

# Sekundäre Nutzung von hausärztlichen Routinedaten ist machbar – Bericht vom RADAR Projekt

## Secondary Use of Electronic Medical Record Data from Primary Health Care is Feasible: Report from RADAR Project

### Autoren

Johannes Hauswaldt<sup>1</sup> , Thomas Bahls<sup>2</sup>, Arne Blumentritt<sup>2</sup>, Iris Demmer<sup>1</sup>, Johannes Drepper<sup>3</sup>, Roland Groh<sup>4</sup>, Stephanie Heinemann<sup>1</sup>, Wolfgang Hoffmann<sup>2</sup>, Valérie Kempter<sup>3</sup>, Johannes Pung<sup>5</sup>, Otto Rienhoff<sup>5</sup>, Falk Schlegelmilch<sup>1</sup>, Philipp Wieder<sup>4</sup>, Ramin Yahyapour<sup>4</sup>, Eva Hummers<sup>1</sup>

### Institute

- 1 Institut für Allgemeinmedizin, Universitätsmedizin Göttingen, Göttingen, Deutschland
- 2 Institut für Community Medicine, Universitätsmedizin Greifswald, Greifswald, Deutschland
- 3 Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V., TMF, Berlin, Deutschland
- 4 Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Göttingen, Deutschland
- 5 Medizinische Informatik, Universitätsmedizin Göttingen, Göttingen, Deutschland

### Schlüsselwörter

Routinedaten, ambulante Versorgung, Hausarzt, Versorgungsforschung, Machbarkeitsstudie, Datenablage

### Key words

electronic medical records, primary health care, general practitioner, health services research, feasibility study, data repository

### Bibliografie

Gesundheitswesen 2021; 83 (Suppl. 2): S130–S138

DOI 10.1055/a-1676-4020

ISSN 0949-7013

© 2021. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

### Korrespondenzadresse

Dr. Johannes Hauswaldt, MD MPH  
Institut für Allgemeinmedizin  
Universitätsmedizin Göttingen  
Humboldtallee 38  
37075 Göttingen  
Deutschland  
johannes.hauswaldt@med.uni-goettingen.de

### ZUSAMMENFASSUNG

**Ziel der Studie** „Real world“-Daten aus der ambulanten Gesundheitsversorgung sind in Deutschland nur schwer systematisch und longitudinal zu erlangen. Unsere Vision ist eine permanente Datenablage mit repräsentativen, de-identifizierten Patienten- und Versorgungsdaten, längsschnittlich, fortwährend aktualisiert und von verschiedenen Versorgern, mit der Möglichkeit zur Verknüpfung mit weiteren Daten, etwa aus Patientenbefragungen oder biologischer Forschung, zugänglich für andere Forscher. Wir berichten methodische Vorgehensweisen und Ergebnisse aus dem RADAR Projekt.

**Methodik** Untersuchung des Rechtsrahmens, Entwicklung prototypischer technischer Abläufe und Lösungen, mit Machbarkeitsstudie zur Evaluation von technischer und inhaltlicher Funktionalität sowie Eignung für Fragestellungen der Versorgungsforschung.

**Ergebnisse** Ab 2016 entwickelte ein interdisziplinäres Wissenschaftlerteam ein Datenschutzkonzept für Exporte von Versorgungsdaten aus elektronischen Praxisverwaltungssystemen. Eine technische und organisatorische Forschungsinfrastruktur im ambulanten Sektor wurden entwickelt und im Anwendungsfall „Orale Antikoagulation“ (OAK) umgesetzt. In 7 niedersächsischen Hausarztpraxen wurden 100 Patienten gewonnen und nach informierter Einwilligung ihre ausgewählten Behandlungsdaten, reduziert auf 40 relevante Datenfelder, über die Behandlungsdatentransfer-Schnittstelle extrahiert, unmittelbar vor Ort in identifizierende bzw. medizinische Daten getrennt und verschlüsselt zur Treuhandstelle (THS) bzw. an den Datenhalter übertragen. 75 Patienten, die die Einschlusskriterien erfüllten (mind. 1 Jahr Behandlung mit OAK), erhielten einen Lebensqualitäts-Fragebogen über die THS per Post. Von 66 Rücksendungen wurden 63 Fragebogenergebnisse mit den Behandlungsdaten in der Datenablage verknüpft.

**Schlussfolgerung** Die rechtskonforme Machbarkeit der Gewinnung von pseudonymisierten hausärztlichen Routinedaten mit expliziter informierter Patienteneinwilligung und deren wissenschaftliche Nutzung einschließlich Re-Kontaktierung und Einbindung von Fragebogendaten konnte nachgewiesen werden. Die Schutzkonzepte Privacy by design und Datenminimie-

run (Artikel 25 mit Erwägungsgrund 78 DSGVO) wurden systematisch in das RADAR Projekt integriert und begründen wesentlich, dass der Machbarkeitsnachweis rechtskonformer Primärdatengewinnung und sekundärer Nutzung für Forschungszwecke gelang. Eine Nutzung hinreichend anonymisierter, aber noch sinnvoller hausärztlicher Gesundheitsdaten ohne individuelle Einwilligung ist im bestehenden Rechtsrahmen in Deutschland schwerlich umsetzbar.

## ABSTRACT

**Objectives** It is difficult to obtain longitudinal 'real world' data from ambulatory medical care in Germany in a systematic way. Our vision is a large German research data repository featuring representative, anonymized patient and outpatient health care data, longitudinal, continuously updated and across different providers, offering a perspective of linking secondary care data or additional data obtained from research cohorts, for example patient reported data or biodata, and will be accessible for other researchers. Here we report specific methods and results from the RADAR project.

**Methods** Survey of legislation, design of technical processes and organisational solutions, with a feasibility study to evaluate technical and content functionality, acceptability and performance fitness for health services research questions.

