

Stellungnahme des TMF-Arbeitskreises Versorgungsdaten zum Thema Informationssystem Versorgungsdaten

Kontext

Die TMF hat im Rahmen ihres „Forums Versorgungsforschung“ Anfang 2014 in Kooperation mit dem DIMDI und der AGENS (Arbeitsgruppe für die Erhebung und Nutzung von Sekundärdaten) der DGSM (Deutsche Gesellschaft für Sozialmedizin und Prävention) und der DGEpi (Deutsche Gesellschaft für Epidemiologie) eine öffentliche Fachveranstaltung ausgerichtet, auf der die Fortschritte in der Umsetzung der Datentransparenzverordnung (DaTraV), die DaTraV-Daten und das Antragsverfahren vorgestellt wurden. Mit 180 Teilnehmern war dieses Forum ausgebucht und wurde Ende März in identischer Form mit knapp 50 Teilnehmern wiederholt.

In diesen beiden Veranstaltungen wurde deutlich, dass aus Sicht der Nutzungsberechtigten noch viele Fragen zum Datenkörper und zum Antragsverfahren ungeklärt sind. Des Weiteren war in der Verordnung vorgesehen, dass sogenannte standardisierte Datensätze durch das DIMDI zur Verfügung gestellt werden. Hier gilt es die Anforderungen der Nutzungsberechtigten auf breiter Basis zu formulieren und abzustimmen. Auch wurde deutlich, dass die Möglichkeit des Datenzugangs gemäß § 303 SGB V über das DIMDI in Forschungskreisen noch zu wenig bekannt ist, sodass dem Verfahren nicht nur eine fachliche Diskussion, sondern auch eine hiermit einhergehende Bewerbung dieser Möglichkeit in Fachkreisen zugute käme.

Vor diesem Hintergrund wurde durch die AGENS-Sprecher - Dr. Holger Gothe (Hall i. T.), Peter Ihle (Köln), Dr. Enno Swart (Magdeburg) - die Einrichtung eines Arbeitskreises zur Datentransparenzverordnung („Arbeitskreis Versorgungsdaten“) bei der TMF beantragt und bewilligt. Das DIMDI hatte zuvor auf Anfrage seine Beteiligung am Arbeitskreis zugesagt. Ziel des Arbeitskreises ist es, die Forschung mit einer bundesweiten GKV-Routinedatenbasis nach wissenschaftlichen Standards, vor allem die GEP (Gute Epidemiologische Praxis) und die GPS (Gute Praxis Sekundärdatenanalyse) zu unterstützen sowie Vorschläge zur Weiterentwicklung des Informationssystems Versorgungsdaten (Datentransparenz) gemeinsam mit dem DIMDI zu diskutieren und zu formulieren.

Der Arbeitskreis traf sich seit seinem Bestehen an drei Terminen: 28. Nov. 2014, 27. Mai 2015 und 4. Nov. 2015. Eingeladen war ein Vertreter aus jeder der im § 303e SGB V aufgeführten Kategorie der Nutzungsberechtigten (siehe beiliegende Sitzungsprotokolle)¹. Alle Treffen fanden unter Beteiligung des DIMDI namentlich Herrn Dr. Jochen Dress (inhaltliche Leitung des Informationssystems Versorgungsdaten) statt, der u. a. über aktuelle Entwicklungen zum Thema Informationssystem berichtete. Weiterhin stellten auf den Treffen Nutzer ihre ersten Erfahrungen mit Anträgen und Auswertungen zur Diskussion, in die auch Erfahrungsberichte aus anderen Antragsverfahren zur Nutzung von Sozialdaten einfließen. Aus der gemeinsamen Diskussion ergaben sich vielfältige Anregungen a) zur Erweiterung der Datenbasis, b) zum Antragsverfahren, c) zur Datennutzung, d) zur Beratung und e) zur Weiterentwicklung des Begleitverfahrens. Zum Teil wurden die Anregungen vom DIMDI bereits aufgegriffen, soweit dies bei geltender Rechtslage und festgeschriebener Verfahrenspraxis möglich war.

