TECHNISCHE UNIVERSITÄT DRESDEN

Medizinische Fakultät Carl Gustav Carus

Institut für Medizinische Informatik und Biometrie
Professur für Medizinische Informatik

Sekundärnutzung deutscher Medikationsdaten in internationalen Studien unter Wahrung der semantischen Bedeutung

Ines Reinecke

Geboren am: 5. Januar 1981 in Freiberg

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum medicinalium (Dr.rer.medic.)

Erstgutachter

Prof. Dr. rer. nat. Martin Sedlmayr

Zweitgutachter

Prof. Dr. med. Mario Menk

Eingereicht am: 11. Juli 2023

1. Gutachter: Prof. Dr. rer. nat. Martin Sedlmayr

2. Gutachter: Prof. Dr. med. Mario Menk

Tag der mündlichen Prüfung: 21.11.2023

gez.: Prof. Dr. Timo Siepmann Vorsitzender der Promotionskommission

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich das vorliegende Dokument mit dem Titel *Sekundärnutzung deutscher Medikationsdaten in internationalen Studien unter Wahrung der semantischen Bedeutung* selbstständig und ohne unzulässige Hilfe Dritter verfasst habe. Es wurden keine anderen als die in diesem Dokument angegebenen Hilfsmittel und Quellen benutzt. Die wörtlichen und sinngemäß übernommenen Zitate habe ich als solche kenntlich gemacht. Es waren keine weiteren Personen an der geistigen Herstellung des vorliegenden Dokumentes beteiligt. Mir ist bekannt, dass die Nichteinhaltung dieser Erklärung zum nachträglichen Entzug des Hochschulabschlusses führen kann.

Dresden, 11. Juli 2023

Ines Reinecke

Zusammenfassung

Elektronisch verfügbare Daten aus der Gesundheitsversorgung, sogenannte Real-World Data (RWD), gewinnen zunehmend an Bedeutung für die Forschung, insbesondere in der Pharmakovigilanz und der Arzneimittelsicherheit. Die Schaffung von kollaborativen Forschungsnetzwerken, wie beispielsweise Observational Health Data Sciences and Informatics (OHDSI) oder European Health Data and Evidence Network (EHDEN) stoßen auf positive Resonanz, um die Potenziale von RWD zu nutzen und die Reproduzierbarkeit und Verlässlichkeit von Forschungsergebnissen retrospektiver Beobachtungsstudien zu verbessern. Eine Beteiligung deutscher Universitätskliniken mit RWD der stationären Versorgung fehlt bisher, vor allem weil die qualitativen Eigenschaften der Medikationsdaten aktuell eine Hürde darstellen. In dieser Arbeit wird daher untersucht, wie die Sekundärnutzung von Medikationsdaten der klinischen Versorgung in retrospektiven Beobachtungsstudien in internationalen Forschungsgemeinschaften am Beispiel von OHDSI unter Wahrung der semantischen Bedeutung ermöglicht werden kann.

Initial wird ein Scoping Review durchgeführt, um zu ermitteln, wo die Schwerpunkte der Nutzung des Datenmodells Observational Medical Outcomes Partnership (OMOP) derzeit liegen. Es werden die Anforderungen an die Daten in OMOP seitens der Forschungsgemeinschaft OHDSI ermittelt und mit dem IST-Zustand der Medikationsverordnungen am Beispiel des Universitätsklinikum Carl Gustav Carus Dresden (UKD) abgeglichen. So werden die Inhibitoren identifiziert, welche im Widerspruch zu den Anforderungen stehen. Korrektive Maßnahmen zur Reduktion der Inhibitoren werden konzipiert, umgesetzt und anschließend quantitativ und qualitativ bewertet. Zudem untersucht die Arbeit, wie eine notwendige Transparenz möglicher, verbleibender Limitierungen gewährleistet werden kann.

Das durchgeführte Scoping Review zeigt eine über die vergangenen Jahre stetig zunehmende Bedeutung des Datenmodells OMOP für die Durchführung von Studien unter Verwendung von Daten aus mehreren Ländern. In Deutschland fokussiert sich die Forschung im Kontext OMOP bislang auf die Betrachtung von Trends, des Datentransfers, Mappings und der Entwicklung von Konzepten. Eine aktive Beteiligung an der Durchführung von

Studien mit medizinischen Fragestellungen unter Nutzung von OMOP findet aktuell nicht statt. Zur Verwendung von Medikationsverordnungen in OMOP müssen die Daten strukturiert und unter Verwendung von internationalen Terminologien wie ATC und RxNorm vorliegen. Allerdings zeigt eine Analyse über mehrere Standorte in Deutschland, dass Medikationsverordnungen überwiegend unstrukturiert und ohne belastbare Zuordnung standardisierter, internationaler Klassifikationen dokumentiert werden. Dieses Ergebnis bestätigt sich auch bei der Untersuchung der Medikationsverordnungen des UKD der Jahre 2016 bis 2020.

Die in dieser Arbeit entwickelten und durchgeführten Maßnahmen wurden abgeleitet aus der Teilnahme an einem Pilotprojekt der European Medicines Agency (EMA) und fokussieren auf der Verbesserung der Datenstruktur sowie der Überführung der Medikationsverordnungen nach RxNorm. So konnte der Grad der Klassifizierung der Medikationsverordnungen des UKD unter Verwendung der Standard-Terminologie RxNorm von initial 0% auf 66,39% erhöht werden. Des Weiteren wird durch eine interaktive Visualisierung der Datenstruktur und des Grades der Überführbarkeit von ATC Codes nach RxNorm eine Transparenz der Ergebnisse geschaffen.

Die Beantwortung aller in dieser Arbeit gestellten Forschungsfragen schafft die Voraussetzung, um zukünftig an retrospektiven Beobachtungsstudien der OHDSI Forschungsgemeinschaft teilzunehmen zu können. Die semantische Bedeutung der Medikationsverordnungen, auch unter Verwendung internationaler Terminologien wie RxNorm, bleibt dabei gewahrt. Zusätzliche Transparenz kann Forschenden und Versorgenden helfen, die Datenqualität im Sinne der Strukturiertheit der Medikationsverordnungen am UKD in Zukunft bereits zum Zeitpunkt der Entstehung zu verbessern.

Abstract

Electronically available healthcare data, known as Real-World Data (RWD), are increasingly gaining importance in research, particularly in pharmacovigilance and drug safety. Development efforts to build collaborative research networks, such as Observational Health Data Sciences and Informatics (OHDSI) or European Health Data and Evidence Network (EHDEN), has received positive feedback in order to harness the potential of RWD and improve the reproducibility and reliability of research findings from retrospective observational studies. However, German university hospitals have not yet participated in retrospective, observational studies on Observational Medical Outcomes Partnership (OMOP) with RWD from inpatient care, mainly due to the current challenges due to the qualitative characteristics of medication data. This doctoral thesis therefore examines how the secondary use of clinical care medication data can be facilitated in retrospective observational studies within international research communities, using OHDSI as an example, while preserving semantic meaning.

Initially, a scoping review is conducted to determine the current focus areas of utilizing OMOP as a data model. The requirements for data within OMOP, as determined by the research community of OHDSI, are assessed and compared to the current state of medication prescriptions at the Universitätsklinikum Carl Gustav Carus Dresden (UKD) as an example. This process identifies any inconsistencies between the requirements and the current state of medication prescriptions. Corrective measures are designed, implemented, and subsequently evaluated quantitatively and qualitatively to reduce these inconsistencies. Furthermore, this doctoral thesis explores how necessary transparency of any remaining limitations can be ensured.

The conducted scoping review demonstrates a steadily increasing importance of OMOP for conducting studies using data from multiple countries over the past years. In Germany, research efforts related to OMOP have primarily focused on exploring trends, data transfer, mappings, and concept development. Currently, there is no active involvement with inpatient RWD in conducting studies addressing medical questions using OMOP. To utilize

medication prescriptions within OMOP, the data must be structured and employ international terminologies such as Anatomisch-Therapeutisch-Chemisches Klassifikationssystem (ATC) and Prescription (Rx) Normalisierung (Norm) (RxNorm). However, an analysis across multiple sites in Germany reveals that medication prescriptions are predominantly documented in an unstructured manner, lacking reliable associations with standardized international classifications. This result is further confirmed by the examination of medication prescriptions at the UKD from 2016 to 2020.

The measures developed and implemented in this doctoral thesis were derived from participating in a pilot project of the European Medicines Agency (EMA) and focused on improving data structure and transitioning medication prescriptions to RxNorm. As a result, the degree of classification of medication prescriptions at the UKD according to the standard terminology RxNorm increased from an initial 0% to 66,39%. Furthermore, an interactive visualization of the data structure and the degree of conversion from ATC codes to RxNorm provides transparency of the results.

The answers to all research questions posed in this doctoral thesis establish the prerequisites for future participation in retrospective observational studies conducted by the OHDSI research community. The semantic meaning of medication prescriptions, including the use of international terminologies like RxNorm, is preserved. Moreover, enhancing transparency can aid researchers and healthcare providers in enhancing data quality, particularly in terms of the structured nature of medication prescriptions at the UKD, right at the time of their origin.

Danksagung

Ich möchte mich bei **Prof. Dr. Martin Sedlmayr** für die fantastische Zusammenarbeit und Betreuung meiner wissenschaftlichen Arbeit und das entgegengebrachte Vertrauen in meine Ideen und mein Handeln bedanken.

Ein großer Dank geht an meine liebe Kollegin **Franzi** für ihre grenzenlose Unterstützung, den Zuspruch und die Geduld, ihre Erreichbarkeit und die zahlreichen, oft stundenlangen, aber sehr konspirativen Gespräche und Spaziergänge.

Ich bedanke mich bei all meinen **Kolleginnen und Kollegen** für das unermüdliche Engagement, die Unterstützung und Mitwirkung an den wissenschaftlichen Publikationen, die im Rahmen meiner Dissertation entstanden und veröffentlicht worden sind. Vor allem bedanke ich mich bei meinem Team auch für die Rücksichtnahme und das Verständnis.

Besonderer Dank gilt meinen lieben **Eltern Gina und Berndt** und meiner **Schwester Anja**, die immer an mich glauben und mich mit aller Kraft unterstützen und stärken und eine besondere und verlässliche Hilfe darstellen - die ich nicht vermissen möchte. Ich danke auch allen **Freundinnen und Freunden** und allen Menschen, die mir Mut gemacht haben.

Der größte Dank geht an vier besondere und geliebte Menschen - mein Mann **Thomas** und meine Kinder **Anna, Max und Ole**. Ihr habt vor allem in den vergangenen Monaten viel auf mich verzichten müssen. Ihr habt mir Ruhe gegeben und Rücksicht genommen, auch wenn es anstrengend war. Ihr habt mich zum Lachen gebracht, wenn mir zum Weinen zumute war. Tom, du hast mich durch die letzten Monate getragen - bedingungslos und in jeder Hinsicht. Du hast mich vom Boden aufgesammelt, wenn ich liegen bleiben wollte. Ich danke dir von ganzem Herzen.

Inhaltsverzeichnis

Symbole und Abkürzungen

MI-I	Medizininformatik-Initiative
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
RWD	Real-World Data
PROM	Patient Reported Outcomes
FDA	Food and Drug Administration
FAERS	FDA Adverse Event Reporting System
CDM	Common Data Model
EHDEN	European Health Data and Evidence Network
HL7	Health Level Seven
BMBF	Bundesministerium für Bildung und Forschung
EMA	European Medicines Agency
FHIR	Fast Healthcare Interoperability Resources
MIRACUM	Medical Informatics in Research and Care in University Medicine
WHO	World Health Organization

ICD10	International Classification of Diseases Version 10
GM	German Modification
SNOMED-CT	Systematized Nomenclature of Medicine - Clinical Terms
OPS	Operationen- und Prozedurenschlüssel
DIZ	Datenintegrationszentrum
CTRSS	Clinical Trial Support System
ETL	Extract-Transform-Load
HGNC	HUGO Gene Nomenclature Committee
PRISMA-ScR	Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
CSV	Comma Separated File
IT	Informationstechnik
DSGVO	Datenschutz-Grundverordnung
UKD	Universitätsklinikum Carl Gustav Carus Dresden
KDS	Kerndatensatz
DQD	Data Quality Dashboard
AMTS	Arzneimitteltherapiesicherheit
KIS	Krankenhausinformationssystem
ATC	Anatomisch-Therapeutisch-Chemisches Klassifikationssystem
DS-Med	Datensatz - Medikationsverordnungen
DS-Katalog	Datensatz - Hauskatalog für Arzneimittel
DS-Gruppiert	Datensatz - unstrukturierte Verordnungen aus DS-Med , gruppiert nach dem Freitext für die Medikation

DS-Top1000	Datensatz - Teilmenge von DS-Gruppiert, die häufigsten 1000 Freitexte für Medikation
DS-ATC	Datensatz - ATC Vokabular aus der Tabelle concept in OMOP
DS-Relation	Datensatz - Beziehungen zwischen allen Konzepten der OMOP Vokabulare ATC und RxNorm
EPHMRA	European Pharmaceutical Market Research Association
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
RxNorm	Prescription (Rx) Normalisierung (Norm)
UMLS	Unified Medical Language System
TTY	Term Types
DFG	Deutsche Forschungsgemeinschaft
MIRACOLIX	Medical Informatics ReusAble eCO-system of open source Linkable and Interoperable software toolbox
NLP	Natural Language Processing
EU PAS	European post-authorisation study
IQVIA	I-Quintiles-Verscend-Information-Assets
TU	Technische Universität
SQL	Structured Query Language
SAP	Systeme, Anwendungen, Produkte
WIdO	Wissenschaftliches Institut der AOK
ATC-GM	ATC-German Modification
REDCap	Research Electronic Data Capture
FDPG	Forschungsdatenportal Gesundheit
PP	Per-Protocol

ITT	Intention-to-Treat
PZN	Pharmazentralnummer
ICM	Integrated Care Manager
APPROVe	Adenomatous Polyp Prevention on Vioxx
NLM	National Library of Medicines
XSLT	Extensible Stylesheet Language Transformation
MIMIC	Medical Information Mart for Intensive Care
NDC	National Drug Code
SNDS	Système National des Données de Santé
UCD	Unité Commune de Dispensation
CIP13	Code Identifiant de Présentation 13
NCTS	National Clinical Terminology Service
NSG	Nationales Steuerungsgremium
LOINC	Logical Observation Identifiers Names and Codes

1 Einleitung

1.1 Motivation

Elektronisch verfügbare Daten aus der Gesundheitsversorgung sogenannte RWD sind gemäß des „Framework of Real-World Evidence Program“ der Food and Drug Administration (FDA) wie folgt beschrieben „data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources“ (US Food and Drug Administration, 2022). RWD der Gesundheitsversorgung sind gekennzeichnet durch eine Vielfalt in der Art als auch der Herkunft der Daten und lassen sich in unterschiedlichste Typen wie elektronische Patientenakten, Abrechnungsdaten, mobile Anwendungen, Sensordaten, Patient Reported Outcomes (PROM), soziale Medien, molekulare Daten, Umweltdaten, Literatur oder Familienhistorie kategorisieren (Swift et al., 2018).

Die Verwendung von RWD gewinnt für die Forschung aufgrund ihrer elektronischen Verfügbarkeit neben randomisierten, kontrollierten Studien zunehmend an Bedeutung (Magalhães et al., 2022; Safran et al., 2007; Schuemie, Cepeda et al., 2020). Die Anwendungsbereiche von RWD sind vielfältig. Einen wichtigen Bereich stellt die Überwachung der Sicherheit von Arzneimitteln in der klinischen Versorgung, die sogenannte Pharmakovigilanz dar (Pitts und Le Louet, 2018; Pitts, Louet et al., 2016). Besondere Aufmerksamkeit erlangte diese Thematik in den frühen 2000er Jahren, als die Anzahl von zugelassenen Medikamenten, die aufgrund von schwerwiegenden Nebenwirkungen zurückgerufen wurden, drastisch anstieg (Sadhna et al., 2015). Als repräsentatives Beispiel sei hier Medikament VIOXX des Herstellers Merck genannt, welches im Jahre 1999 als Schmerzmedikament zugelassen wurde. Der Hersteller hat die Langzeitstudie Adenomatous Polyp Prevention on Vioxx (APPROVe) durchgeführt,

durch die ein erhöhtes Risiko schwerwiegender kardiovaskulärer Nebenwirkungen in Zusammenhang mit der Einnahme von VIOXX aufgezeigt wurde (Baron et al., 2008; Bresalier et al., 2005). Schätzungen zufolge starben 40.000 Menschen an den direkten oder indirekten Folgen der Einnahme von VIOXX. Das Medikament wurde seitens des Herstellers im Jahr 2004 weltweit vom Markt genommen (Sibbald, 2004). Aufgrund einer generellen Zunahme der Rückrufe von bereits zugelassenen Medikamenten durch Hersteller, rief die FDA das Sentinel Programm ins Leben, um die aktive Überwachung von Medikamenten zu stärken (Platt, Wilson et al., 2009). Neben etablierten Systemen zur Dokumentation von unerwünschten Nebenwirkungen, wie beispielsweise dem FDA Adverse Event Reporting System (FAERS) (Platt, Brown et al., 2018) in den USA, EudraVigilance (Postigo et al., 2018) in Europa und Möglichkeiten der Berichterstattung über unerwünschte Reaktionen nach Medikamenteneinnahme in Asien (Biswas, 2013), bietet die Nutzung von RWD großes Potenzial für eine frühzeitige Erkennung von Risiken, Möglichkeiten zur Intervention und damit zur potenziellen Erhöhung der Arzneimittelsicherheit (Burcu et al., 2020; Corrigan-Curay et al., 2018; Hampson et al., 2018).

In der Vergangenheit wurden medizinische Daten jedoch primär in proprietären Systemen isoliert gespeichert, sodass eine Zusammenführung von Daten, ein gesamtheitliches inhaltliches Verständnis und eine Darstellung und Nutzung von Informationen über diese Silos großen Mehraufwand bedurfte (Ganslandt et al., 2015; Miller et al., 2014). Dabei bekommt die Fähigkeit der Zusammenarbeit und Kommunikation über definierte Schnittstellen, zur sicheren und effektiven Übertragung und Zusammenführung von Daten, eine zunehmend hohe Bedeutung in der Medizin und stellt damit einen Schwerpunkt bei der Entwicklung von Software dar. Speziell im Bereich der Forschung und der internationalen, länderübergreifenden Kooperation ist die Gewährung der semantischen Interoperabilität zur Durchführung großangelegter retrospektiver Beobachtungsstudien notwendig (Lehne et al., 2019).

Neben der Harmonisierung und Zusammenführung von Daten, ist auch die Etablierung standardisierter Methoden zur Analyse notwendig. Das zeigt sich in den Ergebnissen der beiden Studien von Cardwell et al. (Cardwell et al., 2010) und Green et al. (Green et al., 2010), die widersprüchliche Ergebnisse liefern, obwohl sie die gleiche Datenbasis nutzten. Die Studien machen gegensätzliche Aussagen zur Erhöhung des Risikos für eine Erkrankung an Speiseröhrenkrebs im Zusammenhang mit der Einnahme von Bisphosphonaten, obwohl sie innerhalb weniger Monate nacheinander im selben Jahr veröffentlicht wurden.

Ähnliche Ergebnisse zeigte auch die OMOP Initiative (Stang, Ryan, Racoosin et al., 2010). Sie wurde durch die FDA als öffentlich-private Partnerschaft im Jahr 2008 ins Leben gerufen. Die OMOP Initiative untersuchte und bewertete zunächst die Verwendung von RWD zur Erkennung von Zusammenhängen von Medikamenten und auftretenden Nebenwirkungen. Sie hatte als weiteres Ziel, standardisierte Methoden zur Durchführung von retrospektiven Beobachtungsstudien zu etablieren, um potenzielle Risiken von retrospektiven Beobachtungsstudien wie beispielsweise eine mögliche Voreingenommenheit (Bias) von Daten vorzubeugen.

Im Rahmen der sogenannten Experimente der OMOP Initiative in den USA wurden Analysen von mehr als 130 Millionen Patient:innendaten, verteilt über mehrere Datenbanken durchgeführt, um die geplante Untersuchung und Bewertung der Verwendung von RWD durchzuführen (Ryan et al., 2012). Zur Auswertung der Daten wurden unterschiedliche Analysemethoden genutzt. Die Ergebnisse des Experiments der OMOP Initiative zeigen eine starke Varianz in den Ergebnissen je nach verwendeter Methode und Datenbank (Ryan et al., 2012). Auch die Ergebnisse anderer Studien wie beispielsweise von David Madigan deuten darauf hin, dass die Ergebnisse von Beobachtungsstudien trotz eines einheitlichen Studiendesigns in Abhängigkeit der Wahl der Datenbank zur Speicherung der RWD zwischen 20 % und 40 % statistisch signifikant schwanken können (Madigan et al., 2013). Das OMOP Experiment wurde in Europa in Großbritannien mit ähnlichen Ergebnissen wiederholt und im Jahr 2013 abgeschlossen (Schuemie, Gini et al., 2013).

Um diese Herausforderungen zu überwinden, ist die Standardisierung von Datenformaten, Inhalten (Kodierung) sowie der Methoden zur Datenanalyse notwendig (Stang, Ryan, Overhage et al., 2013). Im Ergebnis wurde mit der Entwicklung eines standardisierten Common Data Model (CDM) zur Harmonisierung von RWD begonnen. Es soll Forschungsgruppen weltweit ermöglichen, RWD standardisiert zu speichern, auszutauschen und für die Forschung zu verwenden (Overhage et al., 2012) und letztlich die Übertragbarkeit und die Reproduzierbarkeit von Forschungsergebnissen zu verbessern.

Basierend auf den initialen Ergebnissen der OMOP Experimente und der ersten Version des OMOP CDM (im weiteren Verlauf der Arbeit kurz OMOP genannt) wurde 2014 eine gemeinnützige Organisation namens OHDSI gegründet. Das Ziel von OHDSI ist die Verbesserung der Gesundheitsversorgung der Patient:innen durch die Förderung der Verwendung von RWD und die Schaffung einer kollaborativen Forschungsumgebung, in der Forschende zusammenarbeiten und ihr Wissen, ihre Methoden und Daten austauschen können (Hripcsak,

Ryan et al., 2016). Daher stellt OHDSI ein Open-Source-Software-Framework und Methoden zur Standardisierung und Analyse von Daten bereit. OMOP bildet dabei die Grundlage für die Datenspeicherung. Wichtige Komponenten des OHDSI Software Frameworks sind die standardisierten Vokabulare und Terminologien zur Datenharmonisierung, Methoden und Werkzeuge für die Entwicklung von Extract-Transform-Load (ETL) Jobs sowie Werkzeuge für die Datenanalyse. Ein weltweiter Austausch zwischen und die Kooperation von Forschungsteams auf Basis standardisierter und interoperabler Daten und Methoden wird möglich. Heute ist die OHDSI Forschungsgemeinschaft bereits in mehr als 19 Ländern, mit mehr als 200 Millionen Daten von Patient:innen außerhalb der USA und mit mehr als 2.500 mitarbeitenden Menschen, vertreten (OHDSI, 2019).

Die OHDSI Forschungsgemeinschaft gewann in den vergangenen Jahren zunehmend an Bedeutung (Reinecke, Zoch, Reich, Sedlmayr et al., 2021). Beispiele in Europa sind (I) das Forschungsnetzwerk EHDEN (EHDEN, 2022), (II) die Zusammenarbeit mit Health Level Seven (HL7) (HL7, 2021) und (III) die deutsche Medizininformatik-Initiative (MI-I), eine Fördermaßnahme des Bundesministerium für Bildung und Forschung (BMBF) (Semler et al., 2018). Das EHDEN Konsortium arbeitet an einer Harmonisierung von Patientenakten in 22 Ländern Europas auf der Grundlage von OMOP. Im Juni 2020 kündigte die EMA ein Projekt zum Aufbau eines Forschungsrahmens für multizentrische Studien an COVID-19-Patienten an, das die Zusammenarbeit mit EHDEN umfasst und auf OMOP als Grundlage für die Datenharmonisierung aufbaut (European Medicines Agency (EMA), 2020a). Im März 2021 wurde eine Zusammenarbeit zwischen HL7 und der OHDSI Forschungsgemeinschaft mit dem Ziel angekündigt, ein gemeinsames Datenmodell zu entwickeln, das HL7 Fast Healthcare Interoperability Resources (FHIR) und OMOP integriert. Das Konsortium Medical Informatics in Research and Care in University Medicine (MIRACUM) als Teil der MI-I und gefördert durch das BMBF, arbeitet daran, klinische Daten aus der Versorgung in die Forschung zu integrieren. Dabei führt es Methoden und Werkzeuge zur verteilten Datenanalyse ein, die auf Open-Source-Software basieren. Derzeit entwickelt und evaluiert MIRACUM eine IT-gestützte Rekrutierungsinfrastruktur für klinische Studien, die auf OMOP basiert (Reinecke, Gulden et al., 2020).

1.2 Offene Herausforderungen

Auch in Deutschland ist das Forschungsinteresse für die Thematik OHDSI und OMOP im Vergleich zu anderen Ländern sehr groß (Reinecke, Zoch, Reich, Sedlmayr et al., 2021). Die bisherigen Publikationen aus Deutschland decken die folgenden Themengebiete ab:

1. Datentransfer und Mapping von Daten nach OMOP (Maier et al., 2018)
2. Trends und Initiativen im Bereich der Forschung mit RWD (Prokosch, Acker et al., 2018; Tresp et al., 2016)
3. Entwicklung von Tools auf Basis von OMOP (Freitas Da Cruz et al., 2019; Gruendner et al., 2019; H. Spengler et al., 2020; Unberath et al., 2020)
4. Konzepte zur Nutzung von OMOP (Fischer et al., 2020; Gruhl et al., 2020; Reinecke, Gulden et al., 2020)

Keine der genannten Publikationen von Forschungsteams an deutschen Standorten nutzt derzeit klinische Daten in OMOP für die Teilnahme an internationalen medizinischen Studien, die innerhalb der OHDSI Forschungsgemeinschaft durchgeführt werden. Vorarbeiten zur Harmonisierung erfolgten bisher ohne Abstimmung gegenüber den Anforderungen der OHDSI Forschungsgemeinschaft, sondern ausschließlich orientiert an nationalen Anwendungsfällen, wie beispielsweise dem *Use Case 1 - Optimierung der Patient:innenrekrutierung* in MIRACUM. Weitere Arbeiten nutzten die in Deutschland standardisiert verfügbaren Abrechnungsdaten der stationären Versorgung gemäß des §21 Datensatzes des Krankenhausentgeltgesetzes (InEK, 2018)) ohne eine Untersuchung der Nutzbarkeit in Studien und mit eingeschränktem Transfer in international gültige Terminologien (Maier et al., 2018).

Die bisherigen Arbeiten aus Deutschland sind dabei limitiert auf die Demonstration der grundsätzlichen Machbarkeit des Transfers von Diagnosen und Falldaten unter Verwendung existierender Übersetzungen nationaler nach internationaler Terminologien. Entsprechende Aktivitäten zur Harmonisierung von Medikationsdaten bleiben bisher offen, obwohl die OHDSI Forschungsgemeinschaft und die in Kapitel 1.1 vorgestellten Projekte einen großen Schwerpunkt auf dem Thema der Arzneimittelüberwachung und -sicherheit aufweisen. Durch die Ausrichtung von OHDSI ist zu erwarten, dass auf OMOP durchgeführte Studien häufig Medikationsdaten als Einschlusskriterien aufführen werden. Ob diese Vermutung objektiv betrachtet belastbar ist, wurde bisher jedoch nie systematisch geprüft und bewertet. Eine weitere offene Herausforderung ist zudem auch die fehlende Ermittlung der

notwendigen Daten, um an OHDSI Netzwerkstudien teilzunehmen. Um mögliche Lücken hinsichtlich relevanter Daten und notwendiger internationaler Standards zu identifizieren, müsste die Nutzung der Daten in bereits durchgeführten Studien innerhalb der OHDSI Forschungsgemeinschaft im Detail betrachtet und analysiert werden.

Die Datenqualität der RWD gilt insbesondere für die Durchführung von Studien auf Basis RWD als entscheidendes Erfolgskriterium und ist wichtig, um die Zuverlässigkeit und Vertrauenswürdigkeit der Ergebnisse zu sichern (Brown et al., 2013; Reimer et al., 2016; Zozus et al., 2015).

Eine ganzheitliche Analyse der Datenqualität der klinischen Versorgungsdaten des Universitätsklinikums Carl Gustav Carus Dresden am Beispiel der Medikationsverordnungen für deren Nutzbarkeit in diesem Kontext existiert bisher nicht. Bezogen auf RWD aus Deutschland existiert zudem derzeit keine gesamtheitliche Betrachtung der Datenqualität unter Berücksichtigung der Anforderungen der Qualitätskriterien Vollständigkeit, Konformität und Plausibilität gemäß Kahn et al. (Kahn et al., 2016) für das Forschungsdatenrepository OMOP als Zielformat. Eine Transparenz möglicher Einschränkungen und Limitierungen ist jedoch notwendig, um Forschungsteams die Möglichkeit einzuräumen, informierte und gesicherte Entscheidungen über die Frage zu treffen, in welchen Fällen eine Teilnahme an Studien möglich sein kann.

1.3 Ziele und Fragestellungen der Arbeit

Die Herausforderungen führen zur zentralen Fragestellung der vorliegenden Arbeit:

*„ Wie kann eine **Sekundärnutzung von Medikationsdaten** der klinischen Versorgung in retrospektiven Beobachtungsstudien in **internationalen** Forschungsgemeinschaften am Beispiel von OHDSI unter Wahrung der **semantischen Bedeutung** ermöglicht werden?“*

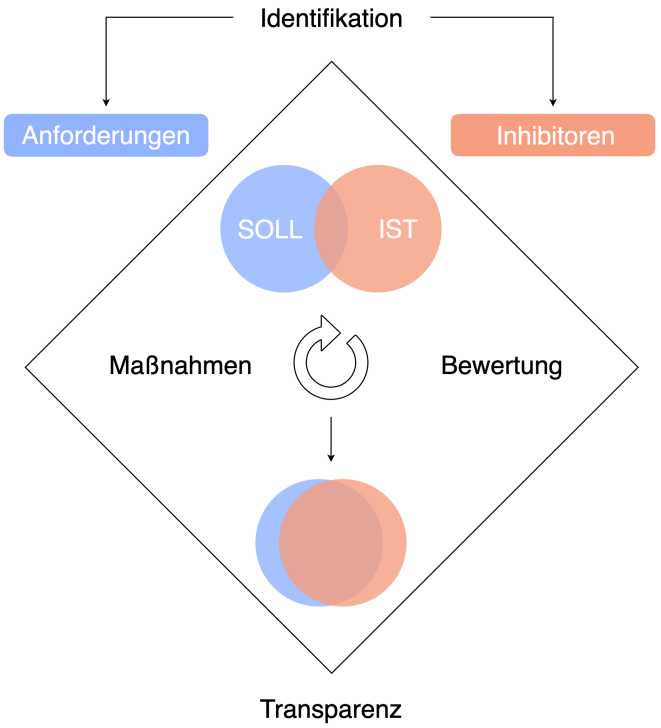


Abbildung 1.1: Überblick dieser Arbeit

Zur Beantwortung dieser zentralen Frage werden folgende Themenschwerpunkte (Abbildung 1.1) und die dazugehörigen Forschungsfragen formuliert, die es zu beantworten gilt:

1. Wo liegen die Schwerpunkte der **Nutzung** von **OMOP** in der Forschung?
2. Welche **Anforderungen** an die Daten seitens der Forschungsgemeinschaft OHDSI unter Verwendung von OMOP existieren?
3. Welche **Inhibitoren in den Medikationsdaten** aus der klinischen Versorgung am Beispiel des UKD stehen im Widerspruch zu den zuvor identifizierten Anforderungen?
4. Welche **Maßnahmen** sind geeignet, um die identifizierten Lücken hervorgerufen durch die Inhibitoren zu reduzieren und so die Anforderungen zu erfüllen?
5. Wie kann eine **Bewertung** der Maßnahmen durchgeführt werden?
6. Wie kann die **Transparenz** verbleibender Limitierungen gewährleistet werden?

1.4 Struktur der Arbeit

Die vorliegende Arbeit ist in die Kapitel Einleitung, Hintergrund, Methoden, Ergebnisse und Diskussion geteilt.

Zunächst wird in der Einleitung (Kapitel 1) die Thematik eingeführt und es erfolgt die Darstellung der offenen Herausforderung sowie die Vorstellung der durch diese Arbeit zu beantwortenden Forschungsfragen.

Im Anschluss erfolgt in Kapitel 2 die Vorstellung notwendiger Informationen zu Konzepten, Datenmodellen, Werkzeugen und Terminologien, die für das Verständnis der Arbeit unverzichtbar sind.

Die Methodik der Arbeit wird in Kapitel 3 vorgestellt und gliedert sich in die Literaturrecherche, Anforderungsanalyse, Identifikation von Inhibitoren, Maßnahmen zur Reduktion von Schwachstellen, Bewertung der Maßnahmen und Schaffung von Transparenz.

Im Anschluss daran werden gemäß der Gliederung der Methoden die Ergebnisse in Kapitel 4 vorgestellt.

Abschließend wird die Arbeit in Kapitel 5 diskutiert. Die Ergebnisse werden dazu in den Kontext der gestellten Forschungsfragen gesetzt. Es erfolgt eine Bewertung, inwieweit die Arbeit die gestellten Fragen beantworten konnte. Außerdem werden die Stärken, aber auch existierende Limitierungen gezeigt. Ein Ausblick soll mögliche anschließende Themen zur Weiterführung dieser Arbeit in neuen Forschungsvorhaben aufzeigen.

2 Hintergrund

Um ein besseres Verständnis der vorliegenden Arbeit zu gewährleisten, sind einige Hintergrundinformationen notwendig. Sie sollen den Lesenden der Arbeit befähigen, relevante Konzepte, Datenmodelle, Werkzeuge und Terminologien zu verstehen und in den folgenden Kapiteln einordnen zu können und ermöglichen das gesamtheitliche Verständnis der Arbeit.

Zunächst wird in Abschnitt 2.1 der Begriff des Datenintegrationszentrum (DIZ) im Kontext der MI-I als landesweite Initiative eingeführt. Im Anschluss daran wird der Kerndatensatz der MI-I in Abschnitt 2.2 vorgestellt. OMOP stellt die Zielarchitektur dieser Arbeit vor und ist daher ein zentrales Element, welches in Abschnitt 2.3 detailliert eingeführt wird. Die Bereitstellung standardisierter Vokabulare in OMOP erfolgt über einen zentralen Dienst namens ATHENA, der in Abschnitt 2.4 beschrieben wird. Für den Transfer von Daten nach OMOP werden durch OHDSI entsprechende Werkzeuge bereitgestellt, welche in Abschnitt 2.5 betrachtet werden. Das OHDSI Data Quality Dashboard (DQD) wird für die Bewertung von Daten in OMOP eingesetzt und wird daher in Abschnitt 2.6 vorgestellt. Abschließend werden notwendige Hintergrundinformationen zu den für diese Arbeit relevanten medizinischen Terminologien ATC und RxNorm gegeben.

2.1 Datenintegrationszentrum

Die MI-I ist eine durch das BMBF bundesweit etablierte Förderrichtlinie, an der alle Universitätskliniken in Deutschland teilnehmen (Semler et al., 2018). Die Initiative fördert in zwei Phasen den Auf- und Ausbau einer Infrastruktur, um die bei der Behandlung von Patient:innen anfallenden Routinedaten in sogenannten DIZ zu konsolidieren und für die

Forschung nutzbar zu machen. Die maßgebliche Vision der MI-I ist die Generierung eines Mehrwertes in vielen Teilen des Gesundheitswesens zum Vorteil der Patient:innen durch optimierte Diagnose- und Behandlungsentscheidungen.

An 32 Standorten in Deutschland wurden Datenintegrationszentren aufgebaut oder befinden sich derzeit im Aufbau (Stand 15.05.2023). Sie sind in der Regel eng an die existierende Informationstechnik (IT) Infrastruktur des jeweiligen Universitätsklinikums mit einer Vielfalt an Datenlieferanten, wie beispielsweise den Krankenhausinformationssystemen, Laborinformationssystemen, intensivmedizinischen Systemen, onkologischen IT Systemen und anderen gekoppelt (TMF – Technologie- und Methodenplattform et al., 2023).

Zu den wichtigen Aufgaben eines DIZ gehören die Aufbereitung der Daten unter Gewährleistung der Interoperabilität auf syntaktischer und semantischer Ebene, die Sicherung der Datenqualität sowie die Einhaltung aller rechtlichen Rahmenbedingungen zum Schutz der Persönlichkeitsrechte der Patient:innen gemäß geltender Datenschutz-Grundverordnung (DSGVO) und die Verwaltung der Datennutzungsanfragen.

Die im Rahmen dieser Arbeit genutzten Routinedaten wurden durch das Datenintegrationszentrum des Universitätsklinikums Carl Gustav Carus Dresden (UKD) vollständig anonymisiert bereitgestellt.

2.2 Medizininformatik Initiative Kerndatensatz

Wie in Abschnitt 2.1 eingeführt, liegt ein Schwerpunkt der Aufbereitung der Daten innerhalb der DIZ auf der Gewährleistung der Interoperabilität der Daten zur standortübergreifenden Nutzung. Deshalb wird auf nationaler Ebene im Rahmen des Nationales Steuerungsgremium (NSG) als Governance Struktur der MI-I, in der Arbeitsgruppe Interoperabilität, der sogenannte Kerndatensatz (KDS) der MI-I als Anforderung an die DIZ erarbeitet.

Der MI-I KDS lässt sich, wie in Abbildung 2.1 ersichtlich, unterteilen in die Basis- und Erweiterungsmodule. Im Rahmen dieser Arbeit sind ausschließlich die Basismodule, im Speziellen das Modul Medikation von Relevanz. Zu den Basismodulen gehören neben der Medikation auch die Module Patient, Fall, Diagnose, Prozedur, Laborbefund. Für

alle Basismodule liegt ein Implementierungsleitfaden unter Verwendung von HL7 FHIR Ressourcen vor.

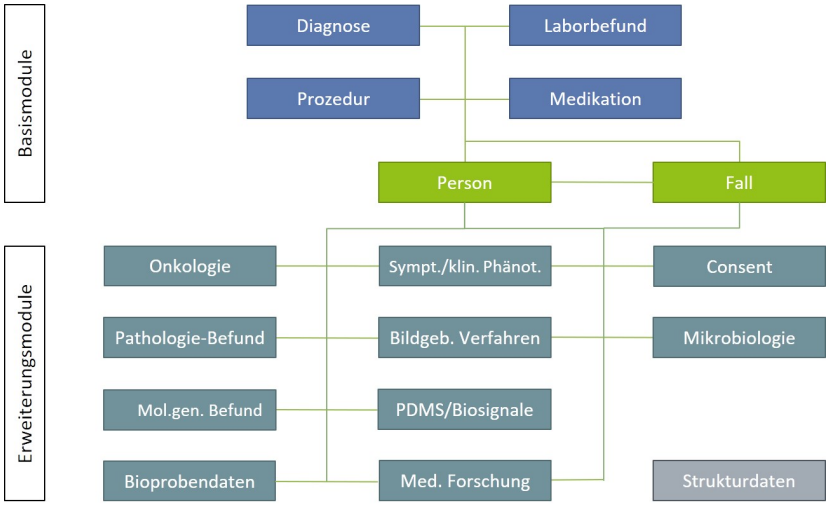


Abbildung 2.1: Blockschema MI-I Kerndatensatz mit Basis- und Erweiterungsmodulen (TMF e.V., 2023)

Gemäß der Modulbeschreibung Medikation Version 1.0.9 mit dem Veröffentlichungsdatum 19.07.2021 soll für die Medikationen der Wirkstoff abrufbar sein (Zautke et al., 2021). Die Abbildung 2.2 zeigt die Definition des MI-I KDS Moduls Medikation in der genannten Version.

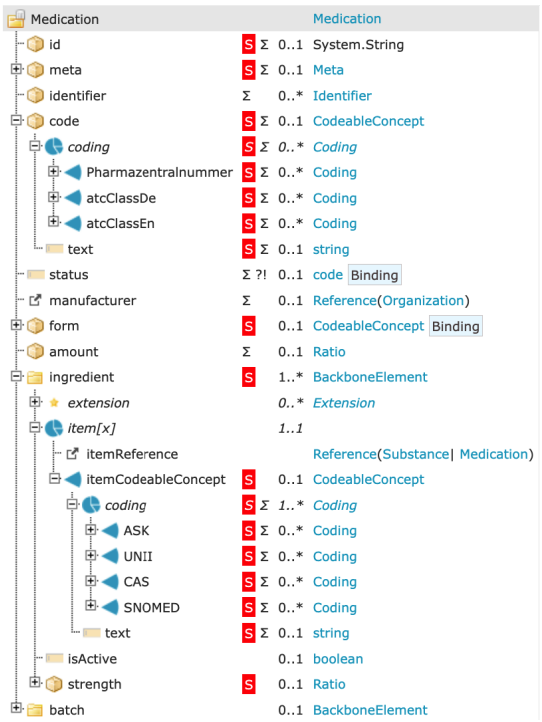


Abbildung 2.2: Snapshot Medication Modul des MI-I KDS Version 1.0.9

2.3 OMOP Common Data Model

Das DIZ des UKD stellt RWD als FHIR Ressourcen gemäß des MI-I KDS bereit. Für eine Teilnahme an internationalen Forschungsvorhaben im Rahmen der OHDSI Forschungsgemeinschaft ist eine Überführung der Daten unter Wahrung der semantischen Bedeutung nach OMOP notwendig. OMOP wird seit dem Jahr 2014 von der OHDSI Community entwickelt und dient als Grundlage für Forschungsprojekte innerhalb dieser Forschungsgemeinschaft (OHDSI, 2019). Ein Datenmodell dient grundsätzlich der Harmonisierung von Daten, gewährt ein gegenseitiges Verständnis von Daten durch entsprechende Standardisierung und ermöglicht die Nutzung und Auswertung in gemeinsamen Forschungsvorhaben unter Wahrung der Interoperabilität. Die Standardisierung von RWD ist insbesondere wichtig, wenn große Datenmengen aus unterschiedlichsten Systemen und Standorten für gemeinsame Forschungsvorhaben genutzt und analysiert werden sollen.

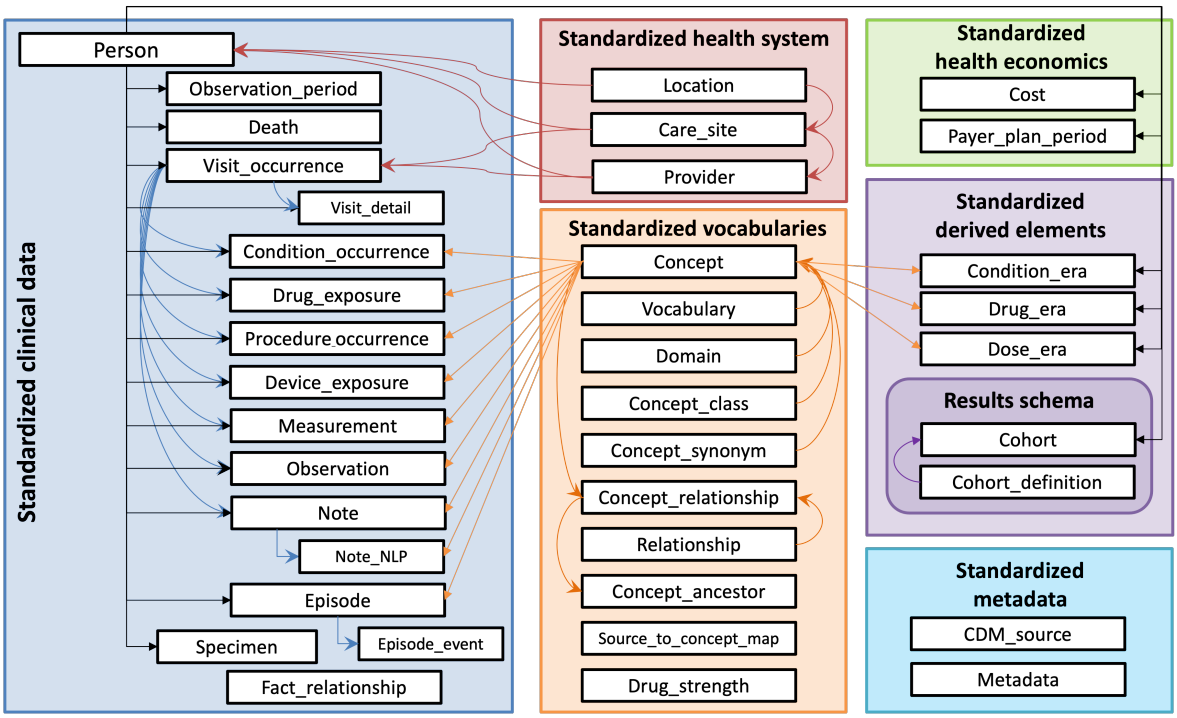


Abbildung 2.3: Datenmodell OMOP Version 5.4 (OHDSI, 2023a)

OMOP legt die Datenstruktur durch die Definition unterschiedlicher Entitäten als relationale Datenbanktabellen fest. Außerdem fordert es die Verwendung von standardisierten medizinischen Terminologien für klinische Fakten, je nach klinischer Domäne. Dabei unterscheidet das Datenmodell OMOP die Daten in klinische Fakten (*Standardized clinical data*) und standardisierte Vokabulare (*Standardized Vocabularies*) wie der Abbildung 2.3 zu entnehmen ist.

Klinische Fakten wie Diagnosen (*condition_occurrence*), Medikamente (*drug_exposure*) und Prozeduren (*procedure_occurrence*) werden in OMOP durch die Verwendung von Konzepten eindeutig abgebildet. Diese Konzepte verweisen auf existierende Vokabulare, die in der *concept* Tabelle von OMOP hinterlegt sind. Eine detaillierte Beschreibung der Spalten der *concept* Tabelle findet sich in Tabelle 2.1. Am Beispiel des Wirkstoffes Pantoprazol, also Konzept der Terminologie ATC sind die Inhalte der Spalten zudem exemplarisch dargestellt.

Tabelle 2.1: Spalten der concept Tabelle des Datenmodells OMOP mit einem Beispieldatensatz

Spaltenname der concept Tabelle	Beschreibung	Beispiel
concept_id	eindeutige ID, Verwendung als Fremdschlüssel in den Tabellen der klinischen Fakten für den Verweis auf das entsprechende Konzept	21600097
concept_name	Bezeichnung, Beschreibung des Konzepts	pantoprazole; systemic
domain_id	Verweis auf die entsprechende Domaine, zuständig für die Entscheidung in welcher Tabelle klinische Fakten innerhalb von OMOP gespeichert werden	Drug
vocabular_id	Verweis auf das entsprechende Vokabular (Terminologie), zu dem ein Konzept gehört	ATC
concept_class_id	Verweis auf die entsprechende Klasse eines Konzeptes, beispielsweise: Wirkstoff, klinischer Befund	ATC 5th
standard_concept	Kennzeichnen, wenn ein Konzept ein Standardkonzept (S) oder eine Klassifikation (C) ist, ansonsten NULL	C
concept_code	Identifikation des Konzepts innerhalb der Terminologie, beispielsweise ICD10GM code	A02BC02
valid_start_date	Datum, seit dem ein Konzept nutzbar ist	1970-01-01
valid_end_date	Datum, bis wann ein Konzept gültig ist, standardmäßig 31.12.2999, nur bei Gültigkeitsablauf das konkrete Datum	2099-12-31
invalid_reason	Kennzeichen, wenn ein Konzept aktuell nicht mehr gültig ist (U), ansonsten NULL	(null)

Da RWD je nach Herkunft und Ursprung nicht immer einem Standardkonzept entsprechen, sind ETL Prozesse zur Harmonisierung gemäß OMOP erforderlich. Studien innerhalb der OHDSI Forschungsgemeinschaft, die auf Basis von OMOP durchgeführt werden, setzen die Harmonisierung der Quelldaten unter Verwendung der Standardkonzepte aus den Vokabularen voraus. Neben der *concept* Tabelle enthält die Tabelle *concept_relationship* Informationen über Beziehungen zwischen verschiedenen Konzepten. Diese Tabelle verknüpft Konzepte über einen bestimmten Beziehungstyp miteinander und ermöglicht die Übersetzung von Konzept von einer Terminologie in eine andere. Zur Beschreibung von Hierarchien innerhalb einer Terminologie werden ebenfalls Beziehungen definiert.

OMOP definiert nicht nur die Struktur der Daten, sondern bietet auch die Harmonisierung der Daten durch die Verwendung standardisierter Terminologien in den *Standardized Vocabularies*. Dadurch wird ein gemeinsames semantisches Verständnis und die notwendige Interoperabilität für standortübergreifende Forschungsprojekte ermöglicht.

2.4 ATHENA und Standardisierte Vokabulare

Wie im Abschnitt 2.3 eingeführt, sind die standardisierten Vokabulare integraler Bestandteil des OMOP CDM. Die Webanwendung ATHENA bietet den Zugriff auf die aktuellste verfügbare Version aller Vokabulare und stellt eine Historisierung der Vokabulare zur Verfügung. So können RWD auch mit den zum Entstehungszeitpunkt gültigen Konzepten in OMOP gespeichert werden. Es ermöglicht die Suche und Filterung nach einzelnen Konzepten und bietet die Möglichkeit, Vokabulare für eine Verwendung in einer OMOP Datenbank zusammenzustellen und abzurufen.

Die verfügbaren Vokabulare werden dabei durch OHDSI nicht immer selbst erstellt. Vorhandene Vokabulare wie beispielsweise Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC), ATC, International Classification of Diseases Version 10 (ICD10) werden lediglich in das notwendige Format für die Verwendung in OMOP überführt. Das OHDSI Team verwaltet zur Nachverfolgung der bereitgestellten Vokabulare ein GitHub Repository „PALLAS“ (Odysseus Data Services Inc., 2023). In diesem Repository befindet sich für alle auf ATHENA verfügbaren Vokabulare eine Dokumentation zu den verwendeten Quellsystemen und Versionen sowie die durch OHDSI durchgeführten Schritte zur Generierung und Aktualisierung der Vokabulare.

2.5 OHDSI ETL Werkzeuge

Als Grundlage für die Entwicklung von ETL Prozessen für den Transfer von RWD nach OMOP ist ein Profiling der Daten von Vorteil. OHDSI stellt mit der Software WhiteRabbit (OHDSI, 2023c) eine Java-basierte Lösung zur Verfügung, die es ermöglicht, Daten als Comma Separated File (CSV) Dateien oder über den Zugriff auf Datenbanken aus Tabellen in Quellsystemen wie beispielsweise aus dem Krankenhausinformationssystem (KIS) einzulesen und zu analysieren.

WhiteRabbit liefert zu den eingelesenen Daten detaillierte Informationen zu den Datenfeldern, wie Datentypen und Mengenverteilung der Daten zurück. Es wird ein Bericht generiert, der als Referenz für die Implementierung des ETL Prozesses genutzt werden kann. Der generierte Bericht kann die Grundlage der Entwicklung des ETL Prozesses bilden.

Rabbit-In-a-Hat (OHDSI, 2023b) ist ein auf dem Bericht von WhiteRabbit aufsetzendes Werkzeug zur Generierung eines Dokuments für das Mapping von Datenelementen der Quellsysteme nach OMOP. Rabbit-In-a-Hat bietet hierzu eine grafische Benutzeroberfläche an, um die Datenelemente pro Tabelle der Quelle mit den Datenelementen der OMOP Tabelle zu verbinden. Rabbit-In-a-Hat generiert abschließend eine Dokumentation, welche das gesamte Mapping beinhaltet, die als Grundlage für die Implementierung des ETL Prozesses dienen kann. Rabbit-In-a-Hat wird zusammen mit WhiteRabbit geliefert und ist konzipiert, WhiteRabbit Dokumente zu verarbeiten.

2.6 OHDSI Data Quality Dashboard

Datenqualität stellt eine wichtige Voraussetzung für die Qualität der Ergebnisse aus retrospektiven Beobachtungsstudien dar. Um die Datenqualität der RWD nach deren Harmonisierung und Transfer nach OMOP zu prüfen, hat die OHDSI Forschungsgemeinschaft in Zusammenarbeit mit dem EHDEN Projekt das OHDSI DQD entwickelt und als Open-Source-Software bereitgestellt (Blacketer, Defalco et al., 2021).

Das DQD ist ein Werkzeug, um Daten in einer OMOP Datenbank transparent und nachvollziehbar gemäß der Dimensionen der Vollständigkeit, Konformität und Plausibilität zu prüfen. Es orientiert sich damit an dem von Kahn et al. entwickelten Framework zur Bewertung von Datenqualität (Kahn et al., 2016). Das OHDSI DQD der Version 2.0.0 implementiert 20 unterschiedliche Typen von Prüfungen (siehe Tabelle 2.2), welche auf alle Tabellen einer OMOP Datenbank angewendet werden können.

Insgesamt erfolgen durch das DQD mehrere tausend einzelne Prüfungen in den oben genannten Dimensionen. In der vorliegenden Arbeit sind ausschließlich die Prüfungen im Kontext der Konformität und Vollständigkeit der Medikationsdaten in der OMOP Tabelle *drug_exposure* relevant, da sich die sieben Prüftypen der Kategorie Plausibilität, wie in Tabelle 2.2 dargestellt, ausschließlich auf den Inhalt der Daten beziehen und im Sinne einer semantischen Interoperabilität keine Bedeutung haben.

Tabelle 2.2: Typen der Prüfungen des OHDSI DQD

Typ der Prüfung	Beschreibung	Level	Dimension
measurePersonCompleteness	Prüft die Anzahl an Personen in der OMOP Datenbank, die nicht mindestens einen Eintrag für einen klinischen Fakt haben.	Tabelle	Vollständigkeit
cdmField	Prüft in allen Tabellen der OMOP Datenbank, ob alle Felder (Spalten) gemäß der Definition des OMOP CDM vorhanden sind.	Datenfeld	Konformität
isRequired	Prüft alle Daten der OMOP Datenbank auf die Existenz von Null-Werten, für Spalten die gemäß OMOP CDM Definition keine Nullwerte enthalten dürfen	Datenfeld	Konformität
cdmDatatype	Prüft alle Daten der OMOP Datenbank auf die Konformität der Datentypen gemäß der OMOP CDM Definition	Datenfeld	Konformität
isPrimaryKey	Prüfung der Eindeutigkeit aller Datenfelder von Spalten, die als Primärschlüssel für eine Tabelle definiert sind	Datenfeld	Konformität
isForeignKey	Prüfung aller Fremdschlüssel gemäß OMOP CDM Definition einen Wert im zugehörigen Primärschlussfeld haben.	Datenfeld	Konformität
fkDomain	Prüfung aller Felder auf korrekte Verwendung von Konzepten, gemäß der Anforderungen hinsichtlich der Domain (domain) im OMOP-CDM.	Datenfeld	Konformität
fkClass	Prüfung aller Felder auf korrekte Verwendung von Konzepten, gemäß der Anforderungen hinsichtlich der Klasse (concept_class) im OMOP-CDM.	Datenfeld	Konformität
isStandardValidConcept	Prüfung der semantischen Korrektheit der Daten. Es wird geprüft, ob gemäß der OMOP CDM Definition die korrekten Standardkonzepte verwendet werden, beispielsweise das Vokabular RxNorm für Medikamente	Datenfeld	Konformität
measureValueCompleteness	Ermittlung aller Felder, in denen ein NULL Wert existiert. Dies Prüfung ist unabhängig davon, ob eine Spalte der Definition nach in OMOP keine Nullwerte enthalten darf (vgl. isRequired Prüfung)	Datenfeld	Vollständigkeit
standardConceptRecordCompleteness	Ermittlung der Anzahl der Einträge, bei denen die concept_id = 0 ist	Datenfeld	Vollständigkeit
sourceConceptRecordCompleteness	Ermittlung der Anzahl der Einträge, bei denen die source_concept_id = 0 ist	Datenfeld	Vollständigkeit
sourceValueCompleteness	Prüfung der Anzahl an eindeutigen Einträgen, bei denen ein Konzept mit der concept_id = 0 existiert.	Datenfeld	Vollständigkeit
plausibleValueLow	Prüft bestimmte Spalten der OMOP Datenbank, ob ein bestimmter Wert kleiner als ein Grenzwert ist.	Datenfeld	Plausibilität
plausibleValueHigh	Prüft bestimmte Spalten der OMOP Datenbank, ob ein bestimmter Wert größer als ein Grenzwert ist.	Datenfeld	Plausibilität
plausibleTemporalAfter	Prüfung, ob es Daten für eine Person in der Datenbank gibt, die zeitlich vor dem Geburtsdatum liegen	Datenfeld	Plausibilität
plausibleDuringLife	Prüfung, ob es Daten für eine Person in der Datenbank gibt, die zeitlich nach dem Todesdatum liegen	Datenfeld	Plausibilität
plausibleValueLow	Prüfung der Plausibilität von Werten, die gemäß ihrer Konzept ID einen bestimmten Schwellenwert nicht unterschreiten sollten, weil dieser klinisch nicht plausibel sein kann (z.B.: Laborwerte)	Konzept	Plausibilität
plausibleValueHigh	Prüfung der Plausibilität von Werten, die gemäß ihrer Konzept ID einen bestimmten Schwellenwert nicht überschreiten sollten, weil dieser klinisch nicht plausibel sein kann (z.B.: Laborwerte)	Konzept	Plausibilität
plausibleGender	Hier wird geprüft, welche Datensätze existieren, bei denen verwendete Konzepte eine falsche Geschlechtszuordnung in den dazugehörigen Personen haben (z.B.: männliche Personen bei denen eine Schwangerschaft vorliegt).	Konzept	Plausibilität

2.7 Relevante Terminologien

Neben den strukturellen Aspekten der Daten, wie ihre Organisation und Formatierung, bedarf es insbesondere für die Gewährleistung der semantischen Interoperabilität einer Harmonisierung der Daten unter Verwendung entsprechender medizinischer Terminologien (De Mello et al., 2022). Die für diese Arbeit relevanten Terminologien ATC und RxNorm werden im Folgenden eingeführt.