**Results** In 2016, a multi-disciplinary scientific team initiated the development of a privacy protection and IT security concept for data exported from the electronic medical records (EMR) of physicians' practices in line with the European General Data Protection Regulation. Technical and organisational requirements for lawful research infrastructure were developed and executed for use in a specific case, namely 'oral anticoagulation'. In 7 Lower Saxonian general practices, 100 patients were selected by their physician and their data – reduced to 40 essential data fields – extracted from EMR via a mandatory software interface after informed consent. Still in the practice, the data were split into identifying or medical data. These were encrypted and transferred either to a trusted third party (TTP) or to a data repository, respectively. 75 patients who met our inclusion criteria (minimum of one year of oral anticoagulation treatment) received a quality-of-life questionnaire via the TTP. Of the 66 returns, 63 responses were then linked to the EMR data in the repository.

**Conclusion** Results from RADAR project proved the technical and organisational feasibility of lawful, pseudonymised data acquisition and the linkage of questionnaires to EMR data. The protecting concepts privacy by design and data minimization (Art. 25 GDPR with Recital 78) were implemented. Without informed consent, secondary use of routine data from ambulatory care which are sufficiently anonymized but still meaningful is all but impossible under current German law.

## Einleitung

„Real world“ – Daten aus dem ambulanten Sektor sind für Versorgungsforschung „eminenter wichtig, um unser Gesundheitssystem zukunftsfähig zu machen“ [1]. Jedoch sind diese aus der ambulanten Gesundheitsversorgung in Deutschland nur schwer systematisch und longitudinal zu erlangen. Bereits für einfache Kennzahlen divergieren die spärlichen Angaben erheblich, bspw. zu Arzt-Patienten-Kontakten: während eine deutsche, häufig angeführte Quelle [2] für die ambulante ärztliche Versorgung 2009 bis 2018 mittlere jährliche Werte zwischen 7,99 und 8,28 Behandlungsfällen angibt, dabei jedoch lediglich die Versicherten einer einzigen gesetzlichen Krankenkasse zugrunde legen kann und zudem methodisch begründet anmerkt, die „... dargestellten Werte aus den Jahren 2009 bis 2018 dürften ... die realen Arztkontaktzahlen erheblich unterschätzen“, gibt ein internationaler Vergleich [3, 4] für denselben Zeitraum zwischen 9,2 und 10,0 jährliche Arztkonsultationen an.

Eine substanzielle Quelle für längsschnittliche, nachhaltige Versorgungsforschung sind medizinische Routinedaten<sup>1</sup>, die primär und zumeist de novo bei der alltäglichen ambulanten ärztlichen Versorgung von Patienten entstehen und zunächst der praxisinternen Dokumentation und Leistungsabrechnung dienen [7]. In elek-

tronischer Form dokumentiert, eröffnen diese Routinedaten weitreichende Erkenntnismöglichkeiten, wie Beispiele aus dem Ausland [8, 9] und Deutschland [10–13] zeigen.

International sind dauerhafte Einrichtungen zur wissenschaftlichen Nutzung von Daten aus der primären Gesundheitsversorgung für die Gesundheitsforschung seit vielen Jahren etabliert. So sammelt Clinical Practice Research Datalink (CPRD) [14] seit mehr als 30 Jahren de-identifizierte Patientendaten aus einem Netzwerk von Hausarztpraxen des Vereinigten Königreiches. Neben [15] finden sich andere Beispiele auch in den Niederlanden [16], Kanada [17], regional in Schottland [18] und in Form von vielen Registern in Dänemark [19]. Untersuchungen auf Basis von Routinedaten zeichnen sich durch Aktualität, Datenumfang und Verallgemeinerbarkeit aus [8].

Unsere Vision ist eine nachhaltige Datenablage mit repräsentativen, de-identifizierten Patienten- und Versorgungsdaten, längsschnittlich aufgebaut, fortwährend aktualisiert und von verschiedenen Versorgern, mit der Möglichkeit zur Verknüpfung mit weiteren Daten, etwa aus Patientenbefragungen oder biologischer Forschung, welche zugänglich für andere Forscher ist.

Auch in Deutschland benötigt man eine dafür dauerhaft eingerichtete Infrastruktur, mittels derer in datenschutzkonformer und transparenter Weise Routinedaten fortlaufend gesammelt, technisch und organisatorisch vorgehalten und Forschern systematisch verfügbar gemacht werden. Auf dem Weg dorthin werden hier methodische Ansätze und Ergebnisse aus dem RADAR Projekt (Routi-

<sup>1</sup> Routinedaten im Gesundheitswesen werden ausführlich verhandelt bei [5]. Allgemein sind „Routinedaten ... prozessproduzierte, umfangreiche Informationssammlungen, die im Rahmen der Verwaltung, Leistungserbringung bzw. Kostenerstattung (z. B. der gesetzlichen Kranken-, Renten- oder Unfallversicherung) anfallen und elektronisch erfasst sind.“ [6].

ne Anonymised Data for Advanced Health Services Research) berichtet.

Das übergeordnete Ziel des RADAR Projekts war, ein nachhaltiges Konzept zu entwickeln, Routinedaten aus der hausärztlichen Versorgung zu verarbeiten, um beobachtende Forschung in großem Maßstab, mit hoher Wertigkeit und von internationaler Sichtbarkeit zu bahnen, zudem um eine unverzerrte Rekrutierung von Patienten für klinische Studien möglich zu machen. Insbesondere wurden folgende Ziele angestrebt:

- Ein Datenschutzkonzept zu entwickeln,
- Festlegung von Arbeitsprozeduren für die Extraktion von Versorgungs- und Metadaten aus (haus-)ärztlichen Praxen auf sicherem Weg und im Einklang mit den Arbeitsabläufen der Praxen
- Prozess- und Auswertungsroutinen zum Sammeln und De-identifizieren von routinemäßig gewonnenen elektronischen Patientendaten zu entwickeln, die mittels der Behandlungsdatentransfer (BDT) Schnittstelle aus ärztlichen Praxen exportiert werden können, sowie Vorkehrungen für die Verwaltung von Identitäten und der informierten Einwilligungen zu treffen
- Eine prototypischen Datenablage zu schaffen, um die Daten dauerhaft zu speichern und für weitere Nutzung zur Verfügung zu halten
- Die entwickelten Konzepte, Werkzeuge und Arbeitsroutinen in die TMF-Sammlung verfügbarer Forschungsinstrumente aufzunehmen und anderen Forschern zugänglich zu machen
- Die Werkzeuge und Routinen als anwendbar zu erproben in zwei Szenarien, (1) in sicher verschlüsselter Weise ohne Einwilligung, (2) mit informierter Einwilligung.