Die vorliegende Stellungnahme fasst die aus Sicht der Vertreter der Nutzungsberechtigten Institutionen relevanten Anregungen in Form von Empfehlungen zusammen. Die Stellungnahme wird dem DIMDI für seinen Evaluationsbericht an das BMG zur Verfügung gestellt.

¹ http://www.tmf-ev.de/Arbeitsgruppen_Foren/ForumVersorgungsforschung/ArbeitskreisVersorgungsdaten.aspx

Erweiterung der Datenbasis

Regionalkennzeichen

Bereits im Laufe des Jahres 2015 wurde das Regionalkennzeichen (Wohnort des Versicherten nach dem amtlichen Gemeindegemeinschaftsschlüssel) für die bestehenden Berichtsjahre 2009 bis 2011 ergänzt, ab dem Berichtsjahr 2016 soll das Regionalkennzeichen (Postleitzahl des Wohnortes) routinemäßig enthalten sein. Diese Ergänzung entspricht einer frühen Anregung des Arbeitskreises und ist sehr zu begrüßen.

Versterbende

In den aktuell beim DIMDI vorliegenden Versorgungsdaten fehlen Angaben der Versterbenden für das Jahr ihres Versterbens und des Vorjahres. Daher können epidemiologisch wichtige und versorgungsrelevante Erkrankungen mit kurzen Überlebenszeiten, z. B. Herzinfarkt, Pankreaskarzinom oder Sepsis, nicht bzw. nur unvollständig ausgewertet werden. Das gilt ebenso für die Abbildung des in der Regel diagnoseunabhängig überdurchschnittlich hohen Versorgungsumfangs in den letzten beiden Lebensjahren.

Die Einbeziehung von Daten der Versterbenden ist für die Analyse des Versorgungsgeschehens unverzichtbar. Die Planung, die bereits in den Morbi-RSA-Lieferungen der gesetzlichen Krankenkassen an das BVA enthaltenen Versterbenden und deren Daten dem DIMDI kurzfristig zur Verfügung zu stellen, wird daher sehr begrüßt.

Weitere Variablen und Sektoren

Im Vergleich mit den bei den Krankenkassen verfügbaren und bisher für die Forschung genutzten versichertenbezogenen Daten stellt der Datensatz im Informationssystem Versorgungsdaten lediglich einen Ausschnitt dar. Beispielsweise fehlen im stationären Sektor tagesgenaue Angaben zu Beginn und Ende des stationären Aufenthalts oder Angaben zu den durchgeführten Leistungen wie OPS. Auch im ambulanten ärztlichen Sektor fehlen die Leistungsangaben (EBM-Ziffern, OPS-Codes). Angaben zu den Leistungserbringern fehlen vollständig, so dass beispielsweise die Einschaltung eines Facharztes im Rahmen von Diagnostik und Therapie nicht analysierbar ist. Andere Versorgungssektoren, wie beispielsweise Heil- und Hilfsmittel fehlen vollständig und sind lediglich als aggregierte jahresbezogene Kosten verfügbar. Auch die Daten der Pflegeversicherung fehlen, stellen doch auch sie eine wichtige Quelle der Versorgungsforschung dar. In diversen Publikationen finden sich hierzu zahlreiche Beispiele, welche Auswertungsmöglichkeiten auf Basis dieser detaillierten Datensätze und Leistungsangaben möglich sind.

Im Sommer 2015 führte das DIMDI eine Online-Befragung zum Thema Informationssystem Versorgungsdaten durch, in der auch Aspekte der Weiterentwicklung abgefragt wurden. Im Unterkapitel Datenanreicherung wurde nach weiteren Daten/Merkmalen gefragt, die aus Sicht der Nutzungsberechtigten zukünftig für Auswertungen zur Verfügung gestellt werden sollten. Erwartungsgemäß fanden sich in den Antworten viele der o. g. fehlenden Variablen. Inwieweit die Befragungsergebnisse eine Priorisierung zulassen, welche Variablen notwendig und welche vielleicht verzichtbar sind, scheint fraglich, da die Einschätzung oftmals vor dem Hintergrund des persönlichen spezifischen wissenschaftlichen Backgrounds und ggf. der spezifischen Fragestellung eines Datennutzers erfolgt. Aus Sicht des Arbeitskreises scheint es daher sinnvoll, die versichertenbezogenen und versorgungsrelevanten Angaben der Krankenkassen möglichst vollständig im Informationssystem bereitzustellen.