2.7.1 Die Anatomisch-Therapeutisch-Chemische (ATC) Klassifikation

ATC wurde erstmals im Jahr 1975 von der European Pharmaceutical Market Research Association (EPHMRA) entwickelt und wird seither jährlich aktualisiert. Seit dem Jahr 1990 stellt die World Health Organization (WHO) die sogenannte ATC WHO Version zur Verfügung. Die beiden ATC Versionen (EPHMRA und WHO) unterscheiden sich dabei nicht nur in ihrem Nutzungszweck und können nicht direkt miteinander verglichen werden (Norwegian Institute of Public Health WHO Collaborating Centre for Drug Statistics Methodology et al., 2021). Die WHO empfehlt die ATC WHO Klassifikation für die Nutzung in der Erforschung der Verwendung von Medikamenten (WHO International Working Group for Drug Statistics Methodology, 2022).

Die ATC WHO Klassifikation unterscheidet dabei Substanzen auch auf Grundlage des Organs oder Systems, auf welches sie wirken. Es ist daher möglich, dass eine Substanz, die an mehreren Organen/Systemen zum Einsatz kommt, durch unterschiedliche ATC Codes repräsentiert wird. Als Beispiel sei hier der Wirkstoff *Clonidin* genannt. Als Mittel gegen Blutdruck wird *Clonidin* als ATC Code C02AC01 dargestellt und unterscheidet sich daher je nach Anwendung beispielsweise im Falle als Mittel gegen Migräne in der Repräsentation durch einen anderen ATC Code N02CX02. Relevant für diese Arbeit sind nur die ATC Versionen der WHO und die angepasste deutsche Version, die in dieser Arbeit durchgängig als ATC-German Modification (ATC-GM) bezeichnet wird.

In Deutschland ist das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) für die Bereitstellung der in Deutschland amtlich gültigen Fassung der ATC-GM Klassifikation verantwortlich. Die Anpassungen der ATC-GM Version tragen den regulatorischen Anforderungen und Gesetzgebungen in Deutschland Rechnung. Ein Beispiel sind zusätzliche Unterkategorien auf nationaler Ebene, die in der WHO Version nicht existieren, wie beispielsweise

der ATC-GM Code A11BA01 für *Multivitamine, rein*, die in der WHO Version nur über den ATC Code A11BA dargestellt werden. Die ATC WHO Version ist als Vokabular im OMOP CDM verfügbar und zu nutzen. Eine ATC-GM Version existiert derzeit nicht als Vokabular in OMOP und wird auch nicht über ATHENA angeboten.

Die Einteilung von Wirkstoffen in ATC WHO und ATC-GM erfolgt basierend auf anatomischen Kriterien, einschließlich des Organs oder Organsystems, an dem sie wirken, sowie unter Berücksichtigung ihrer therapeutischen und chemischen Eigenschaften. Die Einteilung erfolgt dabei in 14 unterschiedliche Gruppen (siehe Tabelle 2.3).

Tabelle 2.3: ATC-GM Gruppen (Level 1) und Bezeichnung

ATC Gruppe (Level 1)	Bezeichnung gemäß ATC-GM
A	Alimentäres System und Stoffwechsel
B	Blut und Blutbildende Organe
C	Kardiovaskuläres System
D	Dermatika
G	Urogenitalystem und Sexualhormone
H	Systemische Hormonpräparate. exkl. Sexualhormone und Insuline
J	Antiinfektiva zur systemischen Anwendung
L	Antineoplastische und Immunmodulierende Mittel
M	Muskel- und Skelettsystem
N	Nervensystem
P	Antiparasitäre Mittel, Insektizide und Repellenzien
R	Respirationstrakt
S	Sinnesorgane
V	Varia

Es gibt innerhalb von ATC WHO und ATC-GM 5 hierarchisch geordnete Ebenen (World Health Organization (WHO), 2023):

- ATC Ebene 1: 14 anatomischen und pharmakologischen Gruppen (vgl. Tabelle 2.3)
- ATC Ebene 2: Pharmakologische und therapeutische Unterkategorien
- ATC Ebene 3 und 4: Chemische, pharmakologische oder therapeutische Unterkategorie
- ATC Ebene 5: chemische Substanz, Wirkstoff

Jedem in Deutschland zugelassenen Fertig-Arzneimittel wird durch den Hersteller ein oder mehrere (bei Kombinationsprodukten möglich) ATC Code(s) auf Wirkstoffebene (ATC Level 5) zugeordnet.

2.7.2 RxNorm

RxNorm ist eine Terminologie für Arzneimittel in den USA und wird von der National Library of Medicines (NLM) entwickelt und bereitgestellt. RxNorm soll die Interoperabilität von Arzneimittelbezeichnungen zwischen verschiedenen Gesundheitssystemen in den USA fördern.

RxNorm ist Teil des Unified Medical Language System (UMLS) Metathesaurus. RxNorm ermöglicht die eindeutige Benennung des Wirstoffes, der Dosis und der Darreichungsform eines Medikamentes. RxNorm beinhaltet verschiedene Term Typen (Term Types (TTY)), die es ermöglichen auf unterschiedlicher fachlicher Ebene Informationen zu Arzneimittel zu erhalten. Zu den wichtigsten RxNorm TTY gehören die Folgenden:

- Wirkstoff (Ingredient)
- Klinische Werkstoffkomponente (Clinical Drug Component), bestehend aus Wirkstoff und Stärke
- Klinische Form eines Medikaments (Clinical Drug Form), bestehend aus Wirkstoff und Darreichungsform
- Kinisches Medikament (Clinical Drug), bestehend aus Wirkstoff, Stärke und Darreichungsform
- Markenname (Brand Name)
- Markenmedikament (Branded Drug), bestehend aus Wirkstoff, Stärke, Darreichungsform und Markenname

Eine vollständige Liste aller verfügbaren TTY kann auf der Webseite des NLM gefunden werden (National Library of Medicine, 2023).

RxNorm enthält seit August 2013 auch ATC als Terminologie und verfügt daher über ein Mapping zwischen den RxNorm und den ATC auf der Ebene der Wirkstoffe. In RxNorm wird, anders als in ATC jeder Wirkstoff durch exakt einen eindeutigen Code abgedeckt. Eine Repräsentation von Wirkstoffen in mehreren Codes existiert in RxNorm nicht. Daher ist ein Mapping zwischen RxNorm und ATC in diesen Fällen nicht über eine eindeutige Verbindung in beide Richtungen abbildbar. Die Abdeckung dieses Mappings ist daher nicht vollumfänglich (Bodenreider et al., 2014).

3 Materialien und Methoden

3.1 Material

In diesem Kapitel wird das für die vorliegende Arbeit relevante Material beschrieben. Dazu gehören die Datensätze (Abschnitt 3.1.1), das Mapping der Datenfelder der Medikationsverordnungen nach OMOP (Abschnitt 3.1.2) und die für diese Arbeit etablierte technische Infrastruktur (Abschnitt 3.1.3).

3.1.1 Verwendete Daten

In diesem Abschnitt werden die für diese Arbeit relevanten Datensätze eingeführt und beschrieben, sodass eine eindeutige Zuordnung der Aussagen in Folgekapiteln zu den verwendeten Daten und Details hergestellt werden kann. Die folgenden Datensätze finden in dieser Arbeit Verwendung:

1. Datensatz - Medikationsverordnungen (DS-Med)
2. Datensatz - Hauskatalog für Arzneimittel (DS-Katalog)
3. Datensatz - unstrukturierte Verordnungen aus DS-Med , gruppiert nach dem Freitext für die Medikation (DS-Gruppiert)
4. Datensatz - Teilmenge von DS-Gruppiert, die häufigsten 1000 Freitexte für Medikation (DS-Top1000)
5. Datensatz - ATC Vokabular aus der Tabelle concept in OMOP (DS-ATC)
6. Datensatz - Beziehungen zwischen allen Konzepten der OMOP Vokabulare ATC und RxNorm (DS-Relation)

Eine tabellarische Übersicht über die Datensätze mit ihren Datenelementen befindet sich in Tabelle 3.1.

Die Verordnungen von Medikamenten (DS-Med) Die Originaldaten wurden wie in Abbildung 3.2 im KIS ORBIS von Dedalus in dem ORBIS-Modul „KURV"" elektronisch dokumentiert. Als Datengrundlage dienen 1.768.153 Medikationsverordnungen des UKD der Jahre 20216 bis 2020 welche durch das DIZ bereits anonymisiert ohne Bezug auf einzelne Behandlungsfälle oder Personen bereitgestellt wurden. Eine Einschränkung auf bestimmte Krankheitsbilder erfolgte nicht. Medikationsverordnungen aus anderen Systemen (z. B. Intensivstationen und Chemotherapie) wurden ausgeschlossen. Das Datenelement *STRUCTURE* zeigt an, ob eine Medikationsverordnung aus den verfügbaren Medikamenten des Hauskatalogs ausgewählt wurde und damit in strukturierter Form inklusive ATC Code vorliegt oder es sich bei dem Inhalt des Datenelements *MEDICATION* um einen unstrukturierten Freitext handelt.

Der Hauskatalog für Arzneimittel (DS-Katalog) wurde am 16. November 2021 aus dem Warenwirtschaftssystem (SAP ERP) des UKD exportiert. Er enthält alle zu diesem Zeitpunkt in der Hausliste des UKD verfügbaren Fertigarzneimittel, die innerhalb des KIS bei der elektronischen Verordnung von Medikamenten aus einer bereitgestellten Liste ausgewählt werden können. Diese Liste enthält den Namen eines Arzneimittels, den Namen des Wirkstoffes, den ATC Level-5 Code und Informationen zu Dosis und Einheit des Arzneimittels. Sofern beide Versionen vorhanden sind, werden Wirkstoffname und der ATC Code in der ATC WHO und ATC-GM Version (vgl. Abschnitt 2.7) im Hauskatalog angegeben.

Aus DS-Med abgeleitete Datensätze (DS-Gruppiert und DS-Top1000) werden ebenfalls für diese Arbeit verwendet. Für DS-Gruppiert und DS-Top1000 wurden ausschließlich die Medikationsverordnungen genutzt, die nicht aus dem Hauskatalog im KIS ausgewählt, sondern als Freitexte eingegeben wurden. Dazu werden die unstrukturierten Freitexte des Datenelements *MEDICATION* aus DS-Med nach ihrer Häufigkeit aggregiert. Ein neues Datenelement *FREQUENCY*, welches die Häufigkeit des Freitextes aufsummiert angibt, wurde dem DS-Gruppiert hinzugefügt. Die Gruppierung wurde in Python mithilfe der Python-Bibliothek Pandas und ihrer groupby-Funktion implementiert.

Bei DS-Top1000 handelt es sich um eine Teilmenge der 1000 am häufigsten vorkommenden Freitexte im Datenelement *MEDICATION* des DS-Gruppiert. DS-Top1000 wird im weiteren Verlauf dieser Arbeit für die manuelle Evaluation der Ergebnisse der automatisierten Zuordnung möglicher ATC-GM Codes verwendet. Die aus dem Hauskatalog des UKD (DS-Katalog) verordneten Medikamente sind als strukturierte Daten gekennzeichnet (z. B. „IBUPROFEN STADA 600 mg Suppositorien | [Ibuprofen-Natrium, Ibuprofen]“), basierend der Kennzeichnung über die Spalte *STRUCTURE* in DS-Med. Verordnungen von Medikamenten, die nicht aus dem Arzneimittelkatalog ausgewählt wurden, werden in der Spalte *STRUCTURE* als unstrukturierte Daten gekennzeichnet (z. B. „Ibuprofen 600“ und „Ibuprofen“).

Das Vokabular ATC aus der OMOP Datenbank (DS-ATC) wurde aus der OMOP Tabelle *concept* als CSV Datei exportiert und besteht aus den Datenelementen wie in Tabelle 3.1 dargestellt. Für diese Arbeit beschränkt sich der Export auf alle im betrachteten Zeitraum der Medikationsverordnungen (2016 bis 2020) gültigen ATC Codes gemäß der ATC Version der WHO.

Quelltext 3.1: SQL Abfrage der OMOP Tabelle concept_relationship

```
SELECT con1.concept_id AS id_atc, con1.concept_code AS code_atc, con1.concept_name AS name_atc,
       rel.relationship_id,
       con2.concept_id AS id_rx, con2.concept_name AS name_rx, con2.concept_class_id AS class_rx
FROM concept con1
JOIN concept_relationship rel ON concept_id = concept_id_1
JOIN concept con2 ON concept_id_2 = con2.concept_id
WHERE con1.vocabulary_id = 'ATC'
  AND (con1.invalid_reason IS NULL
       OR (con1.valid_end_date >= '2020-12-31' AND con1.valid_start_date <= '2016-01-01'))
  AND rel.relationship_id IN ('ATC - RxNorm pr lat', 'ATC - RxNorm pr up',
                              'ATC - RxNorm sec lat', 'ATC - RxNorm sec up', 'Maps to');
```

Die Beziehungen zwischen den Vokabularen ATC und RxNorm wurden unter Verwendung des Structured Query Language (SQL) Quellcode 3.1 als Datensatz DS-Relation aus einer OMOP Datenbank unter Verwendung der Tabellen *concept* und *concept_relationship* generiert. Es werden ausschließlich die Beziehungen vom Typ „ATC - RxNorm pr lat“, „ATC - RxNorm pr up“, „ATC - RxNorm sec lat“, „ATC - RxNorm sec up“ und „Maps to“ inkludiert.

Tabelle 3.1: Datensätze und deren Datenelemente inklusive Beschreibung

Datenelement	Datentyp	Beschreibung
DS-Med: initialer Datensatz, alle Medikationsverordnungen		
MEDICATION	String	Freitext oder vordefinierter Wert, ausgewählt aus einer Liste gemäß Hauskatalog (DS-Katalog). Die Liste enthält den Namen des Medikaments und des Wirkstoffes
start_date	Number	Anonymisiertes Verordnungsdatum, erlaubt nur Rückschluss auf das Jahr
end_date	Number	Anonymisiertes Verordnungsdatum, erlaubt nur Rückschluss auf das Jahr
STRUCTURE	Boolean	TRUE bei Auswahl aus Hauskatalog, ansonsten FALSE bei Freitexteingaben
ATC_L5	String	ATC-GM Code, verfügbar bei Auswahl aus Hauskatalog (STRUCTURE = TRUE)
DS-Katalog: Hauskatalog Arzneimittel		
Product_name	String	Produktname aus dem Warenwirtschaftssystem (SAP ERP)
Ingredient_name	String	Wirkstoffname, wie im Hauskatalog verfügbar, gemäß ATC-GM Version
ATC_Code_DE	String	ATC Code Level 5 gemäß deutscher Version durch BfArM
ATC_Code_WHO	String	ATC Code Level 5 gemäß WHO Version
MEDICATION	String	Gruppierung unstrukturierte Freitexte
FREQUENCY	Number	Häufigkeit des Vorkommens der Texte im Datenelement MEDICATION
Step1	String	Resultat Algorithmus 1, entweder ATC-GM Code oder leer, wenn ergebnislos
Step2	String	Resultat Algorithmus 2, entweder ATC-GM Code oder leer, wenn ergebnislos
Step3	String	Resultat Algorithmus 3, entweder ATC-GM Code oder leer, wenn ergebnislos
DS-Gruppiert: DS-Med gruppiert basierend auf dem Datenelement MEDICATION		
MEDICATION	String	Gruppierung unstrukturierte Freitexte
FREQUENCY	Number	Häufigkeit des Vorkommens der Texte im Datenelement MEDICATION
Step1	String	Resultat Algorithmus 1, entweder ATC-GM Code oder leer, wenn ergebnislos
Step2	String	Resultat Algorithmus 2, entweder ATC-GM Code oder leer, wenn ergebnislos
Step3	String	Resultat Algorithmus 3, entweder ATC-GM Code oder leer, wenn ergebnislos
DS-Top1000: Teilmenge des Datensatzes DS-Gruppiert, die 1000 häufigsten Freitexte in MEDICATION		
MEDICATION	String	Gruppierung unstrukturierte Freitexte
FREQUENCY	Number	Häufigkeit des Vorkommens der Texte im Datenelement MEDICATION
Step1	String	Resultat Algorithmus 1, entweder ATC-GM Code oder leer, wenn ergebnislos
Step2	String	Resultat Algorithmus 2 entweder ATC-GM Code oder leer, wenn ergebnislos
Step3	String	Resultat Algorithmus 3 entweder ATC-GM Code oder leer, wenn ergebnislos
Eval1	Boolean	Algorithm 1 Ergebnis Evaluation
Eval2	Boolean	Algorithm 2 Ergebnis Evaluation
Eval3	Boolean	Algorithm 3 Ergebnis Evaluation
True12	Boolean	TRUE, wenn Algorithmus 1 und 2 das gleiche Ergebnis liefern
True13	Boolean	TRUE, wenn Algorithmus 1 und 3 das gleiche Ergebnis liefern
True23	Boolean	TRUE, wenn Algorithmus 2 und 3 das gleiche Ergebnis liefern
True123	Boolean	TRUE, wenn Algorithmus 1, 2 und 3 das gleiche Ergebnis liefern
CORRECT	String	Korrigierter ATC-GM Code, für alle fehlerhaft ermittelten ATC Codes
COMMENTS	String	Kommentare während Evaluation, wenn notwendig
FINAL	String	Final korrekter ATC Code, entweder auf Basis der korrekten Ergebnisse mindestens eines Algorithmus oder manuell während der Evaluation

DS-ATC: OMOP Vokabular ATC gemäß der Tabelle concept aus der OMOP Datenbank		
concept_id	Number	Eindeutiger Identifier für das Konzept
concept_name	String	Beschreibender Name für das Konzept
domain_id	String	Für ATC Codes ist die Domaine immer „Drug“
vocabulary_id	String	Gibt den Namen des Vokabulars ein, hier „ATC“
concept_class_id	String	Gibt die Klasse innerhalb des Vokabulars an, hier beispielsweise ATC 5th, ATC 4th
standard_concept	String	Zeigt ob das Konzept ein Standard-Konzept ist, für ATC immer „C“; also eine Klassifikation, aber kein Standard-Konzept in der Domäne „Drug“
concept_code	String	Entspricht dem ATC Code
valid_start_date	Date	Datum für den Gültigkeitsbeginn des Konzepts
valid_end_date	Date	Datum für das Gültigkeitsende des Konzepts
invalid_reason	String	Zeigt Gültigkeit eines Konzepts an und einen Grund bei Ungültigkeit, ist bei gültigen Konzepten NULL
DS-Relation: Beziehungen zwischen den Konzepten der OMOP Vokabulare ATC und RxNorm		
id_atc	Number	Eindeutiger Identifier für das Konzept
code_atc	String	Entspricht dem ATC Code
name_atc	String	Beschreibender Name für das Konzept
relationship_id	String	Entspricht einer der Beziehungstypen, die eingeschlossen wurden beim Export
id_rx	Number	Eindeutiger Identifier für das Konzept
name_rx	String	Beschreibender Name für das Konzept
class_rx	String	Konzept Typ innerhalb des Vokabulars RxNorm, zum Beispiel: Ingredient

3.1.2 Datentransfer

Um die Medikationsverordnungen des UKD nach OMOP zu transferieren, wurde der Datensatz DS-Med zunächst unter Verwendung des OHDSI Werkzeugs WhiteRabbit (Version 0.10.7) gescannt, um einen Report zu generieren, der Informationen zur Ausprägung der Daten in den einzelnen Datenelementen (vgl. Tabelle 3.1) von DS-Med enthält. Der Report ermöglicht einen Überblick über Ausprägung der Daten pro Datenelement inklusive deren Datenausprägungen.

Dieser Report wurde für die Erstellung des Mappings der Medikationsdaten (Datensatz DS-Med) nach OMOP in dem Werkzeug OHDSI Rabbit-in-a-Hat (Version 0.10.7) importiert. Das Mapping von DS-Med beschränkt sich auf die Tabelle *drug_exposure* in OMOP. Einen ersten Überblick über die tatsächlich transferierten Datenelemente aus Datensatz DS-Med

in die OMOP Tabelle *drug_exposure* ist in Abbildung 3.1 dargestellt und wird als Ergebnis durch Rabbit-in-a-Hat generiert.

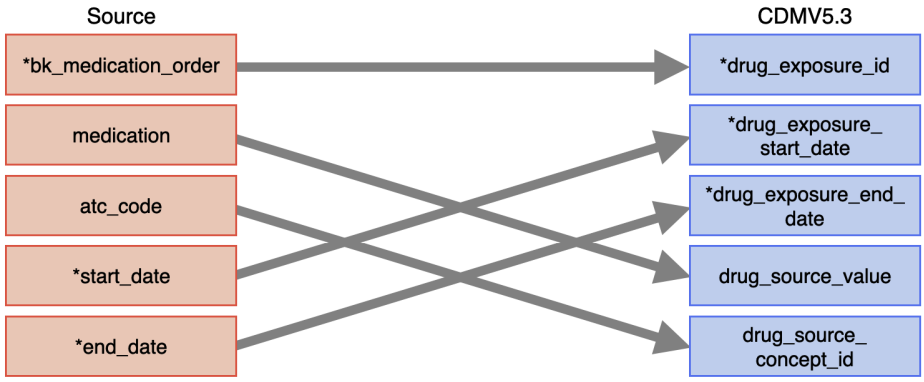


Abbildung 3.1: Mapping Überblick Medikationsverordnungen nach OMOP

Die vollständige Dokumentation mit entsprechenden Kommentaren zum Mapping ist in Tabelle 3.2 dargestellt.

Tabelle 3.2: Beschreibung des Mappings der Medikationsverordnungen nach OMOP Tabelle drug_exposure

Spalte in der OMOP Tabelle drug_exposure	Datenelement in DS-Med	Logik	Kommentar
drug_exposure_id	bk_medication_order	Quelle = Ziel	
person_id		1	Für alle Einträge wird die synthetische Person mit der person_id = 1 genutzt
drug_concept_id		0 (initial)	Erst im späteren Verlauf bei dem Mapping von ATC Codes nach RxNorm wird hier entsprechend die passende *drug_concept_id* für RxNorm auf Wirkstoffebene eingesetzt. Existieren bei ATC Codes mit mehr als einem RxNorm Code mehrere Mappings vom Typ "maps_toßu mehreren RxNorm Wirkstoffen, werden die entsprechenden *drug_exposure* Einträge vervielfältigt.
drug_exposure_start_date	start_date	Quelle = Ziel	
drug_exposure_start_datetime		(null)	Bleibt leer
drug_exposure_end_date	end_date	Quelle = Ziel	
drug_exposure_end_datetime		(null)	Bleibt leer
verbatim_end_date		(null)	Bleibt leer
drug_type_concept_id		32838	Alle Medikationsverordnungen sind vom gleichen Typ (EHR prescription)
stop_reason		(null)	Bleibt leer
refills		(null)	Bleibt leer
quantity		(null)	Bleibt leer
days_supply		(null)	Bleibt leer
sig		(null)	Bleibt leer
route_concept_id		(null)	Bleibt leer
lot_number		(null)	Bleibt leer
provider_id		(null)	Bleibt leer
visit_occurrence_id		1	Es existiert in der OMOP Datenbank exakt ein syntetischer Behandlungsfall mit der *visit_occurrence_id = 1*
visit_detail_id		(null)	Bleibt leer
drug_source_value	medication	Quelle = Ziel	
drug_source_concept_id	atc_code	ATC code *concept_id*	Die *concept_id* wird aus der Tabelle *concept* der OMOP Datenbank extrahiert und wenn für einen ATC Code vorhanden in *drug_source_concept_id* eingefügt, ansonsten wird der Wert auf *drug_source_concept_id = 0* gesetzt.
route_source_value		(null)	Bleibt leer
dose_unit_source_value		(null)	Bleibt leer

Um den Anforderungen gemäß des OMOP CDM an die Pflichtfelder und Referenzierungen zwischen den unterschiedlichen Tabellen gerecht zu werden, wurden neben den Medikationsdaten exakt eine synthetische Person sowie ein dazugehöriger Behandlungsfall in den Tabellen *person* und *visit_occurrence* des OMOP CDM manuell unter Verwendung des SQL Clients DBVisualizer für alle Medikationsverordnungen generiert.

Der Transfer der Medikationsverordnungen aus DS-Med nach OMOP erfolgt in drei Schritten. Zuerst wird der Transfer mit den ursprünglichen Medikationsverordnungen, vor Durchführung der im Kapitel 3.5 beschriebenen Maßnahmen, durchgeführt. Ein zweites Mal werden die Medikationsverordnungen nach der Durchführung der Maßnahmen zur Verbesserung der Datenstruktur (Abschnitt 3.5.2) nach OMOP transferiert. Abschließend werden sie nach der Durchführung der Maßnahmen in Bezug auf die Terminologie (Abschnitt 3.5.3) erneut überprüft. Die in den drei Schritten transferierten Daten werden nun im Rahmen der in Kapitel 3.6 beschriebenen Methodik bewertet.

3.1.3 Infrastruktur

Die für diese Arbeit verwendete Infrastruktur ist wie in Abbildung 3.2 in Form einer virtualisierten Umgebung verfügbar und weil es die Reproduzierbarkeit der Ergebnisse ermöglicht (Cito et al., 2016; Merkel, 2014).

Die Bereitstellung der verwendeten, anonymisierten Medikationsverordnungen erfolgte als CSV Textdatei gemäß der Beschreibung des Datensatzes DS-Med in Abschnitt 3.1.1 durch das DIZ des UKD. Der Datensatz DS-Med wurde unter Verwendung von Jupyter Lab (Version 3.2.9) und Python (Version 3.9.1), wie in Abschnitt 3.1.2 beschrieben, in ein OMOP CDM konformes Datenformat übersetzt. Der Quellcode ist auf Zenodo verfügbar (Reinecke, 2023b). Die Dokumentation zum Quellcode befindet sich in Anhang A.

Die OMOP Datenbank, sowie das OHDSI DQD wurden virtualisiert durch die Verwendung von Docker Containern genutzt. Die Docker Container sind Teil der Medical Informatics ReusAble eCO-system of open source Linkable and Interoperable software toolbox (MIRACOLIX) des MI-I Konsortiums MIRACUM und ermöglichen die Nutzung und Übertragbarkeit der gesamten Infrastruktur auf andere Standorte innerhalb der MI-I (Prokosch und Karg, 2022).

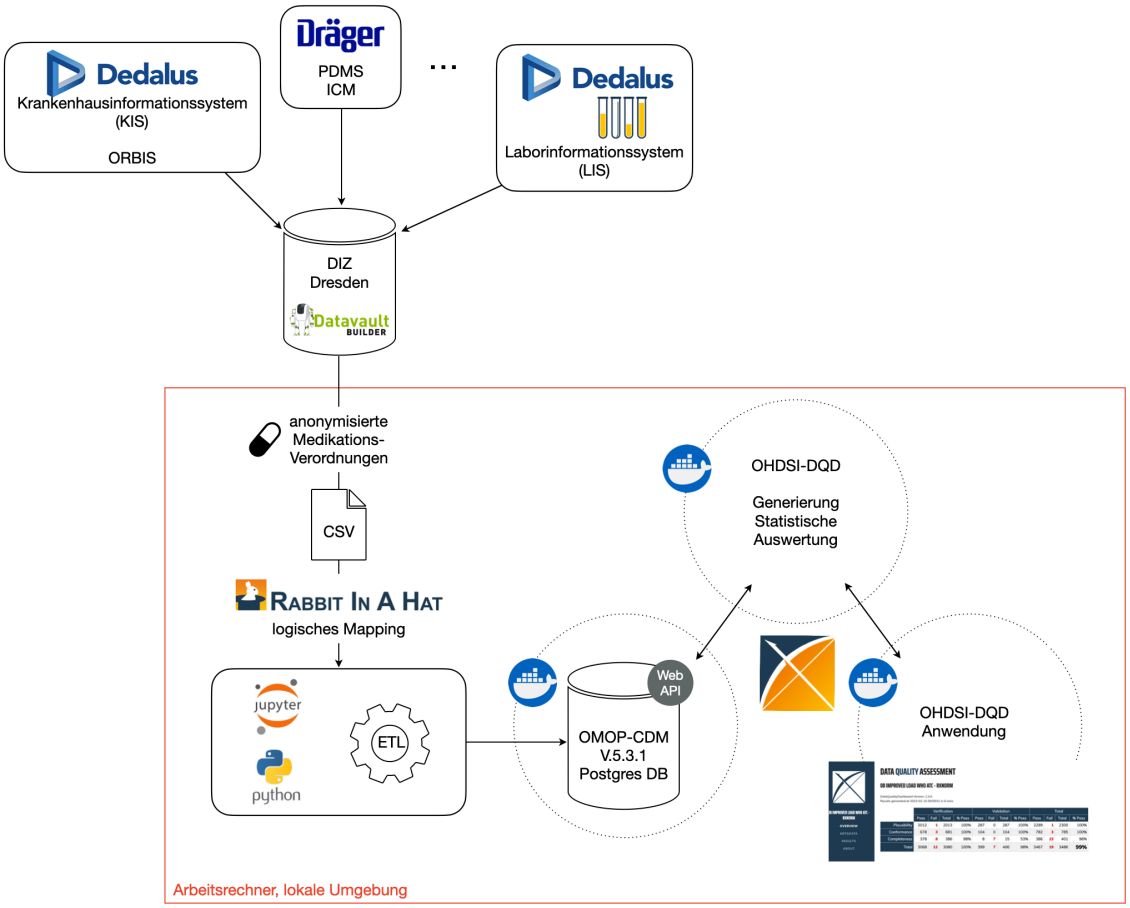


Abbildung 3.2: Überblick der für diese Arbeit verwendete Infrastruktur

3.2 Literaturrecherche

Um den derzeitigen Forschungsstand hinsichtlich der Verwendung von OMOP und zur Verbreitung der OHDSI Forschungsgemeinschaft weltweit zu analysieren und in den Kontext der Arbeit einordnen zu können, wurde initial eine Literaturrecherche als Scoping Review durchgeführt.

Scoping Reviews sind ein hervorragendes Werkzeug, um die Art und Weise zu untersuchen, zu welchen Forschungsschwerpunkten in einem bestimmten Thema durchgeführt wird, um Schlüsselmerkmale in einem abgegrenzten Forschungskontext zu bestimmen und existierende Lücken zu identifizieren (Munn et al., 2018). Der inhaltliche Schwerpunkt bei der Durchführung des Scoping Reviews lag auf der bisherigen Nutzung von OMOP an Standorten in Deutschland im Vergleich zur weltweiten Nutzung.

Um die Entscheidungsfindung für den Ein- und Ausschluss von Publikationen in das Scoping Review transparent zu berichten, wurde die Methode „Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews“ (PRISMA-ScR) genutzt (Tricco et al., 2018). Für die Verwaltung der Literatur wurde die Software Zotero genutzt, um die Identifikation und Auflösung von Duplikaten durchzuführen (Mueen Ahmed et al., 2011).

3.2.1 Identifikation von Publikationen

Die Suche nach existierender Literatur wurde am 22. Februar 2021 durchgeführt und beinhaltet Veröffentlichungen im Zeitraum vom 01.01.2016 bis einschließlich 22.02.2021. Die Suche wurde auf den Plattformen PubMed, IEEExplore und im Web of Science unter Verwendung der in Tabelle 3.3 aufgeführten Suchparameter durchgeführt.

Tabelle 3.3: Übersicht der Suchmaschinen und Suchstrings für die Literaturrecherche

Suchmaschine	Suchbegriff
Pubmed	All Fields: OHDSI or OMOP or „Observational Health Data Sciences and Informatics“ or „Observational Medical Outcomes Partnership“
Web of Science	ALL FIELDS: OHDSI or OMOP or „Observational Health Data Sciences and Informatics“ or „Observational Medical Outcomes Partnership“
IEEEXplore	(„Full Text & Metadata“:OMOP) OR („Full Text & Metadata“:OHDSI) OR („Full Text & Metadata“:Observational Medical Outcomes Partnership) OR („Full Text & Metadata“:Observational Medical Outcomes Partnership)

3.2.2 Einschluss und Ausschluss von Publikationen

Der Prozess des Ein- und Ausschlusses von Publikationen unterteilt sich in (1) die Entfernung von Duplikaten und (2) den Ausschluss von Publikationen, die (a) nicht in englischer Sprache verfügbar waren, (b) Volltext nicht frei verfügbar war oder (c) inhaltlich nicht relevant waren.

Der inhaltliche Ausschluss wurde anhand des Title/Abstract Screening in einem Team von drei Wissenschaftler:innen (IR, MZ, FB) vorgenommen. Eine Liste der verbleibenden Publikation flossen in das Scoping Review ein. Der Export der berücksichtigten Publikation wurde als CSV Datei für die Kategorisierung der eingeschlossenen Literatur (siehe Kapitel 3.2.3) bereitgestellt.

3.2.3 Kategorisierung von Publikationen

Die im Scoping Review eingeschlossenen Publikationen wurden durch ein Team von drei wissenschaftlichen Mitarbeiterinnen (IR, MZ, FB) kategorisiert. Die Kategorisierung erfolgte unabhängig voneinander in einem Verblindungsverfahren. Bei unterschiedlichen Ergebnissen in der Kategorisierung wurde über einen Abstimmungsentscheid durch die einfache Mehrheit ein Konsens herbeigeführt. Es erfolgte eine Einteilung geografisch nach Ländern, dem fachlichen Kontext sowie thematischen Schwerpunkten. Für die geografische Kategorisierung nach Land wurde für jede Publikation das Land der zugehörigen Institution des:r Erstautors:in ermittelt, um so die Verteilung weltweit zu verstehen und einordnen zu können. Zudem wurde eine Analyse der geografischen Herkunft der klinischen Daten, die in der Veröffentlichung Verwendung fanden, durchgeführt. Diese Analyse ist wichtig, um zu verstehen, ob das Ziel der OHDSI Forschungsgemeinschaft, Studien multizentrisch mit Daten aus unterschiedlichen Ländern durchzuführen, tatsächlich erreicht wird. Publikationen, in denen Daten aus mehr als einem Land genutzt wurden, wurden als „Multi-Country“ markiert, Publikationen mit Daten aus einem Land als „Single-Country“.

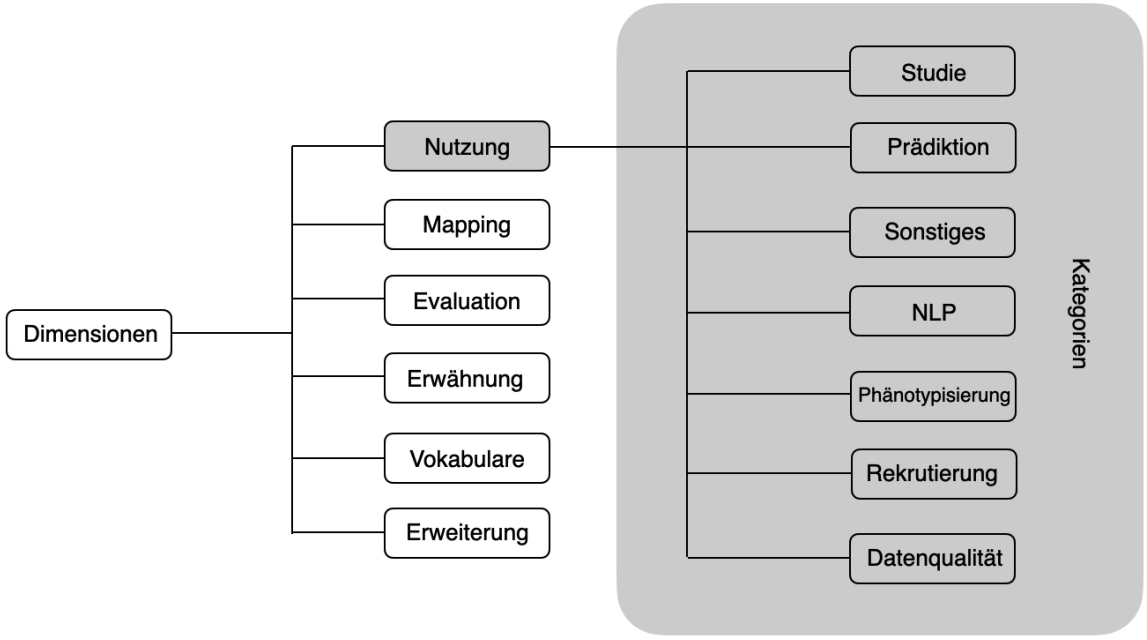


Abbildung 3.3: Inhaltliche Kategorisierung der eingeschlossenen Publikationen

Die Kategorisierung nach dem fachlichen Kontext erfolgte über die Einteilung in die Bereiche Informatik (inf), Medizininformatik (mi), Medizin (med) und Sonstige (other). Diese Einteilung kann helfen, um zu verstehen, ob tatsächlich Studien zur Beantwortung medizinischer Fragen

durchgeführt werden, oder der fachliche Fokus verstärkt auf der Information und Medizininformatik liegt. Zudem wurden die Publikationen unterschieden zwischen Konferenz- und Journalpublikationen. Die Kategorisierung nach dem thematischen Schwerpunkt erfolgte wie in Abbildung 3.3 in 6 Dimensionen. Die Dimension „Nutzung“ wurde aufgrund der Vielfalt der eingeschlossenen Publikationen in 9 weitere Unterkategorien geteilt.

Das Ziel dieser Einteilung war primär die Erfassung tatsächlich durchgeführter retrospektiver Beobachtungsstudien, bei denen das OMOP CDM die Basis darstellt, sowie die Veranschaulichung der vielfältigen Anwendungsbereiche von OMOP. Die Einteilung der eingeschlossenen Publikationen nach Dimension und Kategorien für die Dimension „Nutzung“ ist wichtig, um ein Verständnis für die tatsächliche Verbreitung und den Nutzungskontext des OMOP CDM zu erhalten und somit Aussagen zum aktuellen Forschungsstand auf dieser Ebene treffen zu können.

Die in das Scoping Review eingeschlossenen Veröffentlichungen wurden auch zeitlich nach dem Datum der Veröffentlichung ausgewertet, um die Anzahl der Publikationen pro Jahr und den fachlichen Kontext darstellen zu können.

3.3 Anforderungsanalyse

In diesem Kapitel wird die Methode zur Erhebung der Anforderungen an RWD für die Durchführung von Studien in internationalen Forschungsvorhaben auf Basis von OMOPbeschrieben.

Dazu werden (1) die Anforderungen seitens des Datenmodells OMOP als technische und infrastrukturelle Grundlage vorgestellt und (2) das Vorgehen zur Analyse der bereits veröffentlichten OHDSI Netzwerkstudien (siehe Ergebnisse in Kapitel 4.1.6) erläutert.

Die Untersuchung der tatsächlichen Nutzung und Notwendigkeit von Datengruppen orientiert sich am KDS der MI-I mit den Datengruppen Diagnosen, Medikation, Laborwerte, Prozeduren, Beobachtungen und Scores.

3.3.1 Anforderungen seitens des Datenmodell OMOP

Die Anforderungsanalyse an Medikationsdaten zur Ablage in OMOP wird explorativ durchgeführt und besteht aus folgenden Aktivitäten:

1. Etablieren eines fundierten Verständnisses durch Nutzung existierender Dokumentation und Spezifikationen von OMOP
2. Erweiterung der fachlichen Expertise durch die Nutzung von digitalen Schulungsangeboten und Tutorials
3. Vernetzung und Austausch mit Expert:innen der OHDSI Forschungsgemeinschaft auf Veranstaltungen und in Foren
4. Einbindung in die OHDSI Forschungsgemeinschaft durch aktive Integration und Zusammenarbeit

Die Tabelle 3.4 zeigt eine Übersicht der verwendeten Instrumente für die Durchführung der Aktivitäten, inklusive einer Erläuterung des Zwecks.

Die Anforderungen an die Medikationsdaten werden in die Kategorien (1) syntaktische und (2) semantische Anforderungen unterteilt. Die syntaktischen Anforderungen an Daten beziehen sich auf das Format und die Struktur der Daten, während die semantischen Anforderungen die Bedeutung und das Verständnis der Daten in einem länderübergreifenden Kontext sicherstellen. Beide Kategorien sind für die Interoperabilität der Daten essenziell und stellen ein wichtiges Merkmal der Daten für deren Verwendung in internationalen Forschungsvorhaben dar.

3.3.2 Analyse Studienprotokolle von OHDSI Studien

Die Anforderungen an die Daten seitens OMOP definieren einen idealen Zustand der Daten, der in der Realität in klinischen Routinedaten, die nicht primär für den Zweck der Forschung erhoben werden, häufig nicht vollständig erfüllt wird.

In einem zweiten Schritt sollen auch die bisher durchgeführten retrospektiven Beobachtungsstudien unter Verwendung von OMOP im Hinblick auf die Datenelemente analysiert werden, um ein vollständiges Bild der Anforderungen zu generieren und die Anforderungen seitens OMOP mit den Anforderungen der bereits durchgeführten Studien abzugleichen.

Tabelle 3.4: Übersicht genutzter Instrumente für die Anforderungsanalyse seitens OMOP

Instrumente	Beschreibung	Zweck
Literatur	Das Buch „Book of OHDSI" (OHDSI, 2019) gilt als Standardwerk und ist online verfügbar. Das Buch dient als Einstieg in das Thema oder als Nachschlagewerk. Für die Anforderungsanalyse wurde der Bereich Datenrepräsentation mit den Kapiteln Datenstruktur und standardisierte Vokabulare genutzt.	Verständnis der Datenstruktur in OMOP, verpflichtende Datenelemente und Terminologien
Spezifikation	Das OHDSI GitHub Wiki (OHDSI, 2023a) enthält die Spezifikation von OMOP. Jede Tabelle der OMOP Datenbank wird einführend beschrieben, einzuhaltende Konventionen zur Gewährleistung der Konformität werden definiert.	Verständnis der Datenstruktur in OMOP, verpflichtende Datenelemente und Terminologien
Schulung	Die EHDEN Academy ist eine frei zugängliche online Lernplattform und stellt Tutorials mit praktischen Beispielen zu unterschiedlichen Themenbereichen bereit. Im Kontext der Anforderungsanalyse waren die Tutorials zu den Themen OMOP und Terminologien relevant.	Anwendung des theoretisch erlangten Verständnisses zur Datenstruktur und Terminologien.
Forum	Das OHDSI Forum ist eine Plattform, an dem verschiedenen Interessenvertreter:innen (Forschende, Mediziner:innen, Epidemiolog:innen, Entwickler:innen und Fachleute des Gesundheitswesens) Wissen und Erfahrungen austauschen können. Es dient auch der Vernetzung und der Beantwortung und Diskussion von Fragen.	Adressieren von Herausforderungen der Medikationsdaten der klinischen Versorgung aus Deutschland hinsichtlich der Passfähigkeit im Kontext der geforderten Datenstruktur und Terminologien in OMOP.
Symposium	Die Symposien sind von OHDSI jährlich durchgeführte Veranstaltungen in den USA, in Europa und in Asien. Auf dem Symposium stellen Expert:innen Forschungsergebnisse, Erfahrungen und Neuigkeiten zu OMOP, Terminologien und Werkzeugen der OHDSI Forschungsgemeinschaft vor. Die Symposien dienen dem persönlichen Austausch und der Förderung der überregionalen Zusammenarbeit.	Netzwerkaufbau und Verständnis anderer Forschungsarbeiten, sowie der Herausforderungen anderer Teams.
Studyathons	Die OHDSI Studyathons sind Veranstaltungen, an denen sich interessierte Personen zusammenfinden, um gemeinsam eine Forschungsfrage zu bearbeiten und unter Verwendung der OHDSI Werkzeuge retrospektive Studien als interdisziplinäres Team durchzuführen. Die Ergebnisse der Veranstaltungen werden auf der OHDSI Webseite oder als wissenschaftliche Publikationen veröffentlicht.	Netzwerkaufbau und exemplarische Studiendurchführung
OHDSI Germany	OHDSI Germany ist Teil der weltweiten OHDSI Forschungsgemeinschaft und bringt Expert:innen aus Medizin und Informatik in Deutschland zusammen, um eine sektorübergreifende Arbeitsgemeinschaft zu bilden und deutschen Kliniker:innen und Forschenden eine Plattform zum Austausch von Best Practices und zur Teilnahme an internationalen Studien des OHDSI-Netzwerks zu bieten.	Austausch, Wissensaufbau, Vernetzung

Die Auswahl der zu analysierenden Publikationen basiert auf den in Kapitel 3.2.3 in Abbildung 3.3 dargestellten Dimensionen und Unterkategorien. Für diese Analyse wurden alle Publikationen der Dimension „Nutzung" und die Kategorie „Studie" verwendet.

Dazu wurden die im durchgeführten Scoping Review (siehe Kapitel 3.2) als Studien identifizierten Publikationen in folgenden Schritten analysiert:

1. Volltext Screening aller Publikationen
2. Volltext Screening aller Studienprotokolle, sofern vorhanden
3. Screening aller Kohortendefinitionen, sofern vorhanden

Aus den Studien wurden die relevanten Datengruppen, die genutzten Datenquellen und die Information, ob eine Studie eine offizielle Registrierung in einem Studienregister wie

beispielsweise dem European post-authorisation study (EU PAS) Register besitzt, extrahiert. Tabelle 3.5 gibt einen Überblick über die extrahierten Daten, sortiert nach (A) Metainformationen zur Studie, (B) medizinische Datengruppen und (C) die Datengruppe Medikation im Detail. Die Dokumentation erfolgte strukturiert in Excel und wurde nach Screening Abschluss als CSV Datei extrahiert, um sie für die Auswertung in Python nutzen zu können.

Die unter (C) aufgeführten Datengruppen basieren im Wesentlichen auf den Basismodulen des in Kapitel 2.2 eingeführten KDS der MI-I. Für die Datengruppe Medikation (C) wurden im Detail auch weitere Informationen, wie die Liste der in der jeweiligen Studie relevanten Medikamente oder Wirkstoffe extrahiert. Es wurde abschließend geprüft, ob Informationen über die Dosierung oder die Art der Verabreichung des Medikaments innerhalb der Studie ebenfalls relevant waren.

Tabelle 3.5: Übersicht der extrahierten Informationen aus den OHDSI Netzwerkstudien

Extrahierte Informationen	Beschreibung
(A) Metainformationen zur Studie	
Titel	Titel der Publikation
Publikationsjahr	Jahr der Veröffentlichung
DOI	digital, eindeutiger Identifier
Abstract	kurze Zusammenfassung der Studie
Autor:innen	Liste des Teams von Autor:innen
Registrierung	Information, ob die Studie bei einem Register angemeldet wurde
(B) medizinische Datengruppen	
Diagnosen	Boolescher Wert, der anzeigt ob Diagnosen für die Studie genutzt wurden
Medikation	Boolescher Wert, der anzeigt ob Medikation für die Studie genutzt wurden
Laborwerte	Boolescher Wert, der anzeigt ob Laborwerte für die Studie genutzt wurden
Prozeduren	Boolescher Wert, der anzeigt ob Prozeduren für die Studie genutzt wurden
Beobachtungen	Boolescher Wert, der anzeigt ob Beobachtungen für die Studie genutzt wurden
Medizinische Scores	Boolescher Wert, der anzeigt ob medizinische Scores für die Studie genutzt wurden
(C) Medikationsdaten im Detail	
Medikationsliste	Liste von Medikamenten oder Wirkstoffen, wenn vorhanden auch die entsprechenden Codes relevanter Terminologien
Details	Information, ob die Medikation als Produkt oder Wirkstoff angegeben wurde
Dosis und Darreichungsform	Angaben wenn vorhanden und für Studie relevant

Die extrahierten Daten wurden unter Verwendung von Jupyter Lab und Python (Version 3.9.1) quantitativ untersucht und beschrieben. Der implementierte Quellcode sowie die in Tabelle 3.5 genannten Metainformationen der Studien wurden auf Zenodo bereitgestellt und sind abrufbar(Reinecke, 2023a, 2023b). Ein wichtiger Indikator ist die Häufigkeit der Verwendung der in Tabelle 3.5 unter (B) gelisteten medizinischen Datengruppen. Sie dienen als Kennzahl für deren Wichtigkeit im Kontext retrospektiver Beobachtungsstudien im Forschungsnetzwerk OHDSI.

3.4 Identifikation von Inhibitoren

Ziel dieses Kapitels ist die Beschreibung der Methode zur Feststellung des aktuellen Status von Medikationsdaten aus der klinischen Versorgung deutscher Universitätskliniken, im Besonderen für das UKD, um mögliche Inhibitoren zu identifizieren, die den in Kapitel 4.2.3 genannten Anforderungen widersprechen.

Zunächst wird dazu in Abschnitt 3.4.1 die Methodik beschrieben, um stichprobenartig die Strukturiertheit der klinischen Daten aus den KIS an den Standorten des MI-I Konsortiums MIRACUM zu überprüfen. Anschließend wird in Abschnitt 3.4.2 das Vorgehen beschrieben, um die Medikationsdaten des UKD Dresden aus der klinischen Versorgung systematisch auf ihre Strukturiertheit zu analysieren.

3.4.1 Stichprobenanalyse von Routinedaten an MIRACUM Standorten

Im Rahmen des MIRACUM Projektes wurden in einer vorherigen Arbeit von Gulden et al. (Gulden et al., 2019) für die Nutzung von RWD relevante Datenelemente identifiziert, basierend auf ihrer Häufigkeit in den Ein- und Ausschlusskriterien aus einer Stichprobe von fünfzig klinischen Studien. Die genannte Häufigkeit diente dabei als Kennzahl der Priorisierung deren Wichtigkeit. Von allen identifizierten Datenelementen wurden die 28 am häufigsten verwendeten Datenelemente ausgewählt, um deren Strukturiertheit an den zehn Standorten des MIRACUM Konsortiums zu prüfen. Die 28 Datenelemente wurden jeweils einer von sechs Datengruppen (Demografie, Diagnosen, Prozeduren, Medikation, Scores und Laborwerte) zugeordnet. Die explorative Datenanalyse erfolgte final auf Basis der sechs Datengruppen.

Zur Untersuchung der 28 Datenelemente wurden an den zehn MIRACUM Standorten jeweils neun stationäre, abgeschlossene Behandlungsfälle zufällig ausgewählt und im KIS mit den behandelnden Ärzt:innen anhand definierter Parameter überprüft. Für jedes zu prüfende Datenelement wurden Meta-Informationen wie beispielsweise die Dokumentationsqualität (Vollständigkeit, Strukturiertheit), temporale Informationen (Datum der Ersterfassung, Mehrfacherfassung) oder beispielsweise für die Datenelemente der Datengruppe Medikation, die Unterscheidung zwischen Medikationshistorie, Medikation während Behandlung sowie Entlassmedikation dokumentiert. Insgesamt wurden Behandlungsfälle im Zeitraum vom 01.01.2018 bis 31.12.2018 mit einer Aufenthaltslänge von drei bis 15 Tagen und einer Hauptdiagnose der ICD10 Kapitel C, E, G, oder J stichprobenartig und zufällig an jedem

Standort ausgewählt. In Summe wurden 436 Meta-Informationen zu den Datenelementen (durchschnittlich 45 Behandlungsfälle pro MIRACUM Standort) erfasst und in ein zentral verwaltetes REDCap (Harris et al., 2009) Projekt zur Konsolidierung der Daten überführt.

Für die sechs Datengruppen (Demografie, Diagnosen, Prozeduren, Medikation, klinische Scores und Labor) wurde eine explorative Datenanalyse unter Federführung des Erlanger Kollegen Christian Gulden und des Greifswalder Kollegen Albert Vass durchgeführt. Die Ergebnisse wurden in einem Boxplot-Diagramm dargestellt, um einen schnellen Überblick über die Verteilung der Daten pro Datengruppe hinsichtlich der Strukturiertheit, einschließlich der Darstellung des Median, der Quartile und Ausreißerwerte, zu erhalten. Aus Gründen der Datensicherheit wurde ein Rückschluss auf einzelne Behandlungsfälle oder beteiligte Standorte vermieden, entsprechende Felder wurden daher in der Analyse anonymisiert.

3.4.2 Systematische Analyse der Medikationsdaten am UKD

Eine systematische Untersuchung der Medikationsdaten aus dem KIS des UKD erweitert die stichprobenartig durchgeführte Datenanalyse aus Abschnitt 3.4.1 mit Fokus auf die Medikationsdaten. Die Analyse unterstützt die Beurteilung der Daten, um mögliche Abweichungen von denen in Kapitel 3.3 erhobenen Anforderungen zu erkennen und um mögliche Inhibitoren zu identifizieren.

Für die systematische Analyse wurde der DS-Med, wie in Tabelle 3.1 beschrieben, bestehend aus 1.768.153 Medikationsverordnungen (N), genutzt. Die Strukturiertheit wurde unter Verwendung des Datenelement *STRUCTURE* beurteilt. Für den Datensatz DS-Med wurde das Verhältnis von strukturierten ($DS_{med-strukturiert}$) und unstrukturierten Medikationsverordnungen ($DS_{med-nicht-strukturiert}$) ermittelt. Das Datenelement *STRUCTURE* ist ein boolescher Wert, der ausschließlich einen der beiden Zustände 1 und 0 annehmen kann. Dabei zeigt Zustand 1, dass diese Medikationsverordnungen strukturiert aus dem Hauskatalog-Datensatz DS-Katalog ausgewählt wurden und ein valider ATC Code im Datenelement „ATC_5“ vorliegt:

$$DS_{Med-strukturiert} = \sum_{i=1}^{N} \frac{STRUCTURE_i}{N}$$

$$DS_{Med-nicht-strukturiert} = \sum_{i=1}^{N} \frac{1-STRUCTURE_i}{N}$$

In einem zweiten Schritt wird für $DS_{med-strukturiert}$ zusätzlich geprüft, ob in dem Datenelement „ATC_L5“ tatsächlich ein valider ATC Code vorliegt und somit für alle $DS_{med-strukturiert}$ gilt:

$$\forall i \in [1, length(DS_{Med})].STRUCTURE_i^{DS_{Med}} = 1 \Rightarrow ATC\_5_i^{DS_{Med}} \neq 0$$

Anschließend wurde der unstrukturierte Teil der Medikationsverordnungen nach dem Datenelement *MEDICATION* DS-Gruppiert gruppiert. Die Gruppierung basierte auf der kalkulierten Häufigkeit des Vorkommens der unstrukturierten Freitexte im Datenelement *MEDICATION* und wurde in DS-Gruppiert als neues Datenelement *FREQUENCY* erfasst.

Ein interdisziplinäres Team führte anschließend eine manuelle Prüfung des DS-Gruppiert durch, um mögliche Anweisungen, die andere Anordnungen als Medikationsverordnungen enthalten, zu erkennen. Bei den Anweisungen handelt es sich um Laboranforderungen für Blutbilder, Blutgasanalysen, die Durchführung von Blutentnahmen. Als Resultat wurden eine Reihe von Regeln (siehe Tabelle 3.6) abgeleitet, nach denen automatisiert Freitexte erkennbar sind, die nicht als Medikationsverordnungen gelten.

