

Kommentare zur

Verordnung zur Neufassung der Datentransparenzverordnung und zur Änderung der Datentransparenz-Gebührenordnung

Erstellt durch:

Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi)

und der

Arbeitsgruppe Validierung und Linkage von Sekundärdaten des Deutschen Netzwerks Versorgungsforschung (DNVF).

Ad §3 Absatz (1) Punkt 1

Die Postleitzahl des Wohnortes sollte 5-stellig übermittelt werden, um kleinräumige Analysen durchführen zu können und ggfs. weitere auf Postleitzahlebene oder Gemeindekennzifferenebene vorliegende Angaben zuspitzen zu können. Gemeindekennziffer

Mit den Angaben aus der Datenerfassungs- und -übermittlungsverordnung (DEÜV) liegen bei den Krankenkassen wichtige soziodemographische Angaben wie Tätigkeitsschlüssel und höchster erreichter Schulabschluss für die versicherungspflichtig Beschäftigten vor und sollten ergänzend übermittelt werden.

Ad §3 Absatz (1) Punkt 3.a

Die detaillierte Beschreibung der Leistungs- und Kostendaten bezieht sich fast ausschließlich auf die ambulante ärztliche Versorgung. Die zahnärztlichen Leistungs- und Kostendaten sollten ihrerseits spezifischer aufgeführt werden, da beispielsweise der genannte „Einheitliche Bewertungsmaßstab“ keine zahnärztlichen Leistungen abbildet.

Die Leistungs- und Kostendaten sollten auch die Sonderziffern (EBM) auf Bundesebene sowie der Kassenärztlichen Vereinigungen auf Landesebene mit den KV-spezifischen Verträgen enthalten, beispielsweise die Daten für die spezialisierte ambulante Palliativversorgung (SAPV) und die allgemeine ambulante Palliativversorgung (AAPV).

Ad §3 Absatz (1) Punkt 3.b

Bei den Verordnungsdaten sollten auf jeden Fall auch die so genannten Sonderziffern der Pharmazentralnummern übermittelt werden. Zusätzlich auch die für Rezepturen hinterlegten Angaben aus dem so genannten „Z-Segment“, das Angaben zu den für die Herstellung einer Rezeptur verwendeten Fertigarzneimitteln (mit Pharmazentralnummer) sowie Anzahl und Anteil der verwendeten Packung(en) und Kosten der Zubereitung inkl. verwendeter Hilfsmittel (z. B. Lösungen) enthält.

Ad §3 Absatz (1) Punkt 3.c

Bei der Aufzählung der Kosten und Leistungsdaten sollte im Passus „die abgerechneten Fallpauschalen nach den diagnosebezogenen Fallgruppen (DRG) sowie Angaben zu Entgelten“ die Konkretisierung „inkl. Anzahl Beatmungsstunden“ ergänzt werden. Des Weiteren sollte in der Aufzählung der Passus „bei ärztlicher Verordnung von Krankenhausbehandlung die Arztnummer des einweisenden Arztes, bei Verlegung das Institutionskennzeichen des veranlassenden Krankenhauses, bei Notfallaufnahme die die Aufnahme veranlassende Stelle“ ergänzt werden.

Auch die Angaben aus dem ambulanten Leistungsgeschehen stationärer Einrichten, die so genannte AMBO-Abrechnung gemäß § 301 SGB V, sollte übermittelt werden.

Ad §4

Grundsätzlich fällt auf, dass in der vorliegenden Version der DaTraV-Verordnung die möglichen entstehenden Kosten für verschiedene Institutionen angesprochen sind, jedoch nicht die Kosten für die in §4 aufgeführten Aufgaben für den GKV-Spitzenverband, dem im vorgeschlagenen Konzept jedoch eine zentrale Rolle als Datenannahmestelle zukommt.

Ad §4 Absatz (1)

Eine Prüfung der Daten auf „Vollständigkeit, Plausibilität und Konsistenz“ sollte näher spezifiziert werden.