Die daraufhin entwickelte Forschungsinfrastruktur wurde dann anhand einer klein skalierten Pilotstudie (eines sogenannten „Use Cases“) in hausärztlichen Praxen erprobt. Im Folgenden wird über das RADAR Projekt nach seinem Abschluss berichtet und sein Konzept analysiert. Prozesse, Produkte und Ergebnisse dieser Machbarkeitsstudie von 2016 bis 2021 werden berichtet.

## Methodik und methodische Kernprobleme

Wesentliche Neuerungen der rechtlichen Rahmenbedingungen für das RADAR Projekt waren die Einführung der europäischen Datenschutzgrundverordnung (DSGVO) 2018 [20] und nachfolgende Änderungen des Bundesdatenschutzgesetzes (BDSG) [21]. Ausgehend von einem Datenschutzleitfaden [22] und entsprechend aktuellen Empfehlungen von Experten der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) wurde zunächst spezifisch für das RADAR Projekt eine datenschutzkonforme Strategie entwickelt, mit der die grundlegende Neuforderung nach „Privacy by design“ umfassend umgesetzt werden konnte. Das Konzept wurde für eine sekundäre Datenverarbeitung von hausärztlichen Routinedaten mit Einwilligung der Betroffenen ausgelegt (sog. „pseudonymes Szenario“). Betroffene im Sinne des Artikel 4 DSGVO sind hier Patienten der Hausarztpraxen. Datensparsamkeit wurde erreicht durch eine kleine Zahl von Praxen für die Pilotstudie sowie Beschränkung auf 40 Datenfelder (Va-

riablen) der Behandlungsdatentransfer-Schnittstelle (BDT) [23], einer obligaten Software-Schnittstelle in ambulanten Praxisverwaltungssystemen (PVS). Das detaillierte Datenschutz- und IT-Sicherheitskonzept für das RADAR Projekt wurde einer externen Begutachtung durch die Mitarbeiter der Arbeitsgruppe Datenschutz der TMF und zusammen mit den für die Durchführung erforderlichen Ethik-Unterlagen einem Votum der zuständigen Ethikkommission zugeführt.

Technische und organisatorische Herausforderungen beim Erlangen und Verarbeiten hausärztlicher Routinedaten wie etwa die große Zahl unterschiedlicher PVS, ihre intransparente Datenhaltung in proprietären Anwendungen, die veraltete BDT-Schnittstelle und die zwischen den Praxen sich unterscheidende Datenqualität waren aus früheren Untersuchungen [24, 25] zum Teil bekannt. Jedoch erwachsen aus neuen erheblichen Zugangseinschränkungen und finanziellen Forderungen seitens der Hersteller von Praxisverwaltungssystemen (PVS) zusätzliche Hindernisse [26]. Auf dem Datenschutzkonzept für das RADAR Projekt aufbauend hatten die beteiligten Ärzte, Medizininformatiker, Spezialisten für Datenhaltung und Partner mit datentreuhänderischer sowie juristischer und organisatorischer Kompetenz eine projekteigene Dateninfrastruktur neu zu konzipieren, umzusetzen und als machbar zu erproben.

Die Machbarkeitsuntersuchung im RADAR Projekt hatte zum Ziel, in einer Pilotstudie anhand eines konkreten Anwendungsfalls – Identifizierung und Re-Kontaktierung von Patienten mit oraler Antikoagulation – zu zeigen, dass mittels Datenschutzkonzept und projekteigener Forschungsdateninfrastruktur die rechtskonforme Beantwortung einer exemplarischen, einfachen Versorgungsforschungsfrage technisch und organisatorisch möglich ist.

## Ergebnisse

Das komplexe RADAR Projekt als Untersuchung zum Aufbau und Unterhalt einer Forschungsdateninfrastruktur im ambulanten Versorgungssektor bestand aus zwei aufeinanderfolgenden Teilen. (A) Zunächst galt es, die rechtlichen und faktischen Rahmenbedingungen zu erkunden, ein Datenschutzkonzept zu entwickeln, daraus folgend technische Arbeitsprozeduren festzulegen, Arbeits- und Prozessroutinen zu sammeln, das De-Identifizieren und Aufbewahren von Routinedaten aus Praxisverwaltungssystemen zu entwickeln sowie deren weitere rechtskonforme Verknüpfung und wissenschaftliche Nutzung zu konzipieren. (B) Nachfolgend war das tatsächliche Funktionieren dieser neu geschaffenen Forschungsdateninfrastruktur zu demonstrieren. Dementsprechend werden hier die Ergebnisse in 2 Abschnitten berichtet.

(A) Ab 2016 untersuchte das Konsortium von Wissenschaftlern, ob und wie Routinedaten aus der hausärztlichen Versorgung sekundär für Forschungszwecke genutzt werden können. Schnell war als zwingende Voraussetzung deutlich geworden, dass für das RADAR Projekt zunächst eine datenschutzkonforme Strategie mit einem umfangreichen Datenschutzkonzept einschließlich seiner externen Begutachtung und dem Votum der zuständigen Ethikkommission entwickelt werden musste. Darauf aufbauend mussten die Partner eine projekteigene Dateninfrastruktur neu konzipieren, erstellen und erproben. Die Einigung auf die Zielsetzungen

für die gemeinsame Arbeit (► **Tab. 1**) gelang schnell und erfuhr eine Förderung durch die Deutsche Forschungsgemeinschaft.

Hingegen benötigten wir weit mehr Zeit für das Finden einer gemeinsamen fächerübergreifenden Terminologie sowie eines übereinstimmenden Verständnisses technischer, rechtlicher und ärztlicher Aspekte, bis die Experten aus den unterschiedlichen Fachgebieten einander tatsächlich verstanden und im Detail einheitlich sprachen.

## Datenschutz- und IT-Sicherheitskonzept

Für das RADAR Projekt erarbeiteten die Partner unter Federführung der TMF frühzeitig ein ausführliches Datenschutz- und IT-Sicherheitskonzept. Eine weitergehende Datennutzung, etwa durch eine Weitergabe der Daten an Dritte, war nicht als Teil des RADAR Projekts vorgesehen. Ausgelegt für eine sekundäre Datenverarbeitung von hausärztlichen Routinedaten mit Einwilligung der Betroffenen (sog. „pseudonymes Szenario“) finden sich seine wesentlichen strukturellen und inhaltlichen Elemente in (► **Tab. 2**).