Tabelle 3.6: Regeln zur Erkennung von Einträgen, die keine Medikationsverordnungen sind

Regel	Details
Datenelement MEDIKATION startet mit „BE “	Anordnung einer Blutprobe
MEDICATION startet mit „1 BE“	Anordnung einer Blutprobe
MEDICATION startet mit „BB “	Anforderung für ein Blutbild
MEDICATION startet mit „!“	Einträge die mit „!“ starten, indizieren eine Laboranweisung

3.5 Maßnahmen zur Reduktion der Inhibitoren

Das Kapitel enthält die Beschreibung der Maßnahmen, die dazu dienen, die in Kapitel 3.4 identifizierten Inhibitoren zu reduzieren, um die Anforderungen aus Kapitel 3.3 zu erfüllen.

Dazu wird in Abschnitt 3.5.1 zunächst vorgestellt, wie exemplarisch die Maßnahmen manuell für die Teilnahme an einer Studie durchgeführt werden. Anhand der Ergebnisse dieser exemplarischen Maßnahmen lassen sich generelle Maßnahmen zur Verbesserung der Datenstruktur und der Anpassung an die erforderlichen medizinischen Terminologien in OMOP von Medikationsverordnungen ableiten (Abbildung 3.4). Diese werden im Abschnitt 3.5.2 und 3.5.3 im Detail vorgestellt und beschrieben.

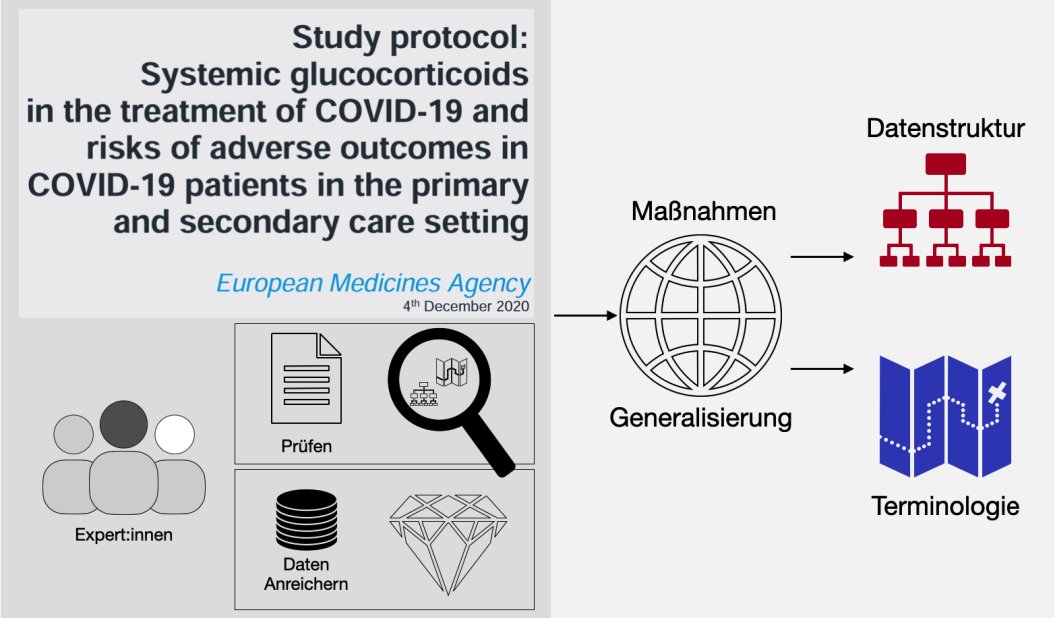


Abbildung 3.4: Überblick der Maßnahmen zur Reduktion der Inhibitoren

3.5.1 Maßnahmen am Beispiel einer EMA Studie

Die in diesem Abschnitt durchgeführten exemplarischen Maßnahmen fanden im Rahmen der Teilnahme an der EMA Studie „Systemic glucocorticoids in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients in the primary and secondary care setting“ (EU PAS Registernummer EUPAS38759) statt. Für die Teilnahme an dieser europäischen Studie war es nötig, die aus dem KIS des UKD bereitgestellten Medikationsverordnungen von Patient:innen mit einem stationären Aufenthalt aufgrund einer COVID-19 Erkrankung OMOP-konform aufzubereiten, um die zentral bereitgestellten Analyseskripte erfolgreich auszuführen und die Ergebnisse korrekt zu liefern. Für Teilnahme an der Studie wurde ein Ethikvotum (Antragsnummer SR-EK-576122020) bei der Ethikkommission der Technische Universität (TU) Dresden eingeholt.

Zunächst wurde das Studienprotokoll Version 1.0 (European Medicines Agency (EMA), 2020b) gesichtet und die notwendigen Medikationsdaten dokumentiert. Anschließend wurde gemeinsam mit einem interdisziplinären Expert:innenteam geprüft, welche der für das Studienprotokoll relevanten Wirkstoffe im angegebenen Zeitraum für die stationäre Behandlung von COVID-19 Patient:innen am UKD Anwendung fanden. Gemeinsam mit dem Team der Apotheke des UKD und unter Verwendung der Software MMI Pharmindex Plus wurden die zur Zeit der Studie verwendeten Medikamente für die relevanten Wirkstoffe identifiziert.

Des Weiteren wurden durch das DIZ des UKD die Medikationsverordnungen aus der Datenbank des KIS ORBIS extrahiert und alle vorhandenen Metainformationen bereitgestellt. Das DIZ führte für alle stationären Behandlungsfälle mit einer Sekundärdiagnose „U.07.1!" (COVID-19, Virus nachgewiesen) einen einfachen Abgleich der Zeichenkette der identifizierten Liste von Medikamenten über die Medikationsverordnungen als SQL Abfrage auf Datenbankebene durch und ordnete die entsprechenden Wirkstoffe als ATC Codes zu.

Außerdem wurden mit der Apotheke des UKD, dem IT Team des KIS und dem Team, welches für die Betreuung des SAP Systems zum Einkauf von Medikamenten zuständig ist, Interviews durchgeführt. Ziel dieser Interviews war die Identifikation notwendiger Daten für die Studie, die Generierung eines Verständnisses für die Datenstrukturen, Meta-Informationen und existierende Verbindungen zwischen SAP und KIS im Kontext von Medikationsverordnungen.

Nach der Überführung der Behandlungsdaten für die relevanten COVID-19-Behandlungsfälle unter Verwendung eines existierenden ETL Prozesses (Peng et al., 2023), wurden initial die zentral bereitgestellten Analyseskripte in RStudio ausgeführt. Die Ergebnisse wurden gemeinsam mit den technischen Ansprechpartner:innen seitens der Studienverwaltung der EMA überprüft und ergaben, dass die Medikationsverordnungen nicht der erwarteten Standardterminologie entsprechen. Daher wurde eine manuelle Anpassung für die relevanten Wirkstoffe und Behandlungsfälle in der OMOP Datenbank durchgeführt. Diese Anpassungen wurden als SQL Statements, wie im Quelltext 3.2 als Pseudocode beispielhaft für den ATC Code H02AB02 dargestellt, mit dem Ziel umgesetzt, die Daten auf die für die Studie notwendigen *concept_id* der Terminologie RxNorm anzupassen.

Quelltext 3.2: SQL Statement - Aktualisierung Standardterminologie Medikationsverordnungen

```
UPDATE drug_exposure
SET drug_concept_id = '1518254'
WHERE visit_occurrence_id IN (
    SELECT DISTINCT(visit_occurrence_id)
    FROM condition_occurrence
    WHERE condition_source_value LIKE 'U07.1%'
)
AND drug_source_value IN ('H02AB02');
```

Alle Maßnahmen wurden exemplarisch im Kontext der Studie manuell auf einem kleinen Datensatz von 721 Behandlungsfällen durchgeführt und dienen im Nachgang der Generali-

sierung der Maßnahmen zur Verbesserung der Datenstruktur (siehe Kapitel 3.5.2) und zur Erfüllung der Anforderungen hinsichtlich standardisierter Terminologien in OMOP.

3.5.2 Maßnahmen - Datenstruktur

In diesem Abschnitt werden die Methoden beschrieben, um die Strukturiertheit der Medikationsverordnungen des Datensatzes DS-Med systematisch zu verbessern und damit eine aus Kapitel 3.5.1 abgeleitete Notwendigkeit der Generalisierung der Maßnahme zur Verbesserung der Datenstruktur umzusetzen. Dabei sollen entsprechende Algorithmen implementiert und auf die Gesamtheit der Medikationsverordnungen angewendet werden. Die Methodik zu den Algorithmen wird in Abschnitt 3.5.2.1 vorgestellt. Die automatisierte Verbesserung der Datenstruktur durch die Algorithmen wird im Anschluss validiert, deren Methodik und Ergebnisse im Abschnitt 3.5.2.2 erläutert wird. Abschließend wird in Abschnitt 3.5.2.3 vorgestellt, wie die Analyse zur Verbesserung der Strukturiertheit durchgeführt wird.

3.5.2.1 Algorithmen

Medikationsverordnungen, die gemäß der Ergebnisse der initialen Analyse in Kapitel 4.3.2 als unstrukturiert identifiziert wurden, dienen als Eingangsgröße für die im Folgenden beschriebene Maßnahme. Die Medikationsverordnungen wurden im original vorliegenden Format verarbeitet, eine Vorverarbeitung der Daten hat nicht stattgefunden. Um die Datenstruktur der Medikationsverordnungen zu verbessern, wurden 3 verschiedene Algorithmen implementiert, um automatisiert ATC Codes auf der Grundlage des Freitextes im Datenelement *MEDICATION* des DS-Med zu identifizieren und zuzuordnen. Die Algorithmen basierten auf unterschiedlichen Mechanismen für den Abgleich des Freitexts gegenüber den Datenelementen *Ingredient_name* und *Product_name* aus DS-Katalog, wie in Tabelle 3.7 detailliert beschrieben.

Tabelle 3.7: Überblick Algorithmen zur ATC Code Identifikation für unstrukturierte Medikationsverordnungen

Algoritmus	Mechanismus	Datenelemente für den Abgleich		Ergebnis
		Datensatz DS1	Datensatz DS2	
1	Vergleich von Zeichenketten	MEDICATION	Ingredient_name	ATC Code
2	Vergleich von Zeichenketten	MEDICATION	Product_name	ATC Code
3	Abgleich von Ähnlichkeiten	MEDICATION	Ingredient_name und Product_name	ATC Code + Levenshtein Score

Die Algorithmen 1 und 2 stützen sich auf einfache Vergleiche von Zeichenketten, um entweder den Namen des Inhaltsstoffs oder den Produktnamen innerhalb der Medikationsverordnung zu erkennen. Algorithmus 3 ist ein Natural Language Processing (NLP), basierend auf einem Abgleich von Ähnlichkeiten zwischen dem Datenelement *MEDICATION* im Datensatz DS-Med und den beiden Datenelementen *Product_name* und *Ingredient_name* im Datensatz DS-Katalog unter Verwendung der Python-Bibliothek FuzzyWuzzy (Cohen, 2020) und der Levenshtein-Distanz, da diese in anderen Forschungsbereichen des Gesundheitswesens zuverlässige Ergebnisse erzielen konnte (Bobroske et al., 2020; Hutchison et al., 2020). Das beste mögliche Ergebnis für den Levenshtein Score ist 100. Der Wert 100 bedeutet, dass die Komponenten der Zeichenkette *MEDICATION* vollständig in *Ingredient_name* oder *Product_name* enthalten sind. Je niedriger der Levenshtein Score, desto weniger ähnlich ist die Zeichenfolge *MEDICATION* im Vergleich zu den Einträgen im Arzneimittelkatalog (Datensatz DS-Katalog). Dieser Algorithmus lieferte bis zu 3 mögliche Ergebnisse als ATC Codes, die in absteigender Reihenfolge nach dem Levenshtein Score sortiert waren.

Die Reihenfolge der Wörter in den Medikationsverordnungen im Datenelement *MEDICATION* sind irrelevant und können von der Reihenfolge der Vergleichswerte der Datenelemente *Ingredient_name* und *Product_name* abweichen, ohne sich in der semantischen Bedeutung zu ändern. Die beiden Datenelemente *Ingredient_name* und *Product_name* werden für den Vergleich in ein Datenelement zusammengefasst. Es ist lediglich erforderlich, dass alle Wörter aus dem Eintrag des Datenelements *MEDICATION* im Eintrag von *Ingredient_name* oder *Product_name* enthalten sind, jedoch nicht umgekehrt. Basierend auf diesen Anforderungen wurde die Methode *token_set_ratio* der FuzzyWuzzy Bibliothek genutzt. Die Methode vergleicht zwei Zeichenketten, indem sie diese in Wörter aufteilt, in Kleinbuchstaben umwandelt und Satzzeichen entfernt. Die Wörter werden dann alphabetisch sortiert. Anschließend werden die gemeinsamen Wörter (Schnittmenge) und die unterschiedlichen Wörter (Restgruppe) identifiziert. Das Verhältnis zwischen der Größe der Schnittmenge und der Gesamtanzahl der Wörter wird berechnet und als Ähnlichkeitswert zwischen 0 und 100 ausgegeben. Ein höherer Wert deutet auf eine größere Ähnlichkeit der Zeichenketten hin, wobei 100 für eine exakte Übereinstimmung steht. Diese Methode ermöglicht einen flexiblen und fehlerverzeihenden Vergleich basierend auf den gemeinsamen Wörtern der Zeichenketten, ohne Berücksichtigung der Reihenfolge der Wörter. Wie das folgende Beispiel mit Quellcode in Python 3.3 zeigt, liefert die Methode *token_set_ratio* die besten Ergebnisse für die gegebenen Anforderungen.

Quelltext 3.3: Beispiel für die unterschiedlichen Methoden der FuzzyWuzzy Bibliothek

```
medikation_eintrag = "Stada paracetamol"
katalog_eintrag = "paracetamol Stada 400 mg"
Print("Ratio: ", fuzz.ratio(d1.lower(),d2.lower()))
Print("Partial Ratio: ", fuzz.partial_ratio(d1.lower(),d2.lower()))
Print("Token Sort Ratio: ", fuzz.token_sort_ratio(d1.lower(),d2.lower()))
Print("Token Set Ratio: ", fuzz.token_set_ratio(d1.lower(),d2.lower()))
Ratio: 54
Partial Ratio: 65
Token Sort Ratio: 83
Token Set Ratio: 100
```

Die 3 Algorithmen wurden auf dem Datensatz DS-Gruppiert unter Nutzung der Datenelemente des Datensatzes DS-Katalog wie in Tabelle 3.7 dargestellt, angewendet. Die Ergebnisse der Algorithmen wurden als ATC Codes in Datensatz DS-Gruppiert in den Spalten Step1, Step2, Step3 (siehe Tabelle 3.1) gespeichert. Die Übereinstimmung zwischen den Ergebnissen für jede Permutation (Algorithmen 1+2, 1+3, 2+3 und 1+2+3) wurde ebenfalls kalkuliert und in den Spalten True12, True 13, True23 und True 123 gespeichert. Eine kurze Beschreibung des Quellcodes der Implementierung in JupyterLab in Python Version 3.9.1 befindet sich in Anhang A und ist online auf Zenodo (Reinecke, 2023b) verfügbar.

3.5.2.2 Validierung der Algorithmen

Um die Ergebnisse der Algorithmen auf ihre Korrektheit, im Sinne der richtigen Zuordnung des ATC Codes, zu prüfen und anhand der Prüfergebnisse eine Aussage zur zukünftigen Verlässlichkeit der Algorithmen geben zu können, wurde ein Validierungsschritt durchgeführt. Die Validierung umfasst eine manuelle Überprüfung der automatisch generierten ATC Codes durch ein interdisziplinäres Team. Die Validierung beschränkt sich auf eine Teilmenge der häufigsten unstrukturierten Medikationsverordnungen. Damit der Aufwand für die Validierung in einem angemessenen Verhältnis zum Nutzen steht, wurde für diesen manuellen Schritt ein Mindestziel gemäß des Pareto Prinzips (Chinchuluun et al., 2008; Harvey et al., 2018) definiert, welches insgesamt mindestens 80 % der Medikationsverordnungen (strukturiert und unstrukturiert, manuell validiert, zusammen) abzudecken fordert.

Bei der Validierung wurden für jeden Algorithmus zusätzliche Informationen ergänzt, welche angeben, ob der richtige ATC Code, ein falscher ATC Code oder kein ATC Code identifiziert

wurde. Konnte durch keinen der drei Algorithmen der ATC korrekt identifiziert werden, wurde manuell die Zuordnung vorgenommen, sofern möglich. Freitexte, bei denen es sich inhaltlich nicht um Medikationsverordnungen handelte, dieser aber nicht durch die Regeln aus Tabelle 3.6 erkannt werden konnten, wurden als zusätzliche Einträge ohne Medikationsverordnungen mit dem Schlüsselwort „nomed“ gekennzeichnet. Bei Medikationsverordnungen, die eine weitere Spezifizierung zur Bestimmung des genauen ATC Level 5 Code erfordern, wurde bei der manuellen Validierung geprüft, ob der ATC Code Level 3 oder 4 anhand des Freitextes der Medikationsverordnung bestimmbar war, andernfalls wurde der Eintrag als unspezifisch mit dem Schlüsselwort „unspec“ gekennzeichnet.

Die Ergebnisse der manuellen Validierung wurden zusammengefasst, um etwaige Muster zu identifizieren, die Schlussfolgerungen erlauben und helfen können, die Robustheit der Ergebnisse algorithmisch, automatisiert ermittelter ATC Codes zukünftig weiter zu verbessern. Dazu wurden die folgenden Kennzahlen kalkuliert: (a) die absolute Zahl der identifizierten ATC Codes, (b) davon korrekt identifizierte Anzahl, (c) der Übereinstimmungsgrad zwischen den Ergebnissen der Algorithmen 1, 2 und 3, sowie (d) der Levenshtein Score für Algorithmus 3. Für Algorithmus 3 wurde mit einem zweiseitigen t-Test geprüft, ob ein signifikanter Unterschied zwischen den Mittelwerten des Levenshtein Scores der korrekten und fehlerhafte Ergebnisse besteht. Die fehlerhaften Ergebnisse wurden abschließend von einem interdisziplinären Team untersucht, um Muster zu identifizieren, die mögliche Gründe und Ähnlichkeiten in Bezug auf die betroffenen Wirkstoffe (ATC Codes) sofern möglich aufzeigen.

3.5.2.3 Analyse der Verbesserung der Strukturiertheit

Basierend auf der Validierung der Algorithmen konnte für die in Kapitel 4.4.2 erzielten 85,15% aller Medikationsverordnungen eine abschließende Analyse der Verbesserung der Strukturiertheit vorgenommen werden. Hierzu wird die Häufigkeit der unterschiedlichen ATC-GM Codes inklusive der mit „nomed“ und „unspec“ markierten Medikationsverordnungen bestimmt.

Dies erfolgt über das Zusammenführen der in Abschnitt 3.5.2.2 validierten und korrigierten Ergebnisse in Datensatz DS-Top1000 mit den ursprünglichen unstrukturierten Medikationsverordnungen des Datensatzes DS-Med. Die Analyse erfolgt auf Basis der 14 verschiedenen ATC Gruppen (siehe Tabelle 2.3) und gibt für jede dieser Gruppen die Gesamtzahl, sowie

den Anteil der strukturierten und unstrukturierten Medikationsverordnungen an. Darüber hinaus werden die Gesamtzahl der eindeutigen ATC-GM Level 5 Codes, einschließlich ihrer Strukturiertheit, sowie die häufigsten ATC-GM Codes, der Medikationsverordnungen aus Datensatz DS-Med dargestellt.

3.5.3 Maßnahmen - Terminologie

Dieser Abschnitt beschreibt die Methodik, um die Anforderungen an die Terminologie gemäß der Ergebnisse aus Abschnitt 4.2.3 zu erfüllen. Die Eingangsgröße für dieses Verfahren besteht aus der in Kapitel 3.5.2 festgelegten Mindestmenge von 80% der Medikationsverordnungen im Datensatz DS-Med, die einen ATC Code enthalten. Diese Medikationsverordnungen wurden entweder bereits strukturiert mit ATC Code erfasst, durch die Algorithmen korrekt einem ATC Code zugeordnet oder durch die Validierung entsprechend korrigiert.

Die Aktivitäten in diesem Kapitel lassen sich in die folgenden Arbeitsschritte unterteilen:

1. Durchführung notwendiger Anpassungen für die deutsche ATC Terminologie in OMOP
2. Überführung von ATC Codes nach RxNorm Konzepten auf Basis der Wirkstoffe

3.5.3.1 Anpassungen für die deutsche ATC Terminologie in OMOP

In Abbildung 3.5 ist der Prozess aller in diesem Abschnitt beschriebenen Aufgaben dargestellt. In einem ersten Schritt (A1) wurden die Medikationsverordnungen (Datensatz DS-Med) mit dem ATC Vokabular aus OMOP (Datensatz DS-ATC) verknüpft. Hierfür wurde geprüft, ob für den ATC-GM Code im Datensatzes ein entsprechendes Pendant im Datensatz DS-ATC existiert. Somit wurde der Datensatz DS-Med um eine zusätzliche Spalte *concept_id* erweitert. Diese Spalte bleibt für ATC Codes leer, wenn kein valides Konzept gemäß des ATC Vokabulars in OMOP existiert.

Für alle ATC Codes ohne valide *concept_id*, wird in Schritt (A2) unter Verwendung des Datensatzes DS-Katalog geprüft, ob neben dem ATC Code der deutschen Version des BfArM ein anderer valider ATC Code der WHO Version existiert. In diesem Fall wird in Schritt (A3) der deutsche ATC Code gegen das entsprechende Äquivalent der WHO Version ersetzt. Dies resultiert in einer erneuten Iteration des ersten Schrittes (A1), bei dem diese ATC Codes nun einer validen *concept_id* zugeordnet werden können.

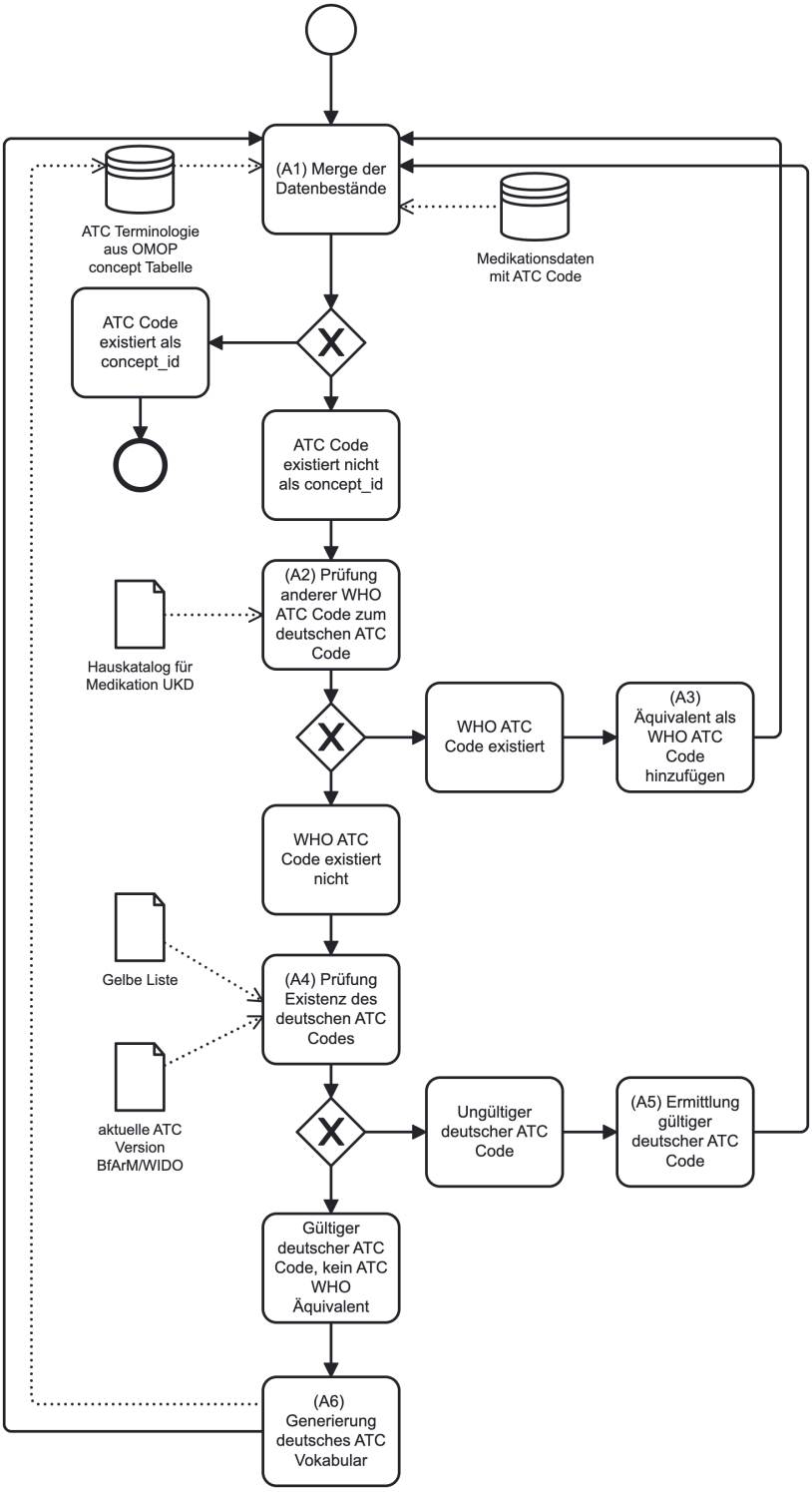


Abbildung 3.5: Prozessdiagramm der ATC Terminologie Aktivitäten

Für alle deutschen ATC Codes ohne entsprechenden WHO ATC Code fand in Schritt (A4) ein Abgleich mit zusätzlichen, online verfügbaren Datenquellen statt, um zu prüfen, ob der ATC Code gemäß der aktuellen deutschen Fassungen gültig ist. Bei den Datenquellen handelt es sich zum einen um die Online-Version der „Gelben Liste“ PHARMINDEX (Vidal MMI Germany GmbH, 2023), einem Arzneimittelverzeichnis, welches eine Suche nach Wirkstoffen

und Medikamenten ermöglicht, zum anderen um die amtliche Fassung der ATC Klassifikation, erstellt vom GKV-Arzneimittelindex im Wissenschaftliches Institut der AOK (WIdO) und herausgegeben vom BfArM (Bundesinstitut für Arzneimittel und Medizinprodukte, 2023).

Für alle als ungültig identifizierten deutschen ATC Codes wurde Schritt (A5 durchgeführt). Dabei handelt es sich um ein exploratives Vorgehen unter Verwendung der bereits genannten sowie Datenquellen, der Änderungshistorie von ATC WHO Codes (WHO Collaborating Centre for Drug Statistics and Methodology, 2022) oder der webbasierten Anwendung Athena (siehe Kapitel 2.4) von OHDSI. Eine Prüfung möglicher Änderungen der betroffenen Codes, um ein gültiges Äquivalent zu identifizieren, wurde durch geführt.

Für alle deutschen ATC Codes ohne gültiges Äquivalent in der WHO Version wird in Schritt (A6) ein angepasstes Vokabular namens ATC-GM in der OMOP Datenbank erstellt und die deutschen ATC-GM Codes hinzugefügt. Dazu werden die Konventionen von OHDSI (OHDSI, 2023a) zur Generierung von eindeutigen *concept_ids* eingehalten und alle Datenelemente des Datensatzes DS-ATC für das benutzerdefinierte Vokabular generiert.

3.5.3.2 Überführung in Zielterminologie (RxNorm)

Die Überführung der Medikationsverordnungen von ATC nach RxNorm erfolgt unter Verwendung aller Medikationsverordnungen des Datensatzes DS-Med, für die ein valider ATC Code nach Abschluss der Anpassungen in Kapitel 3.5.3.1 existiert. Für das semantische Mapping nach RxNorm wurde außerdem der in Kapitel 3.1.1 eingeführte Datensatz DS-Relation genutzt.

Die Überführung in die Zielterminologie RxNorm lässt sich in folgende Schritte unterteilen

1. Identifikation des korrekten semantischen Mappings von ATC nach RxNorm
2. Korrekte Zuordnung von RxNorm Codes für die Medikationsverordnungen des UKD, basierend auf den zuvor identifizieren Mappings
3. Analyse der durchgeführten Zuordnung unter dem Gesichtspunkt des Abdeckungsgrades bezogen auf die Medikationsverordnungen insgesamt

Die Identifikation des korrekten semantischen Mappings zwischen den Vokabularen ATC und RxNorm ist aufgrund der Multidimensionalität des Vokabulars RxNorm komplex. Aufgrund der verschiedenen RxNorm TTY existieren unterschiedlichen Typen von Beziehungen auf mehreren Ebenen, wie beispielsweise das klinische Medikament (RxNorm TTY Clinical Drug), die klinische Wirkstoffgruppe (RxNorm TTY Clinical Drug Component), oder auch der Wirkstoff (RxNorm TTY Ingredient).

Alle existierenden Typen von Beziehungen zwischen den Vokabularen ATC und RxNorm wurden zunächst durch ein exploratives Vorgehen auf der OMOP Datenbank der lokalen Infrastruktur (siehe Kapitel 3.1.3) unter Verwendung des Datenbank-Clients *DBVisualizer* ermittelt. Zudem wurde die vorhandene online Dokumentation im OHDSI Wiki (OHDSI, 2021) und das OHDSI Forum (Ostropolets, 2020) nach Detailinformationen zu den Beziehungstypen im Kontext ATC und RxNorm durchsucht.

Die Ergebnisse dieser beiden Recherchen wurden anschließend mit Expert:innen des OHDSI Vokabular Teams diskutiert und eingeordnet. Es wurde eine Evaluation der verfügbaren Mappings hinsichtlich der Eindeutigkeit, Vollständigkeit, möglicher Datenverluste und Lücken durchgeführt. Im Ergebnis konnten Rückschlüsse auf die korrekten Mappings und die relevanten Beziehungstypen geschlossen werden und ein eindeutiges Mapping zur korrekten Übersetzung wurde als Ergebnis generiert.

Die korrekte Zuordnung von RxNorm Codes für die Medikationsverordnungen des UKD erfolgte unter Verwendung des eindeutigen Mappings aus Schritt 1 in Jupyter Lab in Python (Version 3.9.1) mit der implementierten Funktion *medication_orders_apply_rxnorm*, wie in Quellcode 3.4 abgebildet.

Der im Python Quellcode 3.4 genannte Datensatz „atc_rxnorm_mappings“ wurde durch das SQL Statement 3.5 generiert und aus der OMOP Datenbank exportiert. Dieser Datensatz beinhaltet alle für die Zuordnung der RxNorm Codes relevanten Beziehungstypen. Der vollständige Quellcode ist auf Zenodo verfügbar (Reinecke, 2023b). Eine Dokumentation aller für diese Arbeit relevanten Python Scripte befindet sich in Anhang A.

Quelltext 3.4: Python Funktion - RxNorm Codes Zuordnung

```
def medication_orders_apply_rxnorm(medication_orders_omop_conform,atc_rxnorm_mappings):

    rxnorm_medication_orders_conform = pd.merge(
    medication_orders_omop_conform,
    atc_rxnorm_mappings,
    on="drug_concept_id",
    how="left"
    )
    rxnorm_medication_orders_conform['drug_concept_id'] = np.where(
        rxnorm_medication_orders_conform['concept_id.1'].notnull(),
        rxnorm_medication_orders_conform['concept_id.1'],
        rxnorm_medication_orders_conform['drug_concept_id']
    )

    rxnorm_medication_orders_conform=rxnorm_medication_orders_conform.drop([
        'concept_id.1'],
        axis=1)

    return rxnorm_medication_orders_conform
```

Die Analyse der durchgeführten Zuordnung unter dem Gesichtspunkt des Abdeckungsgrades wurde auf den Medikationsverordnungen des Datensatzes DS-Med durchgeführt.

Es handelt sich um eine quantitative Analyse der Anzahl überführbarer ATC Codes nach RxNorm und einer Analyse der Anzahl der Mappings pro Medikationsverordnung. Dies ist besonders für Medikamente mit mehr als einem Wirkstoff relevant, im Hinblick auf Nutzbarkeit der Daten in Studien und dient der Schaffung von Transparenz von möglichen Grenzen der Überführung nach RxNorm.

Quelltext 3.5: SQL Statement - Export semantisches Mapping von ATC nach RxNorm für „Maps to“

```
SELECT con1.concept_id AS concept_id_atc,
       con1.concept_code AS code_atc,
       con1.concept_name AS name_atc,
       rel.relationship_id AS rel_id,
       con2.concept_id AS concept_id_rx,
       con2.concept_code AS code_rx,
       con2.concept_name AS name_rx,
       con2.concept_class_id AS class_rx,
       con2.standard_concept AS standard_rx
FROM concept con1
JOIN concept_relationship rel
  ON concept_id = concept_id_1
JOIN concept con2
  ON concept_id_2 = con2.concept_id
WHERE con1.vocabulary_id = 'ATC'
  AND (con1.invalid_reason IS NULL
       OR (con1.valid_end_date >= '2020-12-31'
           AND con1.valid_start_date <='2016-01-01'))
  AND rel.relationship_id IN (
                              'Maps to')
  AND con2.concept_class_id = 'Ingredient';
```

3.6 Bewertung der Maßnahmen

Die Anforderungen an Medikationsdaten sind gemäß der Ergebnisse aus Kapitel 4.2.1 sehr divergent und vielfältig. Deshalb erfolgt die Bewertung der in Kapitel 3.5.2 und 3.5.3 durchgeführten Maßnahmen unter Verwendung der "Mixed-Methods“ qualitativ und quantitativ, weil diese sich besonders gut für komplexe Anforderungen eignen (Palinkas et al., 2019).

Die qualitative Bewertung soll anhand der Durchführbarkeit einer ausgewählten Studie aus Kapitel 4.2.2 erfolgen. Dazu wurde eine Studie ausgewählt, bei der die Datengruppe Medikation mit gefordert wurde, die Daten in RxNorm vorliegen mussten, und bei der die Wirkstoffe zudem im Datensatz DS-Med enthalten waren.

Es handelt sich dabei um eine Studie von Duke et al. (Duke et al., 2017) mit dem Titel „Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network“, die innerhalb

der OHDSI Forschungsgemeinschaft als multizentrische, retrospektive Beobachtungsstudie durchgeführt wurde. Bei der Studie handelt es sich um eine Vergleichsstudie, die das Risiko eines Angioödems in Zusammenhang mit der Einnahme des Wirkstoffs Levetiracetam bei Patient:innen mit Krampfanfällen untersuchte. Dazu wurde im Rahmen der Studie eine Kohorte mit Levetiracetam Anwender:innen (n=276.665) mit einer Kohorte von Phenytoin Anwender:innen (n=74.682) verglichen. Die Methoden und Ergebnisse sind im Folgenden kurz zusammengefasst. Mittels des Propensity-Score-Matching wurden Hazard-Ratios für Angioödem Ereignisse durch eine Per-Protocol (PP) Analyse und eine Intention-to-Treat (ITT) Analyse berechnet. Angioödem Ereignisse waren selten sowohl in der Levetiracetam Kohorte als auch in der Phenytoin Kohorte (54 versus 71 in der PP Analyse und 248 versus 435 in der ITT Analyse). In keiner der zehn genutzten Datenbanken wurde ein signifikant erhöhtes Risiko für Angioödeme bei Levetiracetam festgestellt (Hazard-Ratios zwischen 0,43 und 1,31). Eine Metaanalyse zeigte eine Zusammenfassung von Hazard-Ratios von 0,72 (95% Konfidenzintervall 0,39-1,31) und 0,64 (95% Konfidenzintervall 0,52-0,79) für die PP bzw. die ITT Analyse. Die Ergebnisse deuten darauf hin, dass Levetiracetam ein gleiches oder niedrigeres Risiko für Angioödeme aufweist als Phenytoin, das derzeit keine Warnsignale für Angioödeme zeigt. Weitere Studien sind erforderlich, um das Risiko von Angioödemen bei allen Antiepileptika zu bewerten.

Im Rahmen der in Kapitel 3.3 durchgeführten Anforderungsanalyse wurden anhand des Studienprotokolls die erforderlichen Medikationsdaten als RxNorm Codes auf Wirkstoffebene identifiziert. Im Anschluss wurden die relevanten Wirkstoffe Levetiracetam und Phenytoin innerhalb der Medikationsverordnungen des Datensatzes DS-Med ermittelt und ein Vergleich der Datenstruktur in dem Datensatz DS-Med vor und nach Durchführung zur Verbesserung der Datenstruktur durchgeführt. Abschließend wurde geprüft, ob eine eindeutige Zuordnung zu einem RxNorm für die beiden Wirkstoffe möglich war. Durch dieses Vorgehen lässt sich abschließend beurteilen, ob eine Teilnahme an der betrachteten OHDSI Studie vor oder nach Durchführung der Maßnahmen möglich gewesen wäre.

Die quantitative Bewertung der Medikationsverordnungen in der OMOP Datenbank (Datensatz DS-Med) wird unter Verwendung des OHDSI DQD (siehe Kapitel 2.6) durchgeführt. Dabei wird ausschließlich die Tabelle *drug_exposure* in die Auswertung einbezogen, welche die Medikationsverordnungen vorhält.

Die Analyse des OHDSI DQD wurde dreimal für die verschiedenen Entwicklungsstufen der Daten durchgeführt:

1. Entwicklungsstufe 1: Medikationsverordnungen gemäß der Ergebnisse aus Kapitel 4.3.2
2. Entwicklungsstufe 2: Medikationsverordnungen nach Durchführung der Maßnahmen zur Verbesserung der Datenstruktur gemäß der Ergebnisse aus Kapitel 4.4.2.3
3. Entwicklungsstufe 3: Medikationsverordnungen nach Durchführung der Maßnahmen nach der Überführung nach RxNorm gemäß der Ergebnisse aus Kapitel 4.4.3.2

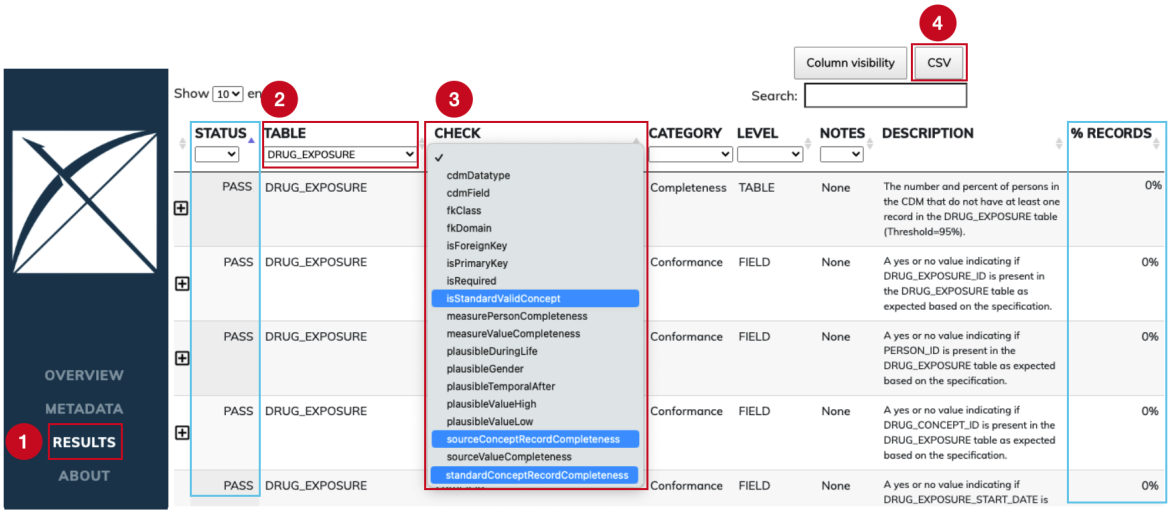


Abbildung 3.6: OHDSI DQD Anwendung - Filterung der quantitativen Bewertung für Tabelle drug_exposure

Die Medikationsverordnungen werden mit dem ATC WHO Code in die OMOP Datenbank abgelegt. Lediglich Medikationsverordnungen, mit einem der in Tabelle 4.7 dargestellten ATC-GM Codes, verbleiben mit den deutschen ATC-GM Codes, weil sie keine Entsprechung in der WHO Version von ATC haben. Für die quantitative Bewertung werden von denen in Kapitel 3.6 genannten Prüfungen (Tabelle 2.2) drei genutzt. Diese prüfen die Medikationsverordnungen in OMOP hinsichtlich Konformität und Vollständigkeit der und unter dem Aspekt der Verwendung von standardisierten Terminologien zur Sicherung der semantischen Bedeutung prüfen. Die relevanten Prüftypen sind:

1. isStandardValidConcept
2. sourceConceptRecordCompletness
3. standardConceptRecordCompletness

Die quantitative Bewertung erfolgt unter Verwendung der vorab genannten drei Prüfungen für die Datenfelder *drug_concept_id* und *drug_source_concept_id* in der Tabelle *drug_exposure*.

Es soll ermittelt werden, wie groß der Anteil der Medikationsverordnungen ist, bei denen die *concept_id* in der Spalte *drug_source_concept_id* (Prüfungstyp: sourceConceptRecordCompleteness) und *drug_concept_id* (Prüfungstyp: standardConceptRecordCompleteness) null entspricht. Außerdem soll ermittelt werden, wie groß der Anteil der Medikationsverordnungen ist, bei denen die genutzte *concept_id* keinem Standard-Konzept für Medikationsdaten entspricht (Prüfungstyp: isStandardValidConcept).

Für jede der drei Stadien der Medikationsverordnung wird die Generierung der statistischen Analysen des OHDSI DQD separat durchgeführt. Anschließend werden jeweils die Ergebnisse aus der Webanwendung des OHDSI DQD (siehe Abbildung 3.6) exportiert. Abbildung 3.6 zeigt rot markiert und nach Reihenfolge nummeriert das Vorgehen. Zunächst wird über das Menü links zu den (1) Ergebnissen navigiert. Aus dem (2) Drop-Down Menü für die Tabellen wird nach *drug_exposure* gefiltert. Bei den (3) Typen der Prüfung (Check) werden nur die blau markierten Elemente für die Bewertung berücksichtigt. Die Ergebnisse können über die (4) Exportmöglichkeit (CSV Button) als CSV Datei gespeichert werden und anschließend durch die Autorin dieser Arbeit in Excel ausgewertet werden.

Durch die 3-Schritt Analyse, basierend auf den drei genannten Entwicklungsstufen der Medikationsverordnungen, lassen sich die durchgeführten Maßnahmen im Einzelnen im Vergleich zum initialen Status der Daten bewerten.

3.7 Schaffung von Transparenz

Die Aktivitäten zur Identifikation von Inhibitoren in RWD sowie die Etablierung von Maßnahmen zu deren Reduktion und deren Bewertung sind wichtige Instrumente, um bereits erhobene Daten aus der Vergangenheit für die Forschung nutzbar zu machen. Sowohl die Etablierung der Maßnahmen als auch das Wissen um fortbestehende Inhibitoren der RWD für Forschungszwecke verlangt ein fachliches Verständnis von Expert:innen. Aus diesem Grund bedarf es eines Feedback Mechanismus, um Inhibitoren und mögliche Auswirkungen mit Kliniker:innen diskutieren zu können (Bradley, 2004; Taggart et al., 2015).

So kann ein besseres beiderseitiges Verständnis geschaffen werden, um:

- die Auswirkungen der identifizierten Inhibitoren einschätzen zu können, und so zukünftig bereits während der Generierung der Daten die Auswirkungen auf die Sekundärnutzung der RWD für die Forschung zu kennen
- die Aufwände für die Nachbearbeitung der Daten während des Prozesses der Integration und Überführung in das Forschungsdatenrepository OMOP zu verringern
- mögliche Gründe für existierende Inhibitoren in einem interdisziplinären Team zu diskutieren, um gemeinsam an der Behebung arbeiten zu können

Auf Basis des beiderseitigen Verständnisses kann eine transparente Darstellung der Defizite von RWD zu einer Entscheidungsunterstützung für die Teilnahme an retrospektiven Beobachtungsstudien innerhalb der OHDSI Forschungsgemeinschaft beitragen. Unter Einbindung medizinischer Expertise wurden folgende Ziele an eine Visualisierung identifiziert: Darstellung von (a) Strukturiertheit und (b) Überführbarkeit in internationale Terminologien von den Medikationsverordnungen, sowie die Möglichkeit der interaktiven Such- und Filterfunktion sowohl nach Wirkstoffname als auch ATC Code.

Zur Wahl der Art der Visualisierung wurde auf entsprechende Literatur zurückgegriffen. Das systematische Review von Weissberger et al. (Weissgerber, Milic et al., 2015) zeigt, dass Balkendiagramme die vorherrschende Art der grafischen Präsentation von Daten und Ergebnissen in wissenschaftlichen Publikationen darstellen. Dabei ist bekannt, dass diese Art von Visualisierung problematisch sein kann, weil Datenverteilungen nicht abbildbar sind (Weissgerber, Milic et al., 2015; Weissgerber, Winham et al., 2019).

Trotz vorhandener Empfehlungen zur Verbesserung der Transparenz und Reproduzierbarkeit in der Forschung und zur Verwendung von interaktiven Visualisierungen, finden diese bisher eine geringe Anwendung (Ostropolets et al., 2023; Weissgerber, Garovic et al., 2016). Die Verwendung von Streudiagrammen als alternative Visualisierungsmethode bietet sich an (Weissgerber, Milic et al., 2015). Die beschriebenen Nachteile in den einfachen Balkendiagrammen wurde anhand von einzelnen Wirkstoffen mit medizinischem Personal bestätigt. Die Diskussion resultierte in der Wahl zur Nutzung von interaktiven Streudiagrammen zur transparenten Darstellung der Informationen. Interaktive Darstellungen der Daten erlauben es zudem, komplexe Beziehungen und Muster zwischen in Daten zu untersuchen.

Daher werden in dieser Arbeit die Medikationsverordnungen auf Basis der ATC Codes in Streudiagrammen wie folgt bereitgestellt:

1. Strukturiertheit der Medikationsverordnungen
2. Überführbarkeit der Medikationsverordnungen nach RxNorm

Durch die interaktive Visualisierung mit Such- und Filterfunktionen können Forschende gezielt nach bestimmten Informationen suchen und relevante Zusammenhänge in den Daten identifizieren. Diese Darstellung als Streudiagramm wurde auf Basis von Python (Version 3.9.1) und der Bokeh-Bibliothek entwickelt und in Zusammenarbeit mit klinischen Expert:innen diskutiert. Der Quellcode zur Visualisierung ist auf Zenodo verfügbar (Reinecke, 2023b). Die Dokumentation zum Quellcode befindet sich in Anhang A.

Aufgrund der Verwendung der logarithmischen Skala für die y-Achse für das Streudiagramm zur Überführbarkeit der Medikationsverordnungen nach RxNorm, kann die Zahl null nicht dargestellt werden. Um die ATC Codes abzubilden, für die keine Überführung nach RxNorm möglich ist (Mapping entspricht null), wurde der Darstellbarkeit halber für alle betreffenden ATC Codes der Wert des Mappings nach RxNorm auf von null auf 0,7 geändert.

4 Ergebnisse

4.1 Ergebnisse Literaturrecherche

Die Literaturrecherche zeigt den aktuellen Stand der Forschung zur Nutzung von OMOP weltweit. Es wird in Abschnitt 4.1.1 zunächst ein allgemeiner Überblick zu den Ergebnissen der Suche und der eingeschlossenen Literatur gegeben. Danach werden die fachlich relevanten Themen der eingeschlossenen Publikationen in Abschnitt 4.1.2 vorgestellt, um ein Verständnis über die relevanten Themen zu erhalten und zu verstehen, ob bereits retrospektive Studien über mehrere Standorte durchgeführt werden. Eine Übersicht über die zeitliche Entwicklung der Publikationen, auch im Kontext der zuvor identifizierten Themen, wird in Abschnitt 4.1.3 gegeben, um zu zeigen, ob die Relevanz des Themas in der Literatur entsprechend quantitativ repräsentiert wird. Abschnitt 4.1.4 widmet sich der geografischen Verteilung der Publikationen, um zu prüfen, ob Studien multizentrisch unter Verwendung von Daten aus mehreren Ländern gemäß der Zielstellung der OHDSI Forschungsgemeinschaft bereits durchgeführt wurden. Abschließend wird in Abschnitt 4.1.5 ein Schwerpunkt auf die Forschungsteams an deutschen Universitäten gelegt, um die Beteiligung an durchgeführten Studien und die allgemeinen Themenschwerpunkte darzulegen. Am Ende dieses Kapitels werden die Ergebnisse in Abschnitt 4.1.6 zusammengefasst dargestellt, um einen ganzheitlichen Überblick über den aktuellen Forschungsstand geben zu können und die initial gestellte Forschungsfrage 1 aus Kapitel 1.3 beantworten zu können.

Die Ergebnisse der durchgeführten Literaturrecherche wurden von der Autorin der vorliegenden Arbeit als Scoping Review veröffentlicht (Reinecke, 2021a).

4.1.1 Allgemeine Übersicht

Das PRISMA Flow Chart der durchgeführten Literaturrecherche wird in Abbildung 4.1 dargestellt und verschafft eine Übersicht über die Mengengerüste der Identifikation, des Ausschlusses sowie des Einschlusses von Publikationen in die Literaturrecherche.

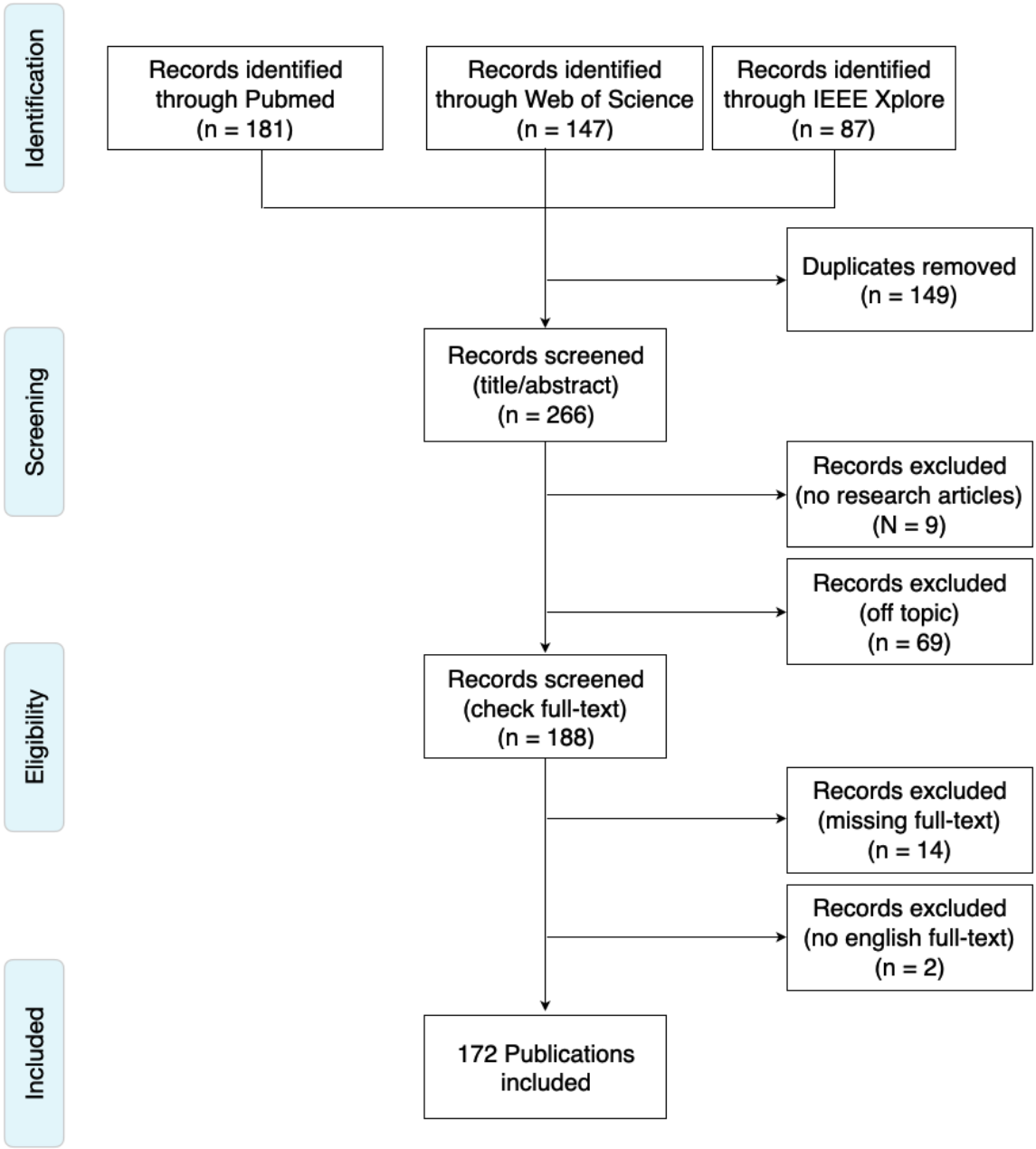


Abbildung 4.1: PRISMA Flow Chart

Im Rahmen der durchgeführten Suche wurden 181 Publikationen in Pubmed, 147 Publikationen in Web of Science und 87 Publikationen in IEEE Xplore eingangs identifiziert. Nach der Entfernung von 149 Duplikaten verblieben 266 Publikationen für das Title/Abstract Screening in Abschnitt 3.2.2. Beim Title/Abstract Screening wurden 9 Publikationen ausgeschlossen, bei denen es sich nicht um Studien handelt, sondern um Jahrbücher von Journalen oder Konferenzen, Programmen von Konferenzen oder Verzeichnisse von Autoren. Außerdem wurden 69 Publikationen thematisch ausgeschlossen.

Weitere 14 Publikationen konnten nicht eingeschlossen werden, weil kein Volltext verfügbar war, 2 Publikationen verfügen nicht über einen Volltext in englischer Sprache. Es wurden 172 Publikationen in das Scoping Review eingeschlossen. Die vollständige Liste der eingeschlossenen Literatur ist auf Zenodo verfügbar (Reinecke, 2021a).

Bei den eingeschlossenen Publikationen handelt es sich in der Mehrheit von 120 Publikationen um Veröffentlichungen in wissenschaftlichen Journalen. Lediglich 52 Publikationen wurden auf wissenschaftlichen Konferenzen veröffentlicht. Die Menge der Konferenzpublikationen bleibt über die Jahre konstant, die der Journalpublikationen steigt kontinuierlich.

4.1.2 Fachliche Themen

Abbildung 4.2 enthält die Aufteilung der Publikationen nach fachlichen Themen gemäß der in Kapitel 3.2 genannten Dimensionen, sowie für die Dimension „ Nutzung“ von OMOP zusätzlich die Unterteilung in sieben Unterkategorien. Die Analyse der Dimensionen zeigt, dass die „Nutzung“ von OMOP mit 105 Publikationen mehr als die Hälfte der insgesamt 172 eingeschlossenen Publikationen ausmacht.

Die Dimensionen „Mapping“ von Daten nach OMOP mit 22 Publikationen und „Evaluation“ von OMOP für eine etwaige Nutzung mit 20 Publikationen sind neben der Nutzung von OMOP ebenfalls sehr relevant. In 16 Publikationen wurde OMOP lediglich als ein mögliches Datenmodell erwähnt. Vokabulare und klinischen Terminologien werden in fünf Publikationen als zentrales Thema betrachtet (Banda, 2019; Gruhl et al., 2020; Jiang, Yu et al., 2019; Warner et al., 2019; Y. Zhang et al., 2020). Warner et al., 2019 und Jiang, Yu et al., 2019 haben neue standardisierte und nicht standardisierte Vokabulare bereitgestellt. Die Arbeit von Gruhl et al. (Gruhl et al., 2020) demonstriert die Verteilung von Vokabularen in

virtualisierten Umgebungen zwischen mehreren Standorten zur Sicherung der Verwendung von einheitlichen Vokabular-Versionen. Die Verknüpfung zwischen nationalen Vokabularen wie dem Normalized Chinese Clinical Drug (NCCD) und dem Standardvokabular RxNorm für Medikationsdaten wurde von Y. Zhang et al., 2020 gezeigt. Die Bereitstellung von zusätzlichen Beziehungen zwischen Vokabularen in OMOP und Vokabularen, die nicht durch OHDSI bereitgestellt werden, wurde von Banda, 2019 umgesetzt.