Antragsverfahren und Inanspruchnahme

Nach aktuellem Stand wurden seit Start des Informationssystems im April 2014 knapp über 30 Anträge gestellt. Von diesen Anträgen wurden bisher etwas mehr als die Hälfte abgeschlossen. Die Bearbeitungszeiten liegen deutlich über den in der Datentransparenzverordnung genannten drei Monaten, teilweise beträgt die Bearbeitungszeit über 12 Monate. Von der in der

Nutzungsverordnung verankerten Möglichkeit, die Bearbeitungszeit um jeweils einen Monat zu verlängern, wurde daher oft Gebrauch gemacht.

Die langen Bearbeitungszeiten halten Antragsteller teilweise davon ab, Folgeanträge zu stellen, da Ergebnismengen nicht innerhalb absehbarer Fristen zur Verfügung gestellt werden können. Dies ist ein Grund für die geringe Anzahl der bisher gestellten Anträge. Weitere Gründe – die Liste erhebt keinen Anspruch auf Vollständigkeit – sind nachfolgend stichwortartig aufgeführt und auch in der Befragung des DIMDI zu finden:

- Unvollständigkeit der Datenbasis (Verstorbene fehlen, Variablen und Sektoren sind nicht enthalten, siehe weiter oben)
- bislang kurzer Beobachtungszeitraum
- keine aktuellen Daten: Zeitverzug von 4 Jahren zwischen Erhebung der Daten und deren Verfügbarkeit für Auswertungen, Anfang 2016 stehen beispielsweise die Versorgungsdaten für die Jahre 2008 bis 2011 zur Verfügung
- nur für die Lauffähigkeitsprüfung von Skripten verwendbarer Beispieldatensatz, keine inhaltliche Prüfung möglich
- fehlende Basisauswertungen / Standardauswertungen / Datencharakterisierungen
- fehlendes Datenmanual
- Inadäquates Analyseverfahren: Bei dem aktuellen Antragsmodus ist lediglich die Übermittlung einer einzigen Ergebnismenge per Bescheid am Antragsende vorgesehen, (mehrmalige) Zwischenergebnisse dürfen nicht übermittelt werden. Notwendige iterative Plausibilisierungsschritte sind daher nicht durchführbar und Auswertungsskripts können nicht auf Basis von Zwischenergebnissen angepasst und optimiert werden. Hieraus folgt, dass zuweilen komplexe und unterschiedliche Modelle parallel programmiert werden, die ihrerseits wieder einen erheblichen Prüfaufwand nach sich ziehen.
- Unkenntnis der Methoden zur Auswertung von Sekundärdaten bei potentiellen Antragstellern
- Mindestfallzahlen in Ergebnismengen: Die derzeitigen Regelungen zur Wahrung der Identität der Versicherten erschweren eine zügige Antragsbearbeitung durch eine teilweise zeitintensive Prüfung der Ergebnismengen. Zudem müssen die Ergebnismengen oft weitreichend modifiziert werden, was ihre Verwendbarkeit entscheidend einschränken kann. Eine Abwägung zwischen den berechtigten Forschungsinteressen und dem Identitätsschutz der Versicherten, wie sie derzeit erfolgen muss, ist komplex. Wünschenswert ist aus Sicht der Antragsteller die gesetzliche Vorgabe einer generellen Mindestfallzahl von 5, wenn nicht doch, wie bei anderen Daten(forschungs)zentren möglich, Einzeldatensatz (Mindestfallzahl = 1) bereitgestellt werden können.