Die Prüfung auf Vollständigkeit sollte vor allem die Aspekte Anzahl und Struktur der übermittelten Tabellen umfassen. Die Anzahl der gelieferten Datensätze, Ausprägungen relevanter Datenfelder (oder Kombinationen von inhaltlich zusammenhängenden Datenfeldern) sowie Anteil fehlender Angaben (missing value) oder offensichtliche Ausreißer (outlier) pro Datenfeld können im Zeitverlauf (vorangegangene Lieferungen, idealerweise mit Quartalsbezug oder wenn notwendig in kürzeren zeitlichen Bezugsperioden) oder anhand von Richtwerten geprüft werden. Inhaltliche Abhängigkeiten von Identifikatoren (Versicherter, Leistungserbringer, Fälle) können auf Konsistenz geprüft werden. Die Aufgabe der „Datenannahmestelle“ beim GKV-Spitzenverband sollte und kann es jedoch nicht sein, die inhaltliche Plausibilität der Daten zu prüfen. Diese qualitative Prüfung kann in der Regel nur vor dem Hintergrund der spezifischen Fragestellung (also im Kontext eines Antrags) erfolgen und bezieht sich auch dann lediglich auf für die Fragestellung herangezogenen Datenfelder und ist durch den Antragsteller auswertungsbegleitend durchzuführen. Eine allgemeine und vollständige Prüfung auf inhaltliche Plausibilität einer Datenlieferung einer Krankenkasse oder der Daten eines Berichtsjahres ist dagegen aus inhaltlichen Gründen und der Vielzahl der zu prüfenden Zusammenhänge grundsätzlich nicht möglich.

Die *versichertenbezogene* Zusammenführung erzwingt den vollständigen jahresbezogenen Aufbau einer Datenbank. Diese Parallelstruktur zum FDZ scheint auch datenschutzrechtlich problematisch. Auch prüftechnisch ist eine versichertenbezogene kassenübergreifende Prüfung (bei Kassenwechsellern) nicht zwingend erforderlich, da zum einen nur ein kleiner Teil der Versicherten von einem Kassenwechsel betroffen ist und zum anderen eine solche inhaltliche Prüfung vollständig und nur vor dem Hintergrund der jeweiligen Fragestellung sinnvoll ist (siehe hierzu auch weiter oben). Ein weiteres Argument ist, dass nur jahresbezogene Prüfungen möglich sind, Inkonsistenzen hingegen sich aber auch über Jahresgrenzen hinweg und damit auch über die Datenlieferungen hinweg ergeben können und sind daher nur auf Basis der Daten im FDZ durch das FDZ selbst oder idealerweise durch den Antragsteller im Rahmen seiner Fragestellung zu leisten.

Ad §4 Absatz (6)

Eine Frist wird entgegen Absatz 5 nicht genannt.

Ad §6 Absatz (2)

Eine Frist wird nicht genannt.

Ad §7 Absatz (4)

Wer legt die benannten „objektiven“ Kriterien fest? Im Sinne einer Transparenz gegenüber dem Antragsteller sollten die Kriterien öffentlich gemacht werden.

Insgesamt wäre die Ergänzung um notwendige Informationen und Tools wünschenswert, mit deren Hilfe die Antragstellung erleichtert wird, z.B. data dictionary, Liste und Hilfe zu Anwendungsprogrammen. Sofern diese Aspekte derzeit in anderen Paragraphen verortet sind, z.B. Testdatensätze in §9 (2), dann sollten sie in diesem Paragraphen als Pflichten des FDZ gebündelt werden.

Ad §8 Absatz 1 Punkt 2

Die Anforderungen an die Beschreibung des methodischen Ansatzes sollten konkretisiert werden.