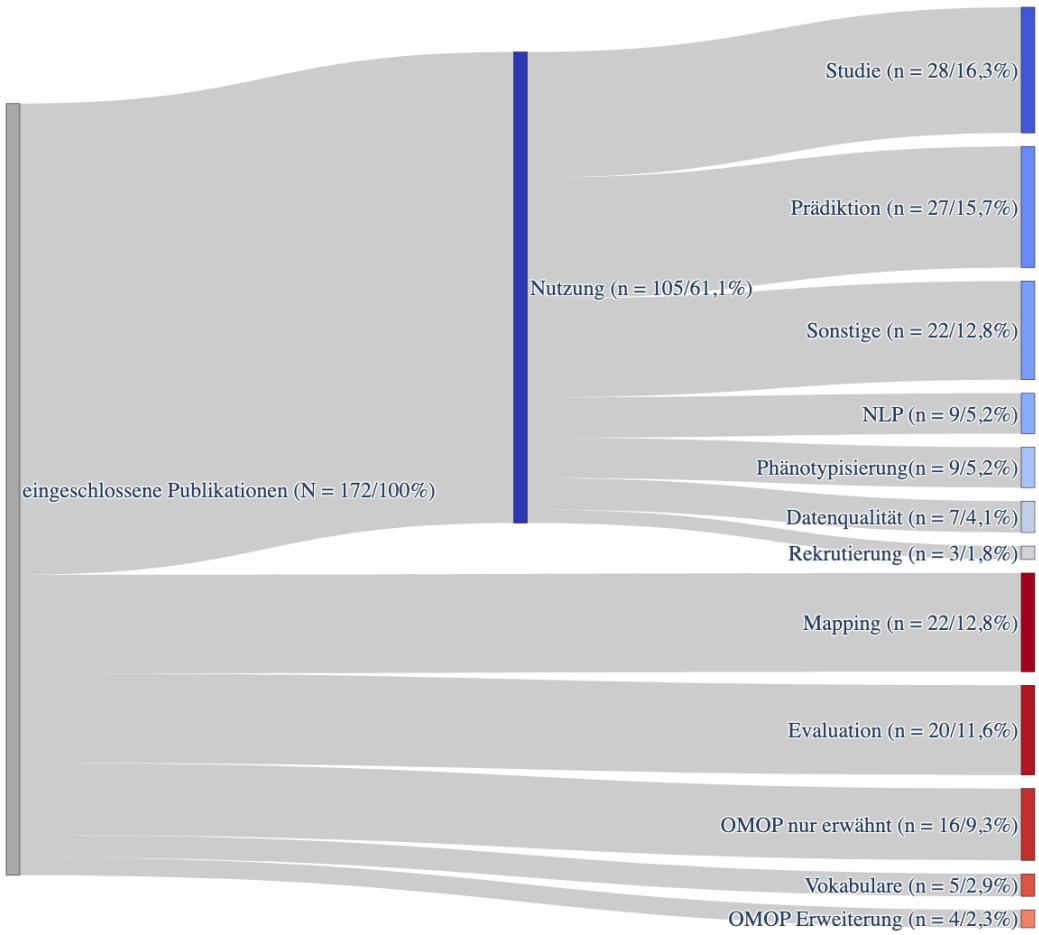


Abbildung 4.2: Anzahl Publikationen nach Dimension und Unterkategorie für die Dimension Nutzung

Die meisten Publikationen der Dimension „Nutzung“ gehören zur Kategorie „Studie“. Das OMOP CDM wird in 28 Publikationen zur Durchführung von retrospektiven Beobachtungsstudien („Studie“) eingesetzt. Die Kategorie „Studie“ nimmt mit 16,3% (n=28/N=172) den größten Anteil der Dimension „Nutzung“ ein. Eine detaillierte Betrachtung der 28 OHDSI Netzwerkstudien erfolgt mit Fokus auf die notwendigen Datengruppen hinsichtlich der Anforderungen in Kapitel 4.2.2.

Neben der Durchführung von retrospektiven Beobachtungsstudien ist die Kategorie „Prädiktion“ mit 27 Publikationen von hoher Relevanz. Dieser Themenbereich ist breit gestreut und beinhaltet beispielsweise die Entwicklung und die Validierung von Vorhersagemodellen, die empirische Überprüfung von Methoden zur Vorhersage, die Durchführung von Prognosen für bestimmte Fragestellungen, aber auch die Entwicklung von Werkzeugen und Software für die Durchführung von Vorhersagen.

Die Kategorie „Sonstige“ stellt mit 22 Publikationen ebenfalls eine große Rubrik dar. In dieser Kategorie befinden sich Publikationen zu diversen Themen, wie beispielsweise zu FHIR von Jiang, Kiefer et al., 2017 und ein Vorschlag zur Prüfung und Durchführung der Anonymisierung von Daten durch Jeon et al., 2020. Viele der Publikationen dieser Kategorie lassen sich nicht eindeutig zu einer der anderen Kategorien zuordnen. Bei keiner der Publikationen der Kategorie „Sonstige“ handelt es sich um eine retrospektive Beobachtungsstudie.

Die überwiegende Anzahl der Publikationen in der Kategorie „NLP“ befassen sich mit der Extraktion von Einschlusskriterien aus Freitexten, um den Rekrutierungsprozess zur Identifizierung potenzieller Teilnehmender für klinische Studien zu optimieren (Butler et al., 2018; J. R. Almeida et al., 2020; Liu et al., 2020; Si et al., 2017; Yuan et al., 2019).

Das Thema „Phänotypisierung“ wird in neun Publikationen adressiert. Diese Publikationen befassen sich mit der Entwicklung und Implementierung von Algorithmen unter Verwendung VON RWD zur Definition von Kohorten für Forschungszwecke. Dabei werden verschiedene Strategien zur Verbesserung der Übertragbarkeit und Skalierbarkeit der Phänotypisierung über verschiedene Standorte und Institutionen unter Verwendung von OMOP hinweg betrachtet. Die Studien zeigen das Potenzial, die Phänotypisierung zu verbessern, den Implementierungsaufwand zu verringern und Übertragbarkeit zu gewährleisten, um die Reproduktion der Ergebnisse auf andere Standorte zu ermöglichen.

In sieben Publikationen wurde zu dem Thema Datenqualität in OMOP geforscht. In diesen Artikeln werden unter anderem verschiedene Methoden und Modelle zur Bewertung der Datenqualität in OMOP diskutiert (Dixon et al., 2020; H. Spengler et al., 2020; Huser et al., 2016; Ta et al., 2019). Die Publikation von Roger et al. (Rogers et al., 2019) vergleicht unterschiedliche Ansätze zur Bewertung von Datenqualität miteinander.

Auch das Thema der Rekrutierung möglicher Personen für klinische Studien unter Verwendung von OMOP wurde in drei Publikationen näher betrachtet. Die Autorin der vorliegenden Arbeit hat dazu ein Konzept einer Rekrutierungsinfrastruktur unter Verwendung von OMOP-konformen und harmonisierten klinischen Daten veröffentlicht (Reinecke, Gulden et al., 2020), welches im Rahmen des MIRACUM Projektes implementiert und evaluiert wurde.

4.1.3 Zeitliche Entwicklung

Abbildung 4.4 zeigt ein stetiges Wachstum der Anzahl der Publikationen insgesamt. Im Jahr 2016 wurden insgesamt 14 Publikationen veröffentlicht. Im Jahr 2020 lag die Anzahl der Publikationen bei 57 und damit um ein Vielfaches höher. Das Jahr 2021 muss gesondert betrachtet werden, da das Scoping Review im Februar 2021 durchgeführt wurde, und daher nur knapp 1/6 des gesamten Jahres repräsentiert.

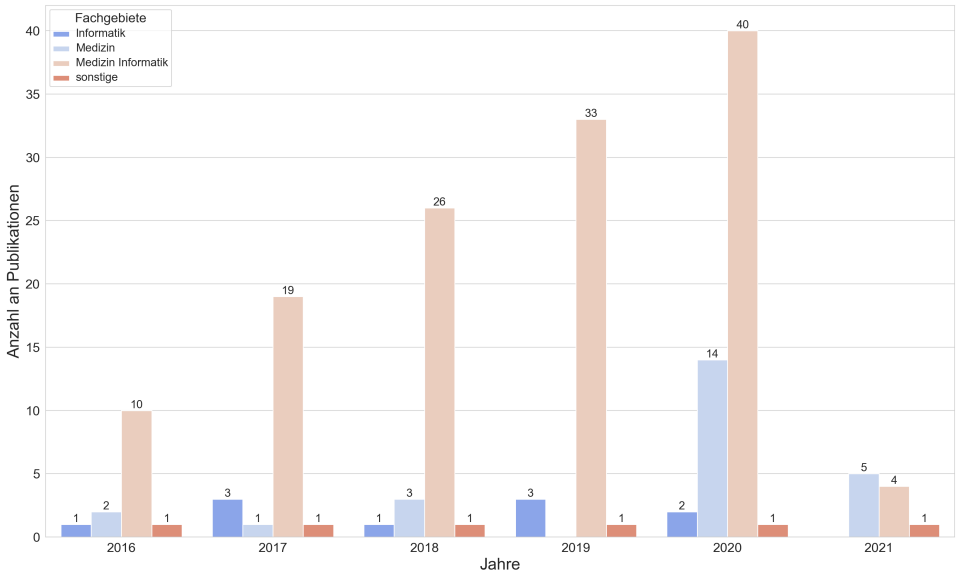


Abbildung 4.3: Publikationen pro Jahr nach fachlichem Kontext

Das Fachgebiet der Medizininformatik zeigt jedes Jahr ein kontinuierliches Wachstum und umfasst thematisch mit insgesamt 128 Publikationen den größten Bereich. Die Publikationen aus dem Fachgebiet der Informatik zeigen über die Jahre kein Wachstum, sondern bleiben stets auf einem geringen Niveau und umfassen insgesamt nur zehn Publikationen. Auch die sonstigen Fachgebiete haben mit fünf Publikationen kaum Relevanz.

Interessant ist die Entwicklung der Publikationen im Fachgebiet der Medizin. In den Jahren 2016 bis 2018 nehmen sie einen sehr geringen Anteil ein. Im Jahr 2020 zeigt sich ein sprunghafter Anstieg auf 14 Publikationen. Auch im Jahr 2021 liegt die Zahl der Publikation im Fachgebiet der Medizin mit fünf bereits im Vergleich zu den anderen Fachgebieten am höchsten.

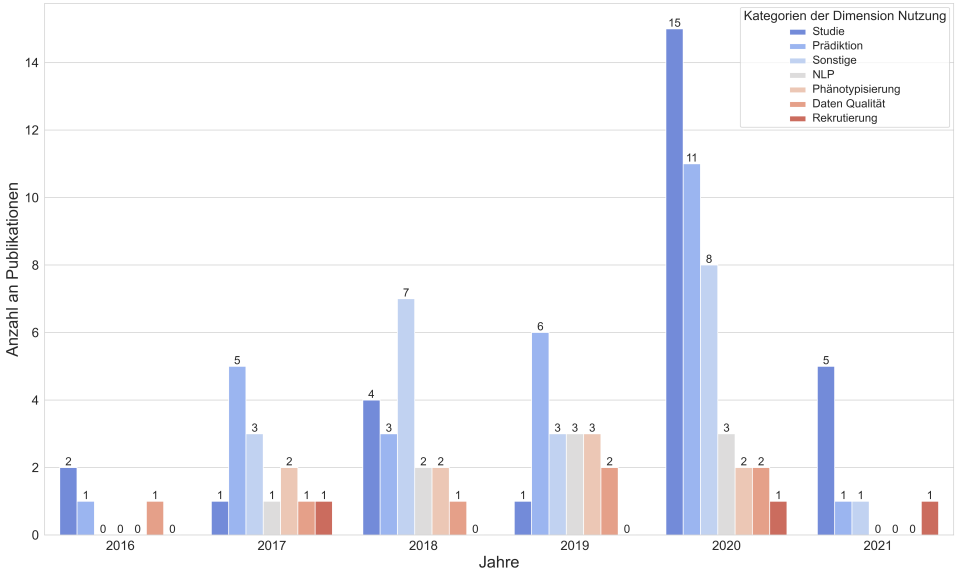


Abbildung 4.4: Publikationen pro Jahr für die Dimension Nutzung, nach fachlicher Kategorie

Der zeitliche Verlauf zeigt in Abbildung 4.4 einen großen Anstieg für die fachlichen Kategorien Studie, Prädiktion und Sonstige im Jahr 2020. Der Anstieg der Nutzung von OMOP, insbesondere auch der Anstieg die durchgeführten retrospektiven Studien im Jahr 2020, zeigt einen direkten Zusammenhang mit dem Anstieg der Publikationen im Jahr 2020 in medizinischen Fachjournalen.

4.1.4 Geografische Verteilung

Die Analyse der Publikationen nach dem Land der Institutionszugehörigkeit der Erstautor:innen ist in Abbildung 4.5 durch die farbliche Kennzeichnung der Länder dargestellt. Die Agenda zeigt das Land, die farbliche Kodierung und die Anzahl der Publikationen in Klammern. Die Mehrheit der Publikationen wurde von Erstautor:innen US-amerikanischer Universitäten veröffentlicht.

Von den 172 eingeschlossenen Publikationen sind insgesamt 111 (64%, 111/172; davon 109 alleinige Erstautorenschaften und zwei geteilte Autorenschaften mit anderen Ländern) den USA zuzuordnen. Die Universitäten in Südkorea haben mit 17 Publikationen vergleichsweise zu den USA einen geringen Anteil an der Gesamtmenge, dennoch ist Südkorea das Land mit der zweithöchsten Anzahl an Publikationen. Die Forschungsteams deutscher Universitäten sind in dem Thema OHDSI und OMOP ebenfalls sehr aktiv, es wurden zehn Publikationen von Erstautor:innen deutscher Universitäten veröffentlicht. Auch in Großbritannien ist das Thema OHDSI und OMOP von großem Interesse, hier wurden immerhin acht Publikationen veröffentlicht. Insgesamt zeigt die Abbildung 4.5 über die Weltkarte, dass das Thema OHDSI OMOP globale Relevanz hat und großes Forschungsinteresse weltweit besteht.

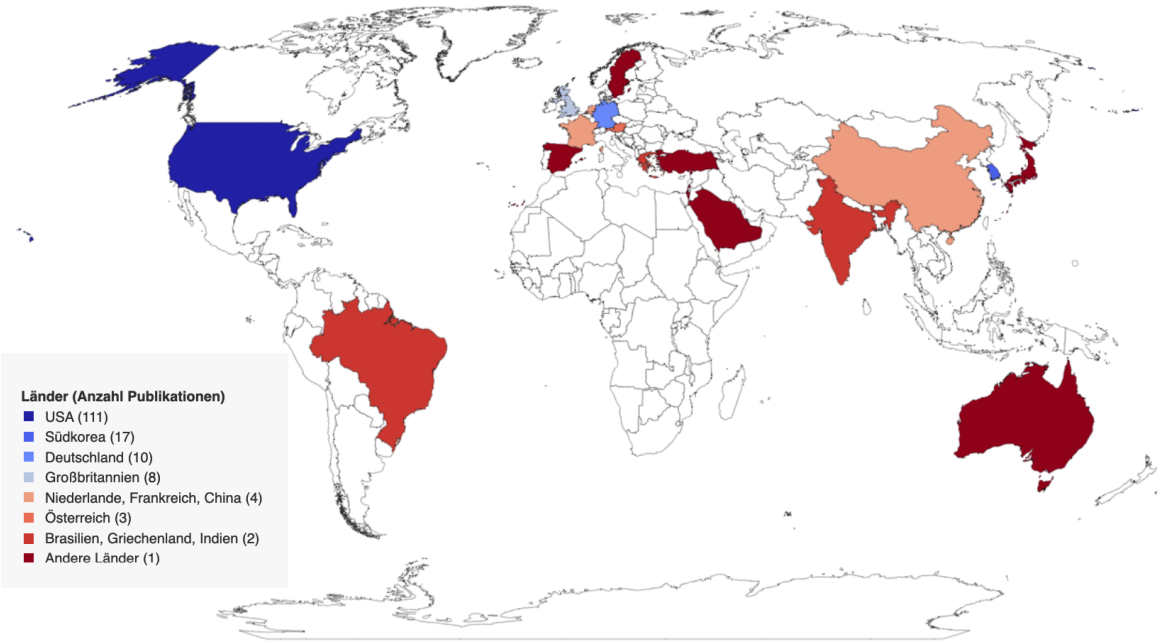


Abbildung 4.5: Anzahl Publikationen pro Land

Die Kennzeichnung der Publikationen, bei denen Daten aus unterschiedlichen Ländern verwendet wurden, erfolgte in der Dimension „Nutzung“. Insgesamt wurden bei 18 Publikationen (von 105 der Dimension „Nutzung") Daten aus mehreren Ländern verwendet.

Abbildung 4.6 enthält eine farbliche Kennung des Anteils der Publikationen mit Daten aus mehreren Ländern. Lediglich die Publikationen der Kategorien „Studie“, „Prädiktion“ und „Sonstige“ verwendeten Daten aus mehreren Ländern. Die Zahl der Publikationen der Kategorie „Studien“ zeigt einen vergleichsweise hohen Anteil mit Daten aus mehreren Ländern.

Die im Review eingeschlossenen Publikationen zeigen, dass die Durchführung von Studien im medizinischen Kontext auf Basis von OMOP weltweit und länderübergreifend bereits stattfindet. Bisher allerdings fehlt eine Beteiligung des Standorts Deutschland mit Daten aus der stationären Krankenhausversorgung.

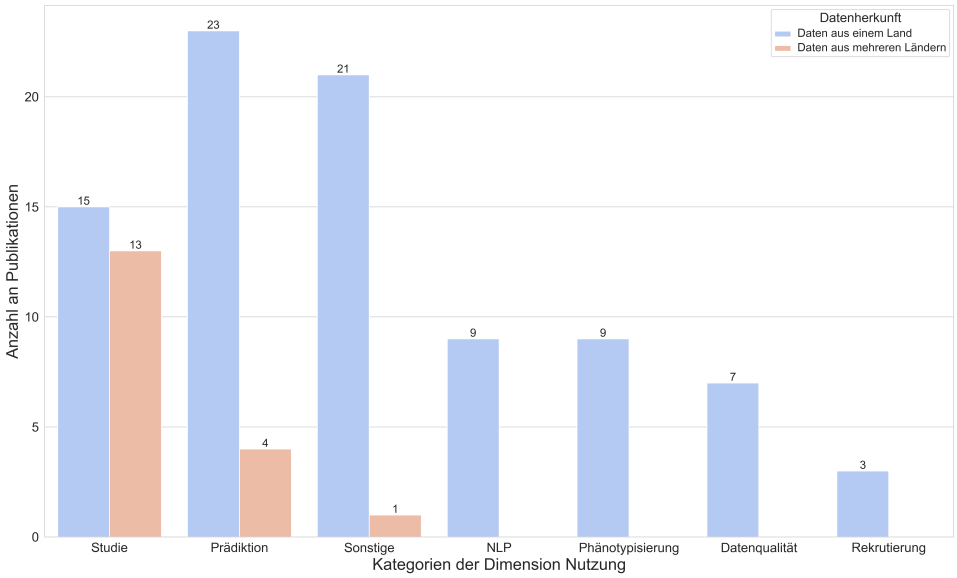


Abbildung 4.6: Herkunft der Daten in Publikationen - Dimension Nutzung, Kategorien

4.1.5 Überblick der Publikationen deutscher Universitäten

Die Literaturrecherche zeigt eine steigende Anzahl an wissenschaftlichen Publikationen zum Thema OMOP und OHDSI im Allgemeinen, insbesondere im Hinblick auf retrospektive Beobachtungsstudien der OHDSI Forschungsgemeinschaft. Trotz der vergleichsweise großen Anzahl an Publikationen deutscher Autoren existiert bisher keine Beteiligung an retrospektiven Beobachtungsstudien im OHDSI Netzwerk auf Basis von OMOP unter Verwendung von stationären Versorgungsdaten aus deutschen Universitätskliniken. Vielmehr fokussieren sich die bisherigen Veröffentlichungen, zusammengefasst in Tabelle 4.1, auf Trends und Initiativen im Bereich der Forschung mit RWD, dem Mapping und Transfer klinischer Daten in das standardisierte Datenformat OMOP, der Erarbeitung von Architekturkonzepten, sowie der Entwicklung von Software basierend auf OMOP.

Trends im Bereich von groß angelegten Analysen von RWD im Gesundheitswesen basierend auf RWD werden von Tresp et al. (Tresp et al., 2016) gezeigt. Das Autor:innenteam erwartet,

dass die vorgestellten Trends in den kommenden Jahren wegweisende Veränderungen hinsichtlich der Organisation des Gesundheitswesens weltweit bestimmen werden. Neben der Diskussion von Initiativen auf politischer Ebene zur Förderung des Paradigmenwechsels von der Dokumentation auf Papier hin zu vollständig digitaler Dokumentation im Gesundheitswesen, werden zudem Netzwerke zur Analyse von Daten durch Zusammenschlüsse von Versorgungs- und Forschungseinrichtungen, aber auch Industriepartnern vorgestellt, die dazu dienen sollen, Entscheidungsunterstützungssysteme zu entwickeln. Erwähnenswert in diesem Zusammenhang ist das kollaborative „Indiana Network for Patient Care" Projekt, bei dem Gesundheitsdaten aus der klinischen Versorgung für Forschungszwecke in Netzwerken zusammengeführt werden. Es stellt ein umfangreiches Datenrepertoire von mehr als 100 Krankenhäusern, niedergelassenen Ärzt:innen, Laboren und anderen Einrichtungen bereit. Insgesamt werden Daten von über 15 Millionen Patient:innen für Forschungsprojekte zur Verfügung gestellt. Diese Daten werden unter anderem innerhalb der OHDSI Forschungsgemeinschaft auf Basis von OMOP genutzt. Zusammenfassend wird in der Publikation darauf hingewiesen, dass die Vollständigkeit und die Genauigkeit von Informationen hochrelevant ist, um kausale Zusammenhänge ableiten zu können und dass eine zeitnahe Integration von RWD in die Forschung notwendig ist, um Erkenntnisse frühzeitig in der klinischen Versorgungen zum Wohle der Patient:innen nutzen zu können.

Tabelle 4.1: Publikationen von Autorenteams deutscher Universitäten

Jahr	Erstautor:in	Titel	Dimension	Unterkategorie	Fachgebiet
2016	Tresp, Volker	Going Digital: A Survey on Digitalization and Large-Scale Data Analytics in Healthcare	mentioned		informatics
2018	Prokosch, Hans-Ulrich	MIRACUM: Medical Informatics in Research and Care in University Medicine.	mentioned		mi
2018	Maier, Christian	Towards Implementation of OMOP in a German University Hospital Consortium.	mapping		mi
2019	Gruendner, Julian	KETOS: Clinical decision support and machine learning as a service - A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services.	usage	prediction	mi
2019	Freitas da Cruz, Harry	MORPHER - A Platform to Support Modeling of Outcome and Risk Prediction in Health Research	usage	prediction	informatics
2020	Gruhl, Mirko	Specification and Distribution of Vocabularies Among Consortial Partners.	vocabulary		mi
2020	Reinecke, Ines	Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP.	usage	recruitment	mi
2020	Fischer, Patrick	Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT.	mapping		mi
2020	Unberath, Philipp	EHR-Independent Predictive Decision Support Architecture Based on OMOP.	usage	prediction	mi
2020	Spengler,Helmut	Improving Data Quality in Medical Research: A Monitoring Architecture for Clinical and Translational Data Warehouses	usage	DQ	informatics

Erstmals wird im Jahr 2018 durch Maier et al.(Maier et al., 2018) über Aktivitäten berichtet, um Gesundheitsdaten aus der klinischen Versorgung von acht deutschen Universitätskliniken als

Partner des MIRACUM Projektes nach OMOP zu überführen. Dabei werden die in Deutschland standardisierten Abrechnungsdaten der beteiligten Universitätskliniken anonymisiert auf Basis der Behandlungsfälle für Patient:innen zusammen mit den gestellten Diagnosen, durchgeführten Operationen und Prozeduren nach OMOP überführt. Die Diagnosen werden in Deutschland mit einer angepassten Version von ICD10 der WHO, der ICD10 German Modification (GM) erfasst. Das Forschungsteam hat dazu die vorhandene Übersetzung der ICD10 WHO Diagnosen in die weltweit umfassendste Ontologie in der Medizin SNOMED-CT genutzt. SNOMED-CT ist innerhalb von OMOP als Standardterminologie für die Ablage von Diagnosen in OMOP zu verwenden. Die Mehrheit der deutschen Diagnosen in ICD10 GM (63,3 %) konnten aufgrund der Übereinstimmung der Kodes zwischen ICD10 GM und ICD10 WHO direkt verwendet werden. Weitere 34,2 % der ICD10 GM Mappings konnten über die Methode der Verallgemeinerung der Diagnosen in SNOMDED-CT über das sogenannte Up-Hill Mapping einer Hierarchiestufe in ICD10 WHO höher zugeordnet werden. Ein weiterer Generalisierungsschritt von zwei Stufen in der ICD10 Hierarchie konnte außerdem eine kleine Menge von 1,3 % der ICD10 GM Codes auf einen generischen Code in ICD10 WHO zuordnen. Eine kleine Minderheit von 1,2 % der Daten konnte keinem validen SNOMED-CT Code zugeordnet werden, weil die ICD10 GM Codes entweder nicht in der ICD10 WHO Version existierten oder aber weil die Übersetzung von ICD10 WHO nach SNOMED-CT nicht existiert. Die Operationen und Prozeduren wurden unter Verwendung der in Deutschland für die Dokumentation und Abrechnung genutzten Terminologie Operationen- und Prozedurenschlüssel (OPS) in OMOP gespeichert. Eine Harmonisierung in ein internationales Format zur Sicherung der semantischen Interoperabilität hat nicht stattgefunden. Diese Arbeit zeigt erstmalig die grundsätzliche Machbarkeit der Überführung von klinischen Daten aus der Versorgung von Universitätskliniken nach OMOP. Nicht enthalten in dieser Arbeit sind jedoch die im Rahmen einer Krankenhausbehandlung verabreichten Medikamente und Laborwerte.

Die Publikation von Prokosch et al. (Prokosch, Acker et al., 2018) stellt das im Rahmen der MI-I durch das BMBF geförderte Konsortium MIRACUM im Detail vor. Das MIRACUM Konsortium ist eine Partnerschaft von Universitätskliniken und Industriepartnern mit dem Ziel, Datenintegrationszentren (DIZ) zu etablieren, die als interoperable Instanzen innovative Lösungen für die Versorgung von Patient:innen, aber auch für die Forschung im Gesundheitswesen bereitstellen. Dabei wird die Idee von MIRACUM mit der Idee von OHDSI als Forschungsnetzwerk in gleich gesetzt. Außerdem wird die Harmonisierung der in den DIZ geschaffenen

Datenmengen auf Basis international etablierter Terminologien und Datenmodellen, wie beispielsweise OMOP, als eines der Ziele des MIRACUM Projektes genannt.

Die Publikation von Gründner et al. (Gruendner et al., 2019) stellt die entwickelte KETOS-Plattform vor, welche es Forschenden ermöglicht, statistische Auswertungen sowie die Entwicklung klinischer Entscheidungsunterstützungssysteme auf Basis von prädiktiven Modellen und der Rückführung von Ergebnissen in die klinische Versorgung durchzuführen. Die Datenhaltung für die KETOS-Plattform erfolgt auf Basis von OMOP. Die Sicherstellung der semantischen Interoperabilität und die Überführung der Daten nach OMOP sind nicht Teil der Arbeit von Gründner et al., vielmehr basiert der Datentransfer auf der zuvor erwähnten Publikation von Maier et al. (Maier et al., 2018), eine Erweiterung der Datenbasis ist nicht durchgeführt worden.

Die Publikation von Freitas da Cruz et al. (Freitas Da Cruz et al., 2019) stellt ein entwickeltes Werkzeug namens MORPHER vor, welches die von den Autor:innen der Publikation identifizierten Anforderungen an die Softwareentwicklung sowie Validierung klinischer Prognosemodelle adressiert. Um die Anforderung nach Unterstützung existierender Standards klinische Prognosemodelle zu erfüllen, werden Standards wie beispielsweise OMOP genutzt. Allerdings gibt die Arbeit keinerlei Aufschluss, inwieweit die Daten nach OMOP überführt wurden.

Die Publikation von Gruhl et al. (Gruhl et al., 2020) stellt eine Infrastruktur mit einer OMOP Datenbank inklusive der minimal notwendigen Terminologien zur Nutzung innerhalb des MIRACUM Konsortiums vor. Die Lösung fokussiert sich dabei auf einer einfachen Verteilung der Infrastruktur zwischen verschiedenen Standorten und basiert auf der Containertechnologie Docker und dem OHDSI Broadsea Toolset.

Die Publikation von Reinecke et al. (Reinecke, Gulden et al., 2020), Autorin der vorliegenden Dissertation, stellt das Konzept eines Clinical Trial Support System (CTRSS) zu der automatisierten Rekrutierung von potenziellen Patient:innen zur Unterstützung des Use Case 1 des MIRACUM Projekts „Alerting in care - IT support for patient recruitment“, vor. Es basiert auf OMOP als Datenbasis, OHDSI ATLAS, dem FHIR HAPI Server und einer Webanwendung für das Screening potenzieller Kandidat:innen für den Einschluss in klinische Studien. Wichtiger Fokus ist der modulare, komponentenbasierte Aufbau des Systems gemäß offener Stan-

dards unter Sicherstellung der semantischen Interoperabilität, um das CTRSS zukünftig auch auf Standorte außerhalb des MIRACUM Konsortiums übertragen zu können. Die Erprobung des vorgestellten CTRSS steht noch aus und wird nachfolgend im Rahmen einer Evaluationsstudie durchgeführt.

Die Publikation von Fischer et al. (Fischer et al., 2020) stellt ein Konzept vor, wie ein Register für den Forschungsbereich der pulmonalen Hypertonie unter Nutzung eines standardisierten Datenformats wie OMOP entwickelt werden kann. Der Datentransfer der für das Register notwendigen Daten stellt eine große Herausforderung dar, deshalb wird in der Arbeit von den Autor:innen evaluiert, inwieweit HL7 FHIR und Extensible Stylesheet Language Transformation (XSLT) genutzt werden können, um einen generischen ETL Prozess zu entwickeln. Die Evaluation des ETL Prozesses erfolgt auf Basis der notwendigen Zeit für den Datentransfer als auch auf Basis der Abdeckungsrate der für das Register relevanten Daten in OMOP CDM.

Die Publikation von Unberath et al. (Unberath et al., 2020) stellt die Entwicklung, den Test und den Einsatz eines Systems zur Diagnose der Wahrscheinlichkeit eines Tumorrezidivs bei Melanompatient:innen, vor. Das System soll die Daten unabhängig von Krankenhausinformationssystemen für die elektronischen Patient:innenakten speichern. Die dazu notwendigen medizinischen Daten werden daher nach OMOP überführt. Es wurde in diesem Kontext eine eingeschränkte, für den Anwendungsfall notwendige Menge von acht Datenelementen, darunter auch OMICS Daten wie die Genexpression basierend auf der Terminologie HUGO Gene Nomenclature Committee (HGNC), verwendet.

Die Publikation von Spengler et al. (H. Spengler et al., 2020) stellt eine Möglichkeit vor, wie Entwicklungsteams von ETL Prozessen die Datenqualität der Quelldaten während des Datentransfers in das Zielformat überprüfen und protokollieren können. Während des Transfers erfasste Probleme können in einem anpassbaren Dashboard angezeigt werden. Das vorgestellte System unterstützt dabei unterschiedliche Dimensionen der Datenqualität wie die „Vollständigkeit“, „Konformität“ und „Plausibilität“. Das vorgestellte System ist kompatibel zu OMOP als Zielformat.

4.1.6 Zusammenfassung der Ergebnisse der Literaturrecherche

Zusammenfassend zeigt die Literaturrecherche, dass die Anzahl der Publikationen, die auf dem Datenmodell OMOP basieren, seit dem Jahr 2016 zugenommen hat. Im Jahr 2020 ist die Zahl der Publikationen auf ein Vierfaches der Publikationen des Jahres 2016 angestiegen. Die Publikationen beleuchten Themen wie die Nutzung von OMOP, das Mapping von Daten nach OMOP, die Evaluation des Datenmodells für spezifische Zwecke, Vokabulare und Terminologien, sowie Erweiterungen von OMOP. Allerdings wird OMOP in der Mehrheit der eingeschlossenen Publikationen genutzt, um retrospektive Beobachtungsstudien mit medizinischer Fragestellung durchzuführen und Prädiktionsmodelle zu entwickeln, maschinelles Lernen durchzuführen, NLP Fragestellungen zu beantworten, Phänotypisierung durchzuführen, Rekrutierung von Patient:innen zu optimieren und Datenqualität zu prüfen. Unter allen Nutzungsthemen von OMOP werden die retrospektiven Beobachtungsstudien der OHDSI Forschungsgemeinschaft am häufigsten durchgeführt.

Deutsche Forschungsteams haben gegenüber anderen Ländern vergleichsweise viele Publikationen veröffentlicht. Allerdings haben sie bisher nicht an den Studien der OHDSI-Forschungsgemeinschaft mit RWD aus der stationären Versorgung in Deutschland teilgenommen. Stattdessen konzentrieren sich diese Veröffentlichungen auf konzeptionelle Arbeiten, die Überführung von Daten nach OMOP oder die Entwicklung von Werkzeugen für eine Zusammenarbeit mit OMOP. Mögliche Ursachen auf Datenebene werden in dieser Arbeit in den folgenden Abschnitten untersucht. Dazu werden im nächsten Schritt in Kapitel 4.2.3 existierende Anforderungen an Daten zur Ablage und Nutzung in OMOP vorgestellt, die im späteren Verlauf der Arbeit den existierenden Inhibitoren in den RWD des UKD gegenüber gestellt werden.

4.2 Soll Zustand gemäß Anforderungsanalyse

Die Ergebnisse der Anforderungsanalyse zeigen auf, welche Daten für die Forschung im Rahmen der OHDSI Forschungsgemeinschaft unter Verwendung von OMOP notwendig sind. Dazu werden zunächst die existierenden Anforderungen seitens OMOP in Abschnitt 4.2.1 vorgestellt. In Kapitel 4.2.2 erfolgt die Darstellung und Zusammenfassung der Ergebnisse aus der Analyse der 28 OHDSI Netzwerkstudien, welche im Rahmen der vorangegangenen Literaturrecherche identifiziert wurden.

4.2.1 Anforderungen seitens OMOP Datenmodell

In Kapitel 2.3 wurde bereits in die Grundsätze von OMOP und der Trennung in klinische Fakten und standardisierte Vokabulare eingeführt. Die Medikationsverordnungen sind klinische Fakten und sind daher in einer OMOP Datenbank in der Tabelle *drug_exposure* abzulegen, sofern das zu verwendende Standard-Konzept aus der Tabelle *concept* der Domäne vom Typ „Drug“ zugeordnet ist.

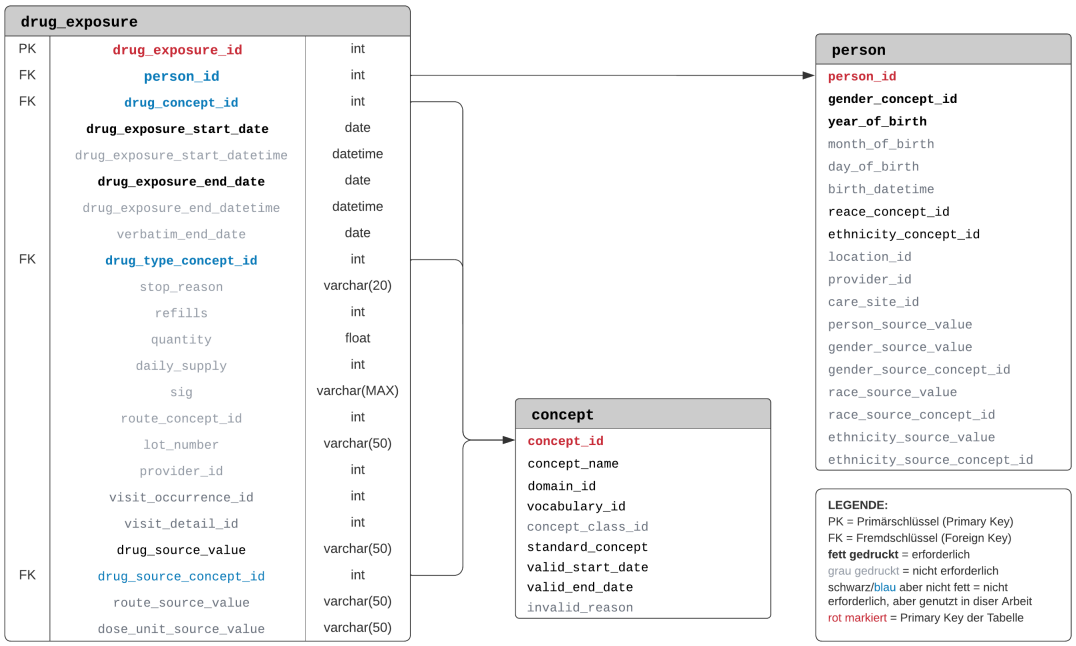


Abbildung 4.7: OMOP drug_exposure Tabelle und Referenzen (Entity Relationship Modell)

Gemäß der Spezifikation des OMOP CDM Version 5.3 (OHDSI und Blacketer, 2021) werden „Drugs“ als aktive Wirkstoffe definiert. Für Medikamente, die mehr als einen Wirkstoff enthalten (sogenannte Kombinationsprodukte) gilt daher, dass aus einer Medikationsverordnung mehr als ein Standardkonzept resultieren kann. Für den Datentransfer (ETL Prozess) muss daher gelten, Kombinationsprodukte gesondert zu beachten und sie, wenn erforderlich, mehr als einmal in der *drug_exposure* Tabelle abzulegen.

Die Spezifikation definiert, dass von den 23 Spalten der Tabelle *drug_exposure* (siehe Tabelle 3.2) lediglich 6 Spalten verpflichtend mit Daten gefüllt werden müssen (*not null* Zwang). Die anderen Spalten sind optional. In Abbildung 4.7 wurde die Tabelle *drug_exposure* anhand der online verfügbaren OMOP (Version 5.3) Spezifikation (OHDSI und Blacketer, 2021) als Entity Relationship Modell mit den Abhängigkeiten zu den Tabellen *concept* und *person* dargestellt

und veranschaulicht damit die syntaktischen Anforderungen an die Medikationsverordnungen in der OMOP Tabelle *drug_exposure*. Die Abhängigkeiten zu anderen OMOP Tabellen sind in der Abbildung limitiert auf die erforderlichen Spalten, die fett gedruckt dargestellt sind. Medikationsverordnungen müssen über die Angabe der entsprechenden *person_id* immer einer Person zugeordnet werden. Die Zuordnung zu einem Behandlungsfall über die *visit_occurrence_id* ist möglich, aber nicht erforderlich. Medikationsverordnungen müssen über die Spalte *drug_concept_id* immer auf ein existierendes Konzept in der Tabelle *concept* unter Verwendung der *concept_id* als Fremdschlüssel verweisen. Gemäß der syntaktischen Mindestanforderungen muss die Information zur verabreichten Dosis pro Tag (Spalte *quantity*) in der *drug_exposure* Tabelle nicht angegeben werden.

Die semantischen Anforderungen an Medikationsverordnungen seitens OMOP sind in dem OHDSI Wiki unter dem Bereich „Domains", „Drug Domain" generell dokumentiert (OHDSI, 2021). Verordnete Medikationen können im Falle von Kombinationsprodukten und auch bei der Kalkulation und Angabe der Dosis sehr komplex werden. Daher gibt es online im OHDSI Forum also auch im Rahmen diverser Veranstaltungen (Studyathons, Symposien) der OHDSI Forschungsgemeinschaft verschiedenste Dokumente und Aufzeichnungen, die zur Klärung von Herausforderungen genutzt werden können. Im Sinne der Mindestanforderungen an die Semantik von Medikationsdaten in OMOP ist das oben bereits erwähnte Wiki (OHDSI, 2021) als Quelle am verlässlichsten und kann als ausreichend angesehen werden. Demnach gilt die in Abschnitt 2.7.2 eingeführte Terminologie RxNorm als das zu verwendende Standard Vokabular für klinische Fakten der Domäne „Drug" in der Tabelle *drug_exposure*. Zur Prüfung der Datenqualität gemäß der Anforderungen seitens des OMOP CDM kann das in Abschnitt 2.6 vorgestellte OHDSI DQD genutzt werden. Es prüft beispielsweise alle Einträge in der Tabelle *drug_exposure* auf die geforderte Verwendung von Standard-Konzepten der Domäne „Drug", die stets zur Terminologie RxNorm gehören (Prüftyp „isStandardValidConcept" in Tabelle 2.2 Abschnitt 2.6). Die Quelldaten der Medikationsverordnungen können optional als Textfeld in der Spalte *drug_source_value* abgelegt werden. Auch hierfür existiert mit der Spalte *drug_source_concept_id* unter Verwendung eines Konzepts der Domäne „Drug", welches nicht zwingend einem Standard-Konzept entsprechen muss. Hier kann also auch auf eine Terminologie, wie beispielsweise ATC als Konzept verwiesen werden. Verfügt eine Medikationsverordnung über kein valides Standard-Konzept in RxNorm oder ist eine eindeutige Zuordnung nicht möglich, so kann gemäß der Dokumentation des OMOP CDM als Konzept in der Spalte *drug_concept_id* auch das sogenannte „No matching concept" mit der *concept_id*

null verwendet werden. Klinische Fakten, die unter Verwendung dieses Konzeptes abgelegt werden, sind für eine Nutzung in Studien jedoch nicht brauchbar.

Die 5 wichtigsten Ergebnisse aus der Anforderungsanalyse gemäß OMOP zur Ablage von Medikationsverordnungen im Hinblick auf die Nutzung für die Forschung unter Wahrung der semantischen Interoperabilität lassen sich wie folgt darstellen:

1. Speicherung erfolgt in der Tabelle *drug_exposure*
2. Zuordnung zu einem gültigen „Drug" Konzept aus der Tabelle *concept* erforderlich
3. Bei Kombinationsprodukten Trennung in mehrere standardisierte Konzepte notwendig
4. RxNorm ist als Standard-Vokabular zu nutzen
5. Konzept „No matching concept" (*concept_id = 0*) als Interimslösung, wenn kein valides Konzept vorhanden
6. Zur Messung der Datenqualität nach Kahn et al. (Kahn et al., 2016) kann das DQD genutzt werden

4.2.2 Anforderungen OHDSI Netzwerkstudien

Die in Kapitel 4.1.6 identifizierte Literatur in Form von 28 Publikationen (Amutha et al., 2021; Boland et al., 2018; Brat et al., 2020; Brauer et al., 2020; Burn, Sena et al., 2020; Burn, You et al., 2020; Chandler, 2020; Chen et al., 2020; Choi et al., 2020; Duke et al., 2017; Hockett et al., 2021; Hripcsak, Ryan et al., 2016; Hripcsak, Suchard et al., 2020; Jensen et al., 2021; H. I. Kim et al., 2021; H. Kim et al., 2020; Y. Kim et al., 2020; Kubota et al., 2018; Lane, Kostka et al., 2020; Lane, Weaver et al., 2021; Morales et al., 2021; Samwald et al., 2016; Seo et al., 2020; Spotnitz et al., 2020; Vashisht et al., 2018; Viernes et al., 2019; You, Rho et al., 2020; Zhang et al., 2018) der Dimension „Nutzung" und der Kategorie „Studie" wurden nach den medizinischen Datengruppen Diagnosen, Medikamente, Laborwerte, Prozeduren, Beobachtungen und medizinische Scores analysiert.

Abbildung 4.8 zeigt, in welchen Studien die medizinischen Datengruppen genutzt wurden. Diagnosen waren in 26 der 28 (93%) OHDSI Netzwerkstudien notwendig. Medikationsdaten wurden in 22 der 28 (79%) Studien verwendet. Die anderen Datengruppen würden in weniger als 30% der Studien verwendet. Am seltensten wurden Beobachtungen und medizinische Scores für die Studien genutzt. In 25 der 26 Studien mit Medikationsdaten, arbeiten auf der Basis von Wirkstoffen oder Wirkstoffgruppen, die entweder als Freitext oder über die

Verwendung der konkreten *concept_id* in den Studienprotokollen definiert werden. Lediglich in einer Studie (Spotnitz et al., 2020) wurde ein RxNorm Konzept des TTY „Branded Drug“ über die Angabe einer *concept_id* konkret spezifiziert. Die komplette Liste von Wirkstoffgruppen oder Wirkstoffen, sofern in den Publikationen enthalten, befinden sich in den extrahierten Metainformationen für alle OHDSI Studien auf Zenodo (Reinecke, 2023a) als Referenz.

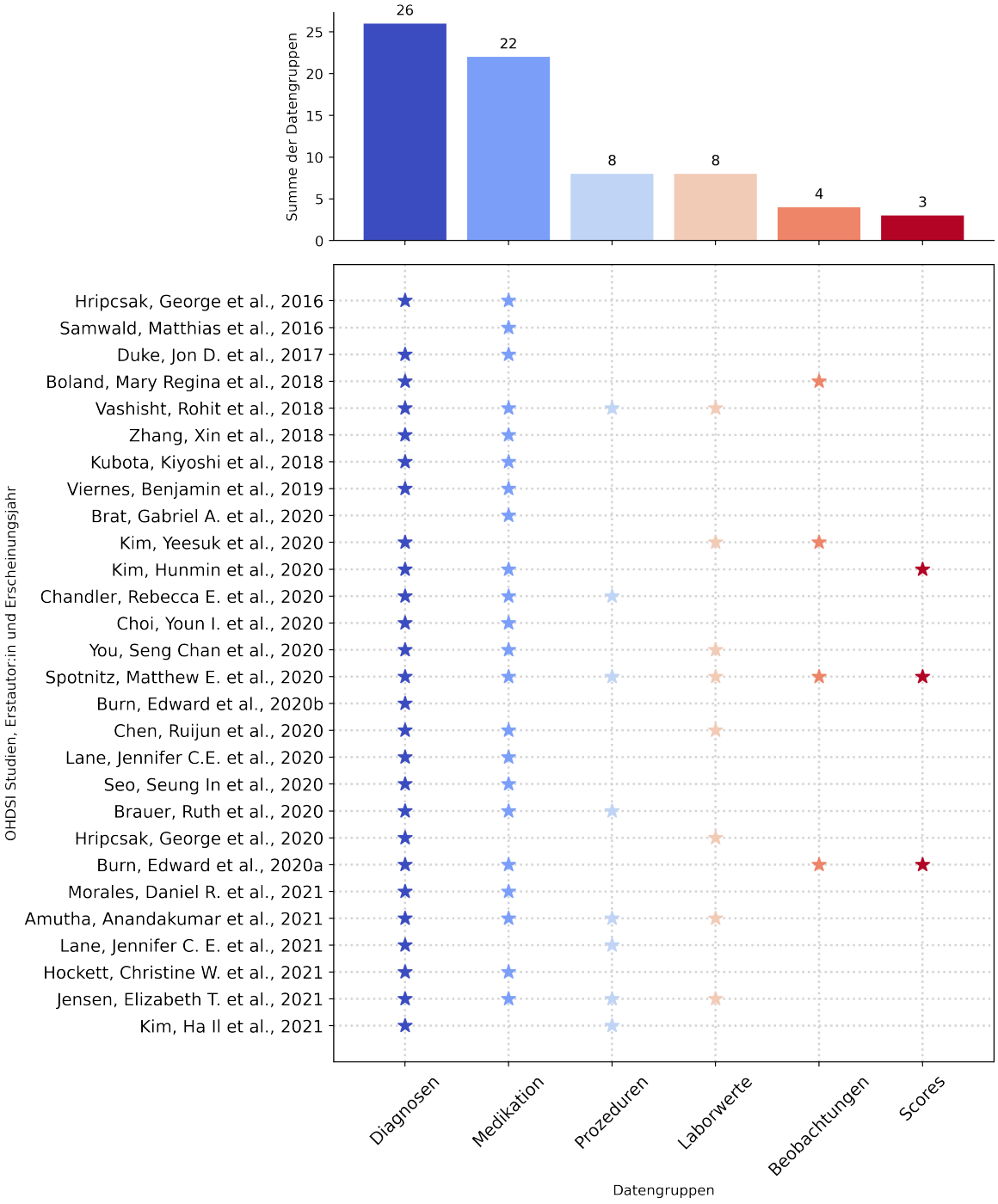


Abbildung 4.8: Verwendung von Daten in OHDSI Netzwerkstudien

Wie aus den Studien hervorgeht, ist die Verfügbarkeit von Medikationsdaten häufig eine Voraussetzung, um an OHDSI Netzwerkstudien teilzunehmen. Dabei ist weniger entscheidend, das konkrete Medikament zu benennen, als vielmehr den konkreten Wirkstoff strukturiert und als valides Konzept zur Verfügung zu stellen.

4.2.3 Zusammenfassung der Ergebnisse der Anforderungsanalyse

Die durchgeführte Anforderungsanalyse zeigt, dass Medikationsverordnungen neben Diagnosen zur Durchführung von Studien innerhalb der Forschungsgemeinschaft OHDSI sehr relevant sind und in der Mehrheit der Studien (22 von 28) verwendet wurden. Grundsätzlich sind alle Medikationsverordnungen in OMOP einem gültigen Standard-Konzept der Terminologie RxNorm zuzuordnen. In den meisten der betrachteten Studien, die Medikationsdaten verwenden, erfolgte die Beschreibung der Medikationsdaten nicht als Produkt, sondern als Wirkstoff oder Wirkstoffgruppe, entweder als Freitext oder durch die Angabe der konkreten *concept_id*.

4.3 Identifizierte Inhibitoren

Aus Kapitel 4.2.3 sind bereits die Anforderungen an Medikationsdaten zur Nutzung im Kontext von OHDSI Netzwerkstudien bekannt. In diesem Kapitel sollen dem gegenüber existierende Abweichungen in Form von möglichen Inhibitoren vorgestellt werden. Dazu werden in Abschnitt 4.3.1 zunächst die Ergebnisse der an den MIRACUM Standorten, deutschlandweiten Stichprobenanalyse hinsichtlich der Verfügbarkeit und Strukturiertheit von 6 Datengruppen in klinischen Versorgungssystemen vorgestellt. Im Anschluss daran werden die Ergebnisse der systematischen Analyse der Strukturiertheit von Medikationsverordnungen des UKD vorgestellt, um daraus abschließend existierende Lücken zur geforderten Struktur und Terminologie der Daten benennen zu können.

4.3.1 Ergebnisse der Stichprobenanalyse

Die Ergebnisse in diesem Abschnitt wurden auch in einer gemeinsamen Publikation mit dem Titel „Availability of Structured Data Elements in Electronic Health Records for Supporting Patient Recruitment in Clinical Trials“ veröffentlicht, in der ich als Autorin beteiligt war (Vass et al., 2022).

Die Verfügbarkeit und Strukturiertheit der sechs Datengruppen in den klinischen Versorgungssystemen an den zehn beteiligten MIRACUM Standorten für die 45 zufällig ausgewählten Behandlungsfälle pro Standort ist in Abbildung 4.9 dargestellt. Zur besseren Übersichtlichkeit wurden die Datengruppen gemäß ihrer Relevanz in den klinischen Studien angeordnet. Die Datenstruktur wird in die beiden Klassen „strukturiert" und „unstrukturiert" unterteilt und durch eine Farbkodierung kenntlich gemacht (weniger als 50 % strukturiert: hellgrau; 50 % oder mehr strukturiert: dunkelgrau).

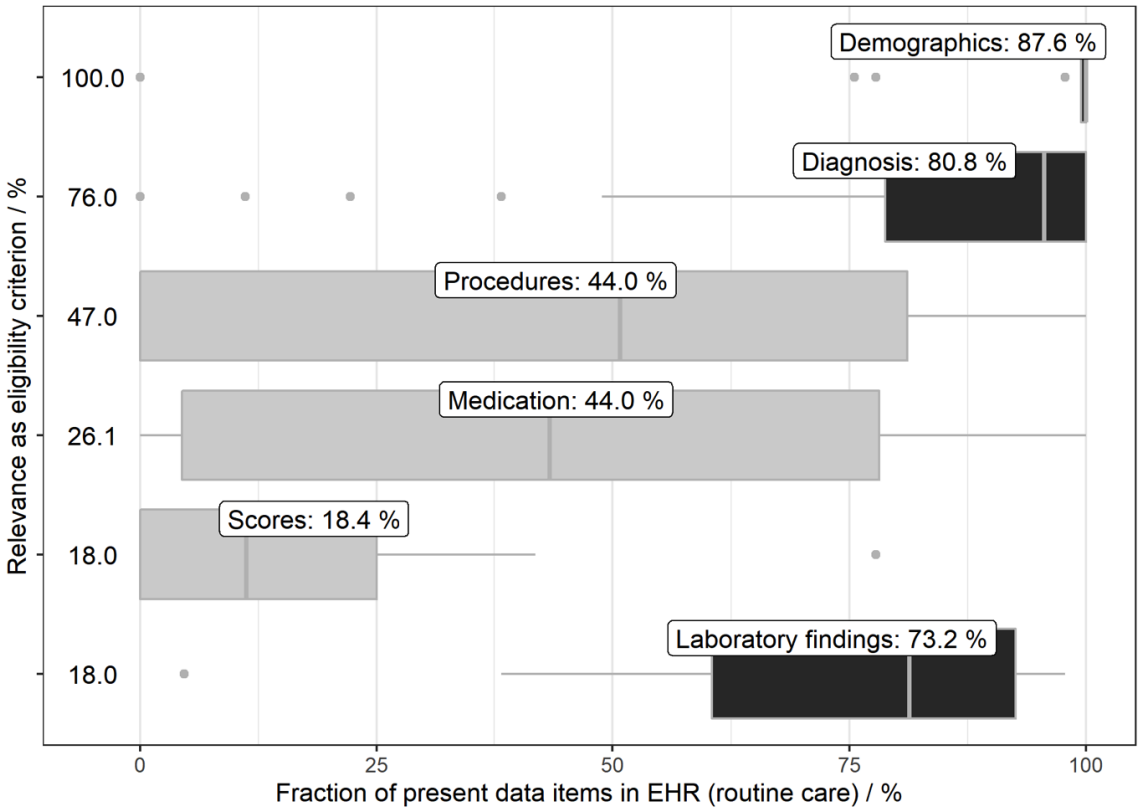


Abbildung 4.9: Strukturiertheit relevanter Datengruppen im Überblick (Vass et al., 2022)

Die Datengruppe „Demografie" ist im Durchschnitt in 87,6% der analysierten Stichproben in den klinischen Informationssystemen vorhanden. Bis auf wenige Ausnahmen ist diese Datengruppe vollständig und strukturiert verfügbar.

Auch die Datengruppe „Diagnose" hat einen sehr hohen Durchschnittswert von 80,8% Vollständigkeit im arithmetischen Mittel und einen sehr hohen Grad der Strukturiertheit mit 97% für die Primärdiagnosen und 91% für die Sekundärdiagnosen. Eine ebenfalls sehr strukturierte Datengruppe sind die „Laborwerte" mit 73,2%.

Die Datengruppe „Prozeduren" hat eine recht breit gefächerte Verteilung hinsichtlich der

Vollständigkeit in den klinischen Systemen zwischen den Standorten, im Durchschnitt liegt die Vollständigkeit bei 44% über alle Behandlungsfälle der Stichprobe. Insgesamt liegen die Prozeduren in den klinischen Systemen eher unstrukturiert vor, da in den meisten Fällen die Anamnese von vergangenen Prozeduren und Operationen überwiegend unstrukturiert und ohne Kodierung erfolgt. Die während eines Krankenhausaufenthaltes durchgeführten Prozeduren werden wiederum in allen Standorten strukturiert erfasst.

Die Datengruppe „Medikation“ beinhaltet Informationen zur Strukturiertheit von Medikationsverordnungen, der Dosis, und dem Enddatum von Medikationsverordnungen. Die Medikationen innerhalb der Stichprobe sind ähnlich heterogen dokumentiert wie die Prozeduren. Die Verfügbarkeit und Strukturiertheit variiert stark zwischen unterschiedlichen Standorten. Die Strukturiertheit der Medikationsdaten liegt ebenfalls bei weniger als 50%, wie durch die hellgraue Visualisierung angezeigt wird.