Vor dem Hintergrund der langen Bearbeitungszeiten regt der Arbeitskreis an, das Antragsverfahren sowie das Serviceangebot zur Unterstützung der Antragsteller auf den Prüfstand zu stellen und die Ressourcenzumessung für das DIMDI zu überdenken. Hierbei sollte dem Abstimmungs- und Beratungsbedarf Rechnung getragen werden. Vom Informationssystem Versorgungsdaten wurde in den Sitzungen des Arbeitskreises berichtet, dass die Anträge sich in ihrer Komplexität stark unterscheiden. Teilweise bestehe ein Antrag aus mehreren Teilprojekten, so dass auch der für die Bearbeitung des Antrags ergebende Ressourcenbedarf deutlich differiere. Insbesondere vor diesem Hintergrund sollte auch eine zweisträngige Bearbeitung von einfachen und aufwändigen bzw. komplexen Anträgen diskutiert werden, da nach dem bisherigen „first in, first out-Prinzip“ ein komplexer Antrag mit hohem Ressourcenbedarf die zeitnahe Bearbeitung nachfolgender (einfacher) Anträge blockiert.

Datennutzung

Gastarbeitsplatz

Nach Informationen des DIMDI soll im Informationssystem Versorgungsdaten ein Gastarbeitsplatz verfügbar sein, den Antragsteller im Rahmen des Antrags nutzen können. Die Inbetriebnahme ist für Anfang 2016 geplant, Erfahrungen hierzu liegen daher noch nicht vor.

Der Gastarbeitsplatz soll es den Nutzern ermöglichen, Skripte vor Ort zu entwickeln, zu testen, auf Performance zu optimieren und interaktiv zu modifizieren. Ebenso soll es möglich sein, die Ergebnismengen im Hinblick auf die Mindestfallzahl zu prüfen und ggfs. Gruppen zusammenzufassen.

Die Frage, inwieweit ein einziger zur Zeit geplanter Gastarbeitsplatz ausreicht, die Wünsche auf interaktive Vor-Ort-Analyse ohne längere Wartezeiten zu erfüllen, muss an dieser Stelle unbeantwortet bleiben. Hierbei sollten die Erfahrungen von anderen Datenzentren berücksichtigt werden, beispielsweise des IAB (Institut für Arbeitsmarkt- und Berufsforschung) oder der Rentenversicherung; über beide wurden in einer Sitzung des AK Versorgungsdaten berichtet.

VPN-Zugang

In anderen Forschungsdatenzentren, z. B. dem FDZ der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) oder dem FDZ der statistischen Ämter sind auch Zugriffe der Nutzer über gesicherte VPN-Verbindungen möglich. Diese Möglichkeit sollte auch für das Informationssystem Versorgungsdaten geprüft werden und könnte eine Entlastung der räumlichen Begrenzungen beim DIMDI selbst herbeiführen. Durch technische Möglichkeiten wie beispielsweise den Einsatz eines personalisierten Token-Generators für einen VPN-gesicherten Remotezugriff kann sichergestellt werden, dass der Zugriff auf die Daten oder Untermengen davon nur durch vertraglich benannte und autorisierte Personen erfolgt.

Beispieldatensatz

Am 5. November 2015 trafen sich 4 Vertreter aus der Forschung sowie eine Vertreterin des GKV-Spitzenverbandes in den Räumen der TMF in Berlin, um über die Möglichkeit eines Beispieldatensatzes zu diskutieren, der es den Antragstellern erlauben würde, Skripte bereits vor Einreichung beim DIMDI auf syntaktische Korrektheit und inhaltliche Plausibilität zu prüfen. In der Diskussion wurde die Notwendigkeit eines solchen Datensatzes betont. Idealerweise sollte er keine – wie etwa bei den Testdatensätzen des FDZ der statistischen Ämter gegeben – künstlichen Daten enthalten, da für die Programmierung *echte* Daten am besten geeignet sind, um auch Fehler und Implausibilitäten im Datenkörper selbst erkennen und programmtechnisch korrigieren zu können. Dies betrifft zum einen inhaltliche Implausibilitäten (z. B. Schwangerschaften bei Männern oder primär geriatrische Diagnosen bei Kindern) oder datentechnische Inkonsistenzen, wie fehlende Variablen oder implausible Ausprägungen von Variablen (z. B. Angabe von mehr als 10.000 abgegebenen Packungen pro Verordnung). Denkbar ist hier eine Zufallsstichprobe mit einem niedrigen Auswahlsatz, beispielsweise zwischen 0,1% und 1,0%. Allerdings wurde auch deutlich, dass noch erhebliche Herausforderungen bestehen, insbesondere sind datenschutzrechtliche Fragen im Hinblick auf die sicherzustellende Anonymität der Versicherten zu klären.

