Thieme

Verknüpfung von Abrechnungsdaten gesetzlicher Krankenkassen und Einsatzprotokollen des Rettungsdienstes: Brückenschlag durch Krankenversichertennummer?

Linking Health Claims Data and Records of Emergency Medical Services: Building a Bridge via Patient's Health Insurance Number?

Autoren
Ludwig Goldhahn[1, 2], Enno Swart[1], Silke Piedmont[1, 3]

Institute
1 Institut für Sozialmedizin und Gesundheitssystemforschung, Otto von Guericke Universität Magdeburg, Magdeburg, Deutschland
2 Medizinische Fakultät, Universitätsklinik für Unfallchirurgie, Otto von Guericke Universität Magdeburg, Magdeburg, Deutschland
3 Medizinische Hochschule Brandenburg Theodor Fontane, Neuruppin, Deutschland

Schlüsselwörter
Notfallrettung, GKV-Daten, Sekundärdaten, Datenlinkage, patientenbezogene Gesundheitsdaten

Key words
Emergency Medical Services, Health Insurance, Claims Data, Data Linkage, Data science

Bibliografie
Gesundheitswesen 2021; 83 (Suppl. 2): S102–S112
DOI 10.1055/a-1630-7398
ISSN 0949-7013
© 2021. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Korrespondenzadresse
Ludwig Goldhahn, M.A.
Institut für Sozialmedizin und Gesundheitssystemforschung
Otto von Guericke Universitat Magdeburg
Leipziger Straße 44
39106 Magdeburg
Deutschland
ludwig.goldhahn@med.ovgu.de

Zusätzliches Material finden Sie unter https://doi.org/10.1055/a-1630-7398

ZUSAMMENFASSUNG

Einleitung In Deutschland wurde der Rettungsdienst im Jahr 2016/2017 bei insgesamt rund 7,3 Mio. Notfallereignissen aktiv [1]. Informationen zu der dabei erfolgenden prähospitalen Versorgung von Patient*innen finden sich in unterschiedlichen Sekundärdatenquellen; jedoch ist die kombinierte Analyse dieser Daten auf Fall-/Einsatz- oder Patient*innen-Ebene selten der Fall. Um dies zu ändern, ist die Erforschung von Merkmalen und Methoden erforderlich, anhand derer Sekundärdaten, welche präklinische Inhalte umfassen, verknüpft werden können.
Methode Es werden Rettungsdienst-Einsatzprotokolle aus 5 bayerischen Rettungsdienstbereichen mit Abrechnungsinformationen 10 gesetzlicher Krankenkassen zusammengeführt (Daten aus 2016). Demonstriert werden 2 Verknüpfungsansätze auf Ebene patient*innenindividueller Einsätze bzw. abgerechneter Leistungen. Zuerst erfolgt die Verknüpfung deterministisch anhand des Personenidentifikators Krankenversichertennummer (KVNR). Der zweite Verknüpfungsansatz beruht auf einem probabilistischen Verfahren, das neben der nur in einem Teil der Einsatzprotokolle verfügbaren KVNR indirekte Schlüsselvariablen berücksichtigt. Dabei handelt es sich um die Kassenzugehörigkeit der Patient*innen, ihr Geschlecht und Geburtsjahr sowie die zurückgelegte Wegstrecke des eingesetzten Rettungsmittels. Zur Prüfung sowohl der deterministischen als auch probabilistischen Verknüpfung werden Übereinstimmungsquoten verschiedener, in beiden Datenquellen vorhandener Variablen ermittelt.
Ergebnisse Ausgangspunkt für die Datenverknüpfungen stellen 106 371 Rettungsdienst-Einsatzprotokolle (unabhängig von der Krankenkasse) sowie 432 693 bei Krankenkassen abgerechnete Leistungen (unabhängig vom Rettungsdienstanbieter) aus dem Bereich der Notfallrettung dar. Von 5921 Einsatzprotokollen, in denen eine an Inno_RD datenliefernde Krankenkasse dokumentiert ist, können 4327 Einsatzprotolle deterministisch unter Rückgriff auf die KVNR mit den Abrechnungsdaten zusammengeführt werden. Im Rahmen der probabilistischen Verknüpfung lässt sich diese Anzahl auf 5379 Einsatzprotolle steigern. Alle Prüfungen deuten sowohl für den deterministischen als auch den probabilistischen Ansatz auf eine hohe Verknüpfungsqualität hin.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Schlussfolgerung Eine personen- und einsatzbezogene Verknüpfung von Einsatzprotokollen des Rettungsdienstes mit Abrechnungsdaten ist möglich. Als Alternative zur deterministischen Verknüpfung anhand der KVNR hat sich in Inno_RD ein probabilistischer Ansatz bewährt, auf dessen Grundlage Einsatzprotokolle auch dann sinnvoll verknüpft werden können, wenn KVNR nicht vorliegen sowie indirekte Schlüsselvariablen im geringen Umfang Nicht-Übereinstimmungen oder Missing-Values aufweisen.

ABSTRACT

Introduction In Germany, Emergency Medical Services (EMS) were involved in a total of 7.3 million emergency cases in 2016/2017. Information on prehospital care is stored in several secondary data sources, yet combined analysis of these data at the level of individual patients or EMS cases happens rarely. Research is needed on which methods and variables are suitable for the linkage of these data sources.

Methods We linked EMS records from five Bavarian emergency service districts to health claims data belonging to ten statutory health insurers (data from 2016). Two linkage approaches at the level of individual patient's EMS case/reimbursement case were demonstrated. First, a deterministic linkage was conducted based on the patient's unique identifying health insurance number. The second linkage was probabilistic. As linkage variables, it comprised the only partially available health insurance number plus several non-unique key variables, the latter being a patient's health insurance provider, sex, year of birth and distance travelled. In order to verify the deterministic and the probabilistic linkages' quality, rates of accordance of several variables present in both data sources were calculated.

Results The starting point for our data linkage were 106,371 EMS records (independent of certain health insurance companies) and 432,693 EMS services reimbursed by health insurers (independent of specific EMS providers). 4,327 EMS records could be linked to health claims data – out of 5,921 EMS records that coded a health insurance company contributing claims data to Inno_RD. With a probabilistic linkage, it was possible to increase this number to a total of 5,379 linked EMS records. All checks carried out indicated a high linkage quality for both the deterministic and the probabilistic approach.

Conclusion A linkage of EMS records with health claims data is possible. In Inno_RD, a probabilistic approach has proven a valuable alternative to deterministic linkage via health insurance number since EMS records can be linked meaningfully even if the health insurance number is unavailable or where a minority of non-unique key variables show non-accordance or missing values.

Einleitung

Forschung und Qualitätssicherung zur Versorgung von Rettungsdienstpatient*innen beziehen sich häufig auf kleine regionale Einheiten (z. B. einzelne Rettungsdienstbereiche), einzelne Erkrankungsbilder oder für die Bewertung der Patient*innenversorgung nur bedingt aussagekräftige Einsatzzeiten und -intervalle der beteiligten Einsatzkräfte. Um Empfehlungen für die Optimierung der präklinischen Notfallversorgung ableiten zu können, benötigt es jedoch überregionale und umfassendere Analysen, die über den Zeitraum des einzelnen Rettungsdiensteinsatzes hinausgehen, zumal wesentliche Outcomes von Patient*innen oft erst nach Einsatzende sichtbar werden (vgl. [2]).

Im vom Innovationsfonds geförderten Projekt „Integrierte Notfallversorgung: Rettungsdienst im Fokus" (kurz: „Inno_RD"; FKZ 01VSF17032) werden über Sektorengrenzen hinweg Versorgungsabläufe untersucht, an welchen der Rettungsdienst beteiligt ist. Hierfür sind Sekundärdatenquellen notwendig, die den gesamten Einsatz abbilden, inklusive der vor- und nachlaufenden patient*innenbezogenen Informationen. In Inno_RD liegen dazu unter anderem Einsatzprotokolle von nicht-ärztlichem Rettungsdienstpersonal sowie Notärzt*innen vor. Komplementiert werden diese durch Abrechnungsdaten gesetzlicher Krankenkassen.

Potenziell ergibt sich durch die Sekundärdatenverknüpfung gegenüber alleinig auf Kassendaten gestützten Analysen, die zu Rettungsdiensteinsätzen im Wesentlichen nur abrechnungsrelevante Informationen enthalten, ein erheblicher Informationszugewinn gerade in Hinblick auf Inhalte zum Notfallgeschehen (z. B. gesundheitliche Beschwerden) oder der Versorgung (z. B. Medikamentengabe) in der Präklinik.

Erforderlich ist für die Verknüpfung von Rettungsdienst-Einsatzprotokollen und Abrechnungsdaten allerdings ein Algorithmus, welcher neben einer möglichst hohen Quote verknüpfter Fälle auch eine hohe Qualität der Verknüpfung gewährleistet. Beschrieben und verglichen werden im vorliegenden Beitrag dazu 2 prototypische Vorgehen zur versichertenbezogenen Verknüpfung von Einsatzprotokollen nicht-ärztlichen Rettungsdienstpersonals mit Abrechnungsdaten gesetzlicher Krankenversicherungen. Neben einer deterministischen Verknüpfung unter Nutzung der Krankenversichertennummer (KVNR) wird ein probabilistischer Ansatz herausgearbeitet, der eine Zusammenführung auch dann ermöglicht, wenn dieser Personenidentifikator nicht in allen Einsatzprotokollen vorliegt.