Die klinischen Scores werden im Durchschnitt in nur 18,4% der Fälle dokumentiert, sie liegen generell in allen Standorten eher selten vor und werden überwiegend unstrukturiert erfasst. Die „Laborwerte“ liegen im Durchschnitt in 73,2% der Stichprobe vor und sind überwiegend strukturiert verfügbar.

4.3.2 Ergebnisse der systematischen Analyse

Die Ergebnisse der durchgeführten Analyse wurden von der Autorin der vorliegenden Arbeit im Rahmen einer Journal-Publikation veröffentlicht (Reinecke, Siebel et al., 2023).

Die durchgeführte Analyse (wie in Kapitel 3.4.2 methodisch beschrieben) zeigte, dass 47,73% Medikationsverordnungen (843.980/1.768.153) im Datensatz DS-Med strukturiert vorliegen. Der Anteil an unstrukturierten Medikationsverordnungen überwiegt mit 52,27% (924.173/1.768.153) geringfügig den Anteil der strukturiert vorliegenden Daten. Wie bereits in Kapitel 3.4.2 in Tabelle 3.6 dargestellt, konnten innerhalb des unstrukturierten Anteils der Medikationsverordnungen 9,1 % aller Medikationsverordnungen (160.896/1.768.153) als Einträge identifiziert werden, bei denen es sich nicht um Medikationsverordnungen, sondern um andere Anordnungen, beispielsweise Blutentnahmen oder Laboranordnungen handelt. Damit reduziert sich der Anteil der unstrukturierten und zu untersuchenden Medikationsverordnungen auf 43,17% (763.277/1.768.153).

Die Gruppierung der unstrukturierten Medikationsverordnungen auf der Grundlage des Datenelements *MEDICATION* des Datensatzes DS-Med führte zu insgesamt 100.004 eindeutigen Freitexten, die als Medikationsverordnung eingegeben und nach Addition der Häufigkeit für jeden Freitext als Datensatz DS-Gruppiert gespeichert wurden. Die Häufigkeit der 100.004 Freitexte wurde in der Spalte *FREQUENCY* des Datensatzes DS-Gruppiert dargestellt.

4.3.3 Zusammenfassung der identifizierten Inhibitoren

Sowohl die stichprobenartige Analyse verteilt über mehrere Standorte, als auch die vollständige und systematische Analyse der Medikationsverordnungen des UKD zeigen eine Strukturiertheit von weniger als 50% in den untersuchten Medikationsdaten. Die ganzheitliche Analyse zeigt zudem auch, dass im Falle von strukturiert vorliegender Medikationsverordnungen, diese nur kodiert durch die Terminologie ATC vorliegen.

Aus den ermittelten Anforderungen in Kapitel 4.2.3 ist die Wichtigkeit der Medikationsdaten für die Forschung im Kontext der OHDSI Forschungsgemeinschaft sowie der Notwendigkeit der Bereitstellung als RxNorm Codes bereits bekannt. Die mangelnde Struktur der Medikationsdaten, als auch die abweichend verwendete Terminologie ATC stellt damit eine Lücke dar, die es in der Arbeit im Folgenden unter Durchführung geeigneter Maßnahmen zu beheben gilt.

4.4 Ergebnisse der Reduktionsmaßnahmen

In diesem Kapitel werden die Ergebnisse der durchgeführten Maßnahmen vorgestellt, die dazu dienen, die in Kapitel 4.3 identifizierten Inhibitoren zu reduzieren, um die in Kapitel 4.2.3 formulierten Anforderungen zu erfüllen. Zunächst werden in diesem Kapitel in Abschnitt 4.4.1 die Ergebnisse der Teilnahme des UKD an einer retrospektiven Beobachtungsstudie der EMA auf Basis des OMOP CDM vorgestellt. Im Anschluss daran, werden die Ergebnisse der generellen Maßnahmen zur Verbesserung der Datenstruktur in Abschnitt 4.4.2 und zur Anpassung an die erforderliche Terminologie RxNorm in Abschnitt 4.4.3 vorgestellt, um die zuvor identifizierten Lücken zwischen den Anforderungen und den existierenden Inhibitoren zu schließen.

4.4.1 Ergebnisse der Maßnahmen am Beispiel einer EMA Studie

In den Methoden dieser Arbeit wurden in Abschnitt 3.5.1 bereits die Details zu der EMA Studie vorgestellt, an der das DIZ des UKD im Rahmen eines europäischen Pilotprojektes teilgenommen hat. Aus dem Studienprotokoll (Version 1.0) konnten alle wichtigen Informationen zur Studie und die für die Studie relevanten Wirkstoffe, die zur Gruppe der Corticosteroide gehören und während einer stationären Behandlung einer COVID-19 Erkrankung verordnet wurden, entnommen werden.
Für diese Studie sind die folgenden vier Wirkstoffe relevant:

- Dexamethason
- Prednison und Prednisolon
- Methylprednisolon
- Hydrocortison

Die Wirkstoffe werden im Studienprotokoll unter Verwendung von ATC WHO Codes Level 5 angegeben. Da die ATC Codes initial nicht in den Daten im DIZ des UKD verfügbar waren, musste eine Identifikation der am UKD verwendeten Medikamente und ein Abgleich mit denen im Studienprotokoll angefragten Wirkstoffen durchgeführt werden. Dazu haben die involvierten Apotheker:innen des UKD anhand der relevanten Wirkstoffe aus dem Studienprotokoll eine Liste mit Produktbezeichnungen der im Jahr 2020 und 2021 durch das UKD eingekauften und verwendeten Medikamente erstellt und eine Liste bereitgestellt, die zehn Produkte für den Wirkstoff *Dexamethason*, acht Produkte für den Wirkstoff *Methylprednisolon* und vier Produkte für den Wirkstoff Hydrocortison enthält. Der angefragte ATC Code A07EA01 für *Prednison* und *Prednisolon* wurde im gesuchten Zeitraum am UKD als Medikament nicht verordnet.

In Vorbereitung auf die Datenauswertung und Überführung der Medikationsdaten nach OMOP wurde durch ein Expert:innenteam festgestellt, dass zum Zeitpunkt der Studienanfrage im Frühjahr 2020 im DIZ des UKD die Medikationsverordnungen ausschließlich als Freitext verfügbar waren. Um die notwendige Strukturiertheit der Medikationsdaten für die Teilnahme an der Studie zu gewährleisten, erfolgte die Einbindung der entsprechenden IT Verantwortlichen des KIS ORBIS. Neue Datenelemente konnten so durch das DIZ im Datensatz DS-Med für die Studie zusätzlich bereitgestellt werden. Die Medikationsdaten des DIZ wurden um die Datenelemente *STRUCTURE* und *ATC-Code* erweitert. Der ATC Code war verfügbar, wenn das Medikament im KIS durch das verordnete Personal aus einer vorde-

finierten Liste (entspricht dem Datensatz DS-Katalog des UKD) ausgewählt wurde. Für die Kohorte der stationär behandelten COVID-19 Patient:innen konnten manuell alle nicht aus der Liste ausgewählten Medikationsverordnungen korrekt zu den gesuchten ATC Codes zugeordnet werden.

Nach der erstmaligen Ausführung des durch die Studienkoordination bereitgestellten Analyseskriptes in R, wurde eine leere Ergebnismenge für die gesuchte Kohorte mit den gesuchten Wirkstoffen generiert. Nach manueller Prüfung der Daten durch die Expert:innen der Apotheke des UKD konnte festgestellt werden, dass hier eine andere Ursache für die leere Ergebnismenge vorliegen musste.

Tabelle 4.2: Wirkstoffe der EMA Studie mit entsprechenden Codes der Terminologien ATC und RxNorm

Wirkstoff	ATC WHO Code	ATC concept_id	RxNorm Code	RxNorm concept_id
Dexamethason	H02AB02	21602730	3264	1518254
Prednison und Prednisolon	A07EA01	Nicht angepasst, weil keine Medikationsverordnungen in den Daten		
Methylprednisolon	H02AB04	21602732	6902	1506270
Hydrocortison	H02AB09	21602737	5492	975125

Durch einen regelmäßigen Austausch mit der zentralen Studienkoordination konnte die Ursache des Problems schnell gefunden werden. Die in OMOP überführten Medikationsverordnungen unter Verwendung des ATC WHO Codes entsprachen nicht den erwarteten Konzepten in RxNorm. Um diese Lücke zu schließen, wurden unter manueller Ausführung der in 4.1 abgebildeten SQL Statements die *concept_ids* für die 4 betroffenen Wirkstoffe in OMOP angepasst. Das abgebildete Statement ist exemplarisch für den Wirkstoff *Dexamethason*. Die entsprechenden *concept_id* Informationen für die Wirkstoffe *Methylprednisolon* und *Hydrocortison* befinden sich für RxNorm in der Tabelle 4.2. Für die Wirkstoffe *Prednison* und *Prednisolon* wurde die Überführung der Terminologie von ATC nach RxNorm nicht durchgeführt, da keine Medikationsverordnungen zu diesem Wirkstoff in den RWD für die Studienkohorte existierten.

Quelltext 4.1: SQL Statement für manuelle Anpassung der Konzepte in OMOP am Beispiel Dexamethason

```
update drug_exposure
set drug_concept_id = '1518254'
WHERE visit_occurrence_id IN
  (SELECT DISTINCT (visit_occurrence_id)
   FROM condition_occurrence
   WHERE condition_source_value LIKE 'U07.1%')
AND drug_source_value IN ('H02AB02');
```

Bezogen auf das erläuterte Vorgehen in Kapitel 3.5, Abbildung 3.4 lassen sich die im Rahmen der Teilnahme an der EMA Studie durchgeführten manuellen Maßnahmen auf die Schwerpunkte Datenstruktur und Terminologie ableiten. Die in Kapitel 4.3 vorgestellten Ergebnisse der systematischen Erfassung der Strukturiertheit der Medikationsverordnungen am UKD zeigen einen großen Bedarf an Verbesserung der Strukturiertheit der Medikationsverordnungen generell. Diese Verbesserung kann für die Gesamtheit der Medikationsverordnungen ohne Einschränkung auf eine kleine Kohorte nicht manuell erfolgen. Vielmehr benötigt es hier ein automatisiertes Vorgehen, unabhängig von den gesuchten Wirkstoffen für eine einzelne Studie. Auch die automatisierte Überführung der Terminologie von ATC nach RxNorm stellt einen offenen Punkt dar, der durch eine geeignete Maßnahme zu schließen ist.

Im Fall der EMA Studie konnte das semantisch korrekte Mapping für die gesuchten Wirkstoffe von ATC nach RxNorm durch die Studienkoordination bereitgestellt werden. Daher lassen sich aus der Teilnahme an der EMA Studie die generellen Maßnahmen zur systematischen Verbesserung der Strukturiertheit und der Überführung in die notwendige Terminologie RxNorm ableiten.

4.4.2 Ergebnisse der Maßnahmen - Datenstruktur

In diesem Kapitel werden die Ergebnisse der Maßnahmen zur Verbesserung der Datenstruktur der Medikationsdaten vorgestellt, mit dem Ziel, die in Abschnitt 4.3.2 als überwiegend unstrukturiert vorliegenden Medikationsverordnungen aus Datensatzes DS-Med zu verbessern und korrekte ATC Codes zuzuordnen. Dazu werden zunächst die Ergebnisse der drei implementierten Algorithmen in Abschnitt 4.4.2.1 vorgestellt. Die Zuverlässigkeit im Sinn der Korrektheit der Algorithmen auch mit Implikationen auf die Korrektheit im Kontext des Maßes der Übereinstimmungsrate der drei Algorithmen wird in Abschnitt 4.4.2.2 dargestellt. Abschließend wird in Abschnitt 4.4.2.3 die Verbesserungsrate der Medikationsdaten nach Anwendung auf die unstrukturierten Medikationsverordnungen dargestellt. Die Ergebnisse der durchgeführten Maßnahmen zur Verbesserung der Datenstruktur wurden von der Autorin der vorliegenden Arbeit im Rahmen einer Journal-Publikation veröffentlicht (Reinecke, Siebel et al., 2023).

4.4.2.1 Ergebnisse der Algorithmen

Die Leistung der drei Algorithmen ist quantitativ als auch qualitativ sehr verschieden. Eine absolute Aussage zur Qualität der einzelnen Algorithmen ist nicht möglich, da neben der Korrektheit der Ergebnisse auch die Anzahl der Ergebnisse einen Indikator für die Performance der Algorithmen darstellen kann. Algorithmus 3 (Abgleich von Ähnlichkeiten) liefert aufgrund seiner Implementierung mit der FuzzyWuzzy Bibliothek (siehe 3.3) für alle unstrukturierten Medikationsverordnungen einen ATC-GM Code.

Abbildung 4.10: Übereinstimmungsraten der Algorithmen, Darstellung auf unterschiedlichen Datensätzen

Algorithmus 1 (basierend auf dem Abgleich von Inhaltsstoffen) konnte ATC-GM Codes für 8.048 eindeutige Freitexte identifizieren. Multipliziert mit der Häufigkeit des Vorkommens der Texte im Datenelement *MEDICATION* ergibt sich eine Gesamtzahl von 244.718 (32,06%) Medikationsverordnungen der insgesamt 763.277 unstrukturierten Medikationsverordnungen. Die quantitative Leistung von Algorithmus 2 (basierend auf dem Produktnamen des Medikament) ist geringer als die von Algorithmus 1, da er ATC Codes für lediglich 6.744 verschiedene Freitexte identifizierte. Dies entspricht insgesamt 126.100 (16,52%) Medikationsverordnungen der insgesamt 763.277 unstrukturierten Medikationsverordnungen. An dieser Stelle kann noch keine Aussage über die Korrektheit der Ergebnisse getroffen werden.

Die Übereinstimmungsraten der Ergebnisse aller Algorithmen ist in Abbildung 4.10 für den Datensatz DS-Top1000 (Abbildung 4.10, *a1* und *a2*) und für den Datensatz DS-Gruppiert (Abbildung 4.10, *b1* und *b2*) dargestellt. Wobei sich die Abbildung 4.10 jeweils auf die gruppierten, eindeutigen Freitexte bezieht (Abbildung 4.10, *a1* und *b1*) und jeweils mit der

Häufigkeit multipliziert wurde (Abbildung 4.10, *a2* und *b2*). Die Ergebnisse aller drei Algorithmen überlappen für 45 Freitexte bezogen auf den Datensatz DS-Top1000 (Abbildung 4.10, *a1*), multipliziert mit den Häufigkeiten sind das 57.264 Medikationsverordnungen (Abbildung 4.10, *a2*). Entsprechend angewandt auf den gesamten Datensatz der unstrukturierten Medikationsverordnungen (Abbildung 4.10, *b2*) sind das 60.906 Medikationsverordnungen (7,98%, n=763.277) bei denen alle drei Algorithmen in den Ergebnissen übereinstimmen.

Die quantitative Leistung der drei Algorithmen zur Zuordnung von ATC Codes basierend auf eindeutigen Freitexten unterscheidet sich sehr. Algorithmus 3 ist aufgrund seiner Implementierung in der Lage, für alle unstrukturierten Medikationsverordnungen einen ATC-GM Code zu liefern. Algorithmus 1 und 2 haben eine weitaus geringere quantitative Leistung gezeigt.

Im nächsten Abschnitt werden die Ergebnisse der Validierung der Algorithmen dargestellt, um auch eine qualitative Bewertung der Algorithmen vornehmen zu können und in Kombination mit der quantitativen Leistung, mögliche Rückschlüsse auf die Korrektheit der identifizierten ATC-GM Code anhand der Übereinstimmungsraten oder weiterer Merkmale vorzunehmen.

4.4.2.2 Ergebnisse der Validierung der Algorithmen

Die Validierung für die häufigsten 1000 Freitexteinträge entspricht 66,56% (615.129/924.173) aller unstrukturierten Medikationsverordnungen im Datensatz DS-Med (siehe Abbildung 4.11). Durch die Validierung einer vergleichsweise geringen Menge an Freitexten kann die Strukturiertheit des Datensatzes DS-Med bereits deutlich verbessert werden. Dies ist notwendig, um den bereits genannten Anforderungen an die Datenstruktur zur Ablage in OMOP gerecht zu werden.

In Abbildung 4.11 wird mit der blauen Linie die kumulative Verteilungskurve dargestellt, startend mit den am häufigsten verwendeten Freitexten auf der x-Achse und dem Anteil an der Gesamtheit der unstrukturierten Freitexte in Datensatz DS-Med auf der y-Achse. Die rote (gestrichelte) senkrechte Linie in Abbildung 4.11 markiert auf der x-Achse die ersten 1000 Freitexte. Die rote (gestrichelte) waagerechte Linie ist die Schnittlinie zwischen der kumulativen Verteilungskurve und der senkrechten roten Linie und liegt auf der y-Achse bei bereits 66,65% der unstrukturierten Medikationsverordnungen.

Zusammen mit dem Anteil der strukturierten Medikationsverordnungen (843.980/1.768.153) und dem Anteil der Einträge ohne Medikation (166.307/1.768.153), die bei der systematischen Analyse der Medikationsverordnungen, wie in Kapitel 4.3.2 dargestellt, identifiziert wurden, kann der Strukturierungsgrad auf 85,18% (1.506.059/1.768.153) aller Medikationsverordnungen aus DS-Med erhöht werden.

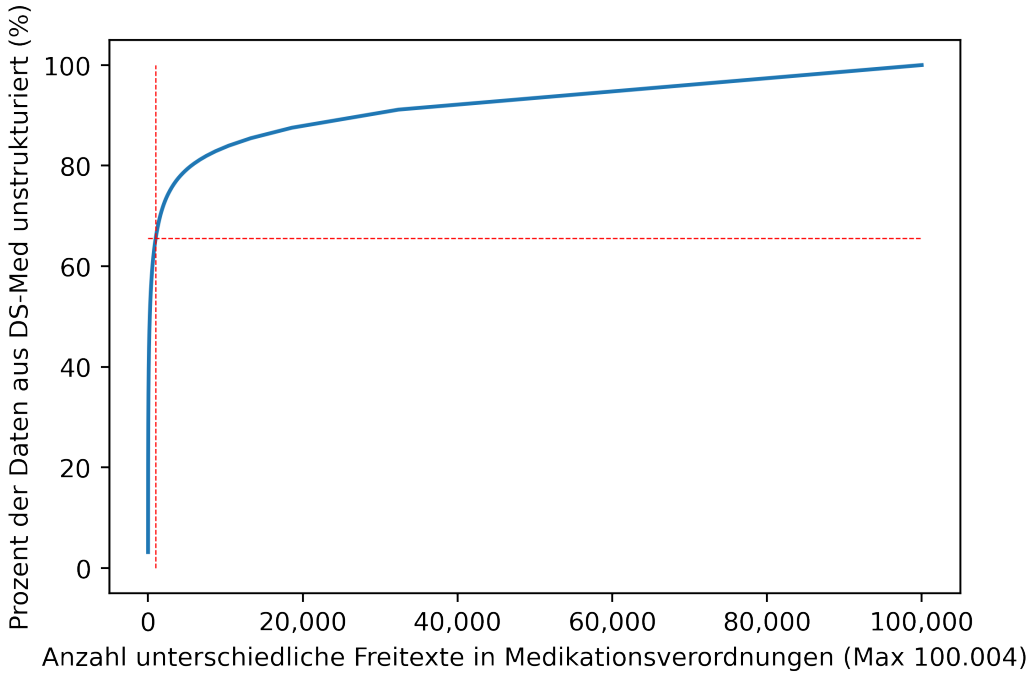


Abbildung 4.11: Kumulative Verteilungskurve der Freitexte der Medikationsverordnungen

Die Fehlerquoten der Algorithmen sind in Tabelle 4.3 dargestellt. Die Darstellung erfolgt einzeln für die Algorithmen und bei Übereinstimmung der Ergebnisse der drei Algorithmen. Die Validierung zeigt, dass Algorithmus 1 in Bezug auf seine Fehlerquote die qualitativ besten Ergebnisse erzielt. Dieser hat von den insgesamt 287 identifizierten ATC Codes für die ersten 1.000 Freitexte aus Datensatz DS-Top1000, 286 richtige Ergebnisse geliefert und lediglich ein falsches Ergebnis. Allerdings hat Algorithmus 1 für 713 (71,3%, 713/1000) Freitexte keine ATC Codes identifizieren können. Im Vergleich dazu identifizierte der Algorithmus 2 nur 148 ATC Codes, von denen 142 korrekt und sechs falsch waren. Dieser hat prozentual weniger Ergebnisse geliefert als Algorithmus 1, weil er für 852 (85,2%, 825/1000) Freitexte aus Datensatz DS-Top1000 keine ATC Codes zuordnen konnte. Algorithmus 3 lieferte für 1000 Freitexte insgesamt 765 richtige und 235 falsche Ergebnisse geliefert und weist damit die höchste Fehlerquote von 23,5% (235/1000) auf.

Die Korrektheit der Ergebnisse wurde auch im Kontext der Übereinstimmungsraten der Ergebnisse zwischen den Algorithmen für Datensatz DS-Top1000 bestimmt (siehe Abbildung 4.10). Die manuelle Validierung des Datensatzes DS-Top1000 zeigte, dass die Ergebnisse immer korrekt waren, wenn alle drei Algorithmen denselben ATC-GM Code identifiziert haben. Gleiches gilt auch bei der Übereinstimmung der Ergebnisse von Algorithmus 1 und 2. Bei den Ergebnissen, die ausschließlich für Algorithmus 2 und 3 übereinstimmen, existiert eine geringe Fehlerzahl von fünf falsch identifizierten ATC-GM Codes der insgesamt 286 übereinstimmenden Ergebnisse (1,74% Fehlerquote, 5/286). Dabei lassen sich vier der fünf falschen Ergebnisse auf Medikationsverordnungen von Kochsalzlösungen zurückführen. Ein weiteres falsches Ergebnis betrifft den Inhaltsstoff *Aciclovir*, der innerhalb der Medikationsverordnung ohne eine weitere Angabe von Details keinen korrekten Rückschluss auf den ATC-GM Code zulässt, da die Form der Darreichungsform (z. B. oral, parenteral und konjunktival) einen unterschiedlichen ATC-GM Code als Resultat hat.

Die Validierung der Ergebnisse bei einer Übereinstimmung der Algorithmen 1 und 3 wurde lediglich ein falsches Ergebnis in Bezug auf den Inhaltsstoff *Telmisartan* ermittelt. Bei dieser Verordnung handelt es sich um ein Kombinationsprodukt mit zwei Wirkstoffen *Telmisartan* und *Diuretika*, beide Algorithmen identifizierten den ATC Code des mono-therapeutischen Wirkstoffs *Telmisartan* stattdessen.

Tabelle 4.3: Quantitative Leistung und Fehlerquoten der Algorithmen

Übereinstimmung der Ergebnisse	Ingesamt von 1000	davon korrekt	davon falsch	Fehlerquote
Algorithmus 1 und 2	45	45	0	0%
Algorithmus 1 und 3	145	144	1	0,69%
Algorithmus 2 und 3	287	282	5	1,74%
Algorithmus 1, 2 und 3	45	45	0	0%
nur Algorithmus 1	287	286	1	0,35%
nur Algorithmus 2	148	143	5	3,37%
nur Algorithmus 3	1000	765	235	23,50%

Die Ergebnisse der deskriptiven Statistik der Mittelwerte des Levenshtein Scores der korrekten und falschen Ergebnisse von Algorithmus 3 sind in Tabelle 4.4 dargestellt. Für den Datensatz DS-Top1000 wurde ein signifikanter Unterschied in den Mittelwerten des Levenshtein Score zwischen den korrekten und den falschen Ergebnissen mit einem P-Wert von $2,4 \times 10^{-47}$, der deutlich unter dem Signifikanzniveau α=0,05 liegt, festgestellt.

Das bedeutet, je höher der Levenshtein Score ist, desto höher ist die Wahrscheinlichkeit, dass das Ergebnis korrekt ist. In absoluten Zahlen ausgedrückt, kann bei Einträgen mit einem Levenshtein Score größer als 84,28 davon ausgegangen werden, dass die Ergebnisse korrekt sind.

Die Untersuchung der falschen Ergebnisse des Algorithmus 3 im Falle eines Levenshtein Score mit dem Wert 80 oder größer, ergab 37 Medikationsverordnungen (dargestellt in Tabelle 4.5), die im Detail von einem Team aus Medizininformatiker:innen und Apotheker:innen begutachtet wurden.

Tabelle 4.4: Deskriptive Statistik des Levenshtein Score von Algorithmus 3

	Deskriptive Statistik	Algorithmus 3, korrekt	Algorithmus 3, falsch
Levenshtein Score	Stichprobengröße	766	234
	Häufigkeit	416.585	84.598
	Arithmetischer Mittelwert (mean)	84,28	67,18
	Standardabweichung (sd)	14,86	15,52
	Kleinster Wert (min)	21	29
	1. Quartil (25. Perzentil)	76	55
	Median (50. Perzentil)	87	63
	3. Quartil (75. Perzentil)	96	75
	Größter Wert (max)	100	100

Das Team identifizierte vier Gründe für die fehlerhaften Ergebnisse des Algorithmus 3: (1) Darreichungsform, (2) Kombinationsprodukte, (3) Ähnlichkeit von Wörtern und (4) zu kurzer Freitext. Am häufigsten lag der fehlerhafte ATC Code in 19 Fällen an der falsch identifizierten Darreichungsform. Als Beispiel sei hier *Prednisolon* genannt, welches als systemisches Präparat (Tabletten, Injektion, Infusion) den ATC Code H02AB06 hat, jedoch auch also topisches Präparat zur Anwendung auf der Haut (ATC Code D07AA03), als Kombinationsprodukt zur Anwendung auf der Haut (ATC Code D07XB02), als nasales Präparat zur Anwendung in der Nase (ATC Code R01AD02) oder als ophthalmisches Präparat zu Anwendungen am Auge (ATC Code S01BA04) verfügbar ist.

Expert:innen der Apotheke des UKD konnten keine ATC Codes für Medikationsverordnungen mit den Freitexten „Aciclovir“, „ASS“, „Magnesium“ und „Vancomycin“ zuordnen. Dies liegt daran, dass es für diese Wirkstoffe je nach Anwendungsgebiet unterschiedliche ATC Codes gibt. Daher bleiben diese Verordnungen ohne spezifischen ATC Code und werden als unspezifisch betrachtet.

Tabelle 4.5: Falsche Ergebnisse von Algorithmus 3, bei Levenshtein Score über 80

Datenelement MEDICATION	Ergebnis Algorithmus 3	Levenshtein Score	korrekter ATC Code	Grund
ASS RATIOPHARM 100 mg TAH Tabletten | (Acetylsalicylsäure)	N02BA01	89	B01AC06	Ähnlichkeiten der Wörter
Prednisolon	S01CA53	100	H02AB06	Darreichungsform
MAGNESIUM VERLA 300 Orange Granulat | (Magnesium-Ion)	A12CC05	100	V06XX02	Ähnlichkeiten der Wörter
ARILIN 500 Filmtabletten |(Metronidazol)	G01AF01	100	P01AB01	Ähnlichkeiten der Wörter
CANDESARTAN HEXAL comp 16mg/12,5 mg Tabletten | (Candesartan)	C09CA06	89	C09DA26	Kombinationsprodukt
Heparin	C05BA03	100	B01AB01	Darreichungsform
PREDNISOLON	S01CA53	100	H02AB06	Darreichungsform
FENISTIL Injektionslösung |(Dimetinden)	D04AA13	100	R06AB03	Darreichungsform
ACIC 250 PI Via Pulver z.Herst.e.Infusionslösg. |(Aciclovir)	D06BB03	100	J05AB01	Darreichungsform
NaCl 0,9%	B05CB01	100	B05BB11	Darreichungsform
VALSARTAN HEXALcomp.160mg/12,5mg Filmtabletten | (Valsartan)	C09CA03	100	C09DA23	Kombinationsprodukt
Prednisolon mg	S01CA53	88	H02AB06	Darreichungsform
NACL 0,9%	B05CB01	100	B05BB11	Darreichungsform
ACIC 200 Tabletten | (Aciclovir)	D06BB03	100	J05AB01	Darreichungsform
ACIC 500 PI Via Pulver z.Herst.e.Infusionslösg. |(Aciclovir)	D06BB03	100	J05AB01	Darreichungsform
Simvastatin	C10BA02	100	C10AA01	Kombinationsprodukt
CANDESARTAN HEXAL comp 8mg/12,5 mg Tabletten | (Candesartan)	C09CA06	89	C09DA26	Kombinationsprodukt
NACL 0,9% | (Natrium-Ion,Chlorid)	B05CB01	100	B05BB11	Darreichungsform
C) FENISTIL 1 Ampulle als Bolus | (Dimetinden)	D04AA13	100	R06AB03	Darreichungsform
HCT	C09DX01	100	C03AA03	Zu kurz
Allopurinol	M04AA51	100	M04AA01	Kombinationsprodukt
Prednisolon 5 mg	S01CA53	81	H02AB06	Darreichungsform
HYDROCORTISON 10 mg Jenapharm Tabletten | (Hydrocortison)	S01BA02	81	H02AB09	Darreichungsform
Simvastatin 20mg	C10BA02	100	C10AA01	Kombinationsprodukt
NaCl 0.9%	B05CB01	100	B05BB11	Darreichungsform
NaCl	B05CB01	100	B05BB11	Darreichungsform
RANITIC Injekt Infusionslösungskonzentrat | (Ranitidin)	J01DH03	83	A02BA02	Ähnlichkeiten der Wörter
Fenistil	R06AB03	100	unspec	Darreichungsform
Methotrexat	L04AX03	100	M01CX01	Zu kurz
Aciclovir	J05AB01	100	unspec	Darreichungsform
ASS	B01AC06	100	unspec	Zu kurz
SURVIMED OPD Easy Bag | (Aminosäuren, essentiell, Isoleucin, Leucin, Lysin, Methionin)	V06DB50	83	nomed	Ähnlichkeiten der Wörter
INFECTOCILLIN parenteral 5 Mega Durchstechflaschen | (Benzylpenicillin, Natrium-Ion)	J01CE02	87	J01CE01	Ähnlichkeiten der Wörter
Aciclovir AS	J05AB01	86	S01AD03	Darreichungsform
LOSARTAN HEXAL 50mg Filmtabletten | (Losartan)	C09DA21	100	C09CA01	Kombinationsprodukt
Magnesium	A12CC05	100	unspec	Zu kurz
Vancomycin	A07AA09	100	unspec	Zu kurz

4.4.2.3 Ergebnisse der Verbesserung der Strukturiertheit

Die Ergebnisse der Analyse der Medikationsverordnungen aus Kapitel 4.3.2 sind in ihrer Aussagefähigkeit begrenzt, weil sie lediglich die Anteile der strukturierten und unstrukturierten Medikationsverordnungen unterscheiden.

Dank der durchgeführten Erhöhung des Grades der Strukturiertheit der Daten und der Validierung der Ergebnisse der durch die Algorithmen identifizierten ATC-GM Codes, kann abschließend eine prozentuale Verteilung der unstrukturierten und strukturierten Medikationsverordnungen auf Basis der 85,18% der Medikationsverordnungen (vgl. Abschnitt 4.4.2.2) für jede der 14 ATC Gruppen und für jeden ATC-GM Code erfolgen.

Tabelle 4.6 gibt die Gesamtzahl und den Anteil der strukturierten und unstrukturierten Medikationsverordnungen jeweils in der absoluten Zahl und mittels des prozentualen Anteils wieder. Der prozentuale Anteil pro Zeile für die Spalten „Strukturierte Medikationsverordnungen“ und „Unstrukturierte Medikationsverordnungen“ bezieht sich dabei auf die jeweilige ATC Gruppe.

Als Beispiel sei hier die ATC Gruppe „N -Nervensystem“ genannt. Der Anteil der strukturierten Medikationsverordnungen dieser ATC Gruppe beträgt 61,38% von insgesamt 322.286 Medikationsverordnungen. Der prozentuale Anteil angegeben in Spalte „Gesamtzahl“ bezieht sich pro Zeile auf die Gesamtzahl der 1.768.153 Medikationsverordnungen. Die ATC Gruppe „N - Nervensystem“ nimmt dabei mit insgesamt 322.286 (24,1%) Medikationsverordnungen den größten Anteil ein. Die am wenigsten häufig verordnete ATC Gruppe ist „P - Antiparasitäre Mittel, Insektizide und Repellenzien’ mit nur 1.461 (0,11%) Medikationsverordnungen.

Der Vollständigkeit halber wurden der Tabelle 4.6 drei weitere Zeilen hinzugefügt. Hierbei handelt sich um die als „nomed“ gekennzeichneten unstrukturierten Verordnungen, die keine Medikationsverordnungen sind. Außerdem werden die mit „unspec“ gekennzeichneten Medikationsverordnungen dargestellt, die aufgrund eines Mangels an Spezifikation keinen eindeutigen ATC-GM Code zugeordnet bekommen können. Und abschließend werden der Anteil der validierten und nicht validierten Medikationsdaten beziffert.

Abbildung 4.12 veranschaulicht den Grad der Strukturiertheit der Medikationsverordnungen für jede der 14 ATC Level-1 Gruppen. Die Abbildung ist nicht nach der Menge der Medikationsverordnungen sortiert, sondern beginnend mit der am meisten strukturiert vorliegenden ATC Gruppe.

Tabelle 4.6: Anteil der Medikationsverordnungen pro ATC Gruppe (strukturierten, unstrukturiert, gesamt)

ATC Gruppe	Strukurierte Medikationsverordnungen (%)	Unstrukurierte Medikationsverordnungen (%)	Gesamtzahl (%)
N - Nervensystem	197.831 (61,38)	124.455 (38,62)	322.286 (24,1)
B - Blut und Blutbildende Organe	164.032 (65,32)	87.088 (34,68)	251.120 (18,77)
A - Alimentäres System und Stoffwechsel	137.988 (55,08)	112.555 (44,92)	250.543 (18,73)
C - Kardiovaskuläres System	170.703 (68,93)	76.926 (31,07)	247.629 (18,51)
J - Antiinfektiva zur systemischen Anwendung	60.844 (68,63)	27.815 (31,37)	88.659 (6,63)
H - Systemische Hormonpräparate exkl. Sexualhormone und Insuline	51.296 (79,9)	12.903 (20,1)	64.199 (4,8)
M - Muskel- und Skelettsystem	12.083 (32,82)	24.736 (67,18)	36.819 (2,75)
R - Respirationstrakt	19.686 (69,94)	8.462 (30,06)	28.148 (2,1)
V - Varia	9.639 (65,7)	5.033 (34,3)	14.672 (1,1)
L - Antineoplastische und Immunmodulierende Mittel	8.670 (59,64)	5.868 (40,36)	14.538 (1,09)
G - Urogenitalystem und Sexualhormone	3.662 (41,71)	5.116 (58,28)	8.778 (0,66)
S - Sinnesograne	5.077 (98,03)	102 (1,97)	5.179 (0,39)
D - Dermatika	2.127 (60,19)	1.407 (39,81)	3.534 (0,26)
P - Antiparasitäre Mittel, Insektizide und Repellenzien	342 (23,41)	1.119 (76,59)	1.461 (0,11)
nomed	0	166.307 (9,4)	166.307 (9,4)
unspec	0	2.187 (0,12)	2.187 (0,12)
Total validiert	843.980 (47,73)	662.079 (37,45)	1.506.059 (85,18)
nicht validiert	0	262.094 (14,82)	262.094 (14,82)

Die ATC Gruppe „S - Sinnesorgane“ ist mit 98,03% strukturierten Daten die mit der höchsten initialen Strukturiertheit, gefolgt von der Gruppe „H - Systemische Hormonpräparate, ausgenommen Sexualhormone und Insuline“ mit 79,90% strukturierten Medikationsverordnungen. Die ATC Gruppen R, C, J, V, B und N, lagen in ihrer Strukturiertheit zwischen 61% und 70% . Die ATC Gruppe P wies mit nur 23,4% den geringsten Anteil an strukturierten Medikationsverordnungen auf.

Insgesamt wurden 739 Wirkstoffe (ATC-GM Codes) in den Medikationsverordnungen identifiziert. Eine vollständige Liste aller vorkommenden ATC-GM Codes, mit der Anzahl der Medikationsverordnungen und den strukturierten und unstrukturierten Anteilen sind in Anhang D vollständig einsehbar. Die Liste ist geordnet nach der Anzahl der Medikationsverordnungen, beginnend mit dem am häufigsten verordneten Wirkstoff. Die 3 am häufigsten in den Medikationsverordnungen (Datensatz DS-Med) vorkommenden ATC-GM Codes sind N02BB02 - *Metamizol-Natrium* (4,57%, 80.866/1.768.153), B05BB01 - *Elektrolyte* (3,86%, 68.299/1.768.153) und A02BC02 - *Pantoprazol* (3,72%, 65.861/1.768.153).

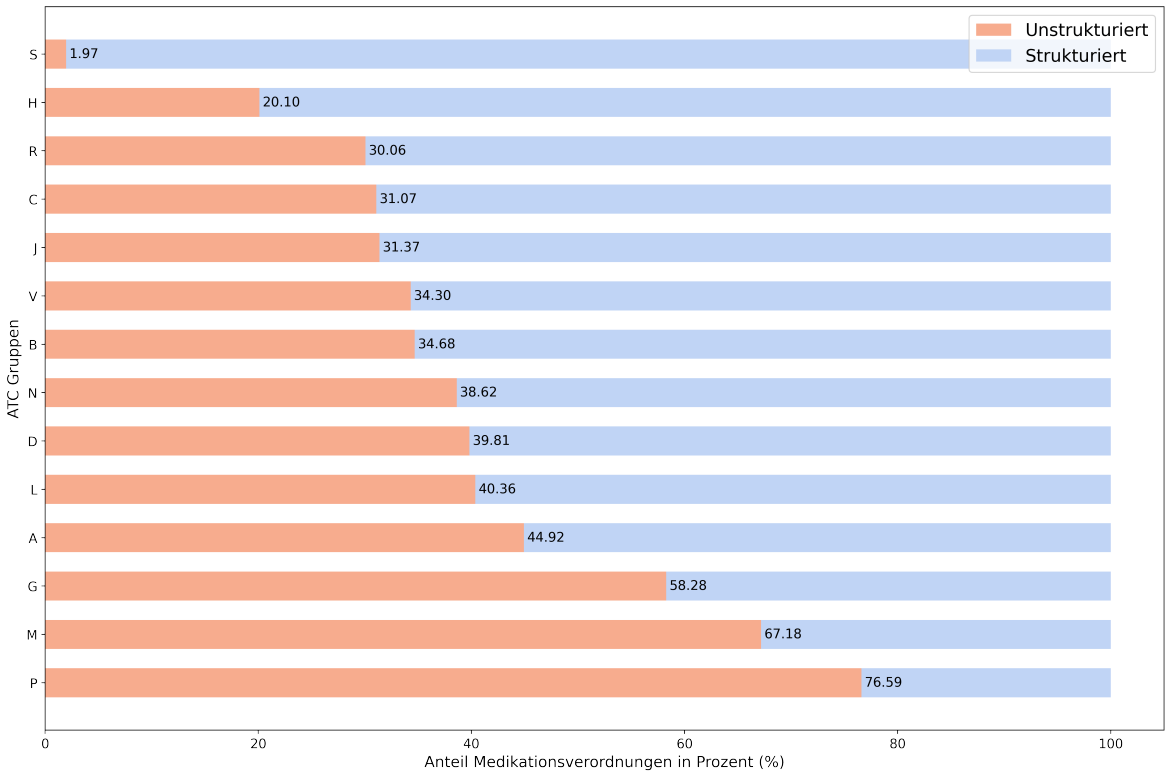


Abbildung 4.12: Strukturiertheit der Medikationsverordnungen für 85,18% des initialen Datensatzes DS-Med

Zusammen machen diese 3 ATC-GM Codes bereits 12,15% der Einträge des Datensatzes DS-Med aus. Während die ATC-GM Codes N02BB02 und B05BB01 zu mehr als 50% strukturiert vorliegen, weicht die Strukturiertheit von A02BC02 stark ab. Hier liegen lediglich nur 15,93% aller Verordnungen strukturiert vor.

4.4.3 Ergebnisse der Maßnahmen - Terminologie

4.4.3.1 Angepasste ATC-GM Codes an die internationale WHO Version

Gemäß des in Kapitel 3.5.3.1 vorgestellten Prozesses (siehe Abbildung 3.5) und der durchgeführten Arbeitsschritte (A1) bis (A6), war es möglich alle 739 ATC-GM Codes als valide Konzepte eindeutig über die *concept_id* in OMOP abzubilden.

In Schritt (A1) konnten unter Verwendung der verbesserten Medikationsverordnungen (DS-Med) 629 der 739 ATC-GM Codes durch ein valides Konzept in OMOP unter Verwendung des aktuellen ATC Vokabulars aus OMOP (Datensatz DS-ATC) basierend auf der international gültigen ATC Version der WHO abgebildet werden. Das entspricht 1.242.986 Medikationsverordnungen des Datensatzes DS-Med von den 1.506.059 Medikationsverordnungen (85,18% des Datensatzes DS-Med) der insgesamt 1.768.153 Medikationsverordnungen. Es gibt also

eine Abweichung der ATC Codes zwischen der deutschen ATC-GM Version und der internationalen ATC WHO Version für 110 ATC-GM Codes, die 263.073 Medikationsverordnungen aus Datensatz DS-Med entsprechen.

Für die betreffenden 110 ATC-GM Codes konnten in Schritt (A2) 95 ATC-GM Codes eindeutig einem ATC WHO Code unter Verwendung des Hauskatalogs für Medikation des UKD (Datensatz DS-Katalog) zugeordnet werden. Die 15 verbleibenden ATC-GM Codes sind in Tabelle 4.7 sortiert nach der Häufigkeit des Vorkommens in den Medikationsverordnungen (Spalte *FREQUENY*) dargestellt.

Tabelle 4.7: ATC-GM Codes ohne ATC WHO Äquivalent nach Schritt (A2)

ATC-GM Code	gültiger ATC-GM Code	Datenelement „Medikation" aus DS-Med	Anzahl	ATC WHO Code	neues Konzept in ATC-GM Vokabular
V06XX02	ja	MAGNESIUM VERLA 300 Orange Granulat . . SCHWEDEN-TABLETTEN 0,25 ... Schwedentabletten	2028	nein	ja
A11EB01	ja	DREISAVIT N Filmtabletten ...	1205	A11EB	nein
D04AB61	ja	Polidocanol-Harnstoff-Creme ...	195	nein	ja
A11BA01	ja	MULTIBIONTA Nutrition Tropfen ...	160	A11BA	nein
L01FA01	ja	Rituximab	157	nein	ja
J01CR21	nein	Unacid (Ampicillin/Sulbactam) i.v. in [g]	142	J01CR01	nein
N05CM27	ja	Sedaplus	117	nein	ja
A11DB03	ja	VITAMIN B Komplex ratiopharm Kapseln ...	115	A11DB	nein
V08EA01	nein	HEXVIX 85 mg Kontrastmittel f.d.PDD ...	51	V04CX06	nein
P03AX10	ja	NYDA gegen Läuse und Nissen Pumplösung ...	34	nein	ja
R04AP30	ja	PULMOTIN Salbe | (Eucalyptusöl)	16	nein	ja
R07AA03	ja	MUCOSOLVAN Infusionslösungskonzentrat...	11	nein	ja
A01AP03	ja	SALBEIBLÄTTER Tee Filterbeutel ...	11	nein	ja
A03AB20	ja	SPASMEX i.v. 1,2 Injektionslösung ...	9	nein	ja
M02AH20	ja	ACONIT Schmerzöl ...	6	nein	ja

Unter Verwendung der online verfügbaren aktuellen ATC WHO Version, konnte für weitere drei ATC-GM Codes aus Tabelle 4.7 ein ATC WHO Code identifiziert werden. Es handelt sich dabei um verschiedene Vitaminkombinationen (ATC-GM Codes A11EB01, A11BA01 und A11DB03), die in der WHO Version bereits auf Level 4 in beiden Versionen eindeutig definiert sind. In der ATC WHO Version existiert jeweils kein hierarchisch darunter liegendes Level 5. Als Beispiel sei hier der ATC-GM Code A11EB01 genannt, der die Bezeichnung „Vitamin-B-Komplex mit Vitamin C" hat und damit derselben Bezeichnung entspricht wie der des höheren Level 4 und des ATC-GM Codes A11EB. Insgesamt konnten also für 98 der 110 ATC-GM Codes entsprechende Äquivalente in der ATC WHO Version identifiziert werden. Nach Durchführung von Schritt (A2) und (A3) verblieben zwölf ATC-GM Codes ohne validen ATC WHO Code.

In Schritt (A3) konnten die identifizierten eindeutigen Äquivalente für die 98 betreffenden ATC-GM Codes in der WHO Version dann im Datensatz DS-Med ersetzt werden. Die erneute Iteration von Schritt (A1) führte in der Folge zur Erhöhung der validen Konzepte in OMOP für die Medikationsdaten auf eine Gesamtheit von 1.332.672 (75,37% von 1.768.153), die bereits 727 der 739 ATC Codes in den Medikationsverordnungen entsprechen.

In Schritt (A4) ergab die Prüfung hinsichtlich Korrektheit der betrachteten zwölf ATC-GM Codes aus Tabelle 4.7 insgesamt zehn korrekte ATC-GM Codes ohne verfügbaren ATC WHO Code. Die anderen 2 ATC-GM Codes (J01CR21 und V08EA01) wurden unter Verwendung der durch das WIdO online ab dem Jahr 2017 bereitgestellten Archiv inklusive Änderungsdateien (Wissenschaftliches Institut der AOK (WIDO), 2023) als ungültig identifiziert. Wie in Tabelle 4.7 konnten die aktuell gültigen ATC Codes zugeordnet werden, die auch weltweit in der WHO Version gültig sind.

In Schritt (A5) wurden für die beiden Codes die aktualisierten und gültigen ATC-GM Codes identifiziert. Die beiden neuen Codes (J01CR01 und V04CX06) sind in der ATC-GM identisch zur ATC WHO Version und konnten daher auch als valide Konzepte durch wiederholte Ausführung von Schritt (A1) mit eindeutiger *concept_id* in OMOP erkannt und zugeordnet werden. Damit erhöhte sich die Anzahl der validen Konzepte in OMOP für die Medikationsdaten um 193 Verordnungen auf eine Gesamtzahl von 1.332.865 (75,38% von 1.768.153)).

In einem abschließenden Schritt (A6) wurde in der OMOP Datenbank ein neues Vokabular mit dem Namen ATC-GM angelegt und die 10 verbleibenden ATC-GM Codes als neue Konzepte in der *concept* Tabelle der OMOP Datenbank angelegt. Diese ATC-GM Codes sind in Tabelle 4.7 in Spalte *Neues Konzept in ATC-GM Vokabular* mit „ja“ gekennzeichnet. Der Vollständigkeit halber wurden zwei zusätzliche Konzepte mit dem *concept_code* „nomed“ und „unspec“ in dem ATC-GM Vokabular in OMOP angelegt. So war es möglich, 100% der Medikationsverordnungen aus Datensatz DS-Med in OMOP über ein entsprechendes Konzept abzudecken. Der Anteil der nicht validierten Medikationsverordnungen erhält in OMOP die *concept_id* Null zugeordnet. So ist es auch in OMOP möglich, die Medikationsverordnungen validiert, aber mit „nomed“ oder „unspec“ gekennzeichneten Verordnungen, von den nicht validierten Verordnungen zu unterscheiden. Eine vollständige Liste der generierten Konzepte in der OMOP Tabelle *concept* mit den Spalten *concept_id* und *concept_name* findet sich in Anhang E.

4.4.3.2 Ergebnisse der Überführung nach RxNorm

In diesem Abschnitt werden Ergebnisse der Überführung nach RxNorm auf Basis der an die ATC Version der WHO angepassten ATC Codes vorgestellt. Um die Überführung nach RxNorm zu realisieren, müssen die als relevant ermittelten Beziehungstypen für das semantisch korrekte Mapping von ATC nach RxNorm bekannt sein. Es wurden gemäß der genutzten Methode aus Abschnitt 3.5.3.2 zunächst fünf relevante Beziehungstypen für das korrekte semantische Mapping auf Wirkstoffebene zwischen ATC und RxNorm Codes der entsprechenden OMOP Vokabulare identifiziert.

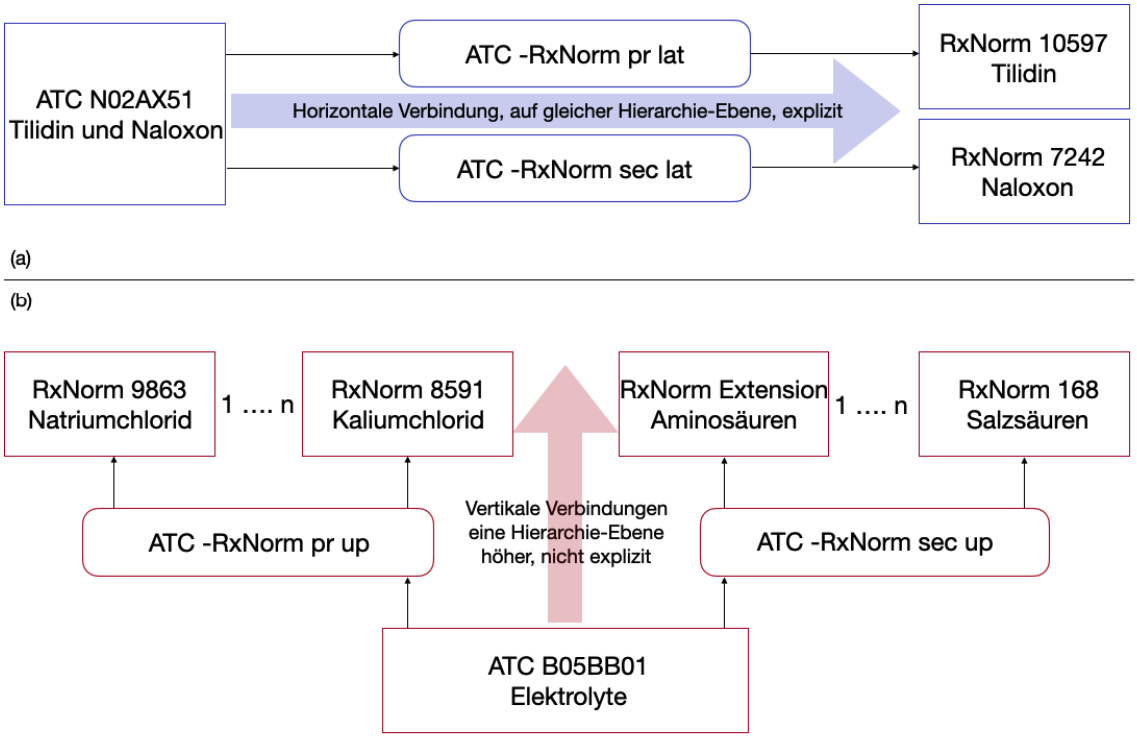


Abbildung 4.13: ATC zu RxNorm Beziehungstypen - exemplarische Darstellung

Zunächst wurden vier der fünf Beziehungstypen „ATC -RxNorm pr lat“, „ATC -RxNorm sec lat“, „ATC -RxNorm pr up“ und „ATC -RxNorm sec up“ gemäß der offiziellen Dokumentation des OHDSI ATC Vokabulars (OHDSI, 2022) als relevant identifiziert. Außerdem wurde anhand eines OHDSI Forum Artikel einer OHDSI Expertin des OHDSI Vokabular Entwicklungsteams mit anschließender Expert:innen Diskussion ein fünfter relevanter Beziehungstyp „Maps to“ identifiziert.

Abbildung 4.13 zeigt die Verwendung der vier genannten Beziehungstypen für das semantische Mapping zwischen ATC und RxNorm Codes exemplarisch an den beiden ATC Codes

N02AX51 und B05BB01. Die beiden Beziehungstypen „ATC -RxNorm pr lat“, „ATC -RxNorm sec lat“ werden für das semantische Mapping auf gleicher Ebene der Hierarchie und für explizit genannte Wirkstoffe in einem ATC Code nach RxNorm verwendet. Im Fall des Beispiels N02AX51 (Abbildung 4.13, a), resultiert das Mapping einer Kombination aus den beiden Wirkstoffen *Tillidin* und *Naloxon* innerhalb eines ATC Codes in einer Darstellung in zwei klinischen Fakten in der OMOP Tabelle *drug_exposure* als zwei separate Einträge.

Ein ATC Code wie beispielsweise B05BB01 lässt sich nach RxNorm nicht auf explizite Wirkstoffe mappen, da es unterschiedliche Zusammensetzungen von elektrolytischen Infusionslösungen gibt, die von dem Medikament mit seinem Zweck und von der Indikation des Arzneimittels abhängen. Daher existieren, wie in Abbildung 4.13 (b) gezeigt, die Beziehungstypen „ATC -RxNorm pr up“ und „ATC -RxNorm sec up“ mit einer Vielzahl von möglichen RxNorm Konzepten. Ein Mapping von Medikationsverordnungen mit dem ATC Code B05BB01 auf die korrekten RxNorm Codes ist ohne zusätzliches Wissen zum konkret verabreichten Medikament nicht möglich. Da die Beziehungstypen eine hohe Komplexität und Dopplung von Mapping-Informationen beinhalten, können sie für die Überführung der Daten von ATC nach RxNorm nicht genutzt werden. Stattdessen sollte für die Datenübertragung ausschließlich der Beziehungstyp „Maps to“ von ATC nach RxNorm genutzt werden, da hier ausschließlich die explizit benennbaren Wirkstoffe eines ATC Codes semantisch gemappt werden.

Es gibt für den ATC Code N02AX51 exakt zwei Einträge in der OMOP Tabelle *concept_relationship* mit den Verbindungstypen „Maps to“, die für dieses Beispiel eine vollständige Überführung der semantischen Bedeutung des ATC Codes nach RxNorm erlauben. Für ATC Codes, bei denen nur ein Wirkstoff von mehreren explizit benannt wird, beispielsweise bei C09BA05 *Ramipril und Diuretika*, existiert auch ausschließlich für diesen expliziten Wirkstoff *Rampiril* ein semantisches Mapping über den Beziehungstyp „Maps to“ nach RxNorm zur Überführung der Daten.

Von den 6.552 im ATC Vokabular existierenden ATC WHO Codes, die im Betrachtungszeitraum dieser Arbeit gültig waren oder aktuell noch gültig sind, verfügen 4.824 (73,63%, 4.824/6.552) Codes über mindestens einen der genannten fünf Beziehungstypen zu mindestens einem RxNorm Code auf Ebene der Wirkstoffe (TTY entspricht „Ingredient“).

Die für die Überführung nach RxNorm notwendigen Informationen zum semantischen Mapping beschränken sich allerdings auf den Beziehungstyp „Maps to“ und sind verfügbar für 4.678 ATC Codes. Nicht für jeden der 4.678 ATC Codes existiert nur ein einzelnes Mapping nach RxNorm. Aus Abschnitt 2.7.1 ist bereits bekannt, dass in der ATC Terminologie mehrere Wirkstoffe durch einen einzelnen Code abgebildet werden können. Daher gibt Tabelle 4.8 einen Überblick über die Anzahl der ATC Codes nach RxNorm mit mehr als einem Mapping, wie beispielsweise der in Abbildung 4.13 illustrierte ATC Code N02AX51 für *Tilidin und Naloxon* durch den zwei Wirkstoffe dargestellt sind. Bei den 13 ATC Codes, mit mehr als 4 Mappings nach RxNorm handelt es sich ausschließlich um Codes aus der ATC Unterkategorie *J07 - Impfstoffe*.

