeitliche Abstand vom Berichtsjahr bis zur Datenlieferung an das DIMDI eben jene vier Jahre. Zur Zeit sind darum auf der DIMDI-Website nur die Daten der Jahre 2008 bis aktuell 2012 verfügbar.

Problem: Datenauszüge

Der durch die DaTraV bedingte Zeitverzug von vier Jahren erklärt sich vor allem aus dem Bestreben, von Beginn an möglichst hochwertige, das heißt möglichst korrekte Daten bereitzustellen. Dabei wurde angenommen, dass die korrektesten Daten erst nach zwei Korrekturen (genauer: zwei Aktualisierungen) durch die GKVen und Plausibilisierungen durch das BVA vorliegen. Das Problem besteht nun darin, dass das BVA (das anders als das DIMDI Daten selbst auswerten darf) seine Analysen auf Basis eben jener plausibilisierten Datensätze durchführt, die wiederum nicht unbedingt zu allen Fragestellungen passen, die man anhand der Daten eigentlich klären könnte oder möchte. Jedoch: Welcher relevante Gewinn an Datenqualität tatsächlich durch die Korrekturmeldungen entsteht, ist unbekannt. Auch konnte bisher nicht geprüft werden, ob der gegebenenfalls erzielte Qualitätsgewinn überhaupt relevant ist.

Ein weiteres Problem liegt aber auch darin, dass dem DIMDI eine Zeit lang nur ein Datenauszug vorlag. Auch diesen Umstand hat das DIMDI nicht zu verantworten, sondern ist einzig und alleine auf eine Interpretation der DaTraV durch das BVA und das BMG zurückzuführen. Diese haben gemeinsam festgelegt, dass die Kölner „Datentankstelle" nur mehr jene Daten bekommen darf, die für den korrigierten Jahresausgleich verwendet worden sind. Tatsächlich steht im SGB X wie auch in der Verordnung selbst, dass an das Kölner Institut nur jene Daten weitergegeben werden dürfen, die für den Morbi-RSA erhoben worden sind. Leider wurde dieses Plazet etwas unglücklich gekoppelt mit einer Formulierung, die den Zeitpunkt der Datenlieferung festlegt.

Nun kann man darüber diskutieren, dass zwar im Gesetz alle für den im M-RSA erhobene Daten steht, wobei anscheinend „erhoben" ein etwas zu schwacher Begriff ist. Letztendlich werden zwar alle Kassendaten aus dem M-RSA, die dem BVA zur Verfügung gestellt werden, irgendwann und irgendwo erhoben, doch BVA und BMG stellen sich auf den Standpunkt, dass damit wohl nur jene gemeint wären, die tatsächlich für den korrigierten Jahresausgleich verwendet werden. Aus diesem Grunde bekam das DIMDI nicht mehr den vollständigen Morbi-RSA-Datensatz, sondern eben nur noch einen Datenauszug. Was sich wie eine bloße Lappalie anhört, hat tiefgreifende Konsequenzen für die Datenqualität: Diese Limitation macht es beispielsweise unmöglich, die Prävalenz und Inzidenz bestimmter Erkrankungen zu berechnen.

Warum dem so ist, kann nur in dem Umstand vermutet werden, dass die vollständigen Daten eventuell in der Lage sein könnten, den Morbi-RSA an sich hinterfragbar zu machen, was aber von verschiedener Seite nicht unbedingt gewünscht wird. Über die Gründe dafür lässt sich sicherlich vortrefflich spekulieren. Interessant ist aber, dass das BMG für 2016 eine Überarbeitung der DaTraV plant, in der dann unter anderem eindeutig festgelegt werden soll, dass die Datenaufbereitungsstelle wieder alle für den Morbi-RSA von den GKVen an das BVA gelieferten Daten erhalten wird. Auch bezüglich der Vorgaben zum Lieferzeitpunkt sind Änderungen geplant.

Ein ähnliches Problemfeld gibt es bei den Regionaldaten. Zwar gab es durchaus einmal die Überlegung, bei der Berechnung des Morbi-RSA auch Regionaldaten hinzuzuziehen, das wurde jedoch aufgrund der Intervention nur eines Bundeslandes verhindert. Mit der Auswirkung, dass das BVA die Gemeindeschlüssel für die Jahre 2009 und 2010 zwar sehr wohl erhalten hat, die aber nicht verwenden und schon gar nicht dem DIMDI weitergegeben darf. Der traurige Nebeneffekt: Damit sind aus den DIMDI-Daten keine regionalen Aussagen möglich.