Ad §8 Absatz 1 Punkt 5

Der Begriff „Zusammenführung“ bedarf der Konkretisierung, um eine Nutzung nicht unnötig einzuschränken. Was meint „Zusammenführung“: individuelles Datenlinkage? Datenlinkage auf dem Niveau aggregierter Entitäten (z. B. PLZ) oder im Sinne einer externen Validierung oder im Rahmen einer wissenschaftlichen Fragestellung. Dieser Punkt gilt entsprechend auch für §9 Absatz (3).

Ad §8 Absatz 3 Punkt 2

Der Nachsatz „die eine bestimmungsgemäße Datenverarbeitung gewährleisten“ ist in diesem Zusammenhang verzichtbar.

Ad §8 Absatz 4

Was genau ist mit dem Begriff „Daten“ gemeint? Personenbezogene Einzeldatensätze oder die Ergebnismengen?

Wenn dies pseudonymisierte Einzeldatensätze sind, dann werden diese (in der Mehrzahl der Anträge vermutlich virtuell und mit personifiziertem Token) dem Antragsteller oder seinen Mitarbeitern zur Verfügung gestellt. Eine Sichtung dieser Datensätze durch den Antragsteller, auch unter Hinzuziehung eines Dritten im Zuge einer Beratung sollte mit Nennung des einzubeziehenden Beraters im Antrag möglich sein.

Wenn es sich dagegen um die übermittelten Ergebnismengen handelt, so sind diese in der Regel auch Bestandteil von Publikationen, die damit öffentlich werden. Eine Beratung durch Dritte sollte dann problemlos möglich sein und bedarf keiner expliziten Nennung in der Verordnung.

Ad §9

Unklar ist, ob sich die genannten Bearbeitungszeiten von 3 Monaten inkl. einer einmaligen Verlängerung um einen Monat auf die Prüfung des Antrags selbst oder auf den gesamten Antragsprozess inkl. Übermittlung der Ergebnismengen bezieht. Die Unklarheit ergibt sich aus der Tatsache, dass ein eingereicherter Antrag auch mit der Beantragung verknüpft ist, personifizierte Einzeldatensätze zur Erstellung des Auswertungsprogramms zur Verfügung gestellt zu bekommen. Hierfür sollte allerdings bereits eine Prüfung des Antrags stattgefunden haben mit dem Ergebnis, dass das beantragte Projekt auf Basis der FDZ-Daten durchgeführt werden darf.

Die genannte Bearbeitungsfrist von maximal 4 Monaten bezieht sich auf den Antragsprozess inkl. der Prüfung eines ggf. einzureichenden Auswertungsprogramms. Aus Forschersicht wäre zu begrüßen, die Bereitstellung der Ergebnismengen in die Frist mit einzubeziehen, da auf diese Weise Planungssicherheit erreicht wird.

Insgesamt sollten alle Bearbeitungsfristen hier explizit festgeschrieben werden, um genau diese Planungssicherheit zu gewährleisten. Genau an dieser Stelle gab es ja in der aktuellen Ausgestaltung Probleme, da die Antragsbearbeitung beliebig oft verlängert wurde. Nur wenn dieses Problem gelöst wird, kann ein FDZ funktionieren.

Ad §9 Absatz Punkt 2

Sind mit „Testdatensätze“ die pseudonymisierten Einzeldatensätze nach §11 Absatz 1 gemeint oder andere vom FDZ zur Verfügung zu stellende Datensätze? Wenn nein, nach welcher Logik werden diese Testdatensätze erstellt?