Tabelle 4.8: Anzahl ATC Codes mit einer Verbindung nach RxNorm durch „Maps to“ mit Beispielen

Anzahl ATC Codes	Anzahl „Maps to“ Verbindungen	Beispiel für die „Maps to Verbindung“	
		ATC Code	RxNorm Code
146	0	B05BB01 (Elektrolyte)	Keiner
4285	1	H02AB02 (Dexamethason)	3264 - Dexamethason
290	2	N02AX51 (Tilidin und Naloxon)	10597 - Tilidin 7242 - Naloxon
77	3	A10BD25 (Metformin, Saxagliptin und Dapagliflozin)	6809 - Metformin 857974 - Saxagliptin 1488564 - Dapagliflozin
13	4	A02BD11 (Pantoprazol, Amoxicillin, Clarithromycin und Metronidazol)	40790 - Pantoprazol 723 - Amoxicillin 21212 - Clarithromycin 6922 - Metronidazol
13	>4	J07CA11 (Diphtherie-Haemophilus influenzae B-Pertussis-Tetanus-Hepatitis B)	798302 - azellulärer Pertussis-Impfstoff 798306- Tetanus-Toxoid-Impfstoff 797752 - HBV-Impfstoff 798279 - Hib-Impfstoff 798304 - Diphtherie-Toxoid-Impfstoff

Die in Tabelle 4.8 abgebildeten Beispiele zeigen ausschließlich ATC Codes, die ein exaktes und explizites Mapping auch mehrerer Wirkstoffe nach RxNorm beinhalten. Es gibt auch andere ATC Codes, wie beispielsweise C09BA04 (*Rampiril und Diuretika*), bei denen ein Wirkstoff über die Verbindung „Maps to“ explizit für den Wirkstoff *Rampiril* von ATC nach RxNorm überführbar ist. Für die *Diuretika* jedoch fehlt eine explizite Defintion des Wirkstoffes und es gibt in RxNorm mehrere Möglichkeiten. Daher existieren zusätzlich zu dem expliziten Mapping noch Beziehungen der vier genannten Beziehungstypen („ATC -RxNorm pr lat“, „ATC -RxNorm sec lat“, „ATC -RxNorm pr up“ und „ATC -RxNorm sec up“). Ohne Kenntnisse zum Produkt des ATC Codes ist daher das Mapping für die „Diuretika“ nicht möglich und bleibt unvollständig. Diese nicht-expliziten Wirkstoffe entfallen nach der Überführung nach RxNorm und sind nicht gekennzeichnet und nicht abbildbar.

In Kapitel 3.5.3.2 wurde im Quellcode 3.4 eine Methode vorgestellt, die zur Zuordnung der richtigen RxNorm Codes zu den Medikationsverordnungen mit ATC Code eingesetzt wurde. Mit diesem Vorgehen konnte für insgesamt 1.169.330 Medikationsverordnungen (abgebildet durch 620 verschiedene ATC Codes) aus DS-Med mindestens ein valider RxNorm Code korrekt zugeordnet werden. Dabei handelt es sich jedoch nicht ausschließlich um Mappings von einem ATC Code zu exakt einem RxNorm Code.

Es existieren auch Medikationsverordnungen, bei denen ein ATC Code zu zwei oder drei RxNorm Codes gemappt wurde. Das betrifft von den genannten 1.169.330 Medikationsverordnungen nur einen kleinen Anteil von 1,17% (13.638/1.169.330). Es handelt sich dabei um 17 verschiedene ATC Codes, verwendet in 13.632 Medikationsverordnungen, bei denen der ATC Code in zwei RxNorm Codes semantisch gemappt wird und einen ATC Code (J07AJ52), verwendet in sechs Medikationsverordnungen, der in drei RxNorm Codes semantisch gemappt wird. Auf dieser Grundlage wurde die Anzahl der Medikationsverordnungen in der OMOP Datenbank für diese betreffenden Einträge vervielfacht. Insgesamt gibt es nun 1.781.797 Einträge (Datensatz DS-Med ursprünglich 1.768.153 Medikationsverordnungen umfassend) in der OMOP Tabelle *drug_exposure*, von denen 1.182.974 Einträge als gültige Standardkonzepte in OMOP mit einem validen RxNorm Code gelten.

Da die Validierung der Maßnahmen zur Erhöhung der Strukturiertheit der Medikationsverordnungen des Datensatzes DS-Med wie in Tabelle 4.6 in Ziele „Total validiert" lediglich auf 85,18% aller Medikationsdaten angewandt wurde, ist die Überführungsrate von ATC nach RxNorm im Verhältnis 1.182.974/1.506.059 Medikationsverordnungen prozentual mit 78,55% anzugeben. Von den 1.506.059 validierten Medikationsverordnungen wurden außerdem 166.307 Medikationsverordnungen als „nomed" und „unspec" gekennzeichnet. Deshalb kann die Berechnungsgrundlage zur Kalkulation der Überführungsrate von ATC nach RxNorm für die Medikationsverordnungen auf eine Gesamtmenge von 1.337.565 (1.506.059 abzüglich 166.307) reduziert werden. Der tatsächliche Abdeckungsgrad der Überführung von ATC nach RxNorm liegt also 88,44% (1.182.974/1.337.565).

4.4.4 Zusammenfassung der Ergebnisse der Maßnahmen

Die im Rahmen der Teilnahme an der EMA Studie exemplarisch durchgeführten manuellen Maßnahmen zur Verbesserung der Strukturiertheit der Medikationsdaten und deren Konformität zur in OMOP erwarteten Terminologie RxNorm, konnten in einem ersten Schritt für eine kleine Menge von vier Wirkstoffen und eine kleine Kohorte, die ausschließlich stationäre COVID-19 Behandlungsfälle aus den Jahren 2020 und 2021 umfasst, umgesetzt werden.

Die generelle Strukturiertheit der Medikationsverordnungen des UKD für den Datensatz DS-Med konnte durch die Entwicklung und Anwendung geeigneter Algorithmen von initial 47,27% auf 85,18% verbessert werden. Die Überführung von den Medikationsverordnungen mit ATC Code nach RxNorm war für 620 ATC Codes möglich, die 1.169.330 Medikationsverordnungen entsprechen.

4.5 Ergebnisse der Bewertung

In diesem Kapitel werden die Bewertung der zuvor durchgeführten Maßnahmen zur Verbesserung der Strukturiertheit der Medikationsverordnungen aus Kapitel 4.4.2 und der Überführung der Terminologie nach RxNorm aus Kapitel 4.4.3 vorgestellt. Dazu werden zunächst in Abschnitt 4.5.1 die Ergebnisse der qualitativen Bewertung und im Anschluss daran in Abschnitt 4.5.2 die Ergebnisse der quantitativen Bewertung vorgestellt. Eine Zusammenfassung der Ergebnisse der Bewertung insgesamt erfolgt am Ende des Kapitels.

4.5.1 Ergebnisse der qualitativen Bewertung

Gemäß der in Kapitel 3.6 beschriebenen Methodik wurden im Rahmen der durchgeführten qualitativen Bewertung für die Wirkstoffe *Levetiracetam* und *Phenytoin* zunächst der gültige RxNorm Code in der Webanwendung Athena (vgl. Kapitel 2.4) ermittelt. Dabei wurde wie in Abbildung 4.14 in Athena nach den Begriffen *Levetiracetam, Phenytoin* gesucht. Außerdem wurde über die Filteroptionen eingeschränkt auf „Drug“ als Domäne (Domain), Standardkonzepte (Standard) und „Ingredient“ als Klasse (Class).

Die für die Studie relevanten Wirkstoffe wurden im Datensatz DS-Med identifiziert. Die Ergebnisse der systematischen Analyse der Medikationsverordnungen am UKD zeigen eine stark unterschiedliche Strukturiertheit der beiden Wirkstoffe.

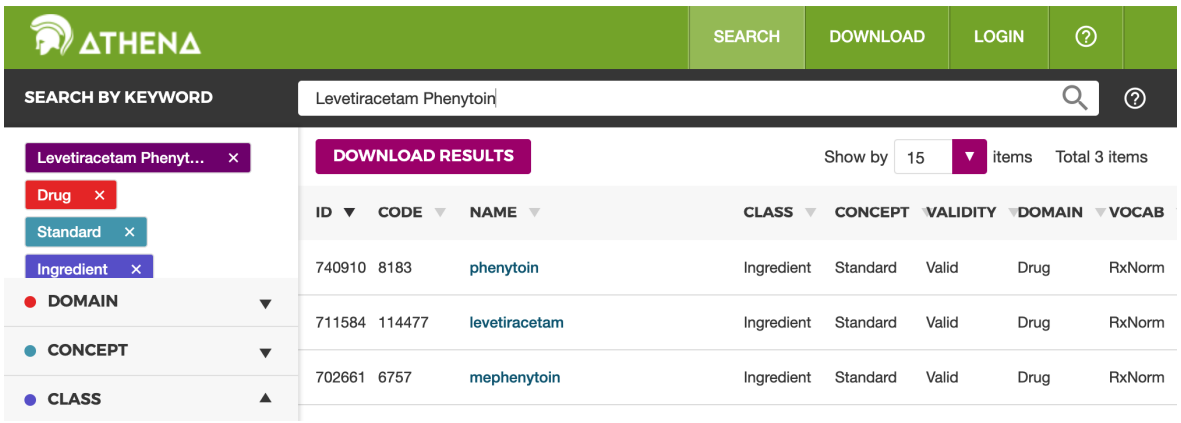


Abbildung 4.14: Athena - Suche nach den Wirkstoffen Levetiracetam und Phenytoin

Der Wirkstoff *Phenytoin* wurde nur 210-mal verordnet. Alle Verordnungen dieses Wirkstoffes erfolgten strukturiert und enthalten den korrekten ATC Code. Der Wirkstoff *Levetiracetam* wurde 6.060-mal verordnet. Allerdings im Gegensatz zum Wirkstoff *Phenytoin* in nur 1.288 Verordnungen tatsächlich aus dem Hauskatalog (DS-Katalog) ausgewählt. Die Strukturiertheit dieses Wirkstoffes in den Medikationsverordnungen liegt bei nur zu 21,25%.

Tabelle 4.9: Wirkstoffe Levetiracetam und Phenytoin, Metainformationen

Wirkstoff	Levetiracetam	Phenytoin
ATC Code	N03AX14	N03AB02
RxNorm Code	114477	8183
Häufigkeit in Medikationsverordnungen (Datensatz DS-Med)	6060	210
Strukturiertheit der Daten, initial	21,25% (1288/6060)	100% (210/210)
Mapping ATC nach RxNorm explizit	ja	ja
Nutzbarkeit in Studie, vor Maßnahmen	0%	0%
Nutzbarkeit in Studie, nach Maßnahmen	100%	100%

Nach der Durchführung der Maßnahme zur Verbesserung der Datenstruktur (siehe Ergebnisse in Kapitel 4.4.2.3) konnten die unstrukturiert vorliegenden Medikationsverordnungen von „Levetiracetam“ durch die Ausführung der Algorithmen automatisiert nach ATC überführt werden und sind folgend für die Nutzung zu Forschungszwecken strukturiert verfügbar. Für beide Wirkstoffe konnte der korrekte ATC WHO Code bestimmt werden (siehe Tabelle 4.9). Die beiden ATC Codes N03AX14 (*Levetiracetam*) und N03AB02 (*Phenytoin*) verfügen über ein explizites semantisches Mapping von ATC nach RxNorm über den Beziehungstypen „Maps to“.

Die Umsetzung der Maßnahmen zur Reduktion der identifizierten Inhibitoren bedeutet eine verbesserte Strukturiertheit der Medikationsverordnungen des UKD und ermöglicht nun eine Teilnahme an der ausgewählten Studie von Duke et al. (Duke et al., 2017) unter Wahrung der semantischen Bedeutung.

4.5.2 Ergebnisse der quantitativen Bewertung

Die Ergebnisse der quantitativen Bewertung unter Verwendung des OHDSI DQD sind in Abbildung 4.15 zusammenfassend für die Schritte 1, 2 und 3 dargestellt. Die vollständigen Ergebnisse des OHDSI DQD befinden sich als Screenshots des DQD für die drei Prüfungen „isStandardValidConcept", „sourceConceptRecordCompletness" und „standardConceptRecordCompletness" in Anhang F.

Prüfungen OHDSI DQD	**Schritt 1** Medikationsverordnungen (DS-Med), original (n = 1.768.153)		**Schritt 2** Medikationsverordnungen (DS-Med), nach Maßnahmen Struktur (n = 1.768.153)		**Schritt 3** Medikationsverordnungen (DS-Med), nach Maßnahmen Terminologie (n = 1.781.797)	
	Status	Anteil n in %	Status	Anteil n in %	Status	Anteil n in %
isStandardValidConcept (Prüfung der Konformität, Tabellenspalte drug_concept_id)	PASS	100	PASS	100	PASS	100
sourceConceptRecordCompletness (Prüfung der Vollständigkeit, Tabellenspalte drug_source_concept_id	FAIL	52,27	FAIL	14,82	FAIL	14,71
standardConceptRecordCompletness (Prüfung der Vollständigkeit, Tabellenspalte drug_concept_id)	FAIL	100	FAIL	100	FAIL	33,61

Abbildung 4.15: DQD Zusammenfassung der Ergebnisse für die 3 Schritte

Die Prüfung „isStandardValidConcept" des Datenfeldes *drug_concept_id* ergab für alle 3 Entwicklungsstufen der Daten eine Konformität von 100% und eine erfolgreiche Absolvierung der Prüfung. Für das Stadium 1 und Stadium 2 wurde in der *drug_concept_id* der Tabelle *drug_exposure* für alle Medikationsverordnungen (n=1.768.153) ausschließlich das „No matching concept" mit der *concept_id* null (*concept_id = 0*) verwendet. Dieses Konzept ist ein gültiges Konzept, jedoch kein Standard-Konzept. Das OHDSI DQD nutzt für die Prüfung „isStandardValidConcept" eine SQL Abfrage, welche das Konzept mit der *concept_id = null* aus der Ergebnismenge der Nicht-Standard-Konzepte ausschließt. Aus diesem Grund kommt es zu dem nicht erwarteten Ergebnis der Konformität für diese Prüfung von 100%.

Die Prüfung „sourceConceptRecordCompletness" des Datenfeldes *drug_source_concept_id* schlägt für alle 3 Schritte und Stadien der Medikationsverordnungen fehl (Status=FAIL). Das OHDSI DQD fordert hier für alle Einträge in der *drug_exposure* Tabelle ein gültiges Konzept aus der *concept* Tabelle, welches nicht zwingend ein Standardkonzept sein muss. In dem Datenfeld *drug_source_concept_id* wird eine valide *concept_id* eingetragen, wenn ein entsprechender ATC Code angegeben wurde. Der ATC Code muss für die korrekte Referenzierung der *concept_id* als Konzept in einem Vokabular in der OMOP Datenbank existieren. Unter Verwendung der initialen Medikationsverordnungen (Schritt 1) des Datensatzes DS-Med

werden 52,27% der Daten mit einem Konzept "No *concept_id*"(*concept_id = 0*) gespeichert. Der Anteil dieser Daten mit der *concept_id= 0* verringert sich nach der Durchführung der Maßnahmen zur Verbesserung der Datenstruktur aus Kapitel 4.4.2 auf 14,82%. Nach den Maßnahmen zur Terminologie aus Kapitel 4.4.3 verringert sich dieser Anteil minimal auf 14,71%. Diese geringfügige Abweichung liegt an der neuen Gesamtmenge von Einträgen der Medikationsverordnungen in der OMOP Datenbank nach der Überführung nach RxNorm aus Kapitel 4.4.3.2 und resultiert aus den Mappings, bei denen ein ATC in zwei oder drei RxNorm Codes überführt wird.

Die Prüfung „standardConceptRecordCompletness“ des Datenfeldes *drug_concept_id* schlägt ebenfalls für alle 3 Schritte und Stadien der Medikationsverordnungen fehl (Status=FAIL). In Schritt 1 und 2 schlägt die Prüfung im DQD für 100% der Medikationsverordnungen fehl, weil alle Einträge in dem Datenfeld *drug_concept_id* kein valides Standardkonzept der Domäne „drug“ gesetzt haben und daher durch die Prüfung als unvollständig gekennzeichnet werden. Nachdem die Überführung nach RxNorm stattgefunden hat, schlägt die Prüfung nur noch für 33,61 % der Medikationsverordnungen fehl.

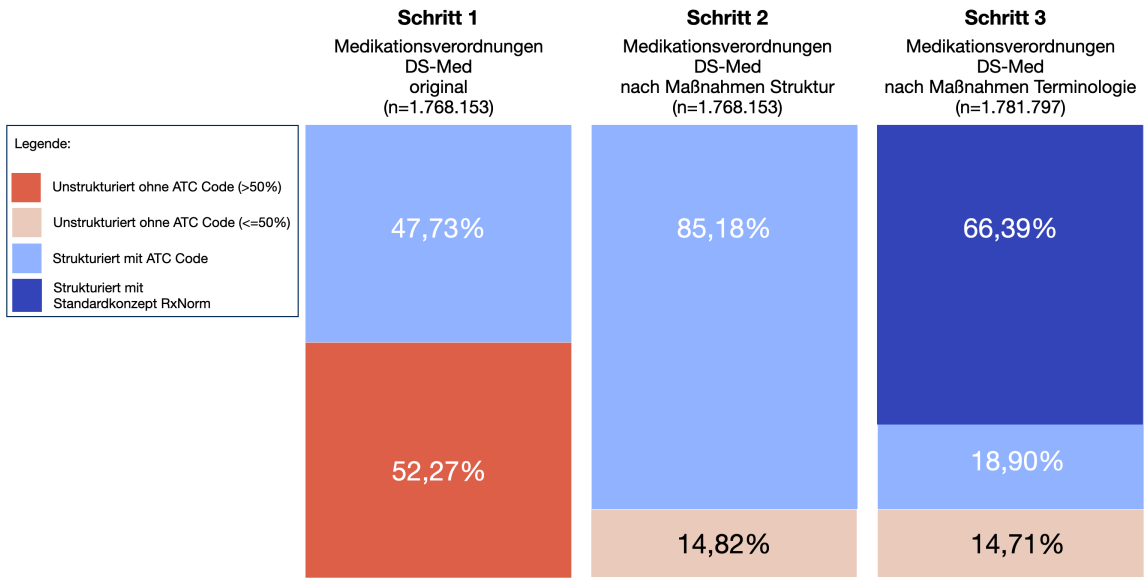


Abbildung 4.16: Quantitative Bewertung der Medikationsverordnungen gemäß DQD

Abbildung 4.16 zeigt eine Zusammenfassung der quantitativen Bewertung nach dem Kriterium der Strukturiertheit der Medikationsverordnungen und der Verfügbarkeit des ATC Codes, beziehungsweise des Standardkonzeptes RxNorm. Korrespondierend zu den Ergebnissen des DQD entspricht der unstrukturierte Anteil der Daten in allen 3 Schritten dem Anteil der Daten, die in der Prüfung „sourceConceptRecordCompletness“ fehlgeschlagen sind.

Nach Schritt 3 liegen für 66,39% der Medikationsverordnungen eine Strukturiertheit mit gültigem RxNorm Code vor. Mit Blick auf Vollständigkeit der Tabellenspalte *drug_concept_id*, kann mit diesen 66,39% Medikationsverordnungen geforscht werden. Im Vergleich dazu waren bei Schritt 1 keine der Medikationsverordnungen über einen RxNorm Code zugreifbar und damit hinsichtlich der Anforderungen an des OMOP Datenmodell nicht nutzbar.

4.5.3 Zusammenfassung der Ergebnisse der Bewertung

Die Ergebnisse der Bewertung der durchgeführten Maßnahmen stellt ein wichtiges Messinstrument dar, um sicherzustellen, dass die zuvor identifizierten Inhibitoren hinsichtlich Strukturiertheit und Mangel an verfügbarer standardisierter Terminologie in den Medikationsverordnungen reduziert wurden. Die Ergebnisse der Bewertung beziffern das Ausmaß der Reduktion der Inhibitoren.

Die Durchführung der qualitativen Bewertung an einem Beispiel hat gezeigt, dass nach den Maßnahmen an dieser Studie dank der verbesserten Strukturiertheit der Medikationsverordnungen und der Überführung nach RxNorm eine erfolgreiche Teilnahme aus Sicht der Daten und unter Wahrung der semantischen Bedeutung möglich gemacht wurde. Die Bewertung konnte qualitativ exemplarisch als auch quantitativ erfolgreich durchgeführt werden.

Die quantitative Bewertung unter Verwendung des DQD hat für die Medikationsverordnungen eine kontinuierliche Verbesserung nach Durchführung der Maßnahmen gezeigt. Im Hinblick auf die Anforderungen seitens der geforderten Qualität an Daten in OMOP hinsichtlich der Konformität und Vollständigkeit konnte zudem eine deutliche Verbesserung nachgewiesen werden. Die Ergebnisse des DQD zeigen, dass die Medikationsverordnungen initial keiner validen Standardterminologie in OMOP entsprachen.

Nach Abschluss der Maßnahmen zeigte sich hier eine Erhöhung der Vollständigkeit, weil eine Verfügbarkeit der Standardterminologie RxNorm für 66,39% der Medikationsverordnungen gemessen werden konnte.

4.6 Ergebnisse zur Transparenz

Wie in Kapitel 3.7 beschrieben, wurde eine interaktive Visualisierung zur Darstellung der Strukturiertheit der Medikationsverordnungen aus DS-Med und das semantische Mapping der Medikationsverordnungen von ATC nach RxNorm umgesetzt. Die Visualisierung wurde im Rahmen eines Konzeptes für eine vorgeschlagene Feedbackschleife von Reinecke et al. (Reinecke, Bathelt et al., 2022) vorab veröffentlicht und kann dazu genutzt werden, gemeinsam mit den forschenden und versorgenden Teams am UKD, bereits während der Entstehung der Daten die Datenqualität zu verbessern.

Die interaktive Visualisierung der Medikationsverordnungen des Datensatzes DS-Med erfolgt auf Basis der ATC Codes als Datenpunkte. Zum einen wird dazu in Abschnitt 4.6.1 die Visualisierung der Datenstruktur pro ATC Code, zum anderen die Visualisierung des semantischen Mappings von ATC nach RxNorm in Abschnitt 4.6.2 vorgestellt.

Die beiden Visualisierungen (siehe Abbildung 4.17 und Abbildung 4.18) sind dabei grundsätzlich gleich aufgebaut und bestehen aus den folgenden 3 Teilbereichen:

1. Filtermöglichkeit auf Basis des ATC Codes oder des aktiven Wirkstoffs (oben links)
2. tabellarische Darstellung der gefilterten ATC Codes (oben rechts)
3. Streudiagramm (unten mittig)

Auf Basis des eingegebenen ATC Codes bzw. Wirkstoffnamens oder Teilen davon erfolgt die Filterung und Darstellung in der Tabelle und im Streudiagramm. Im Streudiagramm werden die gefilterten ATC Codes als größere Punkte dargestellt. Es ist zudem möglich, auf eine Zeile der Tabelle oben links zu klicken, um die Filterung im Streudiagramm auf den entsprechend markierten ATC Code einzuschränken. In diesem Fall wird ausschließlich der in der Tabelle markierte Code als größerer Punkt dargestellt.

Des Weiteren verfügen die Streudiagramme in Abbildung 4.17 und 4.18 über eine *Hover* Funktion. Schwebt der Mauszeiger direkt über einem ATC Code, werden weitere Informationen angezeigt. Diese Informationen unterscheiden sich jedoch pro Visualisierung. Weitere Details hinsichtlich der Unterschiede in Bezug auf die dargestellten Inhalte und Informationen werden in den folgenden beiden Abschnitten 4.6.1 und 4.6.2 anhand entsprechender Beispiele im Detail beschrieben.

4.6.1 Transparenz Datenstruktur

Das in Abbildung 4.17 dargestellte interaktive Streudiagramm zeigt alle in den Medikationsverordnungen vorkommenden 739 ATC Codes. Auf der x-Achse wird die Anzahl der Medikationsverordnungen pro ATC Code auf einer logarithmischen Skala dargestellt. Je weiter links ein Punkt auf der x-Achse liegt, desto weniger häufig kommt er in den Medikationsverordnungen vor. Die y-Achse bildet den Anteil der unstrukturierten Daten pro ATC Code ab.

Liegen die Medikationsverordnungen ausschließlich strukturiert vor, befinden sie sich am unteren Ende der y-Achse bei 0%. ATC Codes, die ausschließlich unstrukturiert vorliegen, befinden sich bei 100% auf der y-Achse.

Bei der gestrichelten grauen Linie handelt es sich um den Durchschnittswert von 52,27 % unstrukturiert vorliegenden Medikationsverordnungen des Datensatzes DS-Med gemäß der in Kapitel 4.3.2 identifizierten Inhibitoren durch die systematische Datenanalyse der Medikationsverordnungen am UKD.

Die Farbkodierung sieht eine rote Markierung für ATC Codes vor, wenn deren Strukturiertheitsgrad unter dem Mittelwert von 52,27% liegt. Alle ATC Codes mit einem höheren Grad an Strukturiertheit sind blau dargestellt.

In Abbildung 4.17 wurde der Mauszeiger über den ATC N03AX14 bewegt. Es werden als Metainformationen der ATC Code, die Anzahl der Verordnungen mit diesem ATC Code, der Anteil der unstrukturiert vorliegenden Medikationsverordnungen mit diesem ATC Code sowie der Wirkstoffname angezeigt. In der Tabelle oben rechts werden anhand des gefilterten Textes „N03A“ alle in den Medikationsverordnungen vorhandenen ATC Codes, die mit diesem Text beginnen, angezeigt.

Je weiter links und oben ein Punkt in dem Streudiagramm liegt, desto größer ist sein Einfluss auf die Unstrukturiertheit der Medikationsverordnungen insgesamt. Durch die Nutzung der oben erwähnten Feedback-Schleife können anhand des Streudiagramms sehr unstrukturiert und häufig vorkommende ATC Codes im Detail diskutiert werden.

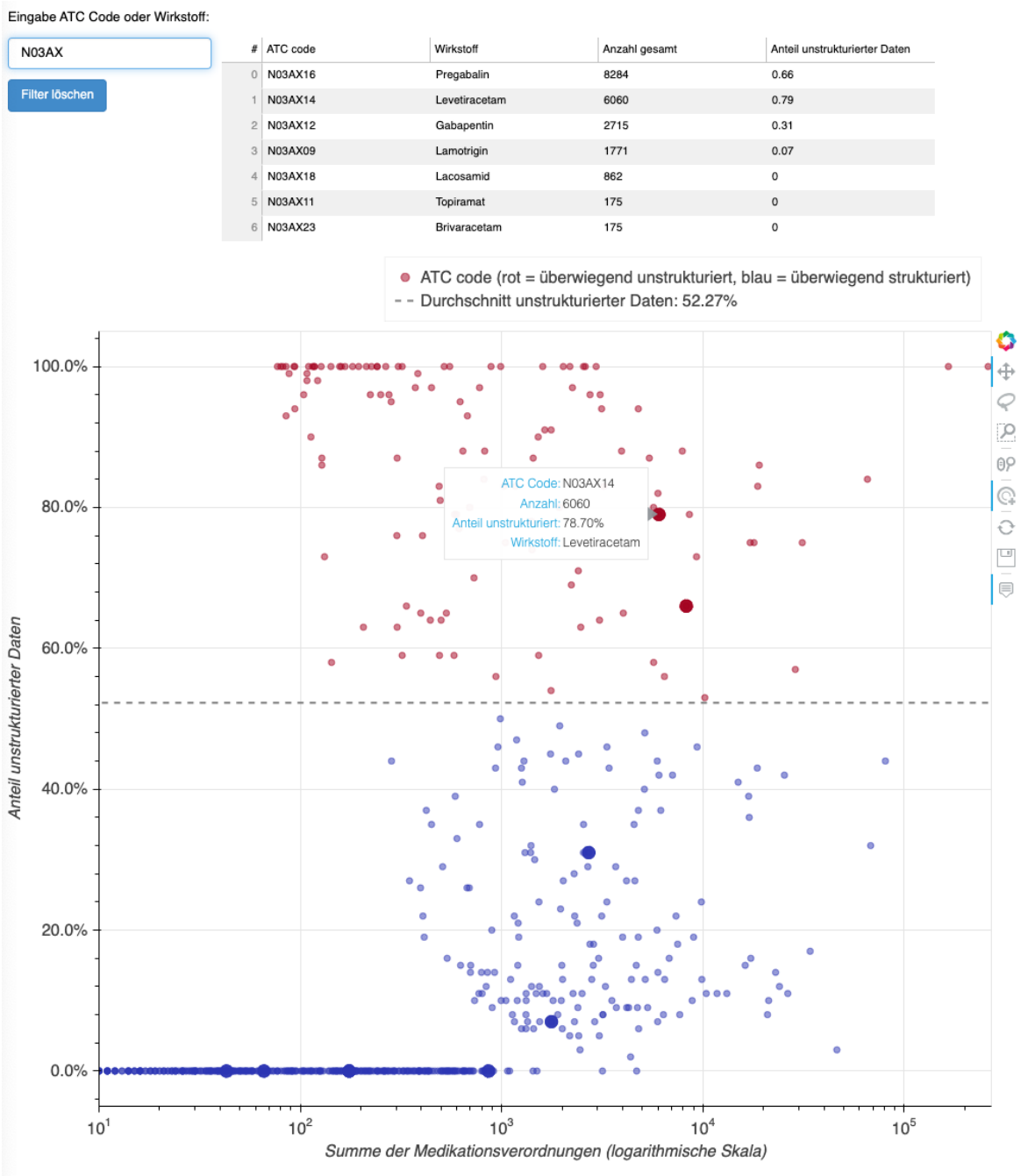


Abbildung 4.17: Interaktive Visualisierung der Strukturiertheit der Medikationsverordnungen pro ATC Code

4.6.2 Transparenz Terminologie

Das in Abbildung 4.18 dargestellte zweite interaktive Streudiagramm erlaubt, alle in den Medikationsverordnungen vorkommenden 739 ATC Codes anzuzeigen.

Die x-Achse gibt die Anzahl der Medikationsverordnungen pro ATC Code auf einer logarithmischen Skala an. Die y-Achse bildet die Anzahl der existierenden Mappings nach RxNorm ebenfalls auf einer logarithmischen Skala ab. Dabei wurde die Anzahl der 4 verschiedenen

Verbindungstypen aus Abbildung 4.13 „ATC -RxNorm pr lat“, „ATC -RxNorm sec lat“, „ATC -RxNorm pr up“ und „ATC -RxNorm sec up“ kumuliert und auf der y-Achse abgebildet.

Auch hier ist die Darstellung der ATC Codes farblich kodiert. Alle blau dargestellten ATC Codes verfügen über mindestens ein Mapping vom Typ „Maps to“, alle rot dargestellten ATC Codes verfügen über kein Mapping vom Typ „Maps to“ und gehören damit zum Anteil der Medikationsverordnungen, die in Abschnitt 4.4.3.2 nicht nach RxNorm überführt werden konnten.

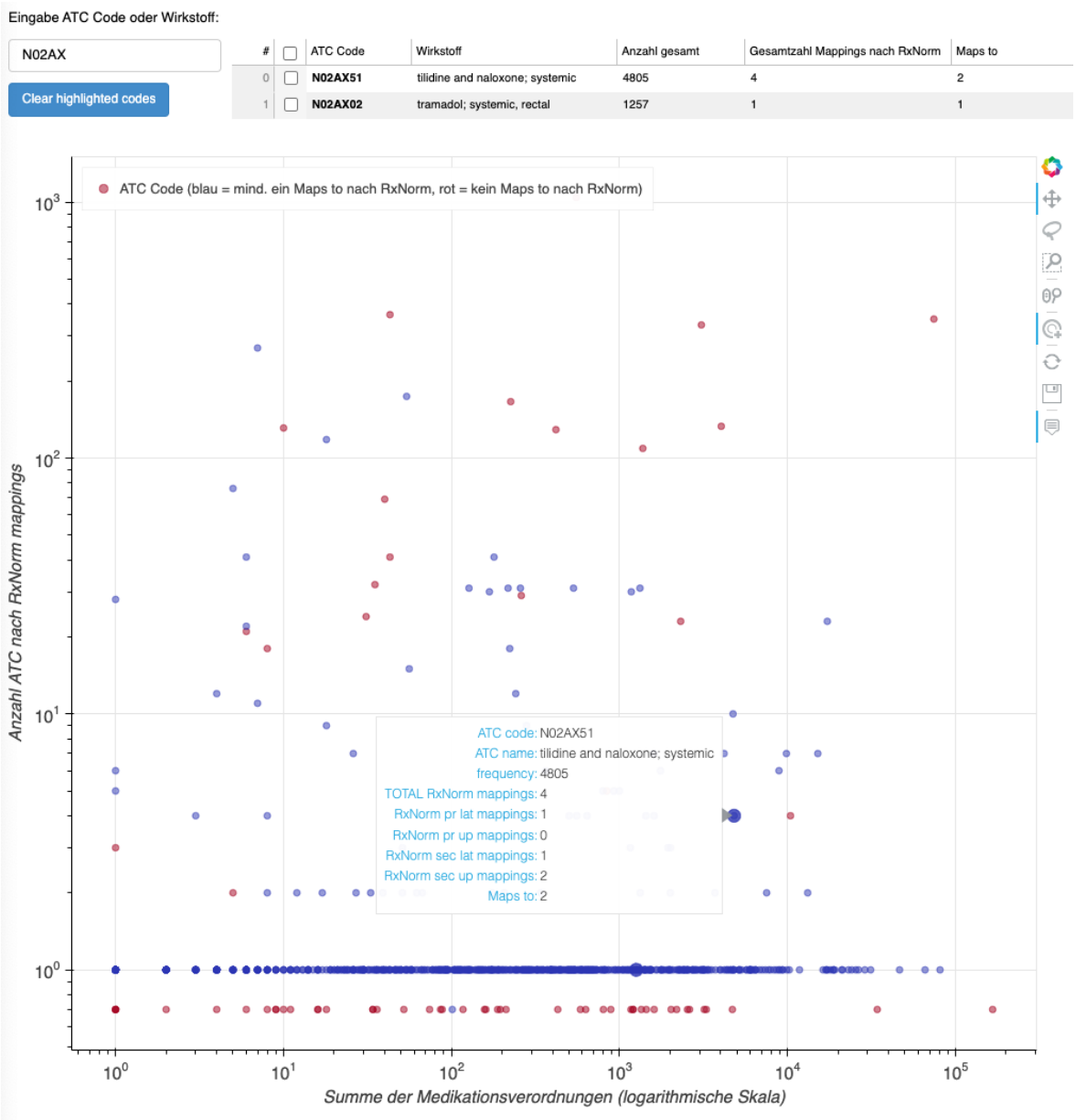


Abbildung 4.18: Interaktive Visualisierung des semantischen Mappings von ATC nach RxNorm

Durch die Bewegung mit dem Mauszeiger über die als Punkte dargestellten ATC Codes werden auch hier zusätzliche Metainformationen angezeigt. In Abbildung 4.18 sind die Me-

tainformationen exemplarisch für den ATC Code N02AX51 *Tilidin und Naloxon* dargestellt. Der ATC Code kommt in den Medikationsverordnungen 4805-mal vor, konnte nach RxNorm in zwei explizite Wirkstoffe durch die beiden „Maps to“ Verbindungen überführt werden. Die Suche nach „N02AX“ resultierte in einer Ergebnismenge von den zwei ATC Codes N02AX51 und N02AX02 in der Tabelle. Beide ATC Codes wurden in dem Streudiagramm als vergrößerte Punkte dargestellt. Damit ist sichtbar, ob ein ATC Code nach RxNorm überführt werden konnte oder nicht und wie häufig der jeweilige ATC Codes in den Medikationsverordnungen verwendet wurde.

Die ATC Codes ohne ein Mapping nach RxNorm werden der Vollständigkeit und Transparenz halber ebenfalls in dem Streudiagramm dargestellt. Es handelt sich dabei um die Punkte mit einem Wert von kleiner als eins auf der y-Achse.

4.6.3 Zusammenfassung der Ergebnisse zur Transparenz

Für die Schaffung der Transparenz der Datenstruktur auf Basis der 739 in den Medikationsverordnungen des DS-Med verwendeten ATC Codes, wurde ein interaktives Streudiagramm generiert, welches die Strukturiertheit für jeden einzelnen ATC Code visualisiert. Darüber hinaus wird die Häufigkeit des Vorkommens jedes ATC Codes im Streudiagramm dargestellt und liefert so ergänzende Informationen zur Relevanz der Strukturiertheit für jeden der abgebildeten ATC Codes.

Das Streudiagramm für die Datenstruktur pro ATC Code visualisiert zudem starke Abweichungen einzelner Codes im Kontext der Strukturiertheit, aber auch in der Häufigkeit des Vorkommens.

Ein zweites Streudiagramm wurde für die Visualisierung des semantischen Mappings von ATC nach RxNorm erstellt. Es zeigt für jeden ATC Code, ob ein Mapping für einen ATC Code vollständig, nur als Teilmenge oder gar nicht möglich war. Es zeigt ebenfalls die Häufigkeit des Vorkommens des ATC Codes in den Daten.

5 Diskussion

5.1 Allgemein

Die Verwendung von RWD eröffnet vielfältige Forschungsmöglichkeiten und kann randomisierte, kontrollierte Studien ergänzen, da Gesundheitsdaten aufgrund der Digitalisierung zunehmend elektronisch verfügbar sind und nach entsprechender Harmonisierung für die Forschung zum Einsatz kommen können. Besonders der Bereich der Pharmakovigilanz kann von großen retrospektiven Beobachtungsstudien basierend auf RWD profitieren und frühzeitig Auffälligkeiten oder unerwünschte Nebenwirkungen erkennen. Dies zeigt sich in diversen Aktivitäten weltweit, so zum Beispiel im Rahmen der OHDSI Forschungsgemeinschaft, die über die letzten Jahre für die Forschung zunehmend an Bedeutung gewinnt. Es wurden bereits eine Vielzahl an retrospektiven Beobachtungsstudien auf Basis von OMOP durchgeführt, bisher allerdings ohne die Beteiligung mit Daten aus der stationären Versorgung von deutschen Universitätskliniken.

Ziel der Arbeit war es zu erforschen, wie die Sekundärnutzung von deutschen Medikationsdaten aus der klinischen Versorgung in retrospektiven Beobachtungsstudien in internationalen Forschungsgemeinschaften am Beispiel von OHDSI unter Wahrung der semantischen Bedeutung ermöglicht werden kann.

Um beantworten zu können, wo die Schwerpunkte der weltweiten Nutzung von OMOP in der Forschung liegen (Forschungsfrage 1), wurde zunächst eine Literaturrecherche in Form eines Scoping Reviews durchgeführt und durch die Autorin der vorliegenden Arbeit veröffentlicht (Reinecke, Zoch, Reich, Sedlmayr et al., 2021). Es zeigt den aktuellen Forschungsstand

zur Nutzung des OMOP. Im Resultat wird ein deutlich steigendes Forschungsinteresse an OMOP und an der Forschungsgemeinschaft OHDSI über die vergangenen fünf Jahre anhand der stetig wachsenden Zahl der Publikationen dargelegt. Besonders auffällig ist eine deutliche Zunahme der Publikationen in medizinischen Journalen im Jahr 2020. Bei der Mehrheit der Publikationen in medizinischen Journalen, handelt es sich um retrospektive Beobachtungsstudien zur Beantwortung medizinischer Fragestellungen, bei denen OMOP als Grundlage für die Datenablage genutzt wird. Im Vergleich zu allen anderen Kategorien (vgl. Abbildung 3.3 auf Seite 30) werden für diese Studien sehr viel häufiger RWD aus mehreren Standorten, zumeist auch aus unterschiedlichen Ländern genutzt. Diese Nutzung von OMOP steht im Einklang mit den Zielen der Forschungsgemeinschaft OHDSI, die darauf abzielt, Studien mit großen Datenmengen über verschiedene Standorte hinweg durchzuführen. Die Literatur zeigt eine aktive Teilhabe von Forschenden deutscher Universitäten im Kontext von OHDSI und OMOP, jedoch beschränkt auf die folgenden Themen: Trends in Initiativen im Bereich der Forschung mit RWD unter Verwendung von OMOP (Prokosch, Acker et al., 2018; Tresp et al., 2016), Datentransfer (Maier et al., 2018), Entwicklung von Software basierend auf OMOP (Freitas Da Cruz et al., 2019; Gruendner et al., 2019; H. Spengler et al., 2020; Unberath et al., 2020) und Entwicklung von Konzepten zur Nutzung von OMOP (Fischer et al., 2020; Gruhl et al., 2020; Reinecke, Gulden et al., 2020).

Der Transfer von RWD aus Deutschland unter Wahrung der semantischen Bedeutung konzentriert sich bisher auf demografische Daten von Patient:innen, Diagnosen, Prozeduren und Laborwerten (Maier et al., 2018). Auch die aktuelle Publikation von Peng et al. (Peng et al., 2023), welche einen ETL Prozess für den Transfer von RWD nach OMOP basierend auf dem MI-I KDS bereitstellt, adressiert das semantische Mapping von Medikationsdaten nicht, sondern fokussiert vielmehr auf den syntaktisch korrekten Transfer der Daten nach OMOP.

Ein inhaltlicher Fokus auf Medikationsdaten sowie Forschungsarbeiten zur Sicherung der semantischen Bedeutung bei der Harmonisierung von Medikationsdaten nach OMOP fehlten bisher. Eine Teilnahme an OHDSI Netzwerkstudien mit Daten deutscher Universitätskliniken existierte zum Zeitpunkt des Reviews nicht, obwohl nationale Aktivitäten im Kontext der MI-I Förderrichtlinie die Schaffung von Infrastrukturen zur Bereitstellung von Daten aus der klinischen Versorgung für Forschende auch im internationalen Kontext anstreben (Bild et al., 2020; Gehring et al., 2018).

Die **Forschungsfrage 2** befasst sich mit den Anforderungen an RWD bei der Speicherung in OMOP als notwendige Grundlage für die Durchführung von retrospektiven Beobachtungsstudien innerhalb der OHDSI Forschungsgemeinschaft. Dazu wurden in dieser Arbeit die anhand der Literaturrecherche identifizierten retrospektiven Beobachtungsstudien im Hinblick auf die erforderlichen Datengruppen „Diagnosen", „Medikamente", „Laborwerte", „Prozeduren", „Beobachtungen" und „medizinische Scores" untersucht. Die überprüften Datengruppen sind angelehnt an die Basismodule des MI-I KDS. Des Weiteren wurden die syntaktischen und semantischen Anforderungen gemäß der Spezifikation von OMOP ermittelt. Anhand der untersuchten 28 retrospektiven Beobachtungsstudien wurde festgestellt, dass Medikationsdaten eine wichtige Voraussetzung darstellen, weil sie in 22 der 28 Studien als Einschlusskriterien als erforderliche Daten definiert waren (Amutha et al., 2021; Brauer et al., 2020; Burn, You et al., 2020; Chandler, 2020; Chen et al., 2020; Choi et al., 2020; Duke et al., 2017; Hripcsak, Ryan et al., 2016; Hripcsak, Suchard et al., 2020; H. Kim et al., 2020; Y. Kim et al., 2020; Kubota et al., 2018; Lane, Kostka et al., 2020; Lane, Weaver et al., 2021; Morales et al., 2021; Samwald et al., 2016; Seo et al., 2020; Spotnitz et al., 2020; Vashisht et al., 2018; Viernes et al., 2019; You, Rho et al., 2020; Zhang et al., 2018).

Für die Speicherung von klinischen Versorgungsdaten in OMOP ist die Strukturiertheit der Daten die Grundvoraussetzung. Daten werden in OMOP als klinische Fakten in unterschiedlichen Tabellen abgelegt. Jeder klinischer Fakt muss dabei einem validen Konzept einer standardisierten medizinischen Terminologie entsprechen. Für Medikationsdaten in OMOP gilt die Terminologie RxNorm als zu verwendender Standard. Als Minimalanforderung ist hier die Angabe des aktiven Wirkstoffes als RxNorm Code gefordert.

Die **Forschungsfrage 3** zielt darauf ab, existierende Inhibitoren in den Medikationsdaten aus der klinischen Versorgung des UKD zu identifizieren, die im Widerspruch zu den Anforderungen stehen. Dazu wurde zunächst eine Stichprobenanalyse der Datenqualität an den Standorten des MIRACUM Konsortiums für die auch im Rahmen der Forschungsfrage 2 verwendeten Datengruppen durchgeführt und veröffentlicht (Vass et al., 2022). Außerdem erfolgte erstmals eine umfassende und systematische Analyse der Strukturiertheit der Medikationsverordnungen des UKD für 1.768.153 Datensätze der Jahre 2016 bis 2020, die ebenfalls veröffentlicht wurde (Reinecke, Siebel et al., 2023). Beide Arbeiten zeigen vergleichbare Ergebnisse für die Strukturiertheit der Medikationsdaten. Die Stichprobenanalyse identifizierte eine sehr heterogene Verfügbarkeit von strukturierten Medikationsdaten

an allen betrachteten Standorten mit einem Strukturiertheitsgrad von weniger als 50% im Durchschnitt. Die Medikationsverordnungen des UKD weisen eine Strukturiertheit von 47,27% auf. Außerdem zeigt die initiale Analyse, dass die Medikationsverordnungen, wenn strukturiert, nicht unter der Verwendung der geforderten Terminologie RxNorm vorliegen, sondern lediglich über einen ATC Code verfügen. Es wurden daher die Datenstruktur und die Terminologie der Medikationsverordnungen als Inhibitoren identifiziert.

Um Forschungsfrage 4 zu beantworten, wurden mögliche Maßnahmen entwickelt und auf die Daten angewendet, um die identifizierten Inhibitoren abzubauen und den Anforderungen gerecht zu werden. Dadurch konnte die Diskrepanz zwischen den Inhibitoren und den Anforderungen reduziert werden, um eine Befähigung zur Teilnahme an internationalen Studien auf Basis von OMOP zu ermöglichen. Die Strukturiertheit der Medikationsverordnungen konnte durch die entwickelten Algorithmen, deren Ergebnisgüte einer Validierung unterlag, von 47,27 % auf 85,18% (1.506.059/1.768.153) erhöht werden (Reinecke, Siebel et al., 2023). Die Überführung der strukturierten Medikationsverordnungen in die, durch das OMOP CDM, geforderte Terminologie RxNorm war für 66,39% (1.182.974/1.768.153) der initialen Datenmenge möglich. Bezogen auf 85,15% der strukturiert vorliegenden Medikationsverordnungen entspricht das 78,55% (1.182.974/1.506.059). Im Vergleich dazu konnte eine Verwendung der Medikationsdaten initial in retrospektiven Beobachtungsstudien auf Basis des OMOP CDM ohne die Überführung nach RxNorm gar nicht stattfinden, da 0 % der Daten in der Terminologie RxNorm vor der Durchführung der Maßnahmen, insbesondere der Überführung nach RxNorm vorlagen. Die durchgeführten Maßnahmen wurden zudem als Konzept von der Autorin der vorliegenden Arbeit veröffentlicht (Reinecke, Zoch, Wilhelm et al., 2021). Eine initiale Untersuchung der Machbarkeit der Überführung von ATC Codes nach RxNorm wurde ebenfalls im Kontext der vorliegenden Arbeit von der Autorin veröffentlicht (Reinecke, Henke et al., 2023).

Die Forschungsfrage 5 setzte sich mit der Bewertung der im Rahmen von Forschungsfrage 4 durchgeführten Maßnahmen auseinander. Es wurde eine umfassende qualitative und quantitative Bewertung durchgeführt. Es konnte so im Rahmen der qualitativen Ansicht gezeigt werden, dass an einer bereits durchgeführten Studie von Duke et al. (Duke et al., 2017) der OHDSI Forschungsgemeinschaft zunächst nicht teilgenommen werden konnte, weil die dafür notwendigen Medikationsdaten mangels fehlender Struktur, also ohne korrekt zugeordneten RxNorm Code, nicht den Anforderungen entsprachen. Nach

der Durchführung der korrektiven Maßnahmen zeigt sich hier in der Datenstruktur eine deutliche Verbesserung. Außerdem konnten die für die Studie relevanten Wirkstoffe in den Medikationsverordnungen des UKD vollständig nach RxNorm überführt werden, sodass eine Teilnahme an einer vergleichbaren Studie möglich wird. Mit der quantitativen Bewertung erfolgte eine Prüfung der untersuchten Medikationsverordnungen in OMOP hinsichtlich der Anforderungen seitens der Spezifikation des Datenmodells unter Verwendung des OHDSI DQD. Die Prüfung erfolgte mit dem Fokus auf Konformität und Vollständigkeit der Daten gemäß des Frameworks zur Prüfung und Sicherung der Datenqualität nach Kahn et al (Kahn et al., 2016).

Die quantitative Bewertung durch das OHDSI DQD erfolgte dreimal, (1) angewandt auf die ursprünglichen Medikationsverordnungen, (2) nach Durchführung der Maßnahmen zur Verbesserung der Datenstruktur und abschließend (3) nach der Überführung der Medikationsverordnungen nach RxNorm. Die quantitative Analyse verdeutlicht, dass die anfänglichen Medikationsverordnungen vor der Durchführung der korrektiven Maßnahmen nicht den Anforderungen an die erforderliche Terminologie RxNorm hinsichtlich ihrer Vollständigkeit genügen. Erst nach der Überführung der Medikationsverordnungen des UKD verringert sich dieser Anteil auf nur noch 33,61%, der nicht den geforderten Terminologiestandards von RxNorm entspricht.

Die Forschungsfrage 6, die sich mit der Gewährleistung der notwendigen Transparenz möglicher verbleibender Limitierungen in den Medikationsverordnungen befasst, wurde in dieser Arbeit ebenfalls beantwortet. Da die Medikationsverordnungen primär nicht für die Forschung erfasst wurden, war es erwartbar, dass die Daten nicht in ausreichender Güte für das geforderte strukturelle und semantische Format vorliegen. Mit der in Kapitel 4.6 vorgestellten interaktiven Visualisierung und entsprechenden Such- und Filtermöglichkeiten konnte die Strukturiertheit der Medikationsverordnungen pro ATC Code in Abhängigkeit der Häufigkeit des Vorkommens in den Daten interaktiv dargestellt werden. Zudem wird auf konzeptioneller Ebene ein Feedback-Mechanismus geschaffen, um Gründe für besonders unstrukturierte Medikationsverordnungen gemeinsam mit Teams aus der klinischen Versorgung zu erörtern. Diese Interaktion stellt eine wichtige Voraussetzung dar, um die Datenstruktur zum Zeitpunkt der Dokumentation potenziell positiv zu beeinflussen und aufwändige Nacharbeiten zu minimieren. Außerdem konnte die Überführung der Medikationsverordnungen vom ATC Code nach RxNorm ebenfalls interaktiv visualisiert werden. So

können Forschende oder das DIZ im Bedarfsfall prüfen, welche ATC Codes einen einzelnen expliziten Wirkstoff benennen, ob es sich um Kombinationen von mehreren expliziten Wirkstoffen oder um einen oder mehrere nicht-explizite Wirkstoffe handelt. Diese Visualisierung bietet den Nutzenden die Möglichkeit zu überprüfen, ob eine vollständige, teilweise oder keine Überführung in die Terminologie RxNorm möglich ist. Das Konzept der Etablierung einer Feedback-Schleife zur Schaffung von Transparenz wurde von der Autorin dieser Arbeit ebenfalls veröffentlicht (Reinecke, Bathelt et al., 2022).

Die Beantwortung aller der in dieser Arbeit gestellten Forschungsfragen schafft die Voraussetzung, um an retrospektiven Beobachtungsstudien der OHDSI Forschungsgemeinschaft teilzunehmen. Die semantische Bedeutung der Medikationsverordnungen, auch unter Verwendung internationaler Terminologien wie RxNorm, bleibt dabei gewahrt. Zusätzliche Transparenz kann Forschenden und Versorgenden helfen, die Datenqualität im Sinne der Strukturiertheit der Medikationsverordnungen am UKD weiter zu verbessern.

5.2 Stärken

Internationale Nutzung von OMOP

Die vorliegende Arbeit leistet einen wichtigen Beitrag zur Einordnung des aktuellen Forschungsstandes im Bereich von OHDSI und OMOP durch die Durchführung eines eigenständigen Scoping Reviews (Reinecke, 2021a). Bisher fehlte eine systematische Literaturrecherche in Form eines Reviews zu diesem Thema.

Das durchgeführte Scoping Review schafft daher eine wichtige Grundlage für Forschungsteams weltweit, um einen umfassenden Überblick über die Verwendung von OMOP zu erhalten. Es bietet einen chronologischen Überblick und zeigt die internationale Entwicklungstendenz auf.

Insbesondere werden die Anwendungsfelder von OMOP beleuchtet, wobei ein besonderer Fokus auf den Arbeiten der deutschen Universitätsteams liegt und ihre fehlende Beteiligung an Studien der OHDSI-Forschungsgemeinschaft deutlich wird. Dadurch eröffnet die vorliegende Arbeit Möglichkeiten zur Identifizierung von Potenzialen für zukünftige Forschung im Zusammenhang mit OHDSI und OMOP.

Anforderungen an Daten für Studien auf Basis von OMOP

Auch wenn die Spezifikation von OMOP ausreichend dokumentiert vorliegt, um die Anforderungen an die medizinischen Daten zur Nutzung in OMOP zu ermitteln, fehlte bisher eine Analyse der Relevanz der Datengruppen in Anlehnung an den in Deutschland standardisierten MI-I KDS.

Die vorliegende Arbeit schafft dank der durchgeführten Analyse der retrospektiven OHDSI Netzwerkstudien im Hinblick auf die Häufigkeit der verwendeten Datengruppen erstmals einen Überblick über deren Relevanz in bereits durchgeführten Studien als Einschlusskriterien. Sie zeigt zudem die Wichtigkeit der Verfügbarkeit von Medikationsdaten unter Verwendung international nutzbarer Terminologien wie RxNorm, um die Teilnahme an internationalen Studien zu ermöglichen.

Datenqualität: Analyse und Verbesserung

Die Herausforderungen bei der Verwendung von RWD aufgrund von Defiziten in der Datenqualität, wie Ungenauigkeiten durch unpräzise Informationen oder nicht standardisierte Benennung sowie Unvollständigkeiten aufgrund fehlender Kodierung, wurden bereits in Arbeiten von Hersh et al. und Botsis et al. als Hemmnisse für die Sekundärnutzung identifiziert (Botsis et al., 2010; Hersh et al., 2013).

Allerdings beschränken sich Arbeiten bei der Analyse von RWD auf Diagnosen und Prozeduren (Von Lucadou et al., 2019). Auch in dem systematischen Review von Otero et al., welches Interventionen zu Verbesserungen von RWD hinsichtlich deren Effektivität und Messbarkeit der Verbesserung untersucht, zeigt sich, dass Medikationsdaten in keiner der 24 eingeschlossenen Publikationen betrachtet und untersucht wurden (Otero Varela et al., 2019).

Daher schafft die im Rahmen dieser Arbeit durchgeführte ganzheitliche Analyse der Strukturiertheit von Medikationsverordnungen einen wertvollen Beitrag zur Identifikation möglicher Inhibitoren, welche der Sekundärnutzung der Daten entgegenstehen. Andere Arbeiten wie beispielsweise von Yang et al. (Yang et al., 2018) haben ebenfalls die Datenqualität von Datenbanken mit elektronischen Gesundheitsdaten untersucht, beschränken sich aber in der Auswertung lediglich auf generische Aussagen, wie der Nutzung nationaler Terminologien

oder Freitexte, ohne eine umfassende Analyse auch auf Ebene der Wirkstoffe durchzuführen und auszuwerten.

Systematische Analysen der Strukturiertheit von Medikationsdaten aus der stationären Versorgungen fehlen auch in Deutschland, sie beschränken sich bislang auf die Auswertung und Nutzung von Diagnosen und Prozeduren (Maier et al., 2018; Von Lucadou et al., 2019). Daher schließt die im Rahmen dieser Arbeit durchgeführte ganzheitliche Analyse der Medikationsverordnungen des UKD auf Basis der ATC Codes eine bisher existierende Lücke.

Die Erhöhung der Datenqualität von Medikationsverordnungen zur Sekundärnutzung in der Forschung im Hinblick auf deren Vollständigkeit durch die Erhöhung des Anteils an verfügbaren Daten mit ATC Codes und einer Überführung nach RxNorm, stellt eine große Stärke dieser Arbeit dar. Dies ist vor allem deshalb relevant, weil die Vollständigkeit von RWD ein besonders häufig einschränkendes Merkmal bei der Sekundärnutzung darstellt (Kohane et al., 2021; Priou et al., 2023).

In Frankreich wurde die Système National des Données de Santé (SNDS) Datenbank von Lamer et al. (Lamer et al., 2020) nach OMOP überführt. Im Vergleich zu dieser Arbeit beantwortet die vorliegende Dissertation die Frage nach den verwendeten Beziehungstypen für die Mappings zwischen ATC und RxNorm, was eine wichtige Voraussetzung für die Nutzung der Ergebnisse in anderen Arbeiten darstellt.

Paris et al. (Paris et al., 2021) berichteten über die Machbarkeit des Mappings von intensivmedizinischen Daten aus dem Medical Information Mart for Intensive Care (MIMIC) Datensatz nach OMOP. In der vorliegenden Dissertation können im Vergleich zu Paris et al. ein wesentlich höherer Anteil von Medikationsverordnungen von ATC nach RxNorm überführt werden. Zusätzlich werden Informationen zur Menge der betreffenden Medikationsverordnungen und zu den nicht überführbaren ATC Codes dargestellt.