Ein entsprechendes Konzept für ein „faktisch anonymes Szenario“ (also eine Datenextraktion mit Entfernung von identifizierenden Daten und erweiterter Behandlung quasi-identifizierender Merkmale zur Vermeidung einer Re-Identifikation anhand spezieller oder kombinierter Patienteneigenschaften) gelang im RADAR Projekt nicht. Hierzu untersuchten wir begleitend die für 2012 bis 2016 BDT-extrahierten Routinedaten von 3809 Patienten einer einzelnen Hausarztpraxis auf k-Anonymität (ein Maß für die Ununter-

scheidbarkeit zu Datensätzen anderer Individuen) und fanden heraus, dass bereits für die Kombination von Patientengeschlecht und Patientenalter (Dekade) ein arbiträr gesetzter Grenzwert von  $k = 30$  mehrfach unterschritten wurde (Alter 0 bis 10 Jahre: 14 weiblich und 9 männlich; 90 Jahre und älter: 14 männlich; 100 Jahre und älter: 9 weiblich), der extrahierte BDT-Datensatz auf Praxisebene also nicht als „faktisch anonym“ angesehen werden kann. Somit ist eine sinnvolle Praxisdatenverarbeitung für einen anderen Zweck (Forschung) als den ursprünglichen (Erfüllung des Behandlungsauftrags) ohne Einwilligung der betroffenen Patienten nach den Anforderungen der Datenschutzgrundverordnung nicht zulässig [23]. Wir konnten deshalb auch die beiden initial geplanten Szenarien bezüglich Machbarkeit und Informationsverlust nicht miteinander vergleichen.

Die DSGVO [20] trat während der Projektlaufzeit in Kraft, mit unmittelbaren und weitreichenden Auswirkungen für das RADAR Projekt. Auch wenn dies zunächst viele und umfassende Veränderungen und Anpassungen im Datenschutz- und Datensicherheitskonzept sowie für die technischen und organisatorischen Maßnahmen und die Prozesskonzipierung erforderlich machte, waren die ausführlich begründeten, klar definierenden und kohärenten Rechtssetzungen durch die DSGVO dem RADAR Projekt letztlich hilfreich. Privacy by design (Art. 25 DSGVO mit Erwägungsgrund 78) wurde durch ausgewählte infrastrukturelle Designelemente und -vorkehrungen, die weiter unten im Einzelnen dargestellt werden, durch eine Softwareanwendung zur Verarbeitung der Routinedaten aus den Praxisverwaltungssystemen (PVS) mit Datensatzanalyse (parsing) der einzelnen Datenfelder und Fehlerdokumentation über Log-Dateien, durch eine Trennung der identifizierenden (IDAT) von den medizinischen Daten (MDAT) [19] noch in der Praxis und durch Handlungsanweisungen für die Praxismitarbeiter mit Hilfemöglichkeiten implementiert. Unter mehrstufigem Pseudonym-Austausch und Transportverschlüsselung wurden separat die IDAT an die Treuhandstelle (THS) der Universitätsmedizin Greifswald bzw. die MDAT an die datenhaltende Einrichtung (Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen, GWDG) übertragen. Datensparsamkeit erreichten wir durch konzeptionelle Beschränkung auf 40 Datenfelder (Variablen), die von den Partnern als wesentlich für die Beantwortung von Versorgungsforschungsfragen angesehen wurden (► **Tab. 3**), von den mehr als 263 Feldern der Behandlungsdatentransfer-Schnittstelle, BDT [24, 25].

Das Datenschutz- und IT-Sicherheitskonzept wurde den Mitgliedern der AG Datenschutz der TMF, Berlin, vorgestellt und erhielt nach Diskussion und Überarbeitung deren positives Votum. Ebenso votierten der Datenschutzbeauftragte und die Ethikkommission der Universitätsmedizin Göttingen.

## Technische und organisatorische Maßnahmen

Um die rechtlichen Voraussetzungen auch aus der DSGVO bezüglich eines sicheren Schutzes personenbezogener Daten nach Möglichkeit wahren zu können, erkannten die Partner als wesentlichen Bestandteil der technischen und organisatorischen Vorkehrungen für das RADAR Projekt, dass nach der Gewinnung von Routinedaten aus einem PVS auch alle wesentlichen Teile ihrer ersten Prüfung und Aufbereitung sowie die De-Identifizierung noch in der jeweiligen Praxis durchgeführt und abgeschlossen werden müssen.

► **Tab. 1** RADAR Projekt, Arbeitsschritte.

<b>Datenschutzkonzepte</b> entwickeln und beschreiben
<b>Ablauf- und Datenverarbeitungsroutinen</b> entwickeln: Patienteninformationen über die BDT-Schnittstelle gewinnen und de-identifizieren, Identität und informierte Einwilligung rechtssicher und datenschutzgerecht verwalten
Ein <b>prototypisches Datenregister</b> aufbauen
<b>Pilotstudie:</b> Daten einer kleinen Zahl von Praxen und Patienten gewinnen und für einen Anwendungsfall auswerten
<b>Zwei Szenarien:</b> Pseudonymes und anonymes Szenario, Machbarkeit und Datenqualität vergleichen

► **Tab. 2** Datenschutz- und IT-Sicherheitskonzept des RADAR Projekts, Kapitel.

1	Einleitung
2	Zweckbestimmung
3	Verantwortlichkeiten im RADAR Projekt
4	Von der Datenverarbeitung betroffene Personen
5	Rechtsgrundlagen
6	Daten und Datenkategorien
7	Beschreibung der datenbezogenen Prozesse
8	Schutzbedarfsanalyse
9	Technische und organisatorische Maßnahmen
10	Speicherfristen und Löschung
11	Anlagen

► **Tab. 3** Medizinische Daten im RADAR Projekt, 40 BDT-Feldkennungen.