Ad §9 Absatz Punkt 4

Hier liegt ein Widerspruch zwischen Gesetzestext und Begründungsteil vor: Das FDZ soll prüfen, ob „der Antrag mit den beim Forschungsdatenzentrum vorliegenden Daten bearbeitet werden kann“. Im Kommentar zu diesem Absatz steht jedoch, dass die Entscheidung, ob die Daten „[...] geeignet und erforderlich sind“ in der Verantwortung der Nutzungsberechtigten liegen. Aus diesem Grunde sollte §9 Abs. 1 Nr. 3 gestrichen werden, Nr. 4 sollte konkretisiert werden:

„4. der Antrag mit den beim Forschungsdatenzentrum vorliegenden Daten *formal* bearbeitet werden kann und“

Ad §10

Soweit aus dem Paragraphen ersichtlich, erfolgt ein verpflichtender Eintrag ins Antragsregister sofort nach Bewilligung und dementsprechend zum frühestmöglichen Zeitpunkt. Dies ist aus Transparenzgründen zu begrüßen, zumal andere Antragsteller dann frühzeitig erkennen können, ob ein vergleichbarer Antrag bereits gestellt wurde. Es sollte auch benannt sein, dass Zeiten bzw. der Bearbeitungsstand dort hinterlegt sind. Bei Registern klinischer Studien oder systematischer Reviews ist das ebenfalls ein Pflichtfeld, an dem erkennbar wäre, ob eine Studie schon seit langem nicht weitergeführt wurde, was für andere Antragsteller eine wichtige Information wäre. Auch die Finanzierung sollte als Pflichtfeld hinterlegt werden, ebenso wie eine Kurzzusammenfassung, da ein Titel allein nicht immer aussagekräftig ist.

Ad §11 Absatz 1

Die Intention dieser Paragraphen ist unseres Erachtens die Bereitstellung der (nach Absatz 3 geprüften) Ergebnismengen. Dies kann entweder durch Bereitstellung standardisierter (bereits im FDZ vorhandener) Datensätze erfolgen (Punkt 1) oder sich auf die mit Hilfe eines Auswertungsprogramms erstellten Ergebnismengen beziehen (Punkt 2).

Die Bereitstellung der pseudonymisierten Einzeldatensätze (Punkt 3) hingegen ist keine Bereitstellung (von Ergebnismengen) im Sinne der Logik dieses Paragraphen, sondern dient vorwiegend der Erstellung des Auswertungsprogramms durch den Antragsteller. Eine Prüfung der Einzeldatensätze hinsichtlich ihres spezifischen Reidentifikationsrisikos ist nicht sinnvoll, da es sich hierbei per definitionem um pseudonymisierte Einzeldatensätze handelt, die dem Antragsteller nur mit Geheimhaltungspflicht und strafbewehrt zur Verfügung gestellt werden. Punkt 3 sollte daher im Sinne der Logik der Datenbereitstellung nicht aufgeführt werden, sondern sinngemäß in Absatz 2 verschoben werden.

Ad §11 Absatz 2

Die Beschränkung auf 30 Arbeitstage unter dem Gesichtspunkt der Gleichbehandlung der Nutzungsberechtigten ist nachvollziehbar. Allerdings sollte eine Verlängerung der Nutzungszeit im Rahmen eines Antrags möglich sein. In der Regel betragen die Auswertungszeiten in Versorgungsforschungsprojekten, hier seien beispielhaft die Innovationsfondprojekte mit Nutzung von Routinedaten genannt, mehrere Monate. Zudem besteht bei Publikationen oft die Notwendigkeit weiterer Auswertungen, wenn Gutachten dies fordern. Die Erstellung der Auswertungsprogramme mit entsprechend aufwändigen statistischen Modellen ist zentraler Bestandteil des Auswertungsprozesses. Die Be-

schränkung auf 30 Arbeitstage scheint vor diesem Hintergrund deutlich zu kurz. Unklar ist auch, ob es sich bei dem genannten Zeitraum von 30 Tagen um eine zusammenhängende Periode handelt oder ob die Arbeitstage über einen längeren Zeitraum hinweg anteilig genutzt werden können.

Nach dem Vorbild anderer FDZ sollten längere Nutzungszeiträume eingeräumt werden (3 Jahre), sofern nicht Kapazitätsprobleme diesem entgegenstehen.