Charakterisierung von Rettungsdienst-Einsatzprotokollen und Abrechnungsdaten gesetzlicher Krankenkassen

An primären Notfalleinsätzen des Rettungsdienstes, d. h. der Versorgung von Patient*innen am Einsatzort sowie ihrem zumeist erfolgenden Transport in eine zur Weiterversorgung geeignete Einrichtung, sind verschiedene Akteure beteiligt. In Folge eines von einer Leitstelle aufgefassten Ereignisses (z. B. ein eingehender Notruf) werden benötigte Ressourcen, insbesondere Transportmittel und Einsatzkräfte, identifiziert und mit der Einsatzdurchführung beauftragt. Die Disposition kann ausschließlich mit nicht-ärztlichem Fachpersonal besetzte Rettungsmittel wie Rettungswagen (RTW) umfassen,

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

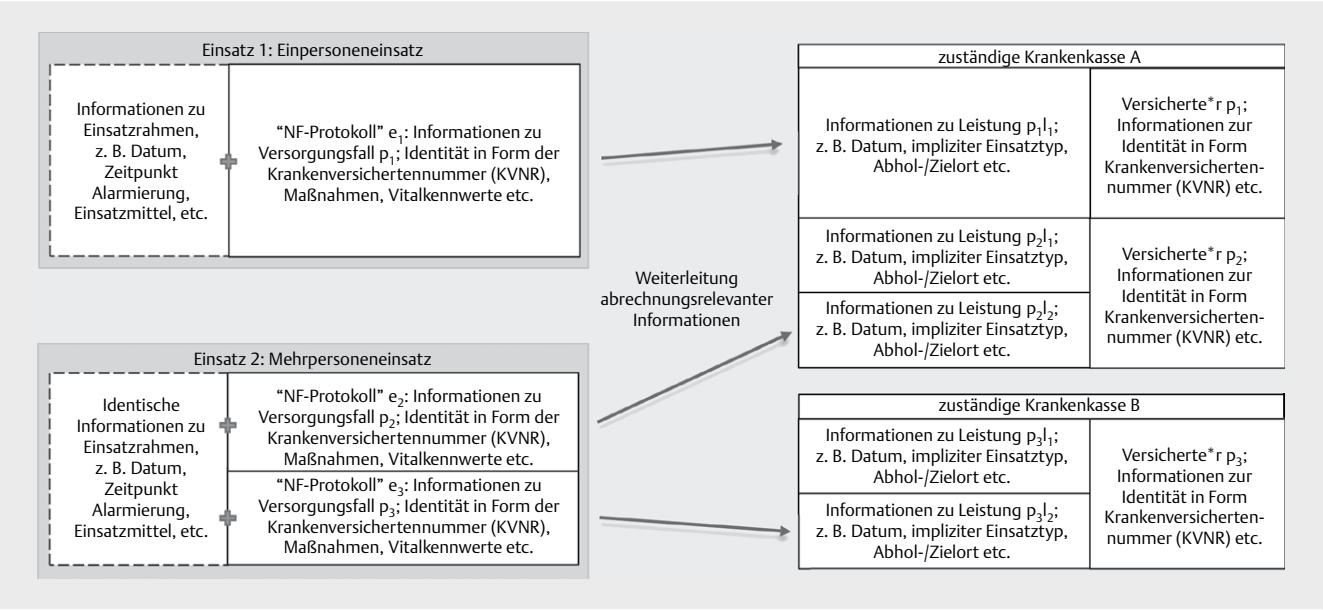
oder im Falle notwendiger ärztlicher Hilfeleistungen Notarzteinsatzfahrzeuge (NEF), Notarztwagen (NAW) oder Rettungshubschrauber (RTH). Mehrere Einsatzmittel können bei einem Notfallereignis zusammenwirken, z. B. patient*innentransportierende RTW sowie NEF, die Notärzt*innen zum Einsatzort bringen. (vgl. [3, 4])

Verpflichtender Bestandteil von Rettungsdiensteinsätzen ist ihre adäquate Dokumentation. In Protokollform werden dabei wichtige Sachverhalte in Hinblick auf die Einsatzdurchführung, Bedarfe zur Weiterbehandlung, Abrechnung und Qualitätssicherung festgehalten. Ein für ganz Deutschland rechtlich bindender Standard in Hinblick auf Inhalte oder Form der Einsatzdokumentation existiert gegenwärtig nicht. Die Fachgesellschaft „Deutsche Interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin" (DIVI) veröffentlichte allerdings den „Minimalen Notfalldatensatz" (MIND) [5, 6], der Mindestanforderungen an eine präklinische Notfall-Dokumentation beschreibt. Dabei wurden inhaltlich relevante Merkmale und Merkmalsbeschreibungen sowie Art und Eigenschaften der zu ihrer Erfassung erforderlichen Datenfelder definiert. Vielerorts werden auf dem MIND basierende Einsatzdokumentationen mit starken Modifikationen genutzt. Charakteristische Inhalte umfassen dabei neben personenidentifizierenden Patient*innendaten (wie Name, Anschrift, Krankenversichertennummer) bspw. Befunde, durchgeführte Assessments und Maßnahmen (inklusive verabreichter Medikamente) sowie Rahmeninformationen zum Einsatz selbst (z. B. Zeitpunkte, genauer Einsatztyp).

In ▶ **Abb. 1** sind Einsatzprotokolle aus struktureller Sicht vereinfacht dargestellt. Im Regelfall spiegelt sich in einem Einsatzprotokoll die medizinische Versorgung und/oder der Transport einer Person wider, siehe dazu „Einsatz 1" mit der Person p_1. Im Folgenden werden entsprechende Einsatzprotokolle als „NF-Protokolle" bezeichnet (Notation: e). Für identische Patient*innen p können dabei mehrere NF-Protokolle vorliegen, z. B. wenn im Rahmen von voneinander unabhängigen Notfallanlässen an unterschiedlichen Tagen ein Kontakt mit dem Rettungsdienst stattfindet. Mitunter erfordern allerdings Rettungsdiensteinsätze auch die Versorgung mehrerer Personen (1 Einsatz: n Patient*innen), was für den Einsatz entsprechend viele NF-Protokolle impliziert (siehe ▶ **Abb. 1**, „Einsatz 2" mit Patient*in p_2 und p_3). Im vorliegenden Paper werden Informationen zu einzelnen Patient*innen auch bei Mehrpersoneneinsätzen jeweils als unabhängige NF-Protokolle betrachtet. In den entsprechenden NF-Protokollen können jedoch Rahmeninformationen zum Einsatz, etwa der Zeitpunkt der Alarmierung oder das individuelle Einsatzmittel, identisch sein.

Auf Grundlage des § 60 SGB V müssen versorgte und/oder transportierte Patient*innen für ihren Rettungsdiensteinsatz in den meisten Fällen nur einen kleinen Eigenbeitrag in der Höhe von 10 Euro leisten. Die Rettungsdienste rechnen den Großteil der Einsatzkosten direkt mit den Krankenkassen der Patient*innen ab. Die Krankenkassen erhalten dazu von den Rettungsdienstorganisationen, gegebenenfalls vermittelt über intermediäre Stellen wie die „Zentrale Abrechnungsstelle für den Rettungsdienst in Bayern" (ZAST), grobe Informationen zu Einsätzen sowie versorgten Patient*innen. Innerhalb dieser Abrechnungsdaten werden allerdings keine medizinischen Daten übermittelt (d. h. bspw. keine Angaben zu Maßnahmen oder Verdachtsdiagnosen). Die Kassen erfahren neben personenidentifizieren Informationen zu den Patient*innen lediglich einsatztechnische Merkmale wie Einsatztag, Einsatzzeiten, Rettungsmitteltyp, Einsatz- und Zielort. Die entsprechenden Informationen können direkt den Rettungsdienst-Einsatzprotokollen entstammen.