Nr.	Feldkennung	Feldinhalt	Verarbeitung in RADAR
1	0202	Praxistyp	
2	0204	Arztgruppe verbal	
3	0206	PLZ und Ort der Praxis	(Feld entbehrlich)
4	0225	Anzahl Ärzte	
5	3103	Geburtsdatum des Patienten	Geburtsjahr des Patienten
6	3110	Geschlecht des Patienten	
7	3649	Dauerdiagnosen ab Datum	
8	3650	Dauerdiagnosen	
9	3651	Dauermedikamente ab Datum	
10	3652	Dauermedikamente	
11	3656	Dauerbemerkungen (Allergien)	
12	4101	Quartal der Abrechnung	
13	4105	Geschäftsstelle	
14	4107	Abrechnungsart (Schein)	
15	4121	Gebührenordnung	
16	5000	Leistungstag	
17	5001	GNR/GNR-Ident	
18	6000	Abrechnungsdiagnose	
19	6001	ICD-Schlüssel	
20	6200	Tag der Speicherung von Behandlungsdaten	
21	6205	Aktuelle Diagnose	
22	6210	Medikament verordnet auf Rezept	
23	6211	Außerhalb Rezepts verordnetes Medikament	
24	6215	Ärztemuster	
25	6220	Befund	
26	6221	Fremdbefund	
27	6222	Laborbefund	
28	6225	Röntgenbefund	
29	6260	Therapie	
30	6265	Physikalische Therapie	
31	6280	Überweisung Inhalt	
32	6285	AU Dauer	
33	6286	AU wegen	
34	6290	Krankenhauseinweisung, Krankenhaus	
35	6291	Krankenhauseinweisung wegen	
36	8401	Befundart	
37	8410	Test-Ident	
38	8411	Testbezeichnung	
39	8420	Ergebnis-Wert	
40	8421	Einheit	

Strikt wurde vermieden, in die Hard- oder Software oder in die primären Daten eines PVS in irgendeiner Weise einzugreifen. Vielmehr wurden über eine obligate Software-Schnittstelle, die Behandlungsdatentransfer-Schnittstelle (BDT), Behandlungsdaten auf einen hardwareverschlüsselten USB-Stick zur Überwindung des in der Regel nicht an das Internet angebundene PVS („air gap“) extrahiert und anschließend, noch in der Praxis, ausschließlich im Arbeitsspeicher eines separaten internetfähigen Rechners mittels der

eigens entwickelten RADAR Software aufbereitet. Die Aufbereitung umfasste eine Strukturprüfung in Form einer Datensatzanalyse (parsing) und mitgeführter Log-Datei, eine Datenselektion allein von Patienten mit vorheriger Einwilligung und aus den ausgewählten 40 BDT-Datenfeldern (Variablen), eine mehrstufig angelegte Pseudonymisierung und die Vorbereitung für den verschlüsselten Versand der MDAT an die Datenhalterin (GWDC) bzw. der IDAT und der Einwilligungserklärung an die Unabhängige Treu-

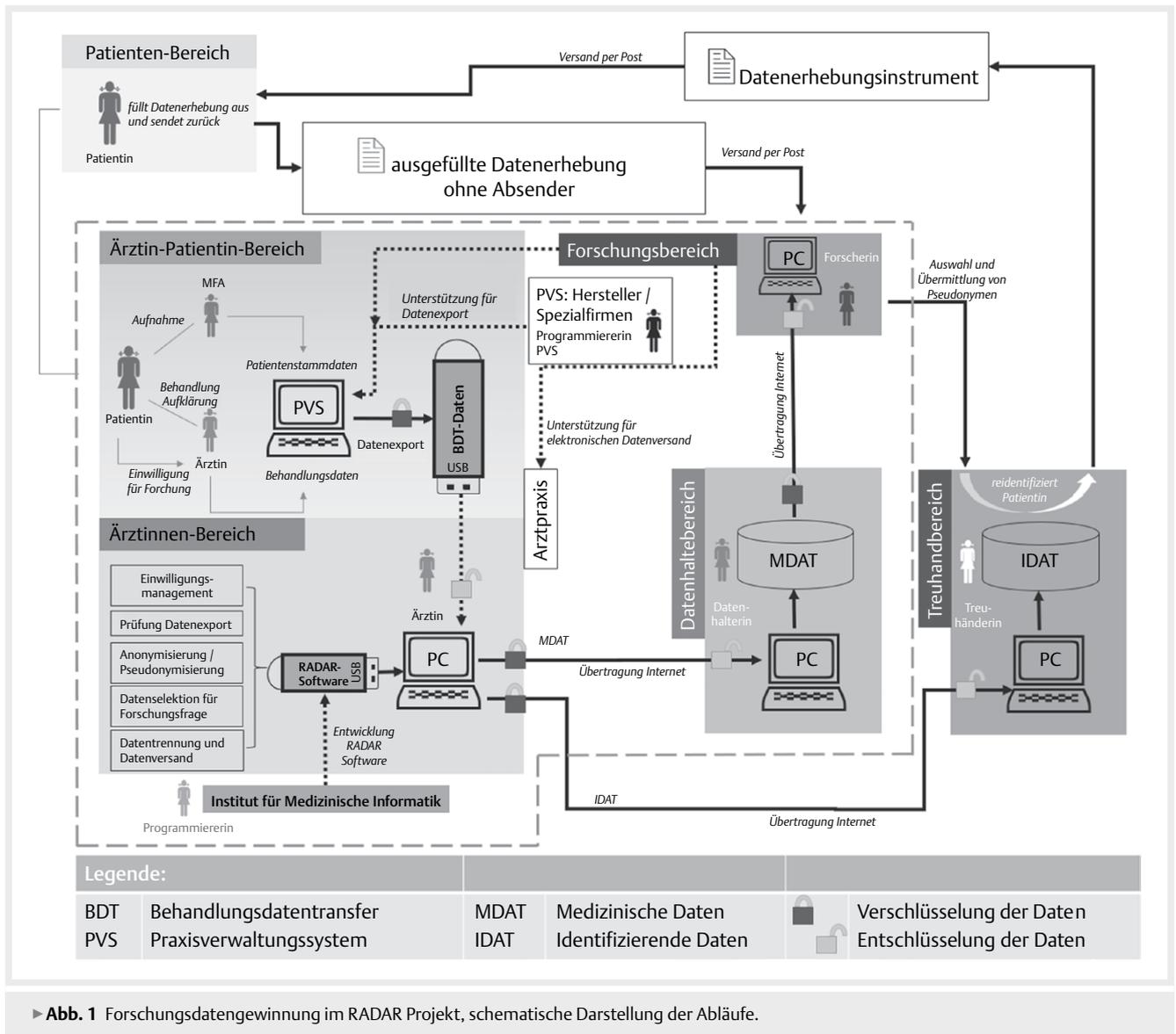
handstelle (THS). Der internetbasierte Versand der verschlüsselten Daten aus der Praxis heraus an die THS bzw. an die GWDG erfolgte mittels des separaten Laptops und eines LTE Surfsticks. Eine weitergehende technische Anbindung an die IT oder an das PVS der Arztpraxis war somit nicht erforderlich. Weitere Details für die Umsetzung des „pseudonymen Szenarios“ finden sich in (► **Abb. 1**) sowie in [27] und [28].