Ad §14

Es wird eine externe Evaluation, wie sie beispielsweise auch im Rahmen von Innovationsfondprojekten Standard ist, vorgeschlagen. Eine externe Evaluation könnte beispielsweise durch den GKV-Spitzenverband erfolgen. Auch eine Einbeziehung externer Expertise der Anspruchsberechtigten ist denkbar, ggfs. unter Einbeziehung des Arbeitskreises Versorgungsdaten der TMF bzw. einschlägiger Fachgesellschaften bzw. Arbeitsgruppen.

Bei den Angaben zum Evaluationsbericht sollte nach dem Passus „einschließlich einer Statistik über die nachgefragten Datensätze“ der Passus „sowie Publikationen“ ergänzt werden.

Das Datum für die erste Evaluation zum 31.12.2023 scheint dabei knapp gewählt, wenn überhaupt erst ein Jahr vorher erstmalig Datensätze an das FDZ übermittelt werden und Datennutzungen erst im Laufe des Jahres 2023 realisiert werden können.

Weitere Anmerkungen

Ad Prüfungen der Ergebnismengen durch das FDZ

Die übermittelten aggregierten Ergebnismengen haben per definitionem keinen Personenbezug mehr. Dies folgt auch der Logik, dass das grundsätzliche Forschungsinteresse nicht darin besteht, Aussagen über Einzelpersonen, sondern über Personengruppen zu treffen. Dass aus den Ergebnismengen keine Einzelpersonen erkennbar sind oder Dritten erkennbar gemacht werden, liegt somit auch in der Verantwortung des Antragstellers. Diese Mitverantwortung sollte auch im Rahmen der Antragstellung explizit schriftlich fixiert werden, z. B. „In dem Fall, dass in den bereitgestellten und aggregierten Ergebnismengen eine natürliche (oder betroffene) Person direkt oder indirekt identifizierbar ist, ist es dem Antragsteller untersagt, mit diesen Angaben ein Datenlinkage zu anderen Informationen herzustellen oder diese Information unabhängig vom Forschungskontext zu nutzen. Auch der Versuch, ein solches Datenlinkage zum Zwecke der Reidentifikation der betroffenen Person durchzuführen, ist strafbewehrt (oder strafbar)“.

Aktuell besteht im Informationssystem Versorgungsdaten trotz vergleichsweise geringer Antragszahlen das Problem, dass die Antragsbearbeitung insgesamt sehr lange dauert. Ein Grund dafür ist die umfassende Prüfung der oftmals komplexen Ergebnismengen auf eine potenzielle Reidentifizierbarkeit. Neben der Übertragung der Mitverantwortung auf den Antragsteller sollte weiterhin eine rechtssichere Handlungsanweisung über den Umfang von Prüfungen durch das FDZ vorliegen. Dies sollte im Rahmen der in § 13, Absatz 4 vor Bereitstellung der Daten an Nutzungsberechtigte durchzuführenden Datenschutzfolgeabschätzung erfolgen. Dabei ist auch der Rat für Sozial- und Wirtschaftswissenschaften (RatSWD) einzubeziehen.

Insgesamt werden die Akzeptanz des FDZ unter den Nutzungsberechtigten sowie der Umfang der tatsächlichen Nutzung ganz entscheidend von der Bearbeitungsdauer der Anträge abhängen. Deshalb sollte ein wesentliches Ziel der neugefassten Verordnung sein, die Grundlagen dafür zu schaffen.

Ad Personalstamm

Für das Forschungsdatenzentrum soll das Personal schrittweise von aktuell 5 auf 15 Stellen aufgestockt werden. Diese Aufstockung ist aus Forschersicht generell zu begrüßen, da die Antragsbearbeitung damit zeitnaher erfolgen kann. Es wird allerdings vorgeschlagen, diese Aufstockung vor dem Hintergrund der zukünftigen Antragszahlen und der damit verbundenen Bearbeitungsaufwände kontinuierlich zu bewerten und ggfs. den Personalstamm daran anzupassen. Ebenso sollten auch Aufwände für Administration und Pflege der aufzubauenden IT-Umgebung eingeplant werden, sowie der reibungslose technische Ablauf, z. B. eine Hotline, an den sich Antragsteller bei technischen Problemen wenden können. Dies kann beispielsweise durch die externe Evaluation erfolgen.