In den Abrechnungsdaten gesetzlicher Krankenkassen ist die Notfallrettung – gemeinsam mit Krankentransporten und Krankenfahrten – unter dem Stichwort „Krankentransportleistungen" im § 302 SGB V für „Sonstige Leistungserbringer" verortet. Strukturell sind die Abrechnungsdaten durch eine Hierarchie gekennzeichnet, bei der pro versicherter Person (Notation: p) eine oder



▶ **Abb. 1** Datenstruktur. Schematisierte Einsatzprotokolle und Abrechnungsdaten der Krankenkassen (vereinfacht). Legende: p_1...p_3: Patient*innen bzw. versicherte Personen; e_1...e_3: NF-Protokolle für Versorgungsfälle p_1...p_3; p_1l_1...p_3l_2: Leistungen in den Abrechnungsdaten.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

mehrere Leistungen (Notation: l) dokumentiert sein können. Exemplarisch weist dazu in ▶ **Abb. 1** (rechte Seite) die/der versicherte Person p_1 eine einzige Leistung auf, während für die Personen p_2 mit p_2l_1 und p_2l_2 sowie p_3 mit p_3l_1 und p_3l_2 jeweils 2 Leistungen vorliegen. Liegen für identische Personen mehrere Leistungen vor, so kann es sich dabei bspw. um die Abrechnung voneinander unabhängiger Rettungsdiensteinsätze handeln, welche zu unterschiedlichen Tagen stattfanden. Entsprechend sind diese durch ein unterschiedliches Leistungsdatum gekennzeichnet. Taggleich vorliegende Leistungen werden auf Ebene identischer Personen hingegen typischerweise durch mehrere, an einem Notfallereignis beteiligte Einsatzmittel (z. B. Einsatz eines RTW sowie eines NEF) bedingt. Ursache können jedoch auch hier mehrere, voneinander völlig unabhängige Versorgungsanlässe am gleichen Tag sein (z. B. ein Einsatz morgens, ein zweiter Einsatz abends). In den vorliegenden Abrechnungsdaten geht die Zusammengehörigkeit personenbezogener, taggleicher Leistungen zu einem Einsatz allerdings nicht immer eindeutig hervor. Der Umgang mit dieser Datenstruktur wird daher jeweils in den folgenden Abschnitten zur Datenverknüpfung näher beschrieben.

Selektion und Aufbereitung der Einsatzprotokolle

In Inno_RD liegen aus 5 Rettungsdienstbereichen in Bayern sowie einem Rettungsdienstbereich in Baden-Württemberg NF-Protokolle von Notfallsanitäter*innen oder sonstigem nicht-ärztlichem Rettungsdienstpersonal vor. NF-Protokolle aus Bayern umfassen dabei den Gesamtjahreszeitraum 2016. Das bayerische Rettungsdienstgesetz sieht in Artikel 46 (3) eine landesweit nach einheitlichen Grundsätzen geartete Dokumentation explizit vor, sodass alle aus dem Bundesland vorhandenen NF-Protokolle einem einheitlichen Dokumentationsstandard folgen. Die Protokollführung erfolgte vollständig mithilfe von mobilen elektronischen Informationssystemen (vgl. [7]). Sofern an der Versorgung von Patient*innen Notfallsanitäter*innen und Notärzt*innen gemeinsam beteiligt sind, dokumentieren in Bayern beide Berufsgruppen in getrennten, eigenständigen Protokollen. Für nicht-ärztliche Einsatzkräfte in Baden-Württemberg befand sich die elektronische Dokumentationsweise 2016 noch im Aufbau, wodurch die in Inno_RD vorliegenden Daten auf das letzte Quartal des Jahres beschränkt sind. Der dabei verwendete Dokumentationsstandard zeigt zudem im Vergleich zu Bayern große Unterschiede. Um die Komplexität zu reduzieren, wird im vorliegenden Beitrag ausschließlich die Verknüpfung von NF-Protokollen aus bayerischen Rettungsdienstbereichen dargestellt.

Das Aufgabengebiet des Rettungsdienstes ist nicht auf die Notfallrettung beschränkt. Krankentransporte, die einer ärztlichen Verordnung bedürfen, umfassen die Beförderungen von nicht akut verletzten oder erkrankten Personen durch qualifiziertes medizinisches Personal (vgl. [4, 8]). Sekundärtransporte zwischen Behandlungseinrichtungen können mit oder ohne dringende medizinische Indikation erfolgen; eine Erstversorgung von Patient*innen am Einsatzort findet typischerweise jedoch nicht statt (vgl. [9]). Es existieren in Bayern Vorgaben, wonach Notfalleinsatzprotokolle nur im Rahmen von Einsätzen der Notfallrettung ausgefüllt werden sollen. Jedoch kann nicht gänzlich ausgeschlossen werden, dass sich im Datenbestand NF-Protokolle befinden, welche Kranken- oder Sekundärtransporte widerspiegeln. Daher erfolgt für die insgesamt n_e = 150 048 NF-Protokolle aus Bayern zunächst anhand der dokumentierten Einsatztypen eine Eingrenzung auf Einsätze der Notfallrettung (siehe ▶ **Abb. 2**, „NF-Protokolle"). Aus den verbliebenen n_e = 142 342 NF-Protokollen werden zudem Fehleinsätze sowie bei Krankenkassen nicht abrechnungsfähige Hilfeleistungen ausgeschlossen (insgesamt n_e = 35 971 NF-Protokolle). Dabei handelt es sich bspw. um Einsätze, bei denen noch im Rahmen der Anfahrt ein Einsatzabbruch erfolgt oder sich am Einsatzort ein Transport bzw. eine Behandlung als nicht notwendig, unerwünscht oder unmöglich herausstellt (vgl. [8, 10]). Es verbleiben n_e = 106 371 NF-Protokolle von abrechnungsfähigen Einsätzen der Notfallrettung. Angaben in den NF-Protokollen zum Geschlecht der Patient*innen umfassen zu 49,9 % (n_e = 46 073) Männer, das Durchschnittsalter liegt bei 59,4 Jahren (siehe ▶ **Tab. 3 im Anhang**).

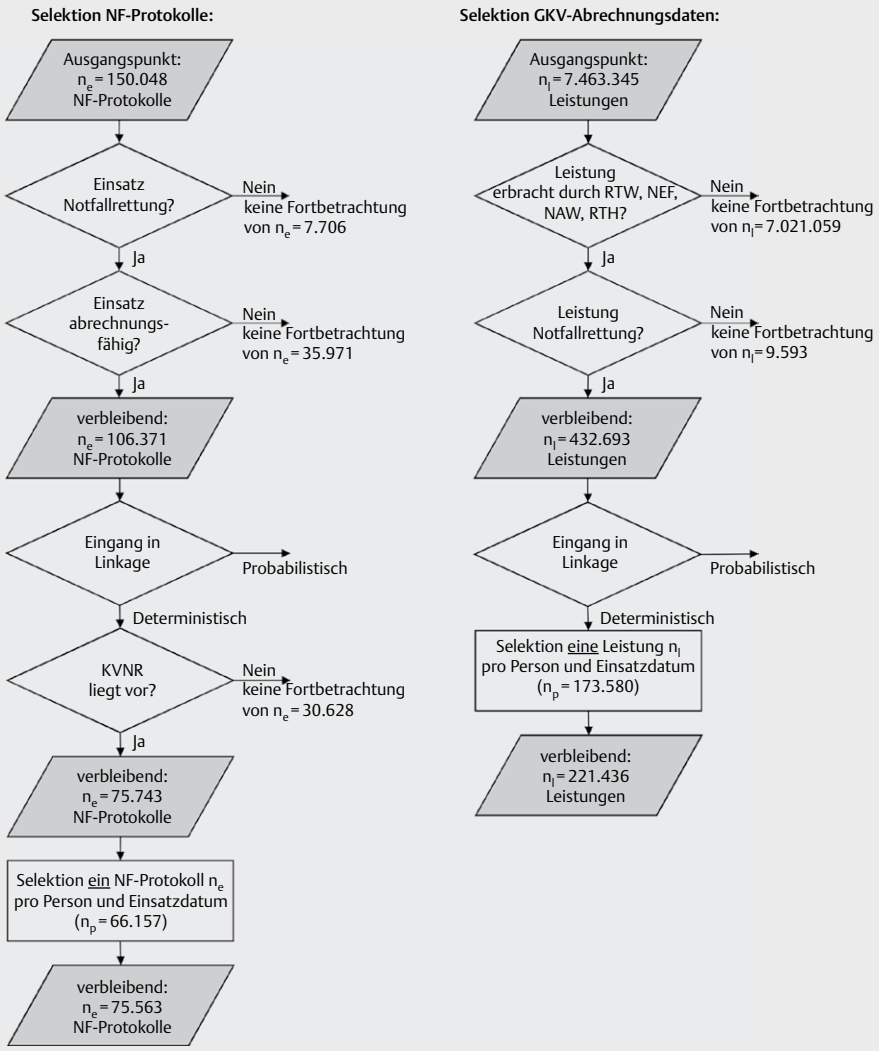
Selektion und Aufbereitung der Abrechnungsdaten

Von n_p = 462 475 bei 10 unterschiedlichen Betriebskrankenkassen (BKK) versicherten Personen liegen insgesamt n_l = 7 463 345 abgerechnete Leistungen aus dem Jahr 2016 vor. Leistungen, welche sich speziell auf den Einsatztyp Notfallrettung beziehen, sind nur anhand der sechsstelligen Abrechnungspositionsnummern für Krankentransportleistungen (vgl. [11]) implizit abschätzbar, welche Angaben zum verwendeten Einsatzmittel enthalten. Im Sinne der G-BA Richtlinie (vgl. [12]) deutet bspw. die Taxi-Nutzung auf eine Krankenfahrt hin, der alleinige Einsatz eines Krankentransportwagens auf einen Krankentransport und die Nutzung eines Rettungswagens auf einen Einsatz zur Notfallrettung (bzw. Rettungsfahrt). Anhand der Abrechnungspositionsnummer werden daher Leistungen selektiert, welche von den für die Notfallrettung charakteristischen Einsatzmitteln RTW, NAW/NEF oder RTH erbracht wurden (Ausschluss: n_l = 7 021 059). Weiter erfolgt ein Ausschluss reiner Sekundärtransportleistungen anhand der Abrechnungspositionsnummer (Ausschluss: n_l = 9593). Eine räumliche Eingrenzung auf Leistungen oder Versicherte aus den Modellregionen erfolgt hingegen nicht. Im Ergebnis liegen Informationen zu n_p = 173 580 Versicherten vor, für die im Jahr 2016 n_l = 432 693 Leistungen abgerechnet wurden (siehe ▶ **Abb. 2**, „GKV-Abrechnungsdaten"). Im Schnitt sind die Versicherten 55,4 Jahre alt; 52,3 % (n_p = 90 697) sind Männer (siehe ▶ **Tab. 3 im Anhang**).