Den Forschern im RADAR Projekt wurde von der Datenhalterin (GWDG) ein geschützter Zugang zu den medizinischen Daten (MDAT) ermöglicht. Sie konnten anhand der Einschlusskriterien des Anwendungsfalls (mindestens 1 Jahr orale Antikoagulation) geeignete Patienten auffinden und eine Liste mit deren Pseudonymen an die THS übergeben. Die THS re-identifizierte mittels dieser Pseudonymliste die ausgewählten Patienten und deren Zustimmung zur Rekontaktierung anhand der allein bei der THS liegenden identifizierenden Daten (IDAT) und leitete an diese Patienten die Datenerhebungsinstrumente (DEI, hier Fragebögen mit Selbstauskunft zur Lebensqualität) weiter. Patienten füllten diese DEI aus und sandten sie

mittels eines vorfrankierten Rückumschlags ohne Absenderangabe an die forschende Einrichtung, das Institut für Allgemeinmedizin der UMG, zurück. Dort wurden anhand der Pseudonyme die medizinischen Daten mit den Daten der DEI zusammengeführt und ausgewertet. Das RADAR-Konzept stellt also sicher, dass IDAT und MDAT zu keinem Zeitpunkt außerhalb der Praxis zusammentreffen.

### Pilotstudie: Rekrutieren von Hausarztpraxen und Patienteneinwilligung

(B) In einem Verfahren, das in 3 Stufen um das Institut für Allgemeinmedizin der UMG, vom Juni bis August 2018 ausgeweitet wurde, baten wir per E-Mail und schriftlicher Einladung 964 niedersächsische Hausärzte, am RADAR Projekt teilzunehmen und partnerschaftlich mitzuarbeiten. Von 923 Praxen erhielten wir keine Rückmeldung, 27 Praxen hatten explizit kein Interesse teilzunehmen. 7 Praxen mit einem Praxisverwaltungssystem, das dem konzipierten Verfahren zugänglich war, waren zur Zusammenarbeit bereit; für Einzelheiten siehe [24]. Dem Studienplan entsprechend



► **Abb. 1** Forschungsdatengewinnung im RADAR Projekt, schematische Darstellung der Abläufe.

wählten die teilnehmenden Hausärzte unter ihren Patienten, die oral antikoaguliert waren, bis zu 20 Patienten zur möglichen Studienteilnahme aus, klärten sie strukturiert auf und holten ihre informierte Einwilligung schriftlich ein.

In den 7 teilnehmenden Praxen wurden von den Hausärzten insgesamt 100 Patienten mit oraler Antikoagulation eingeschlossen, ihre ausgewählten IDAT bzw. MDAT aus 2012 bis 2018 verschlüsselt und übertragen. 14 selbstversicherte Patienten konnten mittels Inhalte der ausgewählten 40 BDT-Felder nicht vollständig auf Erfüllung der Einschlusskriterien (Diagnose „Orale Antikoagulation“, Medikation mit oralen Antikoagulantien, Kennzeichnung in den Dauerbemerkungen des Patienten oder spezifische Laborwertbestimmungen für mindestens ein Jahr) untersucht werden und wurden von der weiteren Auswertung ausgeschlossen. Die Wissenschaftler der forschenden Einrichtung konnten die medizinischen Daten von 86 pseudonymisierten Patienten weiterverarbeiten. 75 erfüllten das Einschlusskriterium, mindestens ein Jahr lang mit oralen Antikoagulanzen behandelt worden zu sein. Sie wurden indirekt mithilfe der THS kontaktiert und erhielten auf diesem Wege per Post einen Lebensqualitäts-Fragenbogen. 66 Rücksendungen erreichten die forschende Einrichtung, von 63 Patienten konnten die Fragebogenergebnisse über das enthaltene Pseudonym mit den medizinischen Behandlungsdaten in der Datenablage verknüpft und ausgewertet werden.

In einer begleitenden Untersuchung zum Risiko der Re-Identifizierung bei der Auswertung von medizinischen Routinedaten [28] wurden weitere Lösungen zur De-Identifizierung vorgeschlagen, außerdem die ausgewählten 40 Variablen zu 11 semantischen Gruppen zusammengefasst und in Bezug zum Kerndatensatz der Medizininformatik-Initiative gesetzt. Bei Untersuchung der medizinischen Daten (MDAT) im Nachgang fanden wir in 2 BDT-Feldern vereinzelt Einträge, die personenidentifizierend sein könnten – dies wird im Folgenden auch mit möglichen Abhilfen diskutiert.

## Diskussion und Schlussfolgerung

Im DFG-geförderten RADAR Projekt (2016–2021) konnten die beteiligten Ärzte, Medizininformatiker, Datenhalter, Treuhänder und Spezialisten für rechtliche, technische und organisatorische Aspekte zeigen, dass hausärztliche Routinedaten mit Einwilligung der Betroffenen (Patienten, Professionelle) datenschutzgerecht gewonnen, sekundär wissenschaftlich genutzt und darüber hinaus auch mit weiteren Daten rechtskonform verknüpft werden können. Dabei standen die forschungsermöglichenden Rahmenbedingungen für die sekundäre Nutzung ambulanter ärztlicher Routinedaten im Vordergrund der interprofessionellen Arbeit.