Ad Datenlinkage

Das Datenlinkage mit anderen Daten (z.B. anderer Sozialleistungsträger wie Pflegeversicherung oder Rentenversicherung) oder anderer Datenquellen sollte grundsätzlich möglich sein. Dies muss nicht zwangsläufig durch den Antragsteller erfolgen, sondern könnte zumindest mittelfristig durch eine offizielle Stelle (z. B. RKI) erfolgen. Die zu etablierende Konstruktion sollte deshalb so gestaltet sein, dass sie solche zukünftigen Entwicklungen ermöglicht. So werden insbesondere Lösungen zunehmend relevanter, um die Daten des Forschungsdatenzentrums personenbezogen mit weiteren Informationen unter Wahrung datenschutzrechtlicher Rahmenbedingungen zu verknüpfen. Dies würde die Perspektiven eines Forschungsdatenzentrums erheblich erweitern.

Ad Forschungsauftrag

Zu diskutieren ist, inwieweit das FDZ selbst mit den Daten forschen sollte, um die Antragsteller qualifiziert beraten zu können oder um ein Datenmanual mit relevanten Basisauswertungen erstellen zu können.

Ad Beirat

Zu erwägen ist, ob für das Forschungsdatenzentrum in der Zukunft ein Beirat gebildet wird, der die Ausgestaltung und die Arbeit des FDZ wissenschaftlich begleitet und den kontinuierlichen Ausbau maßgeblich unterstützt. Über den Beirat könnte außerdem eine kontinuierliche Einbindung der Perspektive der Bürger*innen als letztllicher Nutznießer der Forschung sichergestellt werden.

Dr. Holger Gothe	Sprecher der AGENS, Bereichsleiter Versorgungsforschung, IGES Institut GmbH, Berlin
Prof. Dr. Falk Hoffmann, MPH	Sprecher der AG Datenlinkage, Carl von Ossietzky Universität, Medizin und Gesundheitswissenschaften, Department für Versorgungsforschung, Oldenburg
Peter Ihle	Sprecher der AGENS, wissenschaftlicher Mitarbeiter, PMV forschungsgruppe, Universität zu Köln
Prof. Dr. Stefanie March	Sprecherin der AG Datenlinkage, Hochschule Magdeburg-Stendal, Fachbereich Soziale Arbeit, Gesundheit und Medien, Magdeburg
PD Dr. rer. biol. hum. Enno Swart	Sprecher der AGENS, stv. Institutsleiter, Otto-von-Guericke-Universität, Medizinische Fakultät, Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG), Magdeburg
Unter Mitarbeit von	
Dr. rer. medic. Silke Andrich	Wissenschaftliche Mitarbeiterin, Institut für Versorgungsforschung und Gesundheitsökonomie, Centre for Health and Society, Medizinische Fakultät, Heinrich-Heine-Universität Düsseldorf
Ingo Meyer	Leiter der PMV forschungsgruppe, Universität zu Köln
Dr. Ulrike Nimptsch	wissenschaftlicher Mitarbeiterin, Technische Universität Berlin, Fakultät VII Wirtschaft und Management, Fachgebiet Management im Gesundheitswesen, Berlin
Dr. Nadine Scholten	Stellvertretende Institutsleitung, IMVR - Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft, Universität zu Köln
Dr. Christoph Stallmann	wissenschaftlicher Mitarbeiter, Otto-von-Guericke-Universität, Medizinische Fakultät, Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG), Magdeburg

Magdeburg, Köln, Oldenburg, Berlin; 20. Mai 2020