Insbesondere bitten wir zu prüfen, ob eine Bereitstellung einer Zufallsstichprobe als Testdatensatz über abgesicherte Remotearbeitsplätze dem Forschungsinteresse und dem Recht der Versicherten auf Wahrung ihrer Identität hinreichend Rechnung tragen würde und damit realisiert werden könnte.

Datennutzung für die Gesundheitsberichterstattung (GBE)

Von Seiten der interessierten Öffentlichkeit, von Gesundheitsexperten und -akteuren und von der Politik besteht ein berechtigtes Interesse, regelmäßig und ohne großen Aufwand Zugang zu aussagekräftigen Auswertungen von DaTraV-Daten zu erhalten, welche geeignet sind, die gesundheitliche Lage und gesundheitliche Entwicklung der Bevölkerung in Deutschland und in speziellen Regionen

Deutschlands noch besser einschätzen zu können. Die Bereitstellung solcher Auswertungen von DaTraV-Daten entspricht dem gesellschaftlichen Auftrag der Gesundheitsberichterstattung (GBE) durch den Bund und die Länder. Besonders bedeutsam sind hierbei die DaTraV-Daten aus dem ambulanten Bereich, weil dieser Bereich – obzwar im Krankheitsgeschehen umfassender als alle anderen Quellen - für die GBE bisher sehr schwer zugänglich ist. Die Auswertungen sollten regelmäßig (jährlich) und ohne konkreten Forschungsanlass bereitgestellt werden. Die routinemäßige Bereitstellung von ausgewählten, stark aggregierten DaTraV-Daten ist bisher in der DaTraV nicht vorgesehen und ist im Rahmen der bestehenden Regeln gescheitert. Die Form der erforderlichen periodischen Datenabfrage/Datenlieferung soll in Absprache zwischen den Länder-GBEen und der GBE des Bundes vorbereitet werden ("GBE-Musteranträge") und auch absehbare Anforderungen auf europäischer Ebene berücksichtigen. Sie muss selbstverständlich die Vorgaben des Datenschutzes berücksichtigen.

Datenmanual und Beratung

Die Diskussion im Arbeitskreis hat gezeigt, dass die bisher bereitgestellte Datensatzbeschreibung nicht ausreichend geeignet ist, den Inhalt des Datenkörpers hinreichend gut zu beschreiben. So sollte es von jeder Variable eine Verteilung der vorkommenden Dateninhalte geben, so dass beispielsweise bei ICD-Codes geprüft werden kann, ob nicht vorhandene Schlüsselnummern enthalten sind oder welcher Anteil einer Variable keine Angaben (missings) oder implausible Werte (z. B. beim Geburtsjahr) enthält. Ebenso könnten Basisauswertungen mit Beispielskripten die Datenbasis hinsichtlich geeigneter Zielpopulationen beschreiben, beispielsweise welcher Anteil an Versicherten (stratifiziert nach Geschlecht und Altersgruppen) in einem Jahr durchgängig versichert ist. Ähnliches könnte für die Versicherten erfolgen, die mit Hilfe des jahresübergreifenden Pseudonyms über 2 oder mehr Jahre beobachtet werden können.

Neben der Bereitstellung weiterer Informationen sollten die Möglichkeiten des DIMDI zur Beratung ausgebaut werden. So wäre z.B. die Durchführung von regelmäßigen Treffen von Antragstellern zum Austausch von Erfahrungen sinnvoll. Auch sollten Fortbildungsveranstaltungen zur Vermittlung benötigter Lehrinhalte (SQL, SAS, methodische Aspekte) angeboten werden. Die Einbindung von externen Experten evtl. unter Einbeziehung bestehender Fortbildungsreihen sollte geprüft werden.

Weiterentwicklung des Begleitverfahrens

Die bisherige Arbeit des Arbeitskreises hat gezeigt, dass eine fachliche Diskussion wichtig für die Optimierung des Informationssystems Versorgungsdaten ist. Durch die neutrale Aufhängung und die gemeinsame Gestaltung des Arbeitskreises mit dem DIMDI ist ein kooperativer, fruchtbarer Dialog möglich. Zugleich kann das DIMDI alle interessierten nutzungsberechtigten Kreise gezielt und effizient über die Weiterentwicklung des Verfahrens informieren und den im Arbeitskreis vertretenen Multiplikatorenkreis zur Dissemination und Popularisierung des Verfahrens nutzen.