Als Limitierung ist zum einen anzumerken, dass das umfangreiche Datenschutz- und IT-Sicherheitskonzept des RADAR Projekts weitgehend projektspezifisch und deshalb nicht ohne Weiteres auf andere Projekte übertragbar ist. Mitarbeiter der TMF e.V. haben vom Konzept des RADAR Projekts ausgehend eine verallgemeinerte Fassung [29] erstellt, die auf dem ToolPool Gesundheitsforschung der TMF zugänglich ist und für die Nachnutzung durch andere Forschungseinrichtungen zur Verfügung steht. Ebenso limitierend waren für unser Projekt unbeabsichtigte Selektionseffekte auf Ebene der teilnehmenden Ärzte und Praxen (convenience sam-

ple) und auf der Ebene der Patienten, weil deren Einschluss durch die Auswahl und Ansprache der Hausärzte erfolgte.

Unsere begleitenden Untersuchungen zur Anonymisierung von BDT-extrahierten Daten aus *einer* Praxis ergaben, dass bereits ein Datensatz einer Hausarztpraxis, der lediglich Patientengeschlecht und -alter in Jahren enthält, nicht als anonym im Rechtssinne anzusehen ist und deshalb nicht ohne Patienteneinwilligung im RADAR Projekt „für einen anderen Zweck“ verarbeitet werden darf.

Im RADAR Projekt war die Untersuchung von Patienten mit oraler Antikoagulation exemplarisch als Anwendungsfall (use case) für die Machbarkeit (feasibility study) von Aufbau und Nutzung einer geeigneten Forschungsinfrastruktur vorgesehen. Sein erfolgreicher Abschluss muss durch weitere Beispiele in entsprechendem Untersuchungsrahmen bestätigt werden. Ob für jede neue Forschungsfrage die betroffenen Patienten um Einwilligung gebeten und Daten neu extrahiert werden müssen, ist Gegenstand der laufenden Diskussion beispielsweise im Rahmen der Medizininformatik-Initiative um einen „broad consent“, der ansatzweise bereits im § 363 Abs.(8) SGB V angelegt ist. Ohne Einwilligung der betroffenen Patienten ist jedenfalls im bestehenden Rechtsrahmen eine Verarbeitung und Nutzung personenbezogener Daten nicht zulässig.

Durch Beschränkung auf die gewählten 40 BDT-Felder, deren Auswahl sehr wohl Gegenstand einer zukünftigen Diskussion sein mag, sind im Vergleich zu den Rohdaten in den PVS die sekundären Forschungsdaten eines Patienten sicherlich nicht vollständig. Wir konnten jedoch den prinzipiellen Vorteil zeigen, wenn die Primärdaten mit ihren ursprünglichen personenbezogenen Verknüpfungen untereinander am Ort ihrer Generierung und Speicherung, der Praxis, gewonnen werden. Auch die grundlegende Machbarkeit, Routinedaten personenbezogen mit Daten anderer Provenienz oder Sektoren zu verknüpfen, konnten wir zeigen – hierfür muss jedoch zwingend den aufwendigen Vorgaben eines hier exemplarisch dargestellten Datenschutz- und IT-Sicherheitskonzepts gefolgt werden.

Trotz Datensparsamkeit mit Beschränkung auf Datenfelder, die als wesentlich angesehen wurden, fand unsere systematische Nachuntersuchung der im RADAR Projekt gewonnenen medizinischen Daten – nicht gänzlich unerwartet – in 2 Textfeldern vereinzelt reidentifizierende Informationen, die einen Personenbezug zu Dritten möglich machen. Das BDT-Feld „PLZ und Ort der Praxis“ (FK 0206) muss zukünftig nicht mehr extrahiert werden, weil mittels des Pseudonyms der Praxis erforderlichenfalls diese über die Treuhandstelle in analoger Weise wie oben beschrieben re-kontaktiert werden kann. Im Freitextfeld „Befund“ (FK 6220) hingegen dokumentiert der behandelnde Arzt Beobachtungen und Bemerkungen zum Patienten, die auch aus wissenschaftlicher Sicht wertvoll sein mögen. In Zukunft muss hierfür eine datenschutzgerechte Lösung gefunden werden – für das Ziel hier des Machbarkeitsnachweises ist die Übertragung des Feldes ohne Feldinhalt ausreichend.

Im Datenschutzkonzept für das RADAR Projekt ist eine Weitergabe der gewonnenen Daten bzw. eine Zugriffsmöglichkeit an bzw. für Dritte, also an Forscher zur weiteren Untersuchung, noch nicht vorgesehen. Eine hierfür erforderliches Nutzungs- und Zugangskonzept im Sinne einer umfassenden Forschungsdateninfrastruktur, die die rechtlichen, technischen und organisatorischen, die

steuernden und führenden sowie die akteurspezifischen Voraussetzungen dauerhaft vorhält, muss erst noch geschaffen werden.

In zukünftigen Arbeiten sollten die Nutzungsmöglichkeiten für die extrahierten Daten, die Möglichkeiten der Patienteneinwilligung für unspezifische, noch nicht festgelegte Forschungsfragen („breiter Einsatzzweck“, broad consent) sowie die Anonymisierung für eine breite Datennachnutzung erforscht werden. Die Möglichkeiten zur Aufwandsreduktion (z. B. Digitalisierung der informierten Einwilligung) und die technologische Anpassung an interoperable Schnittstellen sowie der Einbezug der teilnehmenden Hausärzte in die Forschung mit diesen Daten sind ebenso Gegenstand des von der DFG geförderten Fortsetzungsprojektes RADARplus.

## Finanzierung

Das RADAR Projekt wurde gefördert von der Deutsche Forschungsgemeinschaft (DFG), Fördernummern HU 1587/2-1, HO 1937/7-1, RI 1000/7-1, YA 191/8-1 und KR 1093/10-1 vom 29.04.2016.

## Ethikvotum

Ethikkommission der Universitätsmedizin Göttingen, Nr. 23/2/18 vom 24.10.2018.

## Interessenkonflikt

Die Autoren erklären, dass keine Interessenkonflikte bestehen.