Verknüpfung von Einsatzprotokollen und Abrechnungsdaten

In den NF-Protokollen- und den Krankenkassendaten ist das gleiche personenidentifizierende Kriterium in Form der Krankenversichertennummer (KVNR) vorgesehen (siehe dazu auch ▶ **Abb. 1**): Die KVNR ist auf Personenebene eineindeutig, im Regelfall lebenslang gültig und in nur sehr seltenen Ausnahmefällen gesetzlich geregelten Änderungen unterworfen (vgl. [13]). Aufgrund ihrer datenschutzrechtlichen Stellung und den Vorschriften der Sozialgesetzgebung (SGB X § 67ff.) ist ihre Verarbeitung für Forschungszwecke im Klartext allerdings erschwert. In Inno_RD werden daher mithilfe

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.



▶ **Abb. 2** Datenselektion. Selektion NF-Protokolle (linke Seite) sowie GKV-Abrechnungsdaten (rechte Seite). Legende: n_e: Anzahl NF-Protokolle; n_p: Anzahl Personen; n_l: Anzahl Leistungen.

einer von der Projekt-Treuhandstelle entwickelten Software die in den Rettungsdienst- und Kassendaten vorliegenden realen KVNR durch numerische Pseudonyme ersetzt (vgl. [14]). Aufgrund softwareinterner Plausibilitätsprüfungen erfolgt der Pseudonymisierungsprozess nur, wenn eine formal gültige KVNR vorliegt. Die entsprechend erzeugten Pseudonyme werden daher im Folgenden (vereinfachend) als „vorliegende KVNR" (o. Ä.) bezeichnet. Wird hingegen durch die Software in den Originaldaten keine KVNR festgestellt, so wird keine Pseudonymisierung durchgeführt. In diesen Fällen wird von einer „nicht vorliegenden KVNR" (o. Ä.) gesprochen. Andere im Einzelnen oder in ihrer Kombination personenidentifizierende Merkmale wie Name, vollständige Anschrift oder vollständiges Geburtsdatum werden unter Einhaltung des Datenschutzes nicht verarbeitet bzw. nicht von den Dateneignern übermittelt.

Abrechnungsinformationen lassen sich nur mit denjenigen NF-Protokollen sinnvoll zusammenführen, in denen eine Patient*innen-Zugehörigkeit zu einer der 10 an Inno_RD datenliefernden Krankenkassen vermerkt ist. Die Anzahl an NF-Protokollen, in denen eine entsprechende Projekt-Kasse dokumentiert ist, kann als Richtwert für die maximale Anzahl verknüpfbarer NF-Protokolle interpretiert werden. Daher erfolgt eine Abschätzung anhand der durch die Einsatzkräfte notierten Kassennamen bzw. Institutionskennzeichen. Erfasste Kassen, für welche bis zum Jahr 2016 eine Fusionierung mit einer an Inno_RD beteiligten Kasse erfolgte, wurden letzteren jeweils zugeordnet. Laut Eintragungen des Rettungsdienstes deuten n_e = 5921 NF-Protokolle darauf hin, dass es einen entsprechenden Abrechnungsfall bei den an Inno_RD datenliefernden Krankenkassen geben sollte. Eine Selektion dieser Fälle erfolgt im Vorfeld der Datenverknüpfung allerdings bewusst nicht, da von den Einsatzkräften manuell erfasste Kassenzugehörigkeiten Fehler aufweisen könnten, z. B. durch falsche Selbstangaben der Patient*innen. Stattdessen werden alle NF-Protokolle, unabhängig von dem notierten Kassennamen, fortbetrachtet.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Deterministische Verknüpfung – Methode

Auf Grundlage der pseudonymisierten KVNR werden im Folgenden Kassendaten und NF-Protokolle deterministisch verknüpft. Um dabei an unterschiedlichen Tagen erfolgte Rettungsdienstkontakte identischer Patient*innen unterscheiden zu können, wird zudem das Datum der Rettungsdienstleistung als Schlüsselvariable einbezogen. Ausgangslage bilden die insgesamt n_e = 106 371 auf Einsätze der Notfallrettung eingeschränkten NF-Protokolle. Von diesen liegt in 71,2 % (n_e = 75 743) der Fälle eine KVNR vor (siehe ▶**Abb. 2**; Informationen zu Alter (basierend auf Geburtsjahr) und Geschlecht in ▶**Tab. 3 im Anhang**). Verschiedene Gründe sind für diese Füllquote möglicherweise ausschlaggebend, z. B. privat oder im Ausland versicherte Patient*innen, welche schlichtweg keine KVNR besitzen. Möglich ist jedoch auch, dass existierende KVNR im Verlauf der Notfallrettung nicht erfasst werden konnten, bspw. wenn eine Gesundheitskarte am Notfallort nicht vorlag oder nicht ausgelesen werden konnte. Von den n_e = 5921 NF-Protokollen, welche aufgrund des Kassennamens potenziell verknüpfbar mit den vorhanden Krankenkassendaten scheinen, weisen nur 79,4 % (n_e = 4701) eine KVNR auf. Die deterministische Verknüpfung sollte näherungsweise diese Anzahl zusammenführen.

Da anhand der herangezogenen Schlüsselvariablen eine Unterscheidung von taggleichen NF-Protokollen/taggleichen Kassendaten-Leistungen identischer Personen nicht möglich ist, wird jeweils pro Person und Einsatzdatum nur ein NF-Protokoll/eine Leistung für die Verknüpfung herangezogen (Linkage 1:1). Eine entsprechende Selektion ist in beiden Datenquellen erforderlich. In den Kassendaten werden dazu Leistungen priorisiert, welche sich anhand der Positionsnummer auf RTW oder NAW zurückführen lassen, da insbesondere bei diesen von der Anlage eines NF-Protokolls auszugehen ist. Liegen diese nicht vor, erfolgt die Auswahl willkürlich. Es verbleiben im Anschluss n_l = 221 436 Leistungen in den Kassendaten. 75 563 NF-Protokolle liegen nach der Reduktion auf ein Protokoll pro Einsatzdatum und Patient*in vor, wobei als Kriterium jeweils der früheste Alarmierungszeitpunkt fungiert.

Deterministische Verknüpfung – Ergebnisse

In ▶**Tab. 1** sind die Ergebnisse des deterministischen Linkages dargestellt. Insgesamt n_e = 4327 NF-Protokolle werden mit Abrechnungsfällen der Krankenkassendaten zusammengeführt. Anteilig an den NF-Protokollen mit abgeschätzter Zugehörigkeit zu einer der 10 Projekt-Krankenkassen sowie gültig vorliegendem KVNR-Pseudonym können damit 92,0 % der Datenmenge verknüpft werden; bezogen auf alle NF-Protokolle mit entsprechend dokumentierten Krankenkassen (d. h. jene NF-Protokolle mit und ohne KVNR) können 73,0 % verknüpft werden. Im verknüpften Datensatz sind durch Auszählen des Personenidentifikators KVNR n_p = 3810 zugrundeliegende Personen mit einem Durchschnittsalter von 54,7 Jahren und einem Männeranteil von 53,6 % feststellbar (▶**Tab. 3 im Anhang**). 90,2 % dieser Personen (n_p = 3435) wurde genau ein NF-Protokoll aus 2016 zugespielt, das Maximum pro Person hingegen liegt bei 9 NF-Protokollen (bzw. Einsatztagen) pro Jahr.