In Brasilien werden ebenfalls Bestrebungen unternommen, Gesundheitsdaten nach OMOP zu überführen, wobei die Medikationsdaten über eigene Vokabulare in OMOP abgebildet und nur falls möglich nach RxNorm überführt werden (Lima et al., 2019). Allerdings fehlt in der Arbeit von Lima et al. eine transparente Darstellung der verwendeten Mappings und des Abdeckungsgrades der überführten Medikationsdaten.

You et al. (You, Lee et al., 2017) überführten eine nationale Datenbank aus Südkorea mit Informationen zu medizinischen Leistungen, Verschreibungen von Medikamenten, Diagnosen und weiteren Gesundheitsdaten nach OMOP. Im Vergleich zu dieser Arbeit stellt die vorliegende Dissertation vor allem in Hinblick auf die Nachvollziehbarkeit der Mappings und den Abdeckungsgrad der verwendeten Medikationsverordnungen eine Stärke dar, da eindeutige und transparente Rückschlüsse auf mögliche Grenzen der Mappings auf Ebene der ATC Codes und der Häufigkeit der verwendeten ATC Codes in den Medikationsverordnungen möglich sind.

Andere Arbeiten adressieren das Thema der Datenqualität in OMOP bereits (Blacketer, Voss et al., 2021; H. Spengler et al., 2020; Yoon et al., 2016). Das Paper von Yoon et al. prüft die nach OMOP transferierten Daten einmalig und nutzt in ATLAS mitgelieferte Prüfungen der Datenqualität. Diese Bordmittel sind jedoch im Vergleich zum OHDSI DQD limitiert in der Anzahl und der Organisation der Prüfungen (Blacketer, Defalco et al., 2021). Das Paper von Blacketer et al. (Blacketer, Voss et al., 2021) nutzt, wie auch die vorliegende Dissertation, einen iterativen Prozess zur Überprüfung der Datenqualität in OMOP basierend auf dem OHDSI DQD. Allerdings beschränkt sich die Auswertung auf eine quantitative Bewertung der erfolgreichen oder fehlgeschlagenen Prüfungen. Im Gegensatz dazu wurde in der vorliegenden Dissertation eine Methode präsentiert, um die Datenqualität auch auf der Ebene der Prüftypen einzeln und inhaltlich zu bewerten und die Ergebnisse des DQD auch in Relation zu den Anforderungen im Hinblick auf die semantische Bedeutung und Nutzung der geforderten Terminologien auszuwerten. Diese Granularität der Nutzung des OHDSI DQD und der Auswertung der Ergebnisse wird erstmals durch die vorliegende Arbeit erreicht.

Schaffung von Transparenz

Eine weitere Stärke dieser Dissertation zeigt sich in der geschaffenen Transparenz in (a) der vorliegenden Struktur der Medikationsverordnungen auf Basis der Wirkstoffe als ATC Codes und (b) zum anderen in dem Abdeckungsgrad der Überführbarkeit von ATC nach RxNorm. Zusätzlich zur bereits betrachteten Datenqualität mit dem OHDSI DQD, schaffen die generierten Streudiagramme eine ideale Voraussetzung für die Untersuchung der Medikationsverordnungen mit Fokus auf deren Strukturiertheit und Überführbarkeit nach RxNorm.

Aufgrund der Darstellung der einzelnen ATC Codes und der Abbildung der Mengen, ist besonders für unstrukturiert vorliegende Daten der Einfluss auf die Strukturiertheit aller Medikationsverordnungen sofort erkennbar. Diese Information ist eine wichtige Voraussetzung, um in Zusammenarbeit mit den Teams der klinischen Versorgung Gründe und Einflüsse auf die Datenstruktur zu erörtern und entsprechende Aktivitäten zu erarbeiten, um zukünftig die Datenstruktur bereits während der Dokumentation zu verbessern. Die in Kapitel 4.6 geschaffene Visualisierung ermöglicht zudem Aussagen zur Überführbarkeit von Medikationsverordnungen von ATC nach RxNorm. Mögliche Limitierungen und Grenzen der umgesetzten Maßnahmen zur Verbesserung der Datenstruktur und Überführung in die geforderte Terminologie werden durch die Streudiagramme transparent dargestellt.

Insgesamt dient die durch die Streudiagramme geschaffene Transparenz der Förderung der Vertrauenswürdigkeit und Qualität der RWD. Auch die Forschungsergebnisse können von der Transparenz profitieren, da sie dokumentierbar und nachvollziehbar werden. So wird die Schaffung fundierter und nachvollziehbarer Ergebnisse zum Vorteil von Patient:innen unterstützt und informiertere Entscheidungen im Gesundheitswesen werden möglich.

5.3 Limitierungen

Neben den genannten Stärken dieser Arbeit existieren auch Limitierungen, die in diesem Abschnitt diskutiert werden.

Bisher existiert in Deutschland kein nationaler Terminologieserver, wie vergleichsweise der National Clinical Terminology Service (NCTS) in Australien („National Clinical Terminology Service (NCTS) | Digital Health Developer Portal“, 2023), der landesweit genutzte medizinische Terminologien versioniert und für mehrere Jahre inklusive der Gültigkeitszeiträume bereitstellt. Das BfArM stellt zwar die auf Deutschland angepassten Versionen der WHO für ICD10 und ATC sowie OPS zum Download zur Verfügung, allerdings sind diese Daten auf die Nutzung im aktuellen Jahr für die Dokumentation abrechnungsrelevanter Informationen ausgelegt. Eine Historisierung über mehrere Jahre, die eine Nutzung für RWD aus der Vergangenheit unter Sicherstellung der korrekten Versionierung von Codes ermöglicht, wird hiermit nicht gewährleistet. Insbesondere im Hinblick auf die Verwendung von Medikationsdaten auf Basis von Wirkstoffen aus der Vergangenheit fehlt eine durchgängige Historisierung der

ATC-GM Version in einem maschinenlesbaren Format. Das BfArM stellt bisher jedes Jahr nur eine PDF-Version von ATC-GM zur Verfügung. Das WIdO stellt zwar jährlich eine ATC-GM Version in Excel bereit, jedoch zeitverzögert und ohne Informationen zu den Überleitungen von ATC Codes aus der Vorjahresversion. Die durchgeführte Übersetzung von ATC-GM nach ATC der WHO Version ist auf die Medikationsverordnungen des Datensatzes DS-Med des UKD und des am UKD verwendeten Hauskatalogs aus dem Datensatz DS-Katalog beschränkt. Eine Übertragung der in Kapitel 3.5.3.1 beschriebenen Methodik zur Übersetzung aller ATC Codes, die in der nationalen Version ATC-GM von der ATC WHO Version abweichen, konnte wegen des beschriebenen Mangels in dieser Arbeit nicht durchgeführt werden.

Die Verbesserung der Strukturiertheit der Medikationsverordnungen wurde bisher nur für 85,18% aller Medikationsverordnungen durch eine manuelle Validierung geprüft. Da die Validierung der Ergebnisse der Algorithmen gemäß der in Kapitel 3.5.2.2 für die am häufigsten verwendeten Freitexte manuell durchgeführt wurde, steht der Aufwand für die Abdeckung der verbleibenden 14,82 % aufgrund der enorm hohen Anzahl an Freitexten bei einem manuellen Vorgang nicht mehr im vertretbaren Verhältnis zum Ergebnis. Die verbleibenden Medikationsverordnungen wurden jedoch unter Wahrung der originalen Freitexte, allerdings ohne Zuordnung zu einem validen Konzept einer Standardterminologie, in OMOP gesichert. Dies ermöglicht Forschenden die Nutzung der Freitexte zur eigenen Suche und Zuordnung valider Konzepte in OMOP.

Die Überführung der Medikationsverordnungen mit einem ATC Code nach RxNorm ist derzeit auf die Ebene der aktiven Wirkstoffe begrenzt. In dieser Arbeit wurden die RxNorm TTY „Klinische Wirkstoffkomponente“ (Wirkstoff und Stärke), „Klinische Form des Medikaments“ (Wirkstoff und Darreichungsform) sowie „Klinisches Medikament“ (Wirkstoff, Stärke und Darreichungsform) nicht betrachtet. Eine Überführung auf Ebene der Medikamente ist also noch offen. Dazu ist es allerdings notwendig, Medikamenteninformationen in kodierter Form, beispielsweise über die Pharmazentralnummer (PZN) in den Medikationsverordnungen aus dem KIS des UKD zu erhalten. Dies kann aktuell seitens KIS jedoch nicht gewährleistet werden. Auch das Integrated Care Manager (ICM) System des Herstellers Dräger, der für die Dokumentation der intensivmedizinischen Behandlungsfälle inklusive der Dokumentation der Medikationsverordnungen verwendet wird, stellt ausschließlich den ATC Code für Medikamente bereit. Ebenfalls offen bleibt eine vollständige Überführung von ATC Codes, die mehrere Wirkstoffe beinhalten und bei denen Wirkstoffe generisch zusammenfasst werden,

sodass eine explizite Überführung einzelner Wirkstoffe nach RxNorm nicht möglich ist. Diese Limitierung schränkt allerdings die Nutzung der Daten innerhalb der OHDSI Forschungsgemeinschaft nicht ein, da die Mehrheit der bisher durchgeführten Studien auf OMOP mit den Angaben zum Wirkstoff durchführbar sind (vgl. Kapitel 4.2.2, Abbildung 4.8).

Die identifizierten und durchgeführten Maßnahmen wurden im Rahmen dieser Arbeit ausschließlich auf die Medikationsverordnungen des UKD angewendet. Eine Übertragung auf andere Universitätskliniken in Deutschland ist prinzipiell möglich. Da die Universitätskliniken das Austauschformat FHIR interoperabel im Rahmen der MI-I Medikationsdaten als eines der Basismodule des MI-I KDS bereitstellen und bereits ein ETL Prozess zur Überführung nach OMOP entwickelt wurde, können die Daten nach OMOP überführt werden (Peng et al., 2023). Auf dieser Grundlage ist es möglich, die Maßnahmen dieser Arbeit auf andere Standorte zu übertragen und damit die Strukturiertheit von Medikationsverordnungen generell deutschlandweit zu verbessern, sowie die Überführung nach RxNorm zu sichern.

5.4 Ausblick

Im Rahmen dieser Arbeit wurde die Verbesserung der Strukturiertheit der Medikationsverordnungen erfolgreich am UKD implementiert und getestet. Die hierbei gewonnenen Erkenntnisse liefern eine wichtige Grundlage, um strukturierte medizinische Medikationsdaten unter Verwendung internationaler Terminologien für die Nutzung in OMOP zu erlauben. In zukünftigen Arbeiten sollen die entwickelten Algorithmen auch auf andere Universitätskliniken in Deutschland übertragen, und die Ergebnisse validiert werden, um standortübergreifend die Voraussetzungen für die Teilnahme an internationalen Forschungsprojekten auf Basis von OMOP zu schaffen. Die Nutzung von Daten basierend auf dem KDS der MI-I und die Übertragung der Daten nach OMOP sind dabei eine wichtige Voraussetzung, da alle deutschen Universitätskliniken Projektpartner der MI-I sind und sich verpflichtet haben, RWD aus der klinischen Versorgung im Format des MI-I KDS strukturiert bereitzustellen. Die Vorarbeiten für die Harmonisierung dieser Daten nach OMOP wurden bereits von Peng et al. (Peng et al., 2023) geleistet und sind für alle Standorte in Deutschland nutzbar. Außerhalb der MI-I bieten sich für die Datenausweitung auch Kooperationen im Rahmen von OHDSI Germany (Reinecke, Zoch, Reich, Kallfelz et al., 2021) und mit anderen Datenpartnern von EHDEN an, um die Übertragung und den Austausch von Best Practices in Bezug auf die Validierung von Medikationsdaten auf internationaler Ebene zu fördern.

Die Anwendung der Algorithmen ermöglicht die Ausweitung der Überprüfung der Ergebnisqualität der automatisiert zugeordneten ATC Codes für unstrukturierte Medikationsverordnungen auf andere Standorte und kann somit auch zu einer Verbesserung der Algorithmen führen. Bereits in dieser Arbeit wurde ein Zusammenhang der Ergebnisqualität der drei Algorithmen mit den Kriterien (a) Übereinstimmung der Ergebnisse der drei Algorithmen und (b) der Höhe des Levenshtein Scores von Algorithmus 3 untersucht. Mit der Übertragung auf andere Standorte und der Erweiterung der Datenbasis zur Prüfung der Ergebnisqualität der Algorithmen können durch weitere manuelle Validierungen mögliche Muster identifiziert werden. Eine Vergrößerung der Datenbasis und die Ausweitung der manuellen Validierung, kann die Generalisierbarkeit der ersten identifizierten Muster in dieser Arbeit (beispielsweise die signifikant höherer Mittelwert des Levenshtein Scores bei den korrekten Ergebnissen von Algorithmus 3) bestätigen. Dies ist eine wichtige Voraussetzung für eine zukünftige Automatisierung der Validierungsmethoden. Unter Verwendung von maschinellen Lernverfahren und der identifizierten Muster soll die Robustheit der Algorithmen weiter verbessert werden. Auch die Identifikation neuer Muster hinsichtlich der Übereinstimmung der Ergebnisse zwischen unterschiedlichen Algorithmen oder für bestimmte Wirkstoffe, bei denen eine höhere Fehlerquote bei der Zuordnung von ATC Codes existiert, kann in zukünftigen Arbeiten weiter untersucht werden. Dadurch kann es möglich werden, Rückschlüsse auf die bisher nicht validierte Menge an Medikationsverordnungen am UKD zu ziehen. Gleichzeitig sollen die manuellen Aufwände zur Validierung der Ergebnisse der Algorithmen weiter reduziert werden, um letztlich ein vollständig automatisiertes Verfahren zur Zuordnung von ATC Codes für unstrukturierte Medikationsverordnungen bereitzustellen.

Die interaktive Visualisierung zur Darstellung der Strukturiertheit der Medikationsverordnungen wurde als Teil eines Konzeptes zur Etablierung einer Feedbackschleife im Rahmen dieser Arbeit bereits vorgestellt. Mit dem problemorientierten, interdisziplinären Ansatz könnte potenziell ein positiver Einfluss auf die Datenqualität erzeugt werden. Die Feedbackschleife setzt genau zum Zeitpunkt der Generierung der Daten in den klinischen Systemen an. Sie hat das Potenzial, die Datenqualität zum Zeitpunkt der Erstellung zu verbessern und dadurch kostspielige, nachgelagerte Verbesserungen zu reduzieren. In einer anschließenden Forschungsarbeit kann das Konzept der Feedbackschleife in Kooperation mit anderen Teams, wie beispielsweise der Apotheke des UKD, praktisch umgesetzt und genutzt werden, um die Ursachen für unstrukturierte Dateneingaben von Medikationsverordnungen weiter zu

untersuchen. Dabei sollte der Fokus auf der Ausrichtung an den Nutzenden des Verordnungssystems für Medikamente liegen und Dokumentationsdefizite mit ihren Ursachen sichtbar gemacht werden. Dies kann eine Grundlage bilden, um nicht ausschließlich nachträglich, computergestützte Aufwände zur besseren Nutzbarkeit von Medikationsverordnungen für die Forschung betreiben zu müssen. Stattdessen eröffnet es die Möglichkeit, die Ursachen für einen Mangel an Datenstruktur bei Medikationsverordnungen entgegenzuwirken und die Datenqualität auch zum Zweck der Patient:innensicherheit zu fördern.

Literatur

Amutha, A., Praveen, P. A., Hockett, C. W., Ong, T. C., Jensen, E. T., Isom, S. P., D'Agostino, R. B. J., Hamman, R. F., Mayer-Davis, E. J., Wadwa, R. P., Lawrence, J. M., Pihoker, C., Kahn, M. G., Dabelea, D., Tandon, N., & Mohan, V. (2021). Treatment regimens and glycosylated hemoglobin levels in youth with Type 1 and Type 2 diabetes: Data from SEARCH (United States) and YDR (India) registries. *Pediatric Diabetes*, *22*(1), 31–39. https://doi.org/10.1111/pedi.13004

Banda, J. M. (2019). Fully connecting the Observational Health Data Science and Informatics (OHDSI) initiative with the world of linked open data. *Genomics & informatics*, *17*(2), e13. https://doi.org/10.5808/GI.2019.17.2.e13

Baron, J. A., Sandler, R. S., Bresalier, R. S., Lanas, A., Morton, D. G., Riddell, R., Iverson, E. R., & DeMets, D. L. (2008). Cardiovascular events associated with rofecoxib: final analysis of the APPROVe trial. *The Lancet*, *372*(9651), 1756–1764. https://doi.org/10.1016/S0140-6736(08)61490-7

Bild, R., Bialke, M., Buckow, K., Ganslandt, T., Ihrig, K., Jahns, R., Merzweiler, A., Roschka, S., Schreiweis, B., Stäubert, S., Zenker, S., & Prasser, F. (2020). Towards a comprehensive and interoperable representation of consent-based data usage permissions in the German medical informatics initiative. *BMC Medical Informatics and Decision Making*, *20*(1), 103. https://doi.org/10.1186/s12911-020-01138-6

Biswas, P. (2013). Pharmacovigilance in Asia. *Journal of Pharmacology and Pharmacotherapeutics*, *4*(1_suppl), S7–S19. https://doi.org/10.4103/0976-500X.120941

Blacketer, C., Defalco, F. J., Ryan, P. B., & Rijnbeek, P. R. (2021). Increasing trust in real-world evidence through evaluation of observational data quality. *Journal of the American*

Medical Informatics Association, *28*(10), 2251–2257. https://doi.org/10.1093/jamia/ocab132

Blacketer, C., Voss, E. A., DeFalco, F., Hughes, N., Schuemie, M. J., Moinat, M., & Rijnbeek, P. R. (2021). Using the Data Quality Dashboard to Improve the EHDEN Network. *Applied Sciences*, *11*(24), 11920. https://doi.org/10.3390/app112411920

Bobroske, K., Larish, C., Cattrell, A., Bjarnadóttir, M. V., & Huan, L. (2020). The bird's-eye view: A data-driven approach to understanding patient journeys from claims data. *Journal of the American Medical Informatics Association*, *27*(7), 1037–1045. https://doi.org/10.1093/jamia/ocaa052

Bodenreider, O., & Rodriguez, L. M. (2014). Analyzing U.S. prescription lists with RxNorm and the ATC/DDD Index. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2014*, 297–306.

Boland, M. R., Parhi, P., Li, L., Miotto, R., Carroll, R., Iqbal, U., Nguyen, P.-A. A., Schuemie, M., You, S. C., Smith, D., Mooney, S., Ryan, P., Li, Y.-C. J., Park, R. W., Denny, J., Dudley, J. T., Hripcsak, G., Gentine, P., & Tatonetti, N. P. (2018). Uncovering exposures responsible for birth season - disease effects: a global study. *Journal of the American Medical Informatics Association : JAMIA*, *25*(3), 275–288. https://doi.org/10.1093/jamia/ocx105

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *Summit on Translational Bioinformatics*, *2010*, 1–5.

Bradley, E. H. (2004). Data feedback efforts in quality improvement: lessons learned from US hospitals. *Quality and Safety in Health Care*, *13*(1), 26–31. https://doi.org/10.1136/qhc.13.1.26

Brat, G. A., Weber, G. M., Gehlenborg, N., Avillach, P., Palmer, N. P., Chiovato, L., Cimino, J., Waitman, L. R., Omenn, G. S., Malovini, A., Moore, J. H., Beaulieu-Jones, B. K., Tibollo, V., Murphy, S. N., Yi, S. L., Keller, M. S., Bellazzi, R., Hanauer, D. A., Serret-Larmande, A., … Kohane, I. S. (2020). International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *npj Digital Medicine*, *3*(1), 109. https://doi.org/10.1038/s41746-020-00308-0

Brauer, R., Wong, I. C. K., Man, K. K., Pratt, N. L., Park, R. W., Cho, S.-Y., Li, Y.-C. (, Iqbal, U., Nguyen, P.-A. A., & Schuemie, M. (2020). Application of a Common Data Model (CDM) to rank the paediatric user and prescription prevalence of 15 different drug classes in South Korea, Hong Kong, Taiwan, Japan and Australia: an observational, descriptive study. *BMJ Open*, *10*(1), e032426. https://doi.org/10.1136/bmjopen-2019-032426

Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanas, A., Konstam, M. A., & Baron, J. A. (2005). Cardiovascular Events Associated with Rofecoxib in a Colorectal Adenoma Chemoprevention Trial. *New England Journal of Medicine*, *352*(11), 1092–1102. https://doi.org/10.1056/NEJMoa050493

Brown, J. S., Kahn, M., & Toh, S. (2013). Data Quality Assessment for Comparative Effectiveness Research in Distributed Data Networks. *Medical Care*, *51*(Supplement 8Suppl 3), S22–S29. https://doi.org/10.1097/MLR.0b013e31829b1e2c

Bundesinstitut für Arzneimittel und Medizinprodukte. (2023). Amtliche Fassung des ATC-Index mit DDD-Angaben für Deutschland im Jahre 2023. https://www.bfarm.de/SharedDocs/Downloads/DE/Kodiersysteme/ATC/atc-ddd-amtlich-2023.pdf?__blob=publicationFile

Burcu, M., Dreyer, N. A., Franklin, J. M., Blum, M. D., Critchlow, C. W., Perfetto, E. M., & Zhou, W. (2020). Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. *Pharmacoepidemiology and Drug Safety*, *29*(10), 1228–1235. https://doi.org/10.1002/pds.4975

Burn, E., Sena, A. G., Prats-Uribe, A., Spotnitz, M., DuVall, S., Lynch, K. E., Matheny, M. E., Nyberg, F., Ahmed, W.-U.-R., Alser, O., Alghoul, H., Alshammari, T., Zhang, L., Casajust, P., Areia, C., Shah, K., Reich, C., Blacketer, C., Andryc, A., … Duarte-Salles, T. (2020). Use of dialysis, tracheostomy, and extracorporeal membrane oxygenation among 240,392 patients hospitalized with COVID-19 in the United States. *medRxiv : the preprint server for health sciences*. https://doi.org/10.1101/2020.11.25.20229088

Burn, E., You, S. C., Sena, A. G., Kostka, K., Abedtash, H., Abrahão, M. T. F., Alberga, A., Alghoul, H., Alser, O., Alshammari, T. M., Aragon, M., Areia, C., Banda, J. M., Cho, J., Culhane, A. C., Davydov, A., DeFalco, F. J., Duarte-Salles, T., DuVall, S., … Ryan, P. (2020). Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. *Nature communications*, *11*(1), 5009. https://doi.org/10.1038/s41467-020-18849-z

Butler, A., Wei, W., Yuan, C., Kang, T., Si, Y., & Weng, C. (2018). The Data Gap in the EHR for Clinical Research Eligibility Screening. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, *2017*, 320–329.

Cardwell, C. R., Abnet, C. C., Cantwell, M. M., & Murray, L. J. (2010). Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer. *JAMA*, *304*(6), 657. https://doi.org/10.1001/jama.2010.1098

Chandler, R. E. (2020). Nintedanib and ischemic colitis: Signal assessment with the integrated use of two types of real-world evidence, spontaneous reports of suspected adverse drug reactions, and observational data from large health-care databases. *Pharmacoepidemiology and drug safety*, *29*(8), 951–957. https://doi.org/10.1002/pds.5022

Chen, R., Ryan, P., Natarajan, K., Falconer, T., Crew, K. D., Reich, C. G., Vashisht, R., Randhawa, G., Shah, N. H., & Hripcsak, G. (2020). Treatment Patterns for Chronic Comorbid Conditions in Patients With Cancer Using a Large-Scale Observational Data Network. *JCO clinical cancer informatics*, *4*, 171–183. https://doi.org/10.1200/CCI.19.00107

Chinchuluun, A., Pardalos, P. M., Migdalas, A., & Pitsoulis, L. (Hrsg.). (2008). *Pareto optimality, game theory and equilibria*. Springer.

Choi, Y. I., Kim, Y. J., Chung, J.-W., Kim, K. O., Kim, H., Park, R. W., & Park, D. K. (2020). Effect of Age on the Initiation of Biologic Agent Therapy in Patients With Inflammatory Bowel Disease: Korean Common Data Model Cohort Study. *JMIR medical informatics*, *8*(4), e15124. https://doi.org/10.2196/15124

Cito, J., Ferme, V., & Gall, H. C. (2016). Using Docker Containers to Improve Reproducibility in Software and Web Engineering Research [Series Title: Lecture Notes in Computer Science]. In A. Bozzon, P. Cudre-Maroux & C. Pautasso (Hrsg.), *Web Engineering* (S. 609–612). Springer International Publishing. https://doi.org/10.1007/978-3-319-38791-8_58

Cohen, A. (2020). fuzzywuzzy: Fuzzy string matching in python. Verfügbar 26. November 2021 unter https://pypi.org/project/fuzzywuzzy/

Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *JAMA*, *320*(9), 867–868. https://doi.org/10.1001/jama.2018.10136

De Mello, B. H., Rigo, S. J., Da Costa, C. A., Da Rosa Righi, R., Donida, B., Bez, M. R., & Schunke, L. C. (2022). Semantic interoperability in health records standards: a systematic literature review. *Health and Technology*, *12*(2), 255–272. https://doi.org/10.1007/s12553-022-00639-w

Dixon, B. E., Wen, C., French, T., Williams, J. L., Duke, J. D., & Grannis, S. J. (2020). Extending an open-source tool to measure data quality: case report on Observational Health Data Science and Informatics (OHDSI). *BMJ health & care informatics*, *27*(1). https://doi.org/10.1136/bmjhci-2019-100054

Duke, J. D., Ryan, P. B., Suchard, M. A., Hripcsak, G., Jin, P., Reich, C., Schwalm, M.-S., Khoma, Y., Wu, Y., Xu, H., Shah, N. H., Banda, J. M., & Schuemie, M. J. (2017). Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network. *Epilepsia*, *58*(8), e101–e106. https://doi.org/10.1111/epi.13828

EHDEN. (2022). EHDEN - European Health Data & Evidence Network. Verfügbar 2. Januar 2023 unter https://www.ehden.eu/

European Medicines Agency (EMA). (2020a). COVID-19: EMA sets up infrastructure for real-world monitoring of treatments and vaccines. https://www.ema.europa.eu/en/news/covid-19-ema-sets-infrastructure-real-world-monitoring-treatments-vaccines

European Medicines Agency (EMA). (2020b). EMA Study protocol: Systemic glucocorticoids in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients in the primary and secondary care setting. https://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/44049

Fischer, P., Stöhr, M. R., Gall, H., Michel-Backofen, A., & Majeed, R. W. (2020). Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT. *Studies in Health Technology and Informatics*, *270*, 138–142. https://doi.org/10.3233/SHTI200138

Freitas Da Cruz, H., Bergner, B., Konak, O., Schneider, F., Bode, P., Lempert, C., & Schapranow, M.-P. (2019). MORPHER - A Platform to Support Modeling of Outcome and Risk Prediction in Health Research. *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 462–469. https://doi.org/10.1109/BIBE.2019.00090

Ganslandt, T., Hackl, W., & Section Editors for the IMIA Yearbook Section on Clinical Information Systems. (2015). Findings from the Clinical Information Systems Perspective. *Yearbook of Medical Informatics*, *24*(01), 90–94. https://doi.org/10.15265/IY-2015-037

Gehring, S., & Eulenfeld, R. (2018). German Medical Informatics Initiative: Unlocking Data for Research and Health Care. *Methods of Information in Medicine*, *57*(S 01), e46–e49. https://doi.org/10.3414/ME18-13-0001

Green, J., Czanner, G., Reeves, G., Watson, J., Wise, L., & Beral, V. (2010). Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. *BMJ*, *341*(sep01 3), c4444–c4444. https://doi.org/10.1136/bmj.c4444

Gruendner, J., Schwachhofer, T., Sippl, P., Wolf, N., Erpenbeck, M., Gulden, C., Kapsner, L. A., Zierk, J., Mate, S., Stürzl, M., Croner, R., Prokosch, H.-U., & Toddenroth, D. (2019).

KETOS: Clinical decision support and machine learning as a service - A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. *PloS one*, *14*(10), e0223010. https://doi.org/10.1371/journal.pone.0223010

Gruhl, M., Reinecke, I., & Sedlmayr, M. (2020). Specification and Distribution of Vocabularies Among Consortial Partners. *Studies in Health Technology and Informatics*, *270*, 1393–1394. https://doi.org/10.3233/SHTI200458

Gulden, C., Landerer, I., Nassirian, A., Altun, F. B., & Andrae, J. (2019). Extraction and Prevalence of Structured Data Elements in Free-Text Clinical Trial Eligibility Criteria. *Studies in Health Technology and Informatics*, *258*, 226–230.

H. Spengler, I. Gatz, F. Kohlmayer, K. A. Kuhn & F. Prasser. (2020). Improving Data Quality in Medical Research: A Monitoring Architecture for Clinical and Translational Data Warehouses [Journal Abbreviation: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)]. *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 415–420. https://doi.org/10.1109/CBMS49503.2020.00085

Hampson, G., Towse, A., Dreitlein, W. B., Henshall, C., & Pearson, S. D. (2018). Real-world evidence for coverage decisions: opportunities and challenges. *Journal of Comparative Effectiveness Research*, *7*(12), 1133–1143. https://doi.org/10.2217/cer-2018-0066

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Harvey, H. B., & Sotardi, S. T. (2018). The Pareto Principle. *Journal of the American College of Radiology*, *15*(6), 931. https://doi.org/10.1016/j.jacr.2018.02.026

Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., Lehmann, H. P., Hripcsak, G., Hartzog, T. H., Cimino, J. J., & Saltz, J. H. (2013). Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. *Medical Care*, *51*(Supplement 8Suppl 3), S30–S37. https://doi.org/10.1097/MLR.0b013e31829b1dbd

HL7. (2021). Health Level Seven International. http://www.hl7.org/index.cfm

Hockett, C. W., Praveen, P. A., Ong, T. C., Amutha, A., Isom, S. P., Jensen, E. T., D'Agostino, R. B., Hamman, R. F., Mayer-Davis, E. J., Lawrence, J. M., Pihoker, C., Kahn, M. G., Mohan, V., Tandon, N., & Dabelea, D. (2021). Clinical profile at diagnosis with youth-onset type 1

and type 2 diabetes in two pediatric diabetes registries: SEARCH (United States) and YDR (India). *Pediatric Diabetes*, *22*(1), 22–30. https://doi.org/10.1111/pedi.12981

Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., Suchard, M. A., Schuemie, M. J., DeFalco, F. J., Perotte, A., Banda, J. M., Reich, C. G., Schilling, L. M., Matheny, M. E., Meeker, D., Pratt, N., & Madigan, D. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*, *113*(27), 7329–7336. https://doi.org/10.1073/pnas.1510502113

Hripcsak, G., Suchard, M. A., Shea, S., Chen, R., You, S. C., Pratt, N., Madigan, D., Krumholz, H. M., Ryan, P. B., & Schuemie, M. J. (2020). Comparison of Cardiovascular and Safety Outcomes of Chlorthalidone vs Hydrochlorothiazide to Treat Hypertension. *JAMA Internal Medicine*, *180*(4), 542. https://doi.org/10.1001/jamainternmed.2019.7454

Huser, V., DeFalco, F. J., Schuemie, M., Ryan, P. B., Shang, N., Velez, M., Park, R. W., Boyce, R. D., Duke, J., Khare, R., Utidjian, L., & Bailey, C. (2016). Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets. *EGEMS (Washington, DC)*, *4*(1), 1239. https://doi.org/10.13063/2327-9214.1239

Hutchison, E. R., Zhang, Y., Nampally, S., Weatherall, J., Khan, F., & Shameer, K. (2020). Uncovering Machine Learning-Ready Data from Public Clinical Trial Resources: A case-study on normalization across Aggregate Content of ClinicalTrials.gov. *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2965–2967. https://doi.org/10.1109/BIBM49941.2020.9313362

InEK. (2018). Datenlieferung gemäß §21 des KHEntgG - Datenbeschreibung. https://www.g-drg.de/Datenlieferung_gem._21_KHEntgG/Dokumente_zur_Datenlieferung/Datensatzbeschreibung

J. R. Almeida & J. L. Oliveira. (2020). Multi-language Concept Normalisation of Clinical Cohorts [Journal Abbreviation: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)]. *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 261–264. https://doi.org/10.1109/CBMS49503.2020.00056

Jensen, E. T., Dabelea, D. A., Praveen, P. A., Amutha, A., Hockett, C. W., Isom, S. P., Ong, T. C., Mohan, V., D'Agostino, R. J., Kahn, M. G., Hamman, R. F., Wadwa, P., Dolan, L., Lawrence, J. M., Madhu, S. V., Chhokar, R., Goel, K., Tandon, N., & Mayer-Davis, E. (2021). Comparison of the incidence of diabetes in United States and Indian youth: An international harmonization of youth diabetes registries. *Pediatric diabetes*, *22*(1), 8–14. https://doi.org/10.1111/pedi.13009

Jeon, S., Seo, J., Kim, S., Lee, J., Kim, J.-H., Sohn, J. W., Moon, J., & Joo, H. J. (2020). Proposal and Assessment of a De-Identification Strategy to Enhance Anonymity of the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) in a Public Cloud-Computing Environment: Anonymization of Medical Data Using Privacy Models. *Journal of medical Internet research*, *22*(11), e19597. https://doi.org/10.2196/19597

Jiang, G., Kiefer, R. C., Sharma, D. K., Prud'hommeaux, E., & Solbrig, H. R. (2017). A Consensus-Based Approach for Harmonizing the OHDSI Common Data Model with HL7 FHIR [ISSN: 0926-9630 Type: Proceedings Paper]. In Gundlapalli, AV and Jaulent, MC and Zhao, D (Hrsg.), *MEDINFO 2017: PRECISION HEALTHCARE THROUGH INFORMATICS* (S. 887–891). IOS PRESS. https://doi.org/10.3233/978-1-61499-830-3-887

Jiang, G., Yu, Y., Kingsbury, P. R., & Shah, N. (2019). Augmenting Medical Device Evaluation Using a Reusable Unique Device Identifier Interoperability Solution Based on the OHDSI Common Data Model. [Place: Netherlands]. *Studies in health technology and informatics*, *264*, 1502–1503. https://doi.org/10.3233/SHTI190505

Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., & Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*, *4*(1), 18. https://doi.org/10.13063/2327-9214.1244

Kim, H. I., Yoon, J. Y., Kwak, M. S., & Cha, J. M. (2021). Gastrointestinal and Nongastrointestinal Complications of Esophagogastroduodenoscopy and Colonoscopy in the Real World: A Nationwide Standard Cohort Using the Common Data Model Database. [Place: Korea (South)]. *Gut and liver*. https://doi.org/10.5009/gnl20222

Kim, H., Yoo, S., Jeon, Y., Yi, S., Kim, S., Choi, S. A., Hwang, H., & Kim, K. J. (2020). Characterization of Anti-seizure Medication Treatment Pathways in Pediatric Epilepsy Using the Electronic Health Record-Based Common Data Model. *Frontiers in Neurology*, *11*, 409. https://doi.org/10.3389/fneur.2020.00409

Kim, Y., Tian, Y., Yang, J., Huser, V., Jin, P., Lambert, C. G., Park, H., You, S. C., Park, R. W., Rijnbeek, P. R., Van Zandt, M., Reich, C., Vashisht, R., Wu, Y., Duke, J., Hripcsak, G., Madigan, D., Shah, N. H., Ryan, P. B., … Suchard, M. A. (2020). Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. *Scientific Reports*, *10*(1), 11115. https://doi.org/10.1038/s41598-020-68037-8

Kohane, I. S., Aronow, B. J., Avillach, P., Beaulieu-Jones, B. K., Bellazzi, R., Bradford, R. L., Brat, G. A., Cannataro, M., Cimino, J. J., García-Barrio, N., Gehlenborg, N., Ghassemi, M., Gutiérrez-Sacristán, A., Hanauer, D. A., Holmes, J. H., Hong, C., Klann, J. G., Loh, N. H. W., Luo, Y., … Cai, T. (2021). What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask. *Journal of Medical Internet Research*, *23*(3), e22219. https://doi.org/10.2196/22219

Kubota, K., Kamijima, Y., Kao Yang, Y.-H., Kimura, S., Chia-Cheng Lai, E., Man, K. K. C., Ryan, P., Schuemie, M., Stang, P., Su, C.-C., Wong, I. C. K., Zhang, Y., & Setoguchi, S. (2018). Penetration of new antidiabetic medications in East Asian countries and the United States: A cross-national comparative study. *PloS one*, *13*(12), e0208796. https://doi.org/10.1371/journal.pone.0208796

Lamer, A., Depas, N., Doutreligne, M., Parrot, A., Verloop, D., Defebvre, M.-M., Ficheur, G., Chazard, E., & Beuscart, J.-B. (2020). Transforming French Electronic Health Records into the Observational Medical Outcome Partnership's Common Data Model: A Feasibility Study. *Applied clinical informatics*, *11*(1), 13–22. https://doi.org/10.1055/s-0039-3402754

Lane, J. C. E., Kostka, K., Weaver, J., Duarte-Salles, T., Abrahao, M. T. F., Alghoul, H., Alser, O., Alshammari, T. M., Biedermann, P., Banda, J. M., Burn, E., Casajust, P., Conover, M. M., Culhane, A. C., Davydov, A., DuVall, S. L., Dymshyts, D., Fernandez-Bertolin, S., Fišter, K., … Prieto-Alhambra, D. (2020). Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. *The Lancet Rheumatology*, *2*(11), e698–e711. https://doi.org/10.1016/S2665-9913(20)30276-9

Lane, J. C. E., Weaver, J., Kostka, K., Duarte-Salles, T., Abrahao, M. T. F., Alghoul, H., Alser, O., Alshammari, T. M., Areia, C., Biedermann, P., Banda, J. M., Burn, E., Casajust, P., Fister, K., Hardin, J., Hester, L., Hripcsak, G., Kaas-Hansen, B. S., Khosla, S., … for the OHDSI-COVID-19 consortium. (2021). Risk of depression, suicide and psychosis with hydroxychloroquine treatment for rheumatoid arthritis: a multinational network cohort study. *Rheumatology*, *60*(7), 3222–3234. https://doi.org/10.1093/rheumatology/keaa771

Lehne, M., Sass, J., Essenwanger, A., Schepers, J., & Thun, S. (2019). Why digital medicine depends on interoperability. *npj Digital Medicine*, *2*(1), 79. https://doi.org/10.1038/s41746-019-0158-1

Lima, D. M., Rodrigues-Jr, J. F., Traina, A. J. M., Pires, F. A., & Gutierrez, M. A. (2019). Transforming Two Decades of ePR Data to OMOP CDM for Clinical Research [ISSN: 0926-9630 Type: Proceedings Paper]. In OhnoMachado, L and Seroussi, B (Hrsg.), *MEDINFO 2019: HEALTH AND WELLBEING E-NETWORKS FOR ALL* (S. 233–237). IOS PRESS. https://doi.org/10.3233/SHTI190218

Liu, S., Wang, Y., Wen, A., Wang, L., Hong, N., Shen, F., Bedrick, S., Hersh, W., & Liu, H. (2020). Implementation of a Cohort Retrieval System for Clinical Data Repositories Using the Observational Medical Outcomes Partnership Common Data Model: Proof-of-Concept System Validation. *JMIR medical informatics*, *8*(10), e17376. https://doi.org/10.2196/17376

Madigan, D., Ryan, P. B., Schuemie, M., Stang, P. E., Overhage, J. M., Hartzema, A. G., Suchard, M. A., DuMouchel, W., & Berlin, J. A. (2013). Evaluating the Impact of Database Heterogeneity on Observational Study Results [Place: JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA Publisher: OXFORD UNIV PRESS INC Type: Article]. *AMERICAN JOURNAL OF EPIDEMIOLOGY*, *178*(4), 645–651. https://doi.org/10.1093/aje/kwt010

Magalhães, T., Dinis-Oliveira, R. J., & Taveira-Gomes, T. (2022). Digital Health and Big Data Analytics: Implications of Real-World Evidence for Clinicians and Policymakers. *International Journal of Environmental Research and Public Health*, *19*(14), 8364. https://doi.org/10.3390/ijerph19148364

Maier, C., Lang, L., Storf, H., Vormstein, P., Bieber, R., Bernarding, J., Herrmann, T., Haverkamp, C., Horki, P., Laufer, J., Berger, F., Höning, G., Fritsch, H. W., Schüttler, J., Ganslandt, T., Prokosch, H. U., & Sedlmayr, M. (2018). Towards Implementation of OMOP in a German University Hospital Consortium. *Applied clinical informatics*, *9*(1), 54–61. https://doi.org/10.1055/s-0037-1617452

Merkel, D. (2014). Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*, *2*. https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment

Miller, A. R., & Tucker, C. (2014). Health information exchange, system size and information silos. *Journal of Health Economics*, *33*, 28–42. https://doi.org/10.1016/j.jhealeco.2013.10.004

Morales, D. R., Conover, M. M., You, S. C., Pratt, N., Kostka, K., Duarte-Salles, T., Fernández-Bertolín, S., Aragón, M., DuVall, S. L., Lynch, K., Falconer, T., van Bochove, K., Sung, C., Matheny, M. E., Lambert, C. G., Nyberg, F., Alshammari, T. M., Williams, A. E., Park, R. W., … Suchard, M. A. (2021). Renin-angiotensin system blockers and susceptibility

to COVID-19: an international, open science, cohort analysis. *The Lancet. Digital health*, *3*(2), e98–e114. https://doi.org/10.1016/S2589-7500(20)30289-2

Mueen Ahmed, K. K., & Dhubaib, B. E. A. (2011). Zotero: A bibliographic assistant to researcher. *Journal of Pharmacology and Pharmacotherapeutics*, *2*(4), 304–305. https://doi.org/10.4103/0976-500X.85940

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

National Clinical Terminology Service (NCTS) | Digital Health Developer Portal. (2023). Verfügbar 29. Juni 2023 unter https://developer.digitalhealth.gov.au/resources/services/national-clinical-terminology-service-ncts

National Library of Medicine. (2023). Appendix 5 - RxNorm Term Types (TTY). Verfügbar 16. Februar 2023 unter https://www.nlm.nih.gov/research/umls/rxnorm/docs/appendix5.html

Norwegian Institute of Public Health WHO Collaborating Centre for Drug Statistics Methodology, Rogers, B., & Sheffler, S. (2021). COMPARISON OF THE WHO ATC CLASSIFICATION & E PH MRA/ Intellus Worldwide ANATOMICAL CLASSIFICATION. Verfügbar 15. Februar 2023 unter https://www.ephmra.org/sites/default/files/2022-03/WHO%20ATC%202021%20comparison%20Final%202021%20for%20web%20site.pdf

Odysseus Data Services Inc. (2023). PALLAS - GitHub Repo: Build process for OMOP Vocabularies, OMOP-CDM V5. Verfügbar 15. Februar 2023 unter https://github.com/OHDSI/Vocabulary-v5.0

OHDSI. (2019). *The book of OHDSI Observational Health Data Sciences and Informatics* [OCLC: 1141203304].

OHDSI. (2021). OHDSI Wiki - Domain Drug Documentation. Verfügbar 13. April 2023 unter https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:drug

OHDSI. (2022). OHDSI Wiki - Vocabulary ATC. https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:atc

OHDSI. (2023a). OMOP Common Data Model. Verfügbar 16. Mai 2023 unter https://ohdsi.github.io/CommonDataModel/

OHDSI. (2023b). Rabbit in a Hat. Verfügbar 30. Juni 2023 unter https://ohdsi.github.io/WhiteRabbit/RabbitInAHat.html

OHDSI. (2023c). White Rabbit. Verfügbar 30. Juni 2023 unter https://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html

OHDSI & Blacketer, C. (2021). OMOP CDM GitHub Wiki, DRUG_EXPOSURE. Verfügbar 15. April 2023 unter https://ohdsi.github.io/CommonDataModel/cdm53.html#DRUG_EXPOSURE

Ostropolets, A. (2020). Observational Health Data Sciences and Informatics - Forum. Verfügbar 2. Januar 2023 unter https://forums.ohdsi.org/t/atc-release/11182

Ostropolets, A., Albogami, Y., Conover, M., Banda, J. M., Baumgartner, W. A., Blacketer, C., Desai, P., DuVall, S. L., Fortin, S., Gilbert, J. P., Golozar, A., Ide, J., Kanter, A. S., Kern, D. M., Kim, C., Lai, L. Y. H., Li, C., Liu, F., Lynch, K. E., … Ryan, P. B. (2023). Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study. *Journal of the American Medical Informatics Association*, *30*(5), 859–868. https://doi.org/10.1093/jamia/ocad009

Otero Varela, L., Wiebe, N., Niven, D. J., Ronksley, P. E., Iragorri, N., Robertson, H. L., & Quan, H. (2019). Evaluation of interventions to improve electronic health record documentation within the inpatient setting: a protocol for a systematic review. *Systematic Reviews*, *8*(1), 54. https://doi.org/10.1186/s13643-019-0971-2

Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., & Stang, P. E. (2012). Validation of a common data model for active safety surveillance research [Place: BRITISH MED ASSOC HOUSE, TAVISTOCK SQUARE, LONDON WC1H 9JR, ENGLAND Publisher: B M J PUBLISHING GROUP Type: Article]. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*, *19*(1), 54–60. https://doi.org/10.1136/amiajnl-2011-000376

Palinkas, L. A., Mendon, S. J., & Hamilton, A. B. (2019). Innovations in Mixed Methods Evaluations. *Annual Review of Public Health*, *40*(1), 423–442. https://doi.org/10.1146/annurev-publhealth-040218-044215

Paris, N., Lamer, A., & Parrot, A. (2021). Transformation and Evaluation of the MIMIC Database in the OMOP Common Data Model: Development and Usability Study. *JMIR Medical Informatics*, *9*(12), e30970. https://doi.org/10.2196/30970

Peng, Y., Henke, E., Reinecke, I., Zoch, M., Sedlmayr, M., & Bathelt, F. (2023). An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. *International Journal of Medical Informatics*, *169*, 104925. https://doi.org/10.1016/j.ijmedinf.2022.104925

Pitts, P. J., & Le Louet, H. (2018). Advancing Drug Safety Through Prospective Pharmacovigilance. *Therapeutic Innovation & Regulatory Science*, *52*(4), 400–402. https://doi.org/10.1177/2168479018766887

Pitts, P. J., Louet, H. L., Moride, Y., & Conti, R. M. (2016). 21st century pharmacovigilance: efforts, roles, and responsibilities. *The Lancet Oncology*, *17*(11), e486–e492. https://doi.org/10.1016/S1470-2045(16)30312-6

Platt, R., Brown, J. S., Robb, M., McClellan, M., Ball, R., Nguyen, M. D., & Sherman, R. E. (2018). The FDA Sentinel Initiative — An Evolving National Resource. *New England Journal of Medicine*, *379*(22), 2091–2093. https://doi.org/10.1056/NEJMp1809643

Platt, R., Wilson, M., Chan, K. A., Benner, J. S., Marchibroda, J., & McClellan, M. (2009). The New Sentinel Network — Improving the Evidence of Medical-Product Safety. *New England Journal of Medicine*, *361*(7), 645–647. https://doi.org/10.1056/NEJMp0905338

Postigo, R., Brosch, S., Slattery, J., van Haren, A., Dogné, J.-M., Kurz, X., Candore, G., Domergue, F., & Arlett, P. (2018). EudraVigilance Medicines Safety Database: Publicly Accessible Data for Research and Public Health Protection. *Drug Safety*, *41*(7), 665–675. https://doi.org/10.1007/s40264-018-0647-1

Priou, S., Lame, G., Jankovic, M., Chatellier, G., Bey, R., Tournigand, C., Daniel, C., & Kempf, E. (2023). Why Are Data Missing in Clinical Data Warehouses? A Simulation Study of How Data Are Processed (and Can Be Lost). In M. Hägglund, M. Blusi, S. Bonacina, L. Nilsson, I. Cort Madsen, S. Pelayo, A. Moen, A. Benis, L. Lindsköld & P. Gallos (Hrsg.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI230103

Prokosch, H.-U., Acker, T., Bernarding, J., Binder, H., Boeker, M., Boerries, M., Daumke, P., Ganslandt, T., Hesser, J., Höning, G., Neumaier, M., Marquardt, K., Renz, H., Rothkötter, H.-J., Schade-Brittinger, C., Schmücker, P., Schüttler, J., Sedlmayr, M., Serve, H., … Storf, H. (2018). MIRACUM: Medical Informatics in Research and Care in University Medicine: A Large Data Sharing Network to Enhance Translational Research and Medical Care. *Methods of Information in Medicine*, *57*(S 01), e82–e91. https://doi.org/10.3414/ME17-02-0025

Prokosch, H.-U., & Karg, M. (2022). MIRACUM - MIRACOLIX Toolbox. Verfügbar 26. Februar 2023 unter https://www.miracum.org/das-konsortium/datenintegrationszentren/miracolix-tools/

Reimer, A. P., Milinovich, A., & Madigan, E. A. (2016). Data quality assessment framework to assess electronic medical record data for use in research. *International Journal of Medical Informatics*, *90*, 40–47. https://doi.org/10.1016/j.ijmedinf.2016.03.006

Reinecke, I. (2021a). literature list of OHDSI studies. https://doi.org/10.5281/ZENODO.5145048

Reinecke, I. (2021b). The Use of OHDSI OMOP – A Scoping Review: list of included publications. https://doi.org/10.5281/ZENODO.4635599

Reinecke, I. (2023a). Metadaten der OHDSI Studien. https://doi.org/10.5281/ZENODO.8127755

Reinecke, I. (2023b). Quellcode zur Dissertation von Ines Reinecke. https://doi.org/10.5281/ZENODO.8127650

Reinecke, I., Bathelt, F., Sedlmayr, M., & Kühn, A. (2022). Pharmaceutical Feedback Loop – A Concept to Improve Prescription Safety and Data Quality. *Studies in Health Technology and Informatics*. https://doi.org/10.3233/SHTI220910

Reinecke, I., Gulden, C., Kümmel, M., Nassirian, A., Blasini, R., & Sedlmayr, M. (2020). Design for a Modular Clinical Trial Recruitment Support System Based on FHIR and OMOP. *Studies in Health Technology and Informatics*, *270*, 158–162. https://doi.org/10.3233/SHTI200142

Reinecke, I., Henke, E., Peng, Y., Sedlmayr, M., & Bathelt, F. (2023). Fitness for Use of Anatomical Therapeutic Chemical Classification for Real World Data Research. In M. Hägglund, M. Blusi, S. Bonacina, L. Nilsson, I. Cort Madsen, S. Pelayo, A. Moen, A. Benis, L. Lindsköld & P. Gallos (Hrsg.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI230245

Reinecke, I., Siebel, J., Fuhrmann, S., Fischer, A., Sedlmayr, M., Weidner, J., & Bathelt, F. (2023). Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation. *JMIR Medical Informatics*, *11*, e40312. https://doi.org/10.2196/40312

Reinecke, I., Zoch, M., Reich, C., Kallfelz, M., Grewe, N., & Sedlmayr, M. (2021). OHDSI Germany – Join the Journey: A Workshop [Medium: text/html Publisher: German Medical Science GMS Publishing House]. https://doi.org/10.3205/21GMDS031

Reinecke, I., Zoch, M., Reich, C., Sedlmayr, M., & Bathelt, F. (2021). The Usage of OHDSI OMOP - A Scoping Review. [Place: Netherlands]. *Studies in health technology and informatics*, *283*, 95–103. https://doi.org/10.3233/SHTI210546

Reinecke, I., Zoch, M., Wilhelm, M., Sedlmayr, M., & Bathelt, F. (2021). Transfer of Clinical Drug Data to a Research Infrastructure on OMOP - A FAIR Concept. *Studies in Health Technology and Informatics*, *287*, 63–67. https://doi.org/10.3233/SHTI210815

Rogers, J. R., Callahan, T. J., Kang, T., Bauck, A., Khare, R., Brown, J. S., Kahn, M. G., & Weng, C. (2019). A Data Element-Function Conceptual Model for Data Quality Checks. *EGEMS (Washington, DC)*, *7*(1), 17. https://doi.org/10.5334/egems.289

Ryan, P. B., Madigan, D., Stang, P. E., Overhage, J. M., Racoosin, J. A., & Hartzema, A. G. (2012). Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership [Place: 111 RIVER ST, HOBOKEN 07030-5774, NJ USA Publisher: WILEY-BLACKWELL Type: Article; Proceedings Paper]. *STATISTICS IN MEDICINE*, *31*(30, SI), 4401–4415. https://doi.org/10.1002/sim.5620

Sadhna, D., & Nagaich, U. (2015). Drug recall: An incubus for pharmaceutical companies and most serious drug recall of history. *International Journal of Pharmaceutical Investigation*, *5*(1), 13. https://doi.org/10.4103/2230-973X.147222

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., Detmer, D. E., & Expert Panel, n. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association: JAMIA*, *14*(1), 1–9. https://doi.org/10.1197/jamia.M2273

Samwald, M., Xu, H., Blagec, K., Empey, P. E., Malone, D. C., Ahmed, S. M., Ryan, P., Hofer, S., & Boyce, R. D. (2016). Incidence of Exposure of Patients in the United States to Multiple Drugs for Which Pharmacogenomic Guidelines Are Available. *PloS one*, *11*(10), e0164972. https://doi.org/10.1371/journal.pone.0164972

Schuemie, M. J., Cepeda, M. S., Suchard, M. A., Yang, J., Tian, Y., Schuler, A., Ryan, P. B., Madigan, D., & Hripcsak, G. (2020). How Confident Are We about Observational Findings in Healthcare: A Benchmark Study. *Harvard data science review*, *2*(1). https://doi.org/10.1162/99608f92.147cc28e

Schuemie, M. J., Gini, R., Coloma, P. M., Straatman, H., Herings, R. M. C., Pedersen, L., Innocenti, F., Mazzaglia, G., Picelli, G., van der Lei, J., & Sturkenboom, M. C. J. M. (2013). Replication of the OMOP Experiment in Europe: Evaluating Methods for Risk Identification in Electronic Health Record Databases [Place: 5 THE WAREHOUSE WAY, NORTHCOTE 0627, AUCKLAND, NEW ZEALAND Publisher: ADIS INT LTD Type: Article]. *DRUG SAFETY*, *36*(1), S159–S169. https://doi.org/10.1007/s40264-013-0109-8

Semler, S., Wissing, F., & Heyder, R. (2018). German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research. *Methods of Information in Medicine*, *57*(S 01), e50–e56. https://doi.org/10.3414/ME18-03-0003

Seo, S. I., You, S. C., Park, C. H., Kim, T. J., Ko, Y. S., Kim, Y., Yoo, J. J., Kim, J., & Shin, W. G. (2020). Comparative risk of <span style="font-variant:small-caps;» *Clostridium difficile* infection between proton pump inhibitors and histamine-2 receptor antagonists: A 15-year hospital cohort study using a common data model. *Journal of Gastroenterology and Hepatology*, *35*(8), 1325–1330. https://doi.org/10.1111/jgh.14983

Si, Y., & Weng, C. (2017). An OMOP CDM-Based Relational Database of Clinical Research Eligibility Criteria [ISSN: 0926-9630 Type: Proceedings Paper]. In Gundlapalli, AV and Jaulent, MC and Zhao, D (Hrsg.), *MEDINFO 2017: PRECISION HEALTHCARE THROUGH INFORMATICS* (S. 950–954). IOS PRESS. https://doi.org/10.3233/978-1-61499-830-3-950

Sibbald, B. (2004). Rofecoxib (Vioxx) voluntarily withdrawn from market. *Canadian Medical Association Journal*, *171*(9), 1027–1028. https://doi.org/10.1503/cmaj.1041606

Spotnitz, M. E., Natarajan, K., Ryan, P. B., & Westhoff, C. L. (2020). Relative Risk of Cervical Neoplasms Among Copper and Levonorgestrel-Releasing Intrauterine System Users. *Obstetrics and gynecology*, *135*(2), 319–327. https : / / doi . org / 10 . 1097 / AOG . 0000000000003656

Stang, P. E., Ryan, P. B., Overhage, J. M., Schuemie, M. J., Hartzema, A. G., & Welebob, E. (2013). Variation in Choice of Study Design: Findings from the Epidemiology Design Decision Inventory and Evaluation (EDDIE) Survey [Place: 41 CENTORIAN DR, PRIVATE BAG 65901, MAIRANGI BAY, AUCKLAND 1311, NEW ZEALAND Publisher: ADIS INT LTD Type: Article]. *DRUG SAFETY*, *36*(1), S15–S25. https://doi.org/10.1007/s40264-013-0103-1

Stang, P. E., Ryan, P. B., Racoosin, J. A., Overhage, J. M., Hartzema, A. G., Reich, C., Welebob, E., Scarnecchia, T., & Woodcock, J. (2010). Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership [Place: INDEPENDENCE MALL WEST 6TH AND RACE ST, PHILADELPHIA, PA 19106-1572 USA Publisher: AMER COLL PHYSICIANS Type: Article]. *ANNALS OF INTERNAL MEDICINE*, *153*(9), 600–606. https://doi.org/10.7326/0003-4819-153-9-201011020-00010

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018). Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care: Innovation at the Intersection of Clinical Trials. *Clinical and Translational Science*, *11*(5), 450–460. https://doi.org/10.1111/cts.12559

Ta, C. N., & Weng, C. (2019). Detecting Systemic Data Quality Issues in Electronic Health Records [Backup Publisher: French Assoc Med Informat ISSN: 0926-9630 Type: Proceedings Paper]. In OhnoMachado, L and Seroussi, B (Hrsg.), *MEDINFO 2019:*

HEALTH AND WELLBEING E-NETWORKS FOR ALL (S. 383–387). IOS PRESS. https://doi.org/10.3233/SHTI190248

Taggart, J., Liaw, S.-T., & Yu, H. (2015). Structured data quality reports to improve EHR data quality. *International Journal of Medical Informatics*, *84*(12), 1094–1098. https://doi.org/10.1016/j.ijmedinf.2015.09.008

TMF e.V. (2023). Der Kerndatensatz der Medizininformatik-Initiative. Verfügbar 16. Mai 2023 unter https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative

TMF – Technologie- und Methodenplattform & für die vernetzte medizinische Forschung e.V. (2023). Datenintegrationszentren der Medizininformaik Initiative. Verfügbar 13. Februar 2023 unter https://www.medizininformatik-initiative.de/de/konsortien/datenintegrationszentren

Tresp, V., Marc Overhage, J., Bundschus, M., Rabizadeh, S., Fasching, P. A., & Yu, S. (2016). Going Digital: A Survey on Digitalization and Large-Scale Data Analytics in Healthcare. *Proceedings of the IEEE*, *104*(11), 2180–2206. https://doi.org/10.1109/JPROC.2016.2615052

Unberath, P., Prokosch, H. U., Gründner, J., Erpenbeck, M., Maier, C., & Christoph, J. (2020). EHR-Independent Predictive Decision Support Architecture Based on OMOP. *Applied clinical informatics*, *11*(3), 399–404. https://doi.org/10.1055/s-0040-1710393

US Food and Drug Administration. (2022). Real-World Evidence. Verfügbar 12. Mai 2023 unter https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

Vashisht, R., Jung, K., Schuler, A., Banda, J. M., Park, R. W., Jin, S., Li, L., Dudley, J. T., Johnson, K. W., Shervey, M. M., Xu, H., Wu, Y., Natrajan, K., Hripcsak, G., Jin, P., Van Zandt, M., Reckard, A., Reich, C. G., Weaver, J., … Shah, N. H. (2018). Association of Hemoglobin A_{1c} Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Diabetes Treated With Metformin: Analysis From the Observational Health Data Sciences and Informatics Initiative. *JAMA Network Open*, *1*(4), e181755. https://doi.org/10.1001/jamanetworkopen.2018.1755

Vass, A., Reinecke, I., Boeker, M., Prokosch, H.-U., & Gulden, C. (2022). Availability of Structured Data Elements in Electronic Health Records for Supporting Patient Recruitment in Clinical Trials. In P. Otero, P. Scott, S. Z. Martin & E. Huesing (Hrsg.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI220046

Vidal MMI Germany GmbH. (2023). Gelbe Liste. https://www.gelbe-liste.de/

Viernes, B., Lynch, K. E., South, B., Coronado, G., & DuVall, S. L. (2019). Characterizing VA Users with the OMOP Common Data Model [ISSN: 0926-9630 Type: Proceedings Paper]. In OhnoMachado, L and Seroussi, B (Hrsg.), *MEDINFO 2019: HEALTH AND WELLBEING E-NETWORKS FOR ALL* (S. 1614–1615). IOS PRESS. https://doi.org/10.3233/SHTI190561

Von Lucadou, M., Ganslandt, T., Prokosch, H.-U., & Toddenroth, D. (2019). Feasibility analysis of conducting observational studies with the electronic health record. *BMC Medical Informatics and Decision Making*, *19*(1), 202. https://doi.org/10.1186/s12911-019-0939-0

Warner, J. L., Dymshyts, D., Reich, C. G., Gurley, M. J., Hochheiser, H., Moldwin, Z. H., Belenkaya, R., Williams, A. E., & Yang, P. C. (2019). HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *Journal of biomedical informatics*, *96*, 103239. https://doi.org/10.1016/j.jbi.2019.103239

Weissgerber, T. L., Garovic, V. D., Savic, M., Winham, S. J., & Milic, N. M. (2016). From Static to Interactive: Transforming Data Visualization to Improve Transparency. *PLOS Biology*, *14*(6), e1002484. https://doi.org/10.1371/journal.pbio.1002484

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*, *13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128

Weissgerber, T. L., Winham, S. J., Heinzen, E. P., Milin-Lazovic, J. S., Garcia-Valencia, O., Bukumiric, Z., Savic, M. D., Garovic, V. D., & Milic, N. M. (2019). Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency. *Circulation*, *140*(18), 1506–1518. https://doi.org/10.1161/CIRCULATIONAHA.118.037777

WHO Collaborating Centre for Drug Statistics and Methodology. (2022). WHO ATC alterations from 2005-2023. https://www.whocc.no/atc_ddd_alterations__cumulative/atc_alterations/?order_by=3&d=DESC

WHO International Working Group for Drug Statistics Methodology. (2022). *Introduction to drug utilization research*. https://www.who.int/publications/i/item/8280820396

Wissenschaftliches Institut der AOK (WIDO). (2023). ATC-Klassifikation für den deutschen Arzneimittelmarkt. Verfügbar 14. Mai 2023 unter https://www.wido.de/publikationen-produkte/arzneimittel-klassifikation/

World Health Organization (WHO). (2023). World Health Organization (WHO) ATC Classification. Verfügbar 2. Januar 2023 unter https://www.who.int/tools/atc-ddd-toolkit/atc-classification

Y. Zhang, J. Li & M. V. Zandt. (2020). NCCD-RxNorm: Linking Chinese Clinical Drugs to International Drug Vocabulary [Journal Abbreviation: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)]. *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1752–1756. https://doi.org/10.1109/BIBM49941.2020.9313599

Yang, Y., Zhou, X., Gao, S., Lin, H., Xie, Y., Feng, Y., Huang, K., & Zhan, S. (2018). Evaluation of Electronic Healthcare Databases for Post-Marketing Drug Safety Surveillance and Pharmacoepidemiology in China. [Place: New Zealand]. *Drug safety*, *41*(1), 125–137. https://doi.org/10.1007/s40264-017-0589-z

Yoon, D., Ahn, E. K., Park, M. Y., Cho, S. Y., Ryan, P., Schuemie, M. J., Shin, D., Park, H., & Park, R. W. (2016). Conversion and Data Quality Assessment of Electronic Health Record Data at a Korean Tertiary Teaching Hospital to a Common Data Model for Distributed Network Research. *Healthcare Informatics Research*, *22*(1), 54. https://doi.org/10.4258/hir.2016.22.1.54

You, S. C., Lee, S., Cho, S.-Y., Park, H., Jung, S., Cho, J., Yoon, D., & Park, R. W. (2017). Conversion of National Health Insurance Service-National Sample Cohort (NHIS-NSC) Database into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM). *Studies in Health Technology and Informatics*, *245*, 467–470.