## Literatur

- [1] Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin: Die Initiative Deutscher Forschungsnetzwerke – DESAM-ForNet: Sechs Netze und eine Koordinierungsstelle gemeinsam - Bericht zum ersten Jahr. Zeitschrift für Allgemeinmedizin 2021; 97: 187–189
- [2] Grobe TG, Steinmann S, Szecsenyi J. Barmer Arztreport 2020. Rheinbreitbach: Medienhaus Plump GmbH; 2020
- [3] OECD: Health care utilisation. [https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-health-care-utilisation\\_data-00542-en](https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-health-care-utilisation_data-00542-en) (last accessed on 11 September 2021)
- [4] OECD: Doctors' consultations. <https://data.oecd.org/healthcare/doctors-consultations.htm> (last accessed on 11 September 2021)
- [5] Swart E, Ihle P, Gothe H. Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden, und Perspektiven. 2<sup>nd</sup> ed. s.l.: Verlag Hans Huber 2014
- [6] Hoffmann F, Glaeske G. Analyse von Routinedaten. In: Schrappe M, Neugebauer EA, Glaeske G, Pfaff H, Rothmund M, Schwartz W (eds.): Lehrbuch Versorgungsforschung: Systematik – Methodik – Anwendung, 2<sup>nd</sup> ed. Stuttgart: Schattauer Verlag 2017; 122–127
- [7] Richter M, Fleß S, Chenot J-F, Weckmann G, Haase A. Dokumentation in der Hausarztpraxis – eine qualitative Studie. Zeitschrift für Allgemeinmedizin 2018; 94: 223–228
- [8] Hippisley-Cox J, Coupland C. Development and validation of QMortality risk prediction algorithm to estimate short term risk of death and assess frailty: cohort study. BMJ 2017; 358: j4208
- [9] Vinogradova Y, Coupland C, Hippisley-Cox J. Use of combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases. BMJ 2015; 350: may26 h2135
- [10] Hauswaldt J, Himmel W, Hummers-Pradier E. The inter-contact interval: a new measure to define frequent attenders in primary care. BMC Fam Pract 2013; 14: 162
- [11] Hauswaldt J, Hummers-Pradier E, Junius-Walker U. Health service use among patients with chronic or multiple illnesses, and frequent attenders: secondary analysis of routine primary care data from 1996 to 2006. Dtsch Arztebl Int 2012; 109: 814–820
- [12] Hauswaldt J, Kersting M, Hummers-Pradier E. Influenza-Impfungen durch Niedersächsische Hausärzte – eine Sekundäranalyse vertragsärztlicher Versorgungsdaten aus 1995/1996, 2002/2003 und 2005/2006. Gesundheitswesen 2010; 72: 332–339
- [13] Laux G.(ed.) Versorgungsforschung in der Hausarztpraxis: Ergebnisse aus dem CONTENT-Projekt 2006–2009. München: Springer; 2010
- [14] CPRD: Clinical Practice Research Datalink: UK data driving real-world evidence. <https://www.cprd.com/> (last accessed on 11 September 2021)
- [15] QResearch: Generating new knowledge to improve patient care. QResearch is a large consolidated database derived from the anonymised health records of over 35 million patients. <https://www.qresearch.org/> (last accessed on 11 September 2021)
- [16] NIVEL: Nivel Primary Care Database. <https://www.nivel.nl/en/nivel-zorgregistraties-eerste-lijn/nivel-primary-care-database> (last accessed on 11 September 2021)
- [17] CPCSSN: Canadian Primary Care Sentinel Surveillance Network (CPCSSN). <https://cpcssn.ca/> (last accessed on 11 September 2021)
- [18] NHS Scotland: Scottish Primary Care Information Resource (SPIRE) <https://spire.scot/copyright/> (last accessed on 11 September 2021)
- [19] Copenhagen Healthtech Cluster: Explore the Danish Health Data <https://www.danishhealthdata.com/> (last accessed on 11 September 2021)
- [20] DSGVO: Datenschutz-Grundverordnung. Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG <https://dsgvo-gesetz.de/> (last accessed on 11 September 2021)
- [21] DSGVO: Bundesdatenschutzgesetz (BDSG), neu 2018. Teil des Datenschutz-Anpassungs- und -Umsetzungsgesetz EU (DSAnpUG-EU). Zum letzten Mal durch das zweite Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Zweites Datenschutz-Anpassungs- und Umsetzungsgesetz EU – 2. DSAnpUG-EU) geändert, am 26. November 2019 in Kraft getreten. <https://dsgvo-gesetz.de/bdsg/> (last accessed on 11 September 2021)
- [22] Pommerening K, Müller T. Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0. Berlin: MWV Med. Wiss. Verl.-Ges; 2014
- [23] Lichtner F, Sembritzki J. BDT-Satzbeschreibung. Schnittstellenbeschreibung zum systemunabhängigen Datentransfer von Behandlungsdaten. Version 02/94. <https://www.qms-standards.de/standards/standards-zum-download/> (last accessed on 11 September 2021)
- [24] Kersting M, Gierschmann A, Hauswaldt J, H-Pradier E. Routinedaten aus hausärztlichen Arztinformationssystemen – Export, Analyse und Aufbereitung für die Versorgungsforschung. Gesundheitswesen 2010; 72: 323–331

- [25] Himmel W, Hummers-Pradier E, Kochen MM. Medizinische Versorgung in der hausärztlichen Praxis. Ein neuer Forschungsansatz. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2006; 49: 151–159
- [26] Hauswaldt J, Kempster V, Himmel W, Hummers E. Hindernisse bei der sekundären Nutzung hausärztlicher Routinedaten. Gesundheitswesen 2018; 80: 987–993
- [27] Bahls T, Pung J, Heinemann S et al. Designing and piloting a generic research architecture and workflows to unlock German primary care data for secondary use. J Transl Med 2020; 18: 394
- [28] Hauswaldt J, Demmer I, Heinemann S et al. Das Risiko von Re-Identifizierung bei der Auswertung medizinischer Routinedaten – Kritische Bewertung und Lösungsansätze. Z Evid Fortbild Qual Gesundhwes 2019; 149: 22–31
- [29] TMF: Generisches Datenschutz- und IT-Sicherheitskonzept für die Forschung mit Daten aus der ambulanten Versorgung (last accessed on 11 September 2021)