Eine Plausibilisierung der deterministischen Verknüpfung erfolgt durch einen genaueren Blick auf ▶**Tab. 1** (Spalte „Deterministisch"). Verschiedene, jeweils in beiden Datenquellen vorhandene Variablen werden dabei in Hinblick auf (Nicht-)Übereinstimmung untersucht. Geschlecht, Geburtsjahr sowie die nominale Kassenzugehörigkeit weisen eine hohe Übereinstimmung auf, was auf eine hohe Güte der Verknüpfung schließen lässt. Die im geringen Umfang vorhandenen Nicht-Übereinstimmungen können dabei möglicherweise darauf zurückgeführt werden, dass es sich bei Alter und Geschlecht um in beiden Sekundärdatenquellen getrennt erhobene Merkmale handelt. Während die Informationen in den NF-Protokollen von den Einsatzkräften nach Augenschein, Patientenangabe oder durch Einlesen der Gesundheitskarte erfasst wurden, handelt es sich in den Abrechnungsdaten um zugespielte Informationen aus dem Stammdatenbestand der Krankenkassen (siehe ▶**Tab. 4 im Anhang**). Gut 2 % (n_e = 96) Nicht-Übereinstimmungen zeigen sich zudem in Hinblick auf die von den Einsatzkräften notierten Fahrstrecken der Rettungsmittel. Vor dem Hintergrund, dass anhand der Schlüsselvariablen pro Patient*in und Einsatztag nur die Verknüpfung eines NF-Protokolls sowie einer Leistung erfolgt, tatsächlich allerdings mehrere NF-Protokolle wie auch abgerechnete Leistungen existieren können, sind Nicht-Übereinstimmungen hier leicht erklärbar. Insgesamt gelingt mittels des deterministischen Verfahrens auf Grundlage der personengebundenen KVNR und des Einsatzdatums offenbar für einen Großteil der Fälle die Verknüpfung tatsächlich zusammengehöriger NF-Protokolle und abgerechneter Leistungen.

Probabilistische Verknüpfung – Methode

Es liegt nahe, dass Rettungsdienste Informationen über versorgte Patient*innen an die kostenerstattenden Krankenkassen auch dann weiterleiten können, wenn KVNR nicht vorliegen, bspw. anhand anderer personenidentifizierender Informationen wie Vorname, Nachname oder vollständiges Geburtsdatum. Im Inno_RD-Datenbestand dürften sich damit auch NF-Protokolle befinden, welche im deterministischen Linkage allein aufgrund des Fehlens des entsprechenden Pseudonyms keinem Abrechnungsvorgang zugeordnet werden können. Neben einem nicht ausgeschöpften Fallzahlpotenzial sind damit Selektionseffekte denkbar, wenn bspw. durch das deterministische Verfahren vermehrt Einsätze mit spezifischer Soziodemografie, Notfallanlässen oder Outcomes unverknüpft bleiben. Verwiesen sei dazu auf den Umstand, dass sich in NF-Protokollen ohne gültigen Personenidentifikator KVNR ein um knapp 10 Jahre niedrigeres mittleres Lebensalter sowie ein um nahezu 10 Prozentpunkte höherer Anteil von Männern zeigt (siehe ▶**Tab. 3 im Anhang**).

NF-Protokolle ohne vorliegende KVNR sollen daher in den Algorithmus integriert und simultan mit NF-Protokollen, welche den Personenidentifikator aufweisen, verknüpft werden. Die KVNR soll dabei als direkte Schlüsselvariable erhalten bleiben und durch indirekte Schlüsselvariablen ergänzt werden. Letztere verweisen zwar nicht auf explizite Personen, es lässt sich aber insbesondere durch Merkmalskombinationen von einer Verknüpfung zusammengehöriger Einheiten ausgehen (vgl. [15, 16]). Vor dem Hintergrund der hohen Übereinstimmungen im Rahmen der deterministischen Verknüpfung (▶**Tab. 1**) eignen sich als Schlüsselvariablen Geschlecht, Geburtsjahr, nominelle Kassenzugehörigkeit sowie die Fahrstrecke des Rettungsmittels. Speziell die Variablen Geschlecht und Geburtsjahr (respektive Alter) finden sich als Schlüssel in mehreren Studien aus

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

dem englischsprachigen Raum, in denen Rettungsdienst-Einsatzdokumentation deterministisch oder probabilistisch mit verschiedenartigen klinischen Daten verknüpft werden (vgl. [17–22]). Zudem erfolgt die Wahl der indirekten Schlüsselvariablen unter Berücksichtigung ihrer jeweiligen Füllquoten (siehe ▶**Tab. 4 im Anhang**): Für die nominelle Kassenzugehörigkeit sowie das Geburtsjahr ergibt sich in den NF-Protokollen eine Füllquote von 99,9 %, für das Geschlecht 86,8 %. Angaben zur Kilometerleistung des Einsatzmittels liegen in 100,0 % der NF-Protokolle vor.

Für eine simultane Verknüpfung von NF-Protokollen mit vorliegender bzw. nicht vorliegender KVNR ist ein fehlertolerantes Verfahren erforderlich (vgl. [15, 16]): Ist die KVNR in NF-Protokollen und Kassendaten identisch, so muss die Übereinstimmung dieser Schlüsselvariable durch den Algorithmus erkannt werden. Ist allerdings in NF-Protokollen keine KVNR vorhanden, darf dies nicht automatisch die Nicht-Verknüpfung mit den Kassendaten nach sich ziehen. Für die Wahl eines fehlertoleranten Linkageverfahrens spricht außerdem, dass eine Verknüpfung auch dann erfolgen kann, wenn – wie im deterministischen Linkage festgestellt – die indirekten Schlüsselvariablen im gewissen Umfang widersprüchliche Inhalte oder Missing-Values aufweisen.

Die Umsetzung der Datenverknüpfung erfolgt mithilfe des Paketes „reclin“ [23] für die Software RStudio, welches auf die Ende der 1960er Jahre von Fellegi und Sunter entwickelte Idee des probabilistischen Linkages (vgl. [24]) zurückgreift. Im verwendeten Paket werden anhand von Vergleichsfunktionen Wahrscheinlichkeiten für die Übereinstimmung der Schlüsselvariablen unter der Annahme geschätzt, dass diesen identische oder nicht-identische Einheiten zugrunde liegen. Auf Grundlage eines für jede mögliche Verknüpfung abgeleiteten Gewichtes können dann hinreichend schlüssige Linkages identifiziert werden. Dabei erhöht sich mit ansteigendem Gewicht die Wahrscheinlichkeit für die tatsächliche Zusammengehörigkeit. Der Ansatz der Gewichtung beruht auf dem Umstand, dass für Schlüsselvariablen mit nur wenigen Ausprägungen (z. B. Geschlecht) die Wahrscheinlichkeit einer rein zufälligen Übereinstimmung höher ist als im Fall einer Schlüsselvariable, die vielfältige Ausprägungen aufweist (wie die KVNR oder die Fahrtstrecke in km). Stimmen pseudonymisierte KVNR und Fahrtstrecke überein, so ist als Konsequenz mit einem höheren Gewicht zu rechnen als bei Fällen, bei denen nur Geschlecht und Kassenname übereinstimmen.

Ausgangspunkt für das probabilistische Linkage bilden die insgesamt $n_l = 432\,693$ Leistungen (GKV-Daten) sowie $n_e = 106\,371$ NF-Protokolle (▶**Abb. 2**). Für dieses Verfahren ist eine Reduktion auf nur ein NF-Protokoll oder eine abgerechnete Leistung pro Einsatzdatum im Vorfeld der Verknüpfung nicht erforderlich. In Hinblick auf das Einsatzdatum und die ausführende Rettungsdienstorganisation erfolgt ein klassisches Blocking: (vgl. [15, 16]). Die Übereinstimmung von Einsatzdatum und Rettungsdienstorganisation ist damit Voraussetzung für den paarweisen Abgleich der Schlüsselvariablen; eine Nicht-Übereinstimmung schließt eine Verknüpfung aus. Da es sich bei Einsatzdatum sowie ausführender Rettungsdienstorganisation um Informationen handelt, die an die Krankenkassen übermittelt werden und die keine Änderung im Rahmen des Abrechnungsprozesses der Krankenkassen erfahren dürften, wird von einer exakten Entsprechung ausgegangen. Praktisch wird durch das Blocking sichergestellt, dass NF-Protokolle ausschließlich mit Leistungen aus dem Pool der Abrechnungsdaten verknüpft werden können, welche taggleich und vom identischen Leistungserbringer durchgeführt wurden. Eine Verringerung des Risikos falsch-positiver Verknüpfungen ist als Konsequenz zu erwarten. Die für das Blocking erforderlichen Informationen zu Einsatzdatum und ausführender Rettungsdienstorganisation liegen sowohl in NF-Protokollen als auch Abrechnungsdaten zu 100,0 % vor (siehe ▶**Tab. 4 im Anhang**).

Für die indirekten Schlüsselvariablen erfolgt der Vergleich der Wertepaare aus beiden Datenquellen mithilfe eines Abgleichs auf exakte Übereinstimmung. Da speziell die Kilometerleistung des Einsatzmittels die Höhe der durch die Kassen erstatteten Kosten beeinflussen kann, sind im Zuge des Abrechnungsvorgangs bei den Krankenkassen Bearbeitungen des Merkmals denkbar. Um diese bei der Datenverknüpfung zu berücksichtigen, scheint die Verwendung komplexer Funktionen für den Vergleich der Schlüsselvariablen (z. B. Fahrtstrecke ± Toleranzbereich Kilometeranzahl) denkbar. Da allerdings exakte Informationen zu Bearbeitungen in den vorliegenden Anrechnungsdaten nicht enthalten sind, findet auch für die Kilometerleistung ein einfacher Abgleich statt.