Der Arbeitskreis würde es begrüßen, wenn diese Arbeitsform in Kooperation mit dem DIMDI fortgesetzt würde. Der Arbeitskreis bietet an, weiterhin einen Beitrag zur Weiterentwicklung des Systems zu leisten und als breit aufgestelltes Forum den Erfahrungsaustausch zwischen den Nutzungsberechtigten zu fördern. Die TMF als Träger hat eine Bereitschaft zur Fortsetzung der Aktivität bereits zugesichert. Wünschenswert wären ein formaler Auftrag bzw. eine strukturelle Kooperation und eine Kostenbeteiligung. (Bislang werden die Kosten für den Arbeitskreis aus Mitgliedsbeitragseinnahmen der TMF bestritten.) In einem solchen Rahmen wäre es auch möglich, gezielte Fortbildungsmaßnahmen zur optimierten Nutzung der Daten des Informationssystems Versorgungsdaten zu organisieren.

Dringend erforderlich scheint auch, die Erfahrungen bisheriger Antragsteller in einem regelmäßigen strukturierten Prozess unter Einbeziehung des DIMDI zusammenzutragen und Vorschläge für die

Weiterentwicklung des Informationssystems Versorgungsdaten unter Einbeziehung aller Beteiligten transparent zu erarbeiten. In regelmäßigen Nutzertreffen könnten so beispielsweise methodische Erkenntnisse aus bisherigen Projekten diskutiert werden oder Skripte vorgestellt und als Templates partizipativ weiterentwickelt werden. Der AK Versorgungsdaten bietet an, diesen Prozess zu organisieren und zu betreuen.

Zusammenfassung

Insgesamt ist der Aufbau des Informationssystems sehr zu begrüßen, werden dadurch doch erstmalig kassenübergreifend GKV-Routinedaten von allen Versicherten zur Verfügung gestellt, so dass im Gegensatz zu Sekundärdatenanalysen auf Basis einer Krankenkasse(nart) die Ergebnisse nicht mit durch die Krankenkasse(nart) bedingten Verzerrungen behaftet sind. Durch die Auswertung dieser Datenbasis können Analysen für ca. 85% der deutschen Bevölkerung durchgeführt werden.

Allerdings bestehen noch erhebliche Limitationen, die bedingt sind durch den Umfang der vorhandenen Daten sowie den Zugriff darauf. Hier ist sowohl eine Erweiterung und eine höhere Aktualität der Daten zu empfehlen. Modifikationen des Verfahrens auf verschiedenen Ebenen könnten hier Abhilfe schaffen: Zum einen könnte die Bereitstellung eines Testdatensatz förderlich sein. Die Einrichtung mehrerer Gastarbeitsplätze oder sogar die Bereitstellung von VPN-Zugängen scheint vonnöten zu sein, auch um die Bearbeitungsdauer der Anträge zu senken. Zum anderen scheint auch ein beratender Support geboten, hier wäre neben der antragsbezogenen, persönlichen Beratung u.a. an die Bereitstellung eines Manuals mit Beschreibung der Daten sowie ggf. einem Treffen der Nutzer zwecks Erfahrungsaustausch oder auch an Fortbildungs- bzw. Schulungsveranstaltungen zu denken.

Die in dieser Stellungnahme formulierten Ziele sind nach Ansicht des AK Versorgungsdaten geeignet, die Nutzung des Informationssystems Versorgungsdaten zu intensivieren. Damit wird die Intention des Gesetzgebers gefördert, die Daten mit Methoden der Versorgungsforschung zum Nutzen der Gesundheitsversorgung auf breiter Basis zu analysieren.

Berlin, 07.04.2016

Für den AK-Versorgungsdaten:

Peter Ihle, Köln

Enno Swart, Magdeburg

Holger Gothe, Hall in Tirol

Sebastian Claudius Semler, Berlin