Auch hier lässt sich wieder über die möglichen Gründe spekulieren. Das ist aber eigentlich unnötig, da der entsprechende §303 b bereits im letzten Jahr dergestalt angepasst wurde, dass das BVA die Gemeindeschlüssel für die Jahre 2009 und 2010 wieder an die Datenaufbereitungsstelle des DIMDI übermitteln muss. Dies ist bereits geschehen, womit diese seit dem 22. Juni 2015 für Auswertungen zur Verfügung stehen. Künftig müssen übrigens die GKVen der Datenaufbereitungsstelle über das BVA die Postleitzahl des Wohnorts zur Verfügung stellen. Das geschieht aber erst nach Etablierung eines entsprechenden Verfahrens, wahrscheinlich erstmalig im März 2016.

All diese unterschiedlichsten Interessenlagen und Limitationen, die sicher jede einzeln für sich nachvollziehbar oder zumindest erklärbar sind, führen zu Problemen. Die Verzögerungen bei den Antragsbearbeitungen resultieren jedoch vor allem aus

- der Komplexität der Auswertung von Routinedaten der GKVen unter hinreichender Wahrung des Identitätsschutzes der Versicherten, wobei letzteres immer auch eine Abwägung zwischen dem angestrebten Erkenntnisgewinn und dem Recht auf informelle Selbstbestimmung erfordert,
- und den verfügbaren Personalressourcen.

Die Hürde Identitätsschutz

Verkannt werden aber oft auch die hohen Anforderungen des Datenschutzes, vor allem hier des Identitätsschutzes der Versicherten.

Letzteres ist keine Trivialität, die sich mit Pseudonymen aus der Welt schaffen ließe, denn auch pseudonymisierte Daten sind unter Umständen personenbeziehbar – und das nicht nur in kleineren Indikationen, wie oft gesagt wird. So zeigen Untersuchungen aus den USA, dass mit nur zwei bis drei Detailangaben in einem Datensatz eine Person mit einer Wahrscheinlichkeit von immerhin 85 Prozent identifiziert werden könne.

Darum gibt es sogenannte Mindestfallzahlen, wobei die derzeit gültige Standardmindestfallzahl 30 beträgt. Jedoch ist eine Reduzierung auf 5 in begründeten Fällen durchaus möglich. Bei insgesamt 16 von 33 Anträgen wurde entweder eine Mindestfallzahl von < 30 beantragt (8) oder hierzu keinerlei Angaben gemacht (8). Bei den 8

Anträgen mit einer beantragten Mindestfallzahl < 30 wurde beantragt:

- dreimal eine Mindestfallzahl von 1
- viermal eine Mindestfallzahl von 5
- einmal eine Mindestfallzahl von 1

Die Verwendung der Standardmindestfallzahl von 30 kann je nach Fragestellung, Auswertungsmethodik und beantragter Detailtiefe dazu führen, dass mehr als 80 % aller Datenzeilen einer Ergebnismenge modifiziert werden müssen. Dieser Schritt zur Erreichung der primären Geheimhaltung geschieht per Software automatisiert. Allerdings ist im Rahmen der sekundären Geheimhaltung auch immer die Rückrechenbarkeit der modifizierten Angaben zu prüfen und gegebenenfalls unmöglich zu machen. Dies ist in der Regel deutlich aufwendiger und zumindest derzeit nicht automatisierbar.

Im Falle, dass mehr als 20 % aller Datenzeilen einer Ergebnismenge modifiziert werden müssen, erfolgt immer eine Rücksprache mit dem Antragsteller, um

- zu prüfen, ob durch eine Variation der Auswertungsmethodik der Umfang der erforderlichen Modifikationen und damit der Informationsverlust bei Wahrung des Identitätsschutzes der Versicherten reduziert werden kann.
- dem Antragsteller Gelegenheit zu geben, die Verwendung einer kleineren Mindestfallzahl zu beantragen, wobei damit das Prüfergebnis nicht vorweg genommen wird.

All das erklärt wiederum aber auch die mitunter oft lange Bearbeitungszeit eines Datenantrags. Die Arbeit des DIMDI würde es jedenfalls erleichtern, wenn das, was bald im von Dreß und Hellthaler zu schreibenden ersten Evaluationsbericht stehen wird, berücksichtigt würde: Erstens eine Änderung des Datenzuschnitts hin zum gesamten Datenstamm. Zweitens die Lieferung zu einem deutlich früheren Zeitpunkt, verbunden mit dem Auftrag zu prüfen, inwieweit eine Korrekturmeldung tatsächlich mit einem relevanten Qualitätsgewinn einhergeht oder nicht.