Fehlende Angaben bei den Schlüsselvariablen werden als Missing-Values deklariert und gehen als Folge in die Berechnung des Gewichtes nicht ein. Sofern durch das Linkageverfahren identische NF-Protokolle mit mehreren verschiedenen Leistungen aus den Kassendaten gleichzeitig verknüpft werden (und vice versa), wird jeweils nur derjenige Fall mit dem größten errechneten Gewicht fortbetrachtet. Im Anschluss liegen ausschließlich eineindeutige Verknüpfungen vor (insgesamt $n_e = 5660$). Da allerdings noch unplausible Fälle mit einem typischerweise niedrigen Gewicht verbleiben, wird ein Schwellenwert für plausibel akzeptierte Verknüpfungen angesetzt. Bei dessen Wahl ist zu berücksichtigen, dass Verknüpfungen von NF-Protokollen ohne vorliegende KVNR – unter sonst gleichen Umständen – ein per se niedrigeres Gewicht aufweisen als Verknüpfungen, bei denen dieser Personenidentifikator übereinstimmt. Entsprechende Visualisierungen beider Gruppen zeigen die Gewichte unterschiedlich verteilt (siehe ▶**Abb. 3 im Anhang**). Der Schwellenwert wird so gesetzt, dass sich seine Anhebung nicht (mehr) in der verbesserten Übereinstimmung der in beiden Datenquellen vorhandenen indirekten Schlüsselvariablen niederschlägt. Für die Akzeptanz einer Verknüpfung ist ein minimales Gewicht von 4,5 erforderlich. Insgesamt $n_e = 281$ der eineindeutig verknüpften Fälle erreichen dieses Gewicht nicht.

Probabilistische Verknüpfung – Ergebnisse

In ▶**Tab. 1** (Spalte „Probabilistisch“) sind die Ergebnisse der probabilistischen Verknüpfung zusammengefasst. Mit $n_e = 5379$ akzeptierten NF-Protokollen ergibt sich im Vergleich zur deterministischen Variante eine Fallzahlerhöhung um 24,3 %. Dieser Anstieg ist in kleinem Umfang auch darauf zurückzuführen, dass zusätzliche Fälle mit vorliegender KVNR gelinkt werden können: Da im Vorfeld der probabilistischen Verknüpfung keine Reduktion auf ein NF-Protokoll/eine Leistung pro Tag erforderlich ist, steigt die Anzahl der mit (gültig) übereinstimmender KVNR zusammengeführten NF-Protokolle gegenüber der deterministischen Verknüpfung um $n_e = 10$. Insgesamt 90,8 % der auf Projekt-Kassen zurückgeführten 5921 NF-Protokolle können somit verknüpft werden. Die anhand der KVNR in den Kassendaten ausgemachte Anzahl an zugrundeliegenden Personen be-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

▶ **Tab. 1** Ergebnisse der Verknüpfung von NF-Protokollen und GKV-Abrechnungsdaten: Abgleich der Übereinstimmung von Schlüsselvariablen (auf Grund von Missing-Values sind gegenüber der Gesamtzahl der Verknüpfungen Abweichungen möglich).

Indikatoren		**Deterministisch**		**Probabilistisch (sofern vorhanden inklusive KVNR)**	
		gelinkt $n_e = 4327$		**gelinkt und akzeptiert $n_e = 5379$**	
Schlüsselvariablen	**Abgleich**	**n_e**	**Spaltenprozent**	**n_e**	**Spaltenprozent**
Einsatzdatum	stimmt überein	Schlüssel		geblockt	[100,0 %]
	stimmt nicht überein				[0,0 %]
KVNR	stimmt überein	Schlüssel		4337	[80,6 %]
	stimmt nicht überein			20	[0,4 %]
	nicht vorliegend in NF-Daten			1022	[19,0 %]
Kassenname	stimmt überein	4311	[99,6 %]	5361	[99,7 %]
	stimmt nicht überein	16	[0,4 %]	15	[0,3 %]
Geschlecht	stimmt überein	4319	[100,0 %]	4968	[99,4 %]
	stimmt nicht überein	2	[0,0 %]	30	[0,6 %]
Geburtsjahr	stimmt überein	4324	[99,9 %]	5376	[99,9 %]
	stimmt nicht überein	3	[0,1 %]	3	[0,1 %]
Fahrstrecke [km]	stimmt überein	4230	[97,8 %]	5270	[98,0 %]
	stimmt nicht überein	96	[2,2 %]	109	[2,0 %]
Leistungserbringer	stimmt überein	4321	[99,9 %]	geblockt	[100,0 %]
	stimmt nicht überein	6	[0,1 %]		[0,0 %]
Personen n_p		3810		4680	

n_e: Anzahl verknüpfter Fälle (d. h. verknüpfter NF-Protokolle); n_p: Anzahl Personen; Spaltenprozente für (mögliche) Abgleiche in eckigen Klammern.

trägt $n_p = 4680$ (Anstieg um 22,8 %). Stratifiziert nach Vorliegen oder Nicht-Vorliegen der KVNR in den NF-Protokollen zeigt sich für letztere Fälle, wie bereits vor der Verknüpfung, ein niedrigeres Durchschnittsalter (siehe ▶**Tab. 3 im Anhang**).

Im Folgenden ist das Ziel zu überprüfen, ob das probabilistische Linkage auch für Fälle geeignet ist, in denen die KVNR nicht vorliegt. Vertiefend erfolgt daher eine Prüfung der dargestellten probabilistischen Verknüpfung für die Subgruppe derjenigen NF-Protokolle, für welche in beiden Datensätzen KVNR gültig vorliegen ($n_e = 75\,743$; $n_l = 432\,693$). Für diese Subgruppe werden getrennt 2 probabilistische Verknüpfungen durchgeführt: 1.) anhand der KVNR sowie den indirekten Schlüsselvariablen sowie 2.) ausschließlich anhand der indirekten Schlüsselvariablen, d. h. die eigentlich vorhandenen KVNR bleiben für das Linkage unberücksichtigt. Alle weiteren Aspekte der Datenaufbereitung bleiben unverändert. Die jeweils verknüpften Fälle werden allerdings nicht anhand des im vorausgegangenen probabilistischen Linkage genutzten Schwellenwertes für das Gewicht (konkret 4,5) beschnitten, sondern auf dessen Über- bzw. Unterschreitung hin klassifiziert und gemeinsam mit der Information zur (Nicht-)Übereinstimmung des KVNR-Pseudonyms in einer Vier-Felder-Tabelle abgebildet. Die Interpretation erfolgt im Sinne eines binären Klassifikators.

Unter Einbezug der KVNR (siehe ▶**Tab. 2**, Spalte „Linkage inklusive KVNR") fallen insgesamt $n_e = 4357$ verknüpfte Fälle über den für das Gewicht angesetzten Schwellenwert, was exakt der Fallzahl der probabilistischen Verknüpfung in ▶**Tab. 1** entspricht. Davon weisen $n_e = 4337$ eine übereinstimmende KVNR auf; sie werden als richtig-positiv interpretiert. Der positive prädiktive Wert (PPV) [24] liegt damit oberhalb von 0,99. Es verbleiben allerdings $n_e = 20$ akzeptierte Verknüpfungen mit widersprüchlicher KVNR. Vordergründig handelt es sich dabei um falsch-positive Fälle, bei denen unterschiedliche Personen zusammengeführt wurden. Jedoch besteht auch die Möglichkeit, dass – bspw. durch den Rettungsdienst in der Einsatzpraxis – inkorrekt dokumentierte KVNR hier die Nicht-Übereinstimmung der Pseudonyme hervorrufen, tatsächlich allerdings identische Personen zugrunde liegen. Weitere abgeglichene Personenmerkmale, welche für die Verknüpfung nicht berücksichtigt wurden (z. B. Postleitzahl des Wohnorts), deuten darauf hin.

Die Verknüpfung der gleichen Subgruppe anhand der ausschließlich indirekten Schlüsselvariablen, d. h. ohne KVNR, zeigt eine etwas geringere, jedoch nach wie vor hohe Zahl richtig-positiver Verknüpfungen (siehe ▶**Tab. 2**, Spalte „Linkage exklusive KVNR"). Es wird daher davon ausgegangen, dass die indirekten Schlüsselvariablen den Informationsgehalt fehlender KVNR gut ersetzen. Bemerkt sei darüber hinaus die erhöhte Zahl falsch-negativer Verknüpfungen, wenn die KVNR nicht als Schlüssel fungiert ($n_e = 17$): Nach manueller Prüfung handelt es sich hierbei um Fälle, bei denen die im Rahmen des fehlertoleranten Linkages im gewissen Umfang zulässigen Nicht-Übereinstimmungen der indirekten Schlüsselvariablen ein Gewicht unterhalb des Schwellenwertes hervorrufen. Der Einbezug der KVNR als Schlüsselvariable, sofern diese gültig vorliegt, scheint vor diesem Hintergrund auch für das probabilistische Linkage berechtigt.

Schlussfolgerungen und Diskussion

Mit dem in der vorliegenden Arbeit geschilderten Vorgehen wird auf Patient*innen- bzw. Versicherten-Ebene eine belastbare Verknüpfung von Rettungsdienst-Einsatzprotokollen und Abrechnungsdaten der gesetzlichen Krankenversicherung umgesetzt. Durch die kombinierte Betrachtung von Informationen aus beiden Sekundärdatenquellen können sowohl die präklinische Versorgung

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

▶ **Tab. 2** Qualitätsprüfung. Linkage der Abrechnungsfälle (GKV-Daten) und Einsatzprotokolle für die Subgruppe der Fälle, die im Original in beiden Datensätzen eine Krankenversichertennummer (KVNR) aufweisen.

Klassifikation	Probabilistische Verknüpfung für Subgruppe, die im Original eine KVNR enthalten					
	Verknüpfung inklusive KVNR			Verknüpfung exklusive KVNR		
	n_e $KVNR_{NF} = KVNR_{GKV}$	n_e $KVNR_{NF} \neq KVNR_{GKV}$	n_e Gesamt	n_e $KVNR_{NF} = KVNR_{GKV}$	n_e $KVNR_{NF} \neq KVNR_{GKV}$	n_e Gesamt
n_e akzeptierte Verknüpfungen (Gewicht ≥ Schwellenwert)	4337 RP (99,5 %)	20 FP (0,5 %)	4357 (100,0 %)	4319 RP (99,5 %)	22 FP (0,5 %)	4341 (100,0 %)
n_e nicht akzeptierte Verknüpfungen (Gewicht < Schwellenwert)	0 FN (0,0 %)	1091 RN (100,0 %)	1091 (100,0 %)	17 FN (1,5 %)	1091 RN (98,5 %)	1108 (100,0 %)
n_e Gesamt	4337 (79,6 %)	1111 (20,4 %)	5448 (100,0 %)	4336 (79,6 %)	1113 (20,4 %)	5449 (100,0 %)

n_e: Anzahl verknüpfter Fälle (d. h. verknüpfter NF-Protokolle); RP: Richtig-Positiv; FP: Falsch-Positiv; FN: Falsch-Negativ; RN: Richtig-Negativ jeweils bemessen an Übereinstimmung der KVNR; Zeilenprozente in runden Klammern.

von Rettungsdienstpatient*innen als auch deren längerfristige Outcomes im Längsschnitt erforscht werden. Eine systematische prozess- und ergebnisbezogene Evaluation von Rettungsdiensteinsätzen ist damit möglich (siehe exemplarische Umsetzung in Inno_RD [25]).

Auch wenn sich im deutschen Rettungswesen erhebliche Unterschiede, gerade in Hinblick auf die Dokumentation durchgeführter Einsätze zeigen, lässt sich u. a. aufgrund der bundesweit geltenden Abrechnungsvorgaben der GKV von einer grundsätzlichen Übertragbarkeit auf andere Gebiete Deutschlands ausgehen. Zu verweisen ist an dieser Stelle zudem auf den Umstand, dass die in Inno_RD zur Pseudonymisierung der KVNR genutzte Software [26] Open Source verfügbar ist.

Die dargestellte deterministische Verknüpfung von Kassendaten und Rettungsdienst-Einsatzprotokollen auf Grundlage der KVNR führt zu einem hohen Anteil verlinkter Datensätze (73,0 % aller NF-Protokolle, die gemäß der von den Einsatzkräften erfassten Kassennamen zu einer für Inno_RD datenliefernden Krankenkasse gehören). Auch wenn in dieser Veröffentlichung Einsatzprotokolle nicht-ärztlichen Rettungspersonals verknüpft wurden, ist davon auszugehen, dass sich die Einschränkungen und Möglichkeiten auch auf Einsatzprotokolle von Notärzt*innen übertragen lassen (z. B. mittelmäßige Füllquote der KVNR; gute Linkage-Qualität der KVNR und des Einsatzdatums). Es bleibt abzuwarten, ob die zunehmende Verbreitung von mobilen elektronischen Informationssystemen oder Möglichkeiten zum elektronischen Einlesen von Gesundheitskarten zu einer Verbesserung der Datenqualität beitragen. Aus Sicht der Forschung sind alle Maßnahmen begrüßenswert, welche die möglichst häufige Dokumentation der KVNR zum Ziel haben, da somit sowohl die Verknüpfungsquote des vorgestellten deterministischen als auch probabilistischen Verfahrens erhöht werden kann.

Mit hoher Sicherheit konnten auf Grundlage der probabilistischen Verknüpfung 90,8 % aller NF-Protokolle, welche Kassennamen einer datenliefernden Krankenkasse aufweisen, mit Abrechnungsdaten verknüpft werden. Aufgrund methodischer Unterschiede sind Referenzwerte probabilistischer Verknüpfungen aus der Literatur nur bedingt sinnvoll. Die Untersuchungen von Newgard [21] sowie Downing et al. [22] kommen mit 96,1 % (Verknüpfung mit Traumaregisterdaten) bzw. 84,2 % (Verknüpfung mit Notaufnahmedaten) zu vergleichbaren Ergebnissen, wobei allerdings keine expliziten Personenidentifikatoren genutzt wurden.

Die probabilistische Verknüpfung erfolgte zugleich für alle NF-Protokolle, d. h. unter Einschluss von NF-Protokollen mit als auch ohne vorhandener KVNR. Als methodische Alternative dazu besteht die Möglichkeit, für alle NF-Protokolle mit vorliegender KVNR ein deterministisches Linkage anhand dieses Personenidentifikators durchzuführen und den probabilistischen Ansatz ausschließlich auf NF-Protokolle ohne diesen zu beschränken. Erforderlich wäre dabei im Vorfeld der deterministischen Verknüpfung für den entsprechenden Teil der Daten allerdings (wie im Paper beschrieben) wieder die Einschränkung auf ein NF-Protokoll sowie eine abgerechnete Leistung pro Tag und Person. Im Rahmen der probabilistischen Verknüpfung ist diese Beschränkung hingegen nicht notwendig, was die tatsächlichen Datenstrukturen – Patient*innen können zum selben Einsatztag mehrere Rettungsdiensteinsätze aufweisen – besser widerspiegelt. Da zudem alle mit einer übereinstimmenden KVNR probabilistisch zusammengeführten Fälle ohnehin ein Gewicht deutlich über dem für die Akzeptanz erforderlichen Schwellenwert nach sich ziehen (siehe ▶**Abb. 3 im Anhang**), scheint die simultane Verknüpfung von NF-Protokollen mit und ohne vorliegender KVNR zumindest im vorliegenden Fall als praktikablere Lösung.

Durch die probabilistische Verknüpfung wurde zudem eine Kompensation von im gewissen Umfang vorhandenen Nicht-Übereinstimmungen der indirekten Schlüsselvariablen (z. B. widersprüchlichen Geburtsjahren) oder Missing-Values ermöglicht. Grundsätzlich haben sich als indirekte Schlüsselvariablen Geburtsjahr, Geschlecht sowie nominelle Kassenzugehörigkeit bewährt. Speziell letztere erforderte allerdings durch uneinheitliche Benennungen identischer Krankenkassen in den NF-Protokollen einen hohen Aufwand bei der Datenaufbereitung. Mumma et al. [19] berichten von ähnlichen Problemen mit Schlüsselvariablen aus Einsatzdokumentationen des Rettungsdienstes. Voraussetzung für entsprechend geeignete Schlüsselvariablen stellen sowohl eine hohe Füllquote als auch Datenqualität dar. Gerechnet werden kann mit diesen in Sekundärdaten allerdings nicht per se (vgl. [27]).

Bei der manuellen Prüfung wurde eine Reihe von verknüpften Fällen gerade unterhalb des für die Akzeptanz erforderlichen Schwellenwertes identifiziert, welche in den NF-Protokollen durch fehlende Angaben zum Geschlecht gekennzeichnet sind. Höhere

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Füllquoten der Schlüsselvariable würden mutmaßlich ein höheres, akzeptables Gewicht nach sich ziehen. Der angesetzte Schwellenwert ist demnach eher konservativ gewählt mit der Priorität, jene Verknüpfungen zu akzeptieren, bei denen die Fälle mit hoher Wahrscheinlichkeit in der Realität übereinstimmen – anstatt möglichst viele Fälle, darunter auch falsch-positive, zu akzeptieren (Qualität vor Quantität).

In dieser Untersuchung wurden die Krankenkassendaten danach vorselektiert, ob ein Rettungsmittel abgerechnet wurde, das einer „Rettungsfahrt" nach G-BA-Richtlinie [12] entspricht. Aus Gesprächen mit Rettungsdienstmitarbeiter*innen und hier nicht dargestellten Analysen geht hervor, dass es in seltenen Fällen zu unklaren Abgrenzungen zwischen Krankentransporten und Rettungsfahrten kommen kann, d. h. bspw. Einsatzprotokolle aus der Notfallrettung vorliegen, aber ausschließlich KTW abgerechnet wurden. Für die Kassendaten ist daher zukünftig ein Identifikator wünschenswert, der Rettungsfahrten explizit kennzeichnet.