Doch dafür bräuchte das DIMDI mehr Ressourcen. Einerseits. Und andererseits auch den Willen der Versorgungsforscher, sich mit dem hier vorhandenen Datenschatz etwas genauer zu beschäftigen. Der Grund: Die Versorgungsforschung ist es bisher gewohnt, zum Beispiel Kassendaten auf Einzeldatensatzebene zu bekommen. Erst wenn diese Daten vorliegen, wird oft die passende und beste Methodik zur Beantwortung der jeweiligen Fragestellung gewählt. Bei den DIMDI-Daten aber ist das ganz anders. Wer hier Ergebnismengen haben will, muss im Antragsformular nicht nur eine klare wissenschaftliche Fragestellung festlegen, sondern auch die Methodik – möglichst unter Berücksichtigung der Datenschutzaspekte – nachvollziehbar darlegen. Durchaus hilfreich kann die Beifügung eines Auswertungsprogramms sein. Indes: Ein direkt lauffähiges Skript einzureichen hat noch kein einziger Antragsteller geschafft! <<

von:
MVF-Chefredakteur Peter Stegmaier

GKE auch als Antwort auf „zunehmende marktwirtschaftliche Orientierung“

>> Dass in einem Gesundheitssystem der Superlative sinnvolle Medizin nicht die Regel ist, zeigt das Beispiel Kreuzschmerz: Ärzte machen zu viele Röntgenbilder bei akuten Schmerzen, setzen aber zu selten kombinierte, aktivierenden Behandlungsprogramme ein, wenn die Schmerzen chronisch werden. Medizinische Über-, Unter oder Fehlversorgung erzeugt unnötiges Leid und unnötige Kosten. Davor schützen soll die Initiative „Gemeinsam Klug Entscheiden (GKE)“ der AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.) und ihrer derzeit 171 Mitgliedsgesellschaften.

„Weniger ist mehr“ – unter dieser Vorgabe befassen sich seit 2011 international zahlreiche wissenschaftliche Fachgesellschaften mit unnötigen oder sogar schädlichen medizinischen Leistungen. Ihre „Don’t do“-Empfehlungen sind in Top-5-Listen zusammengefasst und kommen zum Beispiel in den USA unter dem Begriff „Choosing Wisely“ zum Einsatz. „Gemeinsam Klug Entscheiden“ setzt umfassender an: „ Die Hinterfragung unnötiger medizinischer Leistungen ist nur eine Seite der Medaille. Zu unserem Verständnis gehört auch die Beschäftigung mit der Frage, welche Leistungen häufiger in Anspruch genommen und stärker unterstützt werden sollten“, sagte AWMF-Präsident Prof. Dr. med. Rolf Kreienberg aus Landshut.

Letzteres beginne bei der geringen Wertschätzung für Gespräche zwischen Arzt und Patient. Auch die international übliche Betrachtung aus der Perspektive einzelner Fachgebiete greife zu kurz. Für eine gute Versorgung ist die gemeinsame Abstimmung verschiedener Fachgebiete, Berufsgruppen und Patientenvertreter unabdingbar. Im Gegensatz zu anderen Ländern können die Fachgesellschaften in Deutschland auf 20 Jahre Erfahrung in Hinblick auf die Entwicklung einer solchen Konsensbildung aufbauen, so Kreienberg. „Als Idee hinter der deutschen Initiative steht, dass es in Deutschland bereits zahlreiche ausgezeichnete evidenzbasierte Leitlinien gibt, an denen die verschiedenen Fachgesellschaften gemeinsam fachübergreifend, mit anderen Berufsgruppen und mit betroffenen Patienten arbeiten“, erläutert Kreienberg, selbst langjähriger Vorsitzender der Ständigen Kommission Leitlinien der AWMF.

GKE ziele darauf ab, die in Deutschland trotz aller Bemühungen für Qualität bestehenden medizinischen Versorgungsprobleme aus Sicht der Fachgesellschaften unter dem Dach der AWMF zu identifizieren, zu benennen und zu beheben. „Die Initiative setzt dort an, wo Leitlinien nicht umgesetzt werden oder fehlen“, sagte Prof. Dr. Ina Kopp, Leiterin des Instituts für Medizinisches Wissensmanagement der AWMF.“

Dabei sei auch die Förderung der objektiven Information von Ärzten und Patienten



Prof. Dr. Ina Kopp, Leiterin des Instituts für Medizinisches Wissensmanagement der AWMF

wichtig, so Kopp – einhergehend mit der Benennung von Fehlanreizen im Vergütungssystem und der Auseinandersetzung mit den Erwartungen von Patienten, die mitunter durch wissenschaftlich nicht haltbare Informationen fehlgeleitet seien. „Am Ende sollen klug ausgewählte Empfehlungen als Wissensgrundlage für wissenschaftlich und ethisch begründete Entscheidungen stehen. Diese gelte es, in verständlicher Form auch in eine öffentliche Diskussion einzubringen“, so Kopp weiter. Denn das Thema Über,- Unter- und Fehlversorgung betreffe nicht allein das ärztliche Sprechzimmer und könne auch nicht allein hier gelöst werden, meint die Wissenschaftlerin: „Gemeinsam klug entscheiden ist auch eine Antwort auf die zunehmende marktwirtschaftliche Orientierung des Gesundheitssystems“. <<