Zu guter Letzt sei auf den Umstand verwiesen, dass zur besseren Nachvollziehbarkeit die Darstellung des Verknüpfungsprozesses ausschließlich für Einsatzprotokolle aus Bayern erfolgte, wenngleich in Inno_RD auch ärztliche und nicht-ärztliche Daten aus Baden-Württemberg verarbeitetet werden. Hintergrund ist hier, dass unterschiedliche räumliche Entitäten oder Rettungsdienstorganisationen sowie uneinheitliche Dokumentationsstandards teils erhebliche Unterschiede in Hinblick auf geeigneten Schlüsselvariablen mit sich bringen. Eine Adaption aller geschilderten Verfahren auf die aus Baden-Württemberg vorliegenden Daten ist jedoch ausdrücklich möglich. Dabei zeigt sich bspw. in den baden-württembergischen NF-Protokollen eine vergleichbare Füllquote der KVNR, sodass der Ansatz der probabilistischen Verknüpfung auch für dieses Bundesland von Bedeutung ist. Die flächendeckendere Etablierung von Standards bei der Rettungsdienst-Einsatzdokumentation, wie bspw. dem Minimalen Notfalldatensatz [5, 6], könnte zukünftig die Verknüpfung von Sekundärdaten aus der Notfallversorgung erleichtern. Entsprechende Forderungen zur bundesweit – oder mindestens landesweit – standardisierten Einsatzdokumentation im Rettungsdienst werden, insbesondere vor dem Hintergrund der Qualitätssicherung, bereits seit einiger Zeit diskutiert und z. T. umgesetzt (vgl. [2, 6, 28–30]).

Danksagung

Für die Abstimmung und Übermittlung der im Projekt verarbeiteten Rettungsdienst- sowie Kassendaten möchten wir den Dateneignern und -lieferanten unseren Dank aussprechen. Aufgrund zahlreicher datentechnischer Hinweise gilt unser besonderer Dank zudem dem Institut für Notfallmedizin und Medizinmanagement am Klinikum der Universität München sowie der Servicestelle Forschungsdatenmanagement an der Fakultät Medizin und Gesundheitswissenschaften der Carl von Ossietzky Universität Oldenburg.

Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

[1] Schmiedel R, Behrendt H. Leistungen des Rettungsdienstes 2016/17. Analyse des Leistungsniveaus im Rettungsdienst für die Jahre 2016 und 2017. Bergisch Gladbach: Bundesanstalt für Straßenwesen

[2] Piedmont S, Brammen D, Branse D et al. Auf dem Weg zur integrierten Qualitätssicherung im Rettungsdienst. Stand – Bedarf – Vision. Notfall & Rettungsmedizin 2018; 21: 682–689

[3] Schmiedel R, Behrendt H, Betzler E. Bedarfsplanung im Rettungsdienst. Standorte – Fahrzeuge – Personal – Kosten. Berlin, Heidelberg: Springer; 2004

[4] Hellmich C. Qualitätsmanagement und Zertifizierung im Rettungsdienst. Grundlagen – Techniken – Modelle – Umsetzung. Berlin, Heidelberg: Springer; 2010

[5] Messelken M, Schlechtriemen T, Arntz H-R et al. Minimaler Notfalldatensatz MIND3. Notfall & Rettungsmedizin 2011; 14: 647–654

[6] Lohs T, Wnent J, Jakisch B. Dokumentation und Qualitätsmanagement im Rettungsdienst. Notfallmedizin up2date 2018; 13: 391–406

[7] Stoschek J. Tablet für bayerische RTW. Ärzte Zeitung online. Veröffentlicht: 26.06.2014. Online: https://www.aerztezeitung.de/Wirtschaft/Tablet-fuer-bayerische-RTW-244135.html

[8] Thümmler C. Das Krankenbeförderungsrecht in der gesetzlichen Krankenversicherung. In: Axer P, Becker U, Bieback K-J, et al., Hrsg. Schriften zum Sozialrecht. Baden-Baden: Nomos Verlag; 2018

[9] Hecker U, Schramm C. Hrsg. Praxis des Intensivtransports. Berlin, Heidelberg: Springer; 2018

[10] Pötsch M. Fehleinsätze im Rettungsdienst – Übertritt der Vorgaben des SGB V durch Landesrecht? Neue Zeitschrift für Sozialrecht 2018; 27: 646–648

[11] GKV-Spitzenverband. Bundeseinheitliches Positionsnummernverzeichnis für Krankentransportleistungen. Fassung vom 15. März 2010. Online: https://www.gkv-datenaustausch.de/media/dokumente/leistungserbringer_1/sonstige_leistungserbringer/positionsnummernverzeichnisse/Krankentransportleistungen-20100315.pdf

[12] Gemeinsamer Bundesausschuss (G-BA). Richtlinie über die Verordnung von Krankenfahrten, Krankentransportleistungen und Rettungsfahrten nach § 92 Absatz 1 Satz 2 Nummer 12 SGB V (Krankentransportrichtlinie). Fassung vom 22.01.2004. Online: https://www.g-ba.de/downloads/62-492-2262/KT-RL_2020-09-17_iK-2020-10-01.pdf

[13] GKV-Spitzenverband. Information des GKV-Spitzenverbandes zur Eindeutigkeit und lebenslangen Gültigkeit der Krankenversichertennummer gemäß § 290 SGB V. Veröffentlicht: 02.10.2015. Online: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/telematik/egk_2/10-2015_Information_GKV-SV_Eineutigkeit_Gueltigkeit_KVNRv06.pdf

[14] Fischer H, Röhrig R, Thiemann VS. A generic IT infrastructure for identity management and pseudonymization in small research projects with heterogeneous and distributed data sources under consideration of the GDPR. Studies in Health Technology and Informatics 2019; 1837–1838

[15] March S, Antoni M, Kieschke J et al. Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. Das Gesundheitswesen 2018; 80: 20–31

[16] March S, Andrich S, Drepper J et al. Gute Praxis Datenlinkage (GDP). Das Gesundheitswesen 2019; 81: 636–650

[17] Crilly JL, O'Dwyer JA, O'Dwyer MA et al. Linking ambulance, emergency department and hospital admissions data: understanding the emergency journey. Medical Journal of Australia 2011; 194: 34–37

[18] Seymour CW, Kahn JM, Cooke CR et al. Prediction of critical illness during out-of-hospital emergency care. Journal of the American Medical Association 2010; 304: 747–754

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

[19] Mumma BE, Diercks DB, Danielsen B et al. Probabilistic linkage of prehospital and outcomes data in out-of-hospital cardiac arrest. Prehospital Emergency Care 2015; 19: 358–364

[20] Oostema JA, Nickles A, Reeves MJ. A comparison of probabilistic and deterministic match strategies for linking prehospital and in-hospital stroke registry data. Journal of Stroke and Cerebrovascular Diseases 2020; 29: 105151

[21] Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. Academic Emergency Medicine 2006; 13: 69–75

[22] Downing A, Wilson R, Cooke M. Linkage of ambulance service and accident and emergency department data: a study of assault patients in the west midlands region of the UK. International Journal of the Care of the Injured 2005; 36: 738–744

[23] van der Laan J. Reclin: Record Linkage Toolkit R package. Version: 0.1.1. Online: https://github.com/djvanderlaan/reclin letzter Zugriff: 25.04.2021

[24] Christen P, Goiser K. Quality and Complexity Measures for Data Linkage and Deduplication. In: Guillet F, Hamilton HJ, Hrsg. Quality Measures in Data Mining. Berlin, Heidelberg: Springer; 2007: 127–151

[25] Piedmont S, Reinhold AK, Bock J-O et al. Which health-related reasons lead to prehospital emergency care and how does subjective emergency status connect to subsequent care? Emergency medical service patients' survey linked to health claims data. Notfall & Rettungsmedizin 2021; 1–11

[26] Fischer H. Simple Batch Record Linkage System (SimBa). Version Client: 2.5.4 - Server: 2.0.4. Online. doi:10.5281/zenodo.4277173 letzter Zugriff: 25.04.2021

[27] Horenkamp-Sonntag D, Lindner R, Wenzel F et al. Prüfung der Datenqualität und Validität von GKV-Routinedaten. In: Swart E, Ihle P, Gothe H, et al., Hrsg. Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Huber; 2014: 314–330

[28] Altemeyer K-H, Schlechtriemen T, Reeb R. Rettungsdienst in Deutschland: Bestandsaufnahme und Perspektiven. Bericht über ein Symposium der Deutschen Gesellschaft für Anästhesiologie und Intensivmedizin (DGAI), Reisensburg 02.–04.10.2002. Notfall & Rettungsmedizin 2003; 6: 89–101

[29] Lohs T. Qualitätssicherung im Rettungsdienst Baden-Württemberg. In: Neumayr A, Baubin M, Schinnerl A, Hrsg. Herausforderung Notfallmedizin. Innovation – Vision – Zukunft. Berlin: Springer; 2018: 233–243

[30] Fischer M, Kehrberger E, Marung H et al. Eckpunktepapier 2016 zur notfallmedizinischen Versorgung der Bevölkerung in der Prähospitalphase und in der Klinik. Notfall & Rettungsmedizin 2016; 19: 387–395

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.