You, S. C., Rho, Y., Bikdeli, B., Kim, J., Siapos, A., Weaver, J., Londhe, A., Cho, J., Park, J., Schuemie, M., Suchard, M. A., Madigan, D., Hripcsak, G., Gupta, A., Reich, C. G., Ryan, P. B., Park, R. W., & Krumholz, H. M. (2020). Association of Ticagrelor vs Clopidogrel With Net Adverse Clinical Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention. *JAMA*, *324*(16), 1640. https://doi.org/10.1001/jama.2020.16167

Yuan, C., Ryan, P. B., Ta, C., Guo, Y., Li, Z., Hardin, J., Makadia, R., Jin, P., Shang, N., Kang, T., & Weng, C. (2019). Criteria2Query: a natural language interface to clinical databases for cohort definition. *Journal of the American Medical Informatics Association : JAMIA*, *26*(4), 294–305. https://doi.org/10.1093/jamia/ocy178

Zautke, A., Bönisch, C., Ammon, D., Räuscher, E., Lautenbacher, H., Saß, J., Buckow, K., Boeker, M., Löbe, M., Semler, S., Wrobel, S., Zabka, S., Thun, S., & Ganslandt, T. (2021). Medizininformatik Initiative - Modul Medikation - ImplementationGuide. Verfügbar 13. Februar 2023 unter https://www.medizininformatik-initiative.de/Kerndatensatz/Modul_Medikation/IGMIIKDSModulMedikation.html

Zhang, X., Wang, L., Miao, S., Xu, H., Yin, Y., Zhu, Y., Dai, Z., Shan, T., Jing, S., Wang, J., Zhang, X., Huang, Z., Wang, Z., Guo, J., & Liu, Y. (2018). Analysis of treatment pathways for three chronic diseases using OMOP CDM. *Journal of Medical Systems*, *42*(12), 260. https://doi.org/10.1007/s10916-018-1076-5

Zozus, M., Hammond, W. E., Green, B. B., Kahn, M. G., Richesson, R. L., Rusincovitch, S. A., Simon, G. E., & Smerek, M. M. (2015). Data Quality Assessment Recommendations for Secondary ise of EHR Data [Publisher: Unpublished]. https://doi.org/10.13140/RG.2.1.4157.7689

Abbildungsverzeichnis

Tabellenverzeichnis

A Anhang: Quellcode Readme

dissertation-code

Dieses Projekt ist Bestandteil der Dissertation von Ines Reinecke, vorgelegt am 11.07.2023 an der Technischen Universität Dresden, Medizinische Fakultät

Struktur des Projektes

Der Ordner *data_in* enthält die Daten, welche von den Jupyter Notebooks gelesen werden. Nicht enthalten in diesem Projekt sind die Dateien:

- DS-Med initiale Version
- DS-Med Version mit verbesserter Datenstruktur nach Durchführung der entsprechenden Maßnahmen (Zuordnung der ATC Codes und Validierung der Ergebnisse für die TOP1000)
- DS-Katalog

Der Ordner *data_results* enthält die Ergebnisse der Jupyter Notebooks. Die Ordner 00*, 01*, 02* und 03* enthalten die Jupyter Notebooks zur Datenanalyse. Im Folgenden sind die Scripte einzeln beschrieben.

Ordner 00_literatur

Hier sind die Analysen der Literatur des Scoping Reviews von Reinecke et al. enthalten. Das Script 00_literatur.ipynb enthält die komplette Analyse und Visualisierung der in das Scoping Review eingeschlossenen Literatur (Methoden Kapitel 3.2, Resultate Kapitel 4.1).
Dieses Script 00_literatur.ipynb umfasst:

- Einlesen der eingeschlossenen Literatur gemäß der Liste auf Zenodo (Reinecke, 2021b) aus dem Scoping Review von Reinecke et al. (Reinecke, Zoch, Reich, Sedlmayr et al., 2021)
- Barplot mit den Jahren, Anzahl der Publikation und die fachliche Kategorie (Medizin, Medizininformatik, Informatik, Sonstiges)
- Weltkarte mit Anzahl Publikationen pro Land
- Barplot, Publikationen der Dimension Nutzung - Kategorien, pro Jahr
- Barplot, Publikationen der Dimension Nutzung, sortiert nach Kategorien und Anzeige der Multi-country bzw. single-country Datennutzung
- Sankey Diagramm - Dimensionen und Kategorien aller Publikation

Das Script 01_ohdsi_studies_analysis.ipynb enthält die Analyse und Visualisieruung der OHDSI Netzwerkstudien (Methoden Kapitel 3.3.2, Resultate Kapitel 4.2.2). Dieses Script enthält die Visualisierung der in OHDSI Studien verwendeten Datengruppen als Scatterplot in Kombination mit einem Histogramm zur Anzeige der Summe der Datengruppen über alle OHDSI Studien hinweg. Dieses Script ist in Anlehnung an die Idee zur Visualisierung von Najia Ahmadi implementiert worden. Vorbild dieser Visualisierung ist das GitHub Projekt von Najia Ahmadi, Release V 1.0, hier:
https://github.com/NajiaAhmadi/VisualisationWithPython/releases/tag/v1.0

Ordner 01_data_structure

Hier sind alle Jupyter Notebooks, die im Rahmen der Durchführung der Maßnahmen zur Verbesserung der Datenstruktur der Medikationsverordnungen (Methode Kapitel 3.5.2) implementiert wurden, abgelegt.
Script 01_Datenstruktur_Algorithmen_Implementierung.ipynb beinhaltet:
Zunächst werden die Rohdaten aus Datensatz DS-Med (Medikationsverordnungen) und Datensatz DS-Katalog (Arzneimittelkatalog des UKD) eingelesen.
Anschließend wird der Datensatz DS-Med wie folgt im Detail geprüft:

- Summe der Arzneimittelverordnungen der Jahre 2016 bis 2020
- Ermittlung der fehlenden Arzneimitteleinträge in den Arzneimittelverordnungsdaten
- Ermittlung der Menge der unstrukturierten Arzneimittelverordnungsdaten
- Gruppierung der unstrukturierten Arzneimittelverschreibungsdaten nach Medikationstext und Berechnung der Häufigkeit
- Überprüfung der Gesamtzahl der verschiedenen unstrukturierten Einträge für den Medikationstext
- Verteilung der Freitext-Arzneimittelverordnungen nach Häufigkeit auswerten
- Prüfung, ob die ersten 1000 häufigsten unstrukturierten Arzneimittelverordnungen für die manuelle Auswertung ausreichen, um das Ziel von 80% aller Arzneimittelverordnungen mit ATC Code zu erreichen
- Ausführen des Algorithmus auf den unstrukturierten Daten zur Bestimmung des ATC-Codes in STEP1 (Regex-Medikamentenprodukt), STEP2 (Inhaltsstoff) und STEP 3 (NLP basierend auf Ähnlichkeit mit Levenshtein-Distanz) - STEP 3 liefert bis zu 3 verschiedene vorgeschlagene ATC-Codes

- Ergebnisse zurückgeben und die häufigsten 1000 Einträge und Ergebnisse für Algorithmus 1, 2 und 3 erstellen
- Vorbereitung der Zahlen für die Visualisierung im Venn Chart, Übereinstimmung der Ergebnisse der Algorithmen

Script 02_Datenstruktur_Ergebnisse_Visualisierung_Übereinstimmung.ipynb beinhaltet:

- Visualisierung der Übereinstimmung der Ergebnisse der Algorithmen auf die Medikationsverordnungen
- Visualisierung Venn-Diagramm, Abbildung 4.9, Kapitel Ergebnisse, Maßnahmen - Datenstruktur, Algorithmen (Kapitel 4.4.2.1)

Script 03_t-test-Algorithmus3-LevenshteinÄhnlichkeit.ipynb beinhaltet:

- Eingelesene Daten - Datensatz DS-Top1000
- Generieren von zwei Dataframes für die korrekten und nicht korrekten Ergebnisse von Algorithmus 3
- statistische Informationen generieren für die beiden neuen Dataframes in Bezug auf Algorithmus 3 und den Ähnlichkeitswert von Levenshtein
- Durchführung eines beidseitigen t-tests um zu prüfen, ob sich die Ähnlichkeitswerte des Algorithmus 3 für die beiden Ergebnismengen signifikant unterscheiden

Script 04_ATC_Codes_Anwenden_DatenVisualisierung.ipynb beinhaltet:

- Einlesen des Datensatzes DS-Med und DS-Top1000
- Generierung einer neuen Spalte „ATC-Correct“ in DS-Top1000
- Zusammenführen (Merge) der beiden Datensätze DS-Med und DS-Top1000 basierend auf der Spalte „MEDICATION“ - nur für die unstrukturierten Medikationsverordnungen
- Generierung eines finalen Datensatzes von DS-Med, bei dem alle ATC Codes, die durch Algorithmus 3 zugeordnet werden, in einer Spalte enthalten sind - nur für die Freitexte, die Teil von DS-Top1000 sind und manuell validiert wurden
- Generierung eines Datensatzes als Eingangsgröße für das Streudiagramm der Visualisierung - Strukturiertheit auf Basis von ATC Code

Ordner 02_data_to_omop+terminology

Script 00_initiale-DS-Med-to-omop.ipynb beinhaltet:

- Laden des Datensatzes DS-Med und des ATC-GM nach ATC WHO Mappings basierend

auf dem Datensatz DS-Katalog (das Mapping wurde basierend auf diesem Datensatz bereitgestellt)

- Ersetzen mit dem ATC WHO Code aller ATC-GM Codes wo möglich und wo der ATC WHO Code im Mapping vorhanden und anders als der ATC-GM Code ist
- Laden von ATC Vokabulars aus OMOP (WHO und deutsche Version)
- Zusammenführen von DS-Med mit dem ATC Vokabular basierend auf dem ATC Code - hinzufügen von der validen concept_id
- Generieren eines OMOP konformen Datenformats des Datensatzes DS-Med für die OMOP Tabelle „drug_exposure“ -> Datenbasis für den Schritt 1 der Bewertung mit dem OHDSI DQD Dashboard
- Speicherung des OMOP konformen Datenformats von Datensatz DS-Med

Script 01_verbesserte-DS-Med-to-omop.ipynb beinhaltet:

- Laden des Datensatzes DS-Med (nach Durchführung der Maßnahmen zur Verbesserung der Datenstruktur) und des ATC-GM nach ATC WHO Mappings basierend auf dem Datensatz DS-Katalog (das Mapping wurde basierend auf diesem Datensatz bereitgestellt)
- Ersetzen mit dem ATC WHO Code aller ATC-GM Codes wo möglich und wo der ATC WHO Code im Mapping vorhanden und anders als der ATC-GM Code ist
- Laden von ATC Vokabulars aus OMOP (WHO und deutsche Version)
- Zusammenführen von DS-Med mit dem ATC Vokabular basierend auf dem ATC Code - hinzufügen von der validen *concept_id*
- Generieren eines OMOP konformen Datenformats des Datensatzes DS-Med für die OMOP Tabelle „drug_exposure“ -> Datenbasis für den Schritt 1 der Bewertung mit dem OHDSI DQD Dashboard
- Speicherung des OMOP konformen Datenformats von Datensatz DS-Med - Schritt2: Verbesserte Datenstruktur der Medikationsverordnungen

Script 02_RxNorm-Transfer-DS-Med-omop.ipynb beinhaltet:

- Laden des Datensatzes DS-Med (nach Durchführung der Maßnahmen zur Verbesserung der Datenstuktur) bereits im OMOP Format (Eingangsgröße hier das Ergebnis von Script 01_verbesserte-DS-Med-to-omop)
- Laden der ATC nach RxNorm Mappings
- Zusammenführen der Medikationsverordnungen mit den Mappings

- Speichern der Medikationsverordnungen, verbessert und mit den *concept_ids* in der Spalte *drug_concept_id* mit RxNorm Standard-Terminologie, wenn möglich - im OMOP Format

Ordner 03_data_transparency

Script 00_Streudiagramm-Struktur.ipynb beinhaltet:

- Einlesen der Daten DS-Med als Eingangsgröße (scatter_input)
- Einlesen der ATC Codes Version 2022 - mit den entsprechenden ATC Beschreibungen in Deutsch
- Zusammenführen der Medikationsverordnungen mit den ATC Codes und den Beschreibungen
- Generieren eines Streudiagramms mit der Bibliothek Bokeh, interaktiv

Script 01_Streudiagramm-Uberfuhrbarkeit-RxNorm.ipynb beinhaltet:

- Einlesen der Daten DS-Med als Eingangsgröße, nach Durchführung der Maßnahmen zur Verbesserung der Datenstruktur, im OMOP Format
- Einlesen der ATC Codes Version 2022 - mit den entsprechenden ATC Beschreibungen in Deutsch
- Zusammenführen der Medikationsverordnungen mit den ATC Codes und den Beschreibungen
- Generieren eines Streudiagramms mit der Bibliothek Bokeh, interaktiv

B Anhang: drug-exposure Tabelle - Wiki Dokumentation

DRUG_EXPOSURE

Table Description

This table captures records about the exposure to a Drug ingested or otherwise introduced into the body. A Drug is a biochemical substance formulated in such a way that when administered to a Person it will exert a certain biochemical effect on the metabolism. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Radiological devices ingested or applied locally do not count as Drugs.

User Guide

The purpose of records in this table is to indicate an exposure to a certain drug as best as possible. In this context a drug is defined as an active ingredient. Drug Exposures are defined by Concepts from the Drug domain, which form a complex hierarchy. As a result, one DRUG_SOURCE_CONCEPT_ID may map to multiple standard concept ids if it is a combination product. Records in this table represent prescriptions written, prescriptions dispensed, and drugs administered by a provider to name a few. The DRUG_TYPE_CONCEPT_ID can be used to find and filter on these types. This table includes additional information about the drug products, the quantity given, and route of administration.

ETL Conventions

Information about quantity and dose is provided in a variety of different ways and it is important for the ETL to provide as much information as possible from the data. Depending on the provenance of the data fields may be captured differently i.e. quantity for drugs administered may have a separate meaning from quantity for prescriptions dispensed. If a patient has multiple records on the same day for the same drug or procedures the ETL should not de-dupe them unless there is probable reason to believe the item is a true data duplicate. Take note on how to handle refills for prescriptions written.

C Anhang: Studienprotokoll EMA Studie

Study protocol: Systemic corticosteroids in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients in the primary and secondary care setting

European Medicines Agency

September 20th , 2020

IQVIA

Protocol Approval and Sign-off
I confirm that I have read the contents of this protocol and its attachments. I approve the protocol in its current form.

Epidemiologist & author	Alexandra Pacurariu		20th September 2020
	Senior Consultant, EMEA Data Science Hub, IQVIA Ltd.	Signature	Date
Principal Investigator	Deborah Layton		20th September 2020
	Director of Drug Safety, EMEA Data Science Hub, IQVIA Ltd.	Signature	Date
Senior QC	Joseph Kim		20th September 2020

PASS Information

Section	Description
Title	Systemic corticosteroids use in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients in the primary and secondary care setting
Protocol version identifier	Version 1.0
Date of last version of protocol	NA
EU PAS register number	To be registered
Active substance	Steroids (H02AB)
Procedure number	EMA/198302/2020
Research questions and objectives	Primary objective To describe utilization of systemic corticosteroids (e.g., dexamethasone, prednisolone, methylprednisolone or hydrocortisone) for treatment of COVID-19 in two types of setting: hospitalized (in hospital care) and ambulatory (any care received outside the hospital) within 90 days following COVID-19 diagnosis. Secondary objectives 1. To describe at COVID-19 diagnosis date the demographic, health and clinical patient characteristics (stratified by setting and systemic corticosteroid user type (naive, prevalent). 2. To quantify the crude and adjusted incidence rates and time to onset of adverse events of interest (i.e., infections, -Hyperglycemia, hypertension, GI bleeding and composite cardiovascular disease events) within 30 and 90 days post treatment index date, by setting, in various treatment groups, systemic corticosteroid user type (naive, prevalent) and sub-populations of special interest. 3. To quantify the crude and adjusted incidence rates of mortality and other disease outcomes within 30- and 90-days post treatment index date, by setting, in various treatment groups, systemic corticosteroid user type (naive, prevalent) and sub-populations of special interest.

	4. To explore and compare different coding definition for COVID-19 over time and in different databases, gathering information on tests used in different countries (differing validity and reliability), to inform the choice of the definition to be used for future project.
Country(-ies) of study	Belgium, France, Germany, UK, Italy, Netherlands, Spain
Author	IQVIA

Section 8.0 Abstract

Section	Description
Title	Systemic corticosteroids use in the treatment of COVID-19 and risks of adverse outcomes in COVID-19 patients, within primary and secondary care settings
Rationale and background	Approximately 10-20% of COVID-19 positive patients, many of whom are older or have co-morbidities, suffer from pneumonia and acute respiratory distress syndrome (ARDS), requiring hospitalization and ventilation support. It has been suggested that this population is also at higher risk of inflammatory immune system disorders. As a result, current treatment recommendations are to combine anti-viral therapy *and* immunosuppressive or immunomodulatory drugs to mitigate these immunologic complications, reducing COVID-19 associated morbidity and mortality. While the search for appropriate anti-viral therapy is ongoing, there have been some positive results with respect to systemic corticosteroid use, such as dexamethasone, which has been associated with reduced mortality in ventilated patients and those on supplemental oxygen therapy. This has mobilised efforts to repurpose some of these steroids for the treatment of severe COVID-19 cases. That said, a lot of information on steroid use in COVID-19 patients is currently missing. Treatment type, dosage, timing of administration, as well as identification of patient risk groups that will benefit most from the treatments, is inadequately explored. To address these research gaps, this protocol describes a study to explore patterns of systemic corticosteroid use and administration in COVID-19 positive patients using healthcare databases from seven European countries.
Research question and objectives	The aim of this study is to describe patterns of systemic corticosteroid use, as well as the risks of adverse events associated with these medications, in COVID-19 positive patients across seven European countries in ambulatory and hospital inpatient care settings. **Primary Objective**: To describe utilization of systemic corticosteroids (e.g., dexamethasone, prednisolone, methylprednisolone or hydrocortisone) for treatment of COVID-19 in two settings: hospitalized (in hospital care) and ambulatory (any care received outside of hospital) within 90 days following COVID-19 diagnosis. The following variables will be described: • Prevalent use or naïve (incident) use of systemic corticosteroid at date of diagnosis of COVID-19 • Concomitant use of other medications (number and type) and invasive/non-invasive respiratory support during follow up • Type of systemic corticosteroid received • Time to systemic corticosteroid initiation from COVID-19 diagnosis, • Route of administration

	• Systemic corticosteroid daily dose at initiation (treatment index date), cumulative duration, distribution of duration of use and cumulative dose of systemic corticosteroid received • For prevalent users: proximity of previous corticosteroid use to COVID-19 diagnosis (current use (concomitant on date of COVID-19 diagnosis, recent use (between 15-30 days before date of COVID-19 diagnosis) or remote use (use ended more than 30 days before date of COVID-19 diagnosis)) **Secondary Objectives:** 1. To describe **at COVID-19 diagnosis date and at treatment index date** the demographic, health and clinical patient characteristics (stratified by setting and systemic corticosteroid user type (naive, prevalent). The following variables will be described: - demographics - comorbidities (number and type) - symptoms (number and type) preceding and/or on the date of diagnosis the diagnosis if captured - time from onset of COVID-19 illness symptoms to confirmed diagnosis date Some of these characteristics will be stratified by subgroups of special interest (chronic cardiac and pulmonary disease, diabetes, renal insufficiency). 2. To quantify the crude and adjusted incidence rates and time to onset of adverse events of interest (e.g., infections, hyperglycaemia, composite GI bleeding, total cardiovascular disease events) within **30 and 90 days post treatment index date**, by setting, in **various treatment groups**, systemic corticosteroid user type (naive, prevalent) and sub-populations of special interest. 3. To quantify the crude and adjusted incidence rates of mortality and other disease outcomes within **30 and 90 days post treatment index date**, by setting, in **various treatment groups**, systemic corticosteroid user type (naive, prevalent) and sub-populations of special interest. 4. To explore and compare **different coding definitions for COVID-19** over time and in different databases, gathering information on tests used in different countries (differing validity and reliability), to inform the choice of the definition to be used for future project.
Study design	This is a PASS study using a descriptive cohort study design using secondary data sources (electronic medical records).

Setting	Patients diagnosed with COVID-19 across primary and secondary care settings in seven European Countries (Belgium, France, Italy, Netherlands, Germany, United Kingdom, Spain), with the study time period from 1st January 2020 to 1st January 2021 (at the latest), will be considered for analysis. Cut-off dates for data inclusion i.e. data lock points will vary based on the country and database used. Four different cohorts based on healthcare setting (ambulatory or hospital setting) and systemic corticosteroid user type (naïve or prevalent) will be created. Index dates, outcomes and follow-up censoring vary per objective and will be appropriately applied in the analyses.

Variables	• **COVID-19 Case Definition**
	  ○ Main analysis: Catch-all definition based on the earliest of a medical code or SARS-CoV-2 positive test
	  ○ Secondary Objective alternative definitions: medical code only, SARS-CoV-2 positive test only, and medical code and SARS-CoV-2 positive test (where available)
	• **Exposure** (based on prescription data)
	  ○ COVID-19 specific corticosteroids (dexamethasone, prednisolone, prednisone, methylprednisolone or hydrocortisone)
	  ○ Corticosteroids for pre-existing conditions, described by metrics such as proportion of days covered in lookback period and recent use of medication based on prescription records.
	  ○ Other COVID-19 treatments e.g. antiviral, antibiotic, statin therapy.
	  ○ Respiratory support
	These will be further characterized using specific criteria and formulae for dose and duration.
	For secondary objectives 2 and 3 where the index date is based on treatment, the treatment exposure groups will be categorized as follows:
	  - Use of systemic corticosteroids without other treatments for COVID-19
	  - Use of systemic corticosteroids plus other treatments for COVID-19
	  - Only other treatments and respiratory support with no systemic corticosteroids
	  - No specific treatments for COVID-19 infection
	• **Outcomes** (Secondary objectives 2 and 3)
	  ○ Adverse events: Bacterial/Fungal infections, Pneumonia, Hyperglycaemia, Hypertension, GI bleeding, cardiovascular disease events
	  ○ Disease:

	▪ Ambulatory: Hospital admission, venous thromboembolism, death of any cause ▪ Hospital: Intensive services as an outcome in inpatient cohorts (including mechanical ventilation and ECMO), venous thromboembolism, discharge from hospital, death of any cause • Demographic and clinical variables: age, sex, month of diagnosis, comorbidities, records of respiratory support.
Data sources	• Belgium, France, Italy – Longitudinal Patient Database (IQVIA) • Germany – Disease Analyser (IQVIA) • United Kingdom – IQVIA Medical Research Data (IMRD) UK • The Netherlands – Integrated Primary Care Information (IPCI) • Spain – Information System for Research in Primary Care (SIDIAP) and Hospital de Madrid Hospitales providing hospital inpatient data Details on each database can be found in Section 7.4.
Data analysis	Continuous variables will be described using mean, standard deviation, median, first and third quartiles, minimum, and maximum. Categorical variables will be described by the number and percentage of patients in each category. Patterns of missingness will be reported and 95% confidence Intervals will be presented. The results for each country and database will be presented separately. • **Primary Objective:** o Descriptive analysis for systemic corticosteroid use patterns will be carried out across the four primary patient cohorts and subgroups of special interest. o Kaplan-Meier methods will be used to estimate time to systemic corticosteroid initiation from COVID-19 diagnosis, stratified by route of administration. • **Secondary Objective 1**: o Cohort-specific descriptive statistics summarizing demographic, health and clinical patient characteristics, stratified by setting, systemic corticosteroid user type (naive, prevalent), and subgroups of special interest will be presented. • **Secondary Objectives 2 and 3:** o Crude incidence risks (presented as both proportions and rates) for the relevant outcomes for each of the treatment exposure groups will be calculated. o The cumulative incidence rates will be reported at the end of follow-up (30 and 90 days). o Data will be stratified by subgroups of special interest to examine risk factors for the outcomes being investigated. o Cox regression models (using multivariable models and using PS adjustment) will be used to compute the adjusted incidence rates.

	• **Secondary Objective 4:** ○ To explore and compare the concordance of different COVID-19 disease definitions, we will use concordance statistics.
Plans for Disseminating and Communicating Study Results	IQVIA will produce a study report in accordance with the GVP guidelines VIII (EMA/813938/2011). Study information (including protocol and final report) will be added in the EU PAS register.

13.5.4 Exposures

The following exposures of interest are captured in this study

	Cohort	When is measured
Corticosteroids specific for COVID-19	All	At diagnosis index date and follow-up
Corticosteroids for pre-existing conditions	Only in prevalent users' cohort	Medical history before diagnosis index date
Other COVID-19 treatments	All	and follow-up
Oxygen therapies	All	At diagnosis index date and follow-up*

* During follow-up this will be considered as an outcome

7.3.3.1 Corticosteroids

Corticosteroids specific for COVID-19

The primary treatment group of interest is based on exposure to systemic corticosteroid. Patients will be considered exposed to a systemic corticosteroid for treatment of COVID-19 if available prescription data are available for the following drugs on or post COVID-19 diagnosis

- Dexamethasone (H02AB02)
- Prednisone/Prednisolone (A07EA01)
- Methylprednisolone (H02AB04)
- Hydrocortisone (H02AB09)

Corticosteroids for pre-existing conditions

For the prevalent users, any steroid in the ATC class H02 corticosteroids for systemic use will be considered. A patient is defined as a systemic corticosteroid prevalent user if it has at least one prior exposure to systemic[1] corticosteroid in the 365 days prior to diagnosis index date.

For the description of prevalent users of corticosteroids, categories of use will be created based on proportion of days covered (PDC) during the lookback period. PDC is calculated as the number of days in period covered by prescriptions divided by number of days in the period.

- PDC > 85% as heavy user
- PDC 50-85% as moderate user
- less than 50% PDC as light user

Another categorization of prevalent corticosteroids user will be based on how recently in the past the last prescription was observed:

- current use (concomitant on date of COVID-19 diagnosis),

[1] Systemic effects of inhaled and topical use of corticosteroids is considered low and therefore use will not be considered for this study

D Anhang: Medikationsverordnungen ATC Codes Strukturiertheit

ATC-GM Code	Unstrukturiert	Strukturiert	Gesamt	Anteil strukturiert (%)	Anteil unstrukturiert (%)
N02BB02	35961	44905	80866	55.53	44.47
B05BB01	21954	46345	68299	67.86	32.14
A02BC02	55371	10490	65861	15.93	84.07
N02AA05	1482	44952	46434	96.81	3.19
B01AB13	5862	28329	34191	82.86	17.14
B01AB05	23583	7658	31241	24.51	75.49
C03CA04	16319	12535	28854	43.44	56.56
C07AB02	2870	23602	26472	89.16	10.84
C09AA05	10649	14817	25466	58.18	41.82
A06AD11	2952	21113	24065	87.73	12.27
C08CA01	3122	19903	23025	86.44	13.56
B01AC06	2187	19079	21266	89.72	10.28
H02AB02	1682	19304	20986	91.99	8.01
N05CF01	16373	2723	19096	14.26	85.74
M01AE01	15549	3220	18769	17.16	82.84
N02BE01	8004	10661	18665	57.12	42.88
C03CA01	13400	4583	17983	25.49	74.51
H03AA01	2778	14559	17337	83.98	16.02
A06AD65	12850	4357	17207	25.32	74.68
H02AB06	6171	10859	17030	63.76	36.24
N05BA06	6614	10302	16916	60.9	39.1
C10AA01	2371	13890	16261	85.42	14.58
J01CR01	6122	8878	15000	59.19	40.81
A11CC05	1400	11778	13178	89.38	10.62
C07AB07	1322	10458	11780	88.78	11.22
A12BA30	1187	9224	10411	88.6	11.4
C10AA05	5470	4774	10244	46.6	53.4
B01AX05	1270	8630	9900	87.17	12.83
J01CR05	2380	7465	9845	75.83	24.17
C09CA06	4278	5097	9375	54.37	45.63
M04AA01	6764	2551	9315	27.39	72.61
B01AB01	1705	7302	9007	81.07	18.93
C03AA03	895	7971	8866	89.91	10.09
J01DD04	6757	1830	8587	21.31	78.69
N03AX16	5454	2830	8284	34.16	65.84
B01AB06	7004	913	7917	11.53	88.47
A10BA02	603	7087	7690	92.16	7.84
B03BB01	1318	6188	7506	82.44	17.56
N05AD03	1637	5730	7367	77.78	22.22
A12BA01	2998	4082	7080	57.66	42.34
C03DA01	1090	5724	6814	84.0	16.0
A10AE04	851	5619	6470	86.85	13.15
J01DC02	3604	2848	6452	44.14	55.86
B01AF02	489	5889	6378	92.33	7.67
N05AX08	2282	3907	6189	63.13	36.87
A04AB02	2552	3510	6062	57.9	42.1
N03AX14	4772	1288	6060	21.25	78.75

N06AX11	4919	1072	5991	17.89	82.11
N04BA10	854	5136	5990	85.74	14.26
B01AF01	418	5554	5972	93.0	7.0
N05AH04	2607	3337	5944	56.14	43.86
N02AA03	1171	4756	5927	80.24	19.76
A03FA01	3289	2414	5703	42.33	57.67
N02AB03	4566	1130	5696	19.84	80.16
B05BB11	4746	682	5428	12.56	87.44
C02AC05	479	4846	5325	91.0	9.0
A02BC05	671	4509	5180	87.05	12.95
B05CB01	2492	2662	5154	51.65	48.35
N02AA01	2059	3071	5130	59.86	40.14
N02AX51	286	4519	4805	94.05	5.95
N05BA01	1779	3011	4790	62.86	37.14
G04CA02	4488	298	4786	6.23	93.77
J05AB01	924	3860	4784	80.69	19.31
J01EE01	409	4329	4738	91.37	8.63
V06DB50	0	4695	4695	100.0	0.0
B01AC04	709	3965	4674	84.83	15.17
C09CA03	1256	3344	4600	72.7	27.3
N05AH03	1583	2976	4559	65.28	34.72
C02CA06	563	3860	4423	87.27	12.73
C09CA01	97	4283	4380	97.79	2.21
A10BH01	388	3899	4287	90.95	9.05
J01CR02	361	3829	4190	91.38	8.62
A04AA02	1143	3038	4181	72.66	27.34
B05BA10	2600	1431	4031	35.5	64.5
R05CB01	773	3227	4000	80.68	19.32
C03BA10	3477	475	3952	12.02	87.98
J01MA02	325	3398	3723	91.27	8.73
A10AB01	1086	2609	3695	70.61	29.39
J01MA12	362	3175	3537	89.77	10.23
A11DA01	1485	1940	3425	56.64	43.36
J01XD01	1526	1816	3342	54.34	45.66
N06AB06	808	2532	3340	75.81	24.19
A02AH01	393	2892	3285	88.04	11.96
A06AB08	2676	553	3229	17.13	82.87
A12AX01	240	2952	3192	92.48	7.52
B03AA01	264	2925	3189	91.72	8.28
A01AB04	0	3176	3176	100.0	0.0
C01AA04	686	2462	3148	78.21	21.79
N05AH02	2962	183	3145	5.82	94.18
N01BB09	2966	126	3092	4.08	95.92
B05BB02	1955	1115	3070	36.32	63.68
J01XX01	167	2895	3062	94.55	5.45
N02AC03	476	2563	3039	84.34	15.66
A02BA02	2953	0	2953	0.0	100.0
N04BA11	191	2714	2905	93.43	6.57
C08CA08	529	2338	2867	81.55	18.45

C08CA13	425	2432	2857	85.12	14.88
L04AD02	357	2452	2809	87.29	12.71
B01AB10	2647	107	2754	3.89	96.11
A11CC04	493	2259	2752	82.09	17.91
N03AX12	837	1878	2715	69.17	30.83
B01AA04	769	1917	2686	71.37	28.63
A10A	2610	0	2610	0.0	100.0
N03AG01	887	1670	2557	65.31	34.69
A12CC04	2555	0	2555	0.0	100.0
B01AF03	789	1764	2553	69.1	30.9
J02AC01	266	2249	2515	89.42	10.58
J01DH02	1552	925	2477	37.34	62.66
C07AG02	78	2379	2457	96.83	3.17
N07XB52	132	2289	2421	94.55	5.45
R06AB03	1079	1338	2417	55.36	44.64
B03BA01	1713	696	2409	28.89	71.11
B02BA01	208	2189	2397	91.32	8.68
J01FA09	493	1883	2376	79.25	20.75
A09AA02	511	1800	2311	77.89	22.11
C07AB12	158	2147	2305	93.15	6.85
N05AD01	637	1659	2296	72.26	27.74
V03AE01	247	2021	2268	89.11	10.89
N06AB04	2181	71	2252	3.15	96.85
R03AC02	1525	699	2224	31.43	68.57
A05AA02	106	2077	2183	95.14	4.86
J01XA01	911	1175	2086	56.33	43.67
V06XX02	2028	0	2028	0.0	100.0
L03AA10	539	1488	2027	73.41	26.59
N05AN01	268	1744	2012	86.68	13.32
C09AA02	126	1879	2005	93.72	6.28
J01FF01	293	1702	1995	85.31	14.69
N06AX16	198	1784	1982	90.01	9.99
H02AB09	452	1514	1966	77.01	22.99
C10AA04	959	989	1948	50.77	49.23
L04AA06	1557	382	1939	19.7	80.3
A12CC11	158	1746	1904	91.7	8.3
H02AB04	740	1093	1833	59.63	40.37
M01AH04	183	1622	1805	89.86	10.14
N03AX09	118	1653	1771	93.34	6.66
A07DA03	1600	165	1765	9.35	90.65
R03BB01	951	812	1763	46.06	53.94
C09DB01	793	960	1753	54.76	45.24
A10AB04	181	1497	1678	89.21	10.79
N05CD02	1498	143	1641	8.71	91.29
N04BA13	180	1426	1606	88.79	11.21
A06AD15	1603	0	1603	0.0	100.0
C02CA04	102	1446	1548	93.41	6.59
N06AX21	182	1360	1542	88.2	11.8
A03BB01	376	1159	1535	75.5	24.5

R05DA04	906	624	1530	40.78	59.22
V03AE02	1375	147	1522	9.66	90.34
R03BB04	0	1501	1501	100.0	0.0
B03XA02	159	1328	1487	89.31	10.69
L04AD01	436	1026	1462	70.18	29.82
A12CB05	85	1359	1444	94.11	5.89
M04AA03	1248	189	1437	13.15	86.85
A10BD07	0	1435	1435	100.0	0.0
N05AD05	1046	372	1418	26.23	73.77
C09AA03	165	1246	1411	88.31	11.69
D07AC18	1102	303	1405	21.57	78.43
R05CB06	453	949	1402	67.69	32.31
A06AB02	435	961	1396	68.84	31.16
A05BA17	99	1249	1348	92.66	7.34
A10AB05	133	1193	1326	89.97	10.03
A04AA01	148	1178	1326	88.84	11.16
A02AD02	81	1242	1323	93.88	6.12
C09BA25	102	1221	1323	92.29	7.71
H03BB02	407	902	1309	68.91	31.09
B05XA01	572	721	1293	55.76	44.24
R06AE07	520	749	1269	59.02	40.98
B05AA01	75	1182	1257	94.03	5.97
N02AX02	537	720	1257	57.28	42.72
N06AB10	236	982	1218	80.62	19.38
J01CF05	258	952	1210	78.68	21.32
A11EB01	181	1024	1205	84.98	15.02
A10AE05	121	1077	1198	89.9	10.1
C10AX09	559	630	1189	52.99	47.01
C09DA26	972	202	1174	17.21	82.79
C09CA07	79	1081	1160	93.19	6.81
A06AX01	257	901	1158	77.81	22.19
N02AE01	88	1045	1133	92.23	7.77
L03AA13	140	970	1110	87.39	12.61
G04BD09	0	1098	1098	100.0	0.0
M03BX01	0	1072	1072	100.0	0.0
R03CC03	109	943	1052	89.64	10.36
C10AA03	780	266	1046	25.43	74.57
P01AB01	853	176	1029	17.1	82.9
R03AL04	101	895	996	89.86	10.14
A02BC01	991	0	991	0.0	100.0
N05AX12	495	492	987	49.85	50.15
A10AC01	445	515	960	53.65	46.35
R01AA07	527	411	938	43.82	56.18
J01CE01	402	532	934	56.96	43.04
R03AK08	128	796	924	86.15	13.85
R04AX01	0	913	913	100.0	0.0
N03AF01	84	815	899	90.66	9.34
B03AC01	181	714	895	79.78	20.22
S01XC05	0	891	891	100.0	0.0

B05BB	887	0	887	0.0	100.0
N03AX18	0	862	862	100.0	0.0
H05BX01	120	732	852	85.92	14.08
J06BA02	100	739	839	88.08	11.92
A04AD12	722	103	825	12.48	87.52
A03FA03	685	133	818	16.26	83.74
B05XA02	0	814	814	100.0	0.0
A06AC01	0	807	807	100.0	0.0
H03BC02	90	711	801	88.76	11.24
R03AK07	0	796	796	100.0	0.0
C08CA05	113	682	795	85.79	14.21
V07AB10	270	508	778	65.3	34.7
L01AA01	751	27	778	3.47	96.53
N06DX01	84	688	772	89.12	10.88
G04CB01	77	658	735	89.52	10.48
R02AA56	513	217	730	29.73	70.27
A07EC02	0	728	728	100.0	0.0
A10BK03	0	718	718	100.0	0.0
C07AA05	102	601	703	85.49	14.51
A03AX13	0	703	703	100.0	0.0
N04BC05	96	605	701	86.31	13.69
A12AA04	558	137	695	19.71	80.29
R04AX03	0	694	694	100.0	0.0
M05BA04	179	513	692	74.13	25.87
V03AE07	0	682	682	100.0	0.0
V03AF01	629	48	677	7.09	92.91
J02AC03	178	496	674	73.59	26.41
J04AB02	0	664	664	100.0	0.0
N06DA03	0	660	660	100.0	0.0
B01AE07	0	652	652	100.0	0.0
M01AH01	0	650	650	100.0	0.0
N06AB03	565	78	643	12.13	87.87
D07AC14	0	641	641	100.0	0.0
R03AL06	0	640	640	100.0	0.0
C01DA02	92	534	626	85.3	14.7
A07AA07	590	33	623	5.3	94.7
N05CH01	472	143	615	23.25	76.75
C02AC01	0	614	614	100.0	0.0
J02AC04	198	403	601	67.05	32.95
C08CA02	0	600	600	100.0	0.0
N06AX05	472	126	598	21.07	78.93
R03BA02	0	590	590	100.0	0.0
N06AA05	227	362	589	61.46	38.54
A07FA02	0	588	588	100.0	0.0
H03CA01	463	122	585	20.85	79.15
A03FA05	345	236	581	40.62	59.38
B05BB13	0	579	579	100.0	0.0
N05BA09	0	575	575	100.0	0.0
C08DA01	0	567	567	100.0	0.0

R03AK06	0	556	556	100.0	0.0
S01XA12	0	555	555	100.0	0.0
A12AA20	553	0	553	0.0	100.0
S01BA01	0	543	543	100.0	0.0
J05AB14	0	542	542	100.0	0.0
A10BB12	0	541	541	100.0	0.0
N06AA09	87	450	537	83.8	16.2
C09DA23	345	187	532	35.15	64.85
A12CE02	0	532	532	100.0	0.0
H01AB01	0	527	527	100.0	0.0
N07AA02	0	523	523	100.0	0.0
N05CF02	519	0	519	0.0	100.0
A10AB06	0	511	511	100.0	0.0
C02DB01	0	510	510	100.0	0.0
N05CM22	147	363	510	71.18	28.82
C10BA02	322	179	501	35.73	64.27
C04AD03	0	500	500	100.0	0.0
L04AX01	0	500	500	100.0	0.0
N06DA02	400	96	496	19.35	80.65
M01CX01	290	202	492	41.06	58.94
C01EB17	407	83	490	16.94	83.06
L04AA18	0	483	483	100.0	0.0
A12CC02	0	465	465	100.0	0.0
C09CA08	0	451	451	100.0	0.0
B01AD02	437	12	449	2.67	97.33
R06AX26	158	291	449	64.81	35.19
N06AA12	284	159	443	35.89	64.11
A04AD01	0	438	438	100.0	0.0
B02AA02	0	437	437	100.0	0.0
C02DC01	0	433	433	100.0	0.0
H03AA51	0	429	429	100.0	0.0
V04CZ09	0	429	429	100.0	0.0
R03AC13	157	266	423	62.88	37.12
B05BA01	0	418	418	100.0	0.0
M01AB05	0	416	416	100.0	0.0
J01CA04	0	413	413	100.0	0.0
G04BC50	78	335	413	81.11	18.89
N06AX22	88	319	407	78.38	21.62
C03DA04	307	98	405	24.2	75.8
S01XA20	0	403	403	100.0	0.0
M04AC01	258	139	397	35.01	64.99
V03AF03	104	292	396	73.74	26.26
A07EA06	0	393	393	100.0	0.0
N05CD08	0	391	391	100.0	0.0
A06AG01	0	390	390	100.0	0.0
S01EE01	0	390	390	100.0	0.0
S01ED01	0	388	388	100.0	0.0
G04BD08	0	385	385	100.0	0.0
V07AB01	380	4	384	1.04	98.96

G01AF01	362	11	373	2.95	97.05
N03AF02	0	373	373	100.0	0.0
S01AE01	0	368	368	100.0	0.0
J01DH51	0	364	364	100.0	0.0
J01CA01	0	356	356	100.0	0.0
B05BC01	95	254	349	72.78	27.22
N06DX18	0	348	348	100.0	0.0
V03AE03	0	344	344	100.0	0.0
L01XX05	0	339	339	100.0	0.0
M02AA15	0	338	338	100.0	0.0
N01BB02	221	116	337	34.42	65.58
N03AE01	0	327	327	100.0	0.0
N04BB01	0	326	326	100.0	0.0
A02BX03	0	323	323	100.0	0.0
C01BD01	188	133	321	41.43	58.57
R06AX13	321	0	321	0.0	100.0
N04BC04	0	311	311	100.0	0.0
B03AC07	0	308	308	100.0	0.0
C09DX04	307	0	307	0.0	100.0
N05AL05	191	112	303	36.96	63.04
M05BA08	265	38	303	12.54	87.46
N02BA01	229	73	302	24.17	75.83
C09CA04	0	296	296	100.0	0.0
H01BA02	0	294	294	100.0	0.0
J01AA02	0	292	292	100.0	0.0
J01DD13	0	290	290	100.0	0.0
J01GB03	0	284	284	100.0	0.0
C09AA01	124	160	284	56.34	43.66
L01XC02	270	13	283	4.59	95.41
A02BA03	0	277	277	100.0	0.0
L01BC01	264	12	276	4.35	95.65
N05CM02	0	272	272	100.0	0.0
A12AA03	0	272	272	100.0	0.0
P01BA02	266	0	266	0.0	100.0
R03BA01	0	266	266	100.0	0.0
M03BX04	0	262	262	100.0	0.0
N02AC06	0	262	262	100.0	0.0
B03XA01	0	261	261	100.0	0.0
R01AX30	0	260	260	100.0	0.0
C09BA54	0	257	257	100.0	0.0
C01DA08	242	9	251	3.59	96.41
C01DX12	0	250	250	100.0	0.0
J01DD02	0	243	243	100.0	0.0
M03BX02	0	243	243	100.0	0.0
R03AL02	241	0	241	0.0	100.0
S01CA55	0	241	241	100.0	0.0
N02AF02	241	0	241	0.0	100.0
J01FA10	0	240	240	100.0	0.0
G04BD06	0	232	232	100.0	0.0

A07DA02	226	0	226	0.0	100.0
B05BA11	0	225	225	100.0	0.0
A07AA11	214	9	223	4.04	95.96
N07CA22	0	222	222	100.0	0.0
A11CC03	0	221	221	100.0	0.0
C09DA21	0	217	217	100.0	0.0
G04BC01	0	216	216	100.0	0.0
L01XA01	212	1	213	0.47	99.53
N03AB02	0	211	211	100.0	0.0
R03AL01	0	210	210	100.0	0.0
L02BA01	0	207	207	100.0	0.0
A10AD01	130	76	206	36.89	63.11
M01AH05	0	203	203	100.0	0.0
N06AX12	0	200	200	100.0	0.0
J01CE02	0	199	199	100.0	0.0
D04AB61	195	0	195	0.0	100.0
S01EC03	0	193	193	100.0	0.0
N04BD02	0	191	191	100.0	0.0
N01AX14	0	190	190	100.0	0.0
C05AD01	0	190	190	100.0	0.0
B05XC30	0	188	188	100.0	0.0
S01EC04	0	187	187	100.0	0.0
J01XX08	0	182	182	100.0	0.0
N05BB01	182	0	182	0.0	100.0
C10AX06	0	179	179	100.0	0.0
N03AX11	0	175	175	100.0	0.0
N03AX23	0	175	175	100.0	0.0
L02AE02	0	175	175	100.0	0.0
N06AA06	0	174	174	100.0	0.0
M05BA07	0	174	174	100.0	0.0
N05AA02	0	174	174	100.0	0.0
L02BB03	0	171	171	100.0	0.0
G04CB02	0	170	170	100.0	0.0
N02BA13	0	169	169	100.0	0.0
C09DA24	0	168	168	100.0	0.0
B01AC24	0	167	167	100.0	0.0
N07BB04	167	0	167	0.0	100.0
R03DC03	0	163	163	100.0	0.0
R03AL05	0	161	161	100.0	0.0
A11BA01	160	0	160	0.0	100.0
S01EE03	0	159	159	100.0	0.0
D04AA13	0	158	158	100.0	0.0
L01FA01	157	0	157	0.0	100.0
N04BC09	0	155	155	100.0	0.0
N06BX03	0	153	153	100.0	0.0
A01AD11	0	152	152	100.0	0.0
J01FA01	0	151	151	100.0	0.0
D06BB03	0	151	151	100.0	0.0
B05XA09	0	148	148	100.0	0.0

N03AA03	0	147	147	100.0	0.0
R03AC04	0	145	145	100.0	0.0
J05AH02	0	144	144	100.0	0.0
N06AF04	83	60	143	41.96	58.04
J01CR21	142	0	142	0.0	100.0
R05CB13	0	140	140	100.0	0.0
N06AX17	0	138	138	100.0	0.0
N05AL03	96	36	132	27.27	72.73
V08AA20	0	130	130	100.0	0.0
G04BA04	111	17	128	13.28	86.72
D08AC05	110	18	128	14.06	85.94
C09DA07	127	0	127	0.0	100.0
J01CA12	0	127	127	100.0	0.0
D01AA01	0	126	126	100.0	0.0
N07CA01	0	125	125	100.0	0.0
S01ED66	0	125	125	100.0	0.0
J01MA14	0	124	124	100.0	0.0
L01AA02	120	2	122	1.64	98.36
A10BF01	0	121	121	100.0	0.0
N04BX04	0	119	119	100.0	0.0
L01DB03	118	0	118	0.0	100.0
N05CM27	117	0	117	0.0	100.0
A11DB03	115	0	115	0.0	100.0
A10BB01	0	115	115	100.0	0.0
M05BX04	0	114	114	100.0	0.0
N06AB05	0	113	113	100.0	0.0
S01AD03	102	11	113	9.73	90.27
C10AC01	0	112	112	100.0	0.0
C02KX06	0	110	110	100.0	0.0
L04AA04	110	0	110	0.0	100.0
J05AB06	0	108	108	100.0	0.0
L01XA02	106	2	108	1.85	98.15
L01AA03	107	1	108	0.93	99.07
D01AC01	0	105	105	100.0	0.0
N02AB02	0	104	104	100.0	0.0
N06BC01	0	104	104	100.0	0.0
L01BA01	100	4	104	3.85	96.15
L02BG04	0	102	102	100.0	0.0
P01CX01	0	101	101	100.0	0.0
A02BX02	0	96	96	100.0	0.0
H01BA04	0	95	95	100.0	0.0
L01CA02	94	0	94	0.0	100.0
L01CB01	88	6	94	6.38	93.62
N05AF05	0	94	94	100.0	0.0
L01XD04	93	0	93	0.0	100.0
N04BD03	0	92	92	100.0	0.0
J01CA08	0	91	91	100.0	0.0
B01AC22	0	91	91	100.0	0.0
C03CA03	0	91	91	100.0	0.0

S01ED62	0	90	90	100.0	0.0
A06AH01	0	89	89	100.0	0.0
L01BC02	87	1	88	1.14	98.86
V06BA50	0	88	88	100.0	0.0
H02AA02	0	87	87	100.0	0.0
A07EF01	0	86	86	100.0	0.0
R05CA01	0	86	86	100.0	0.0
C10AB02	79	6	85	7.06	92.94
J01DD14	85	0	85	0.0	100.0
A11GA01	0	84	84	100.0	0.0
J05AB11	0	83	83	100.0	0.0
L04AA34	82	0	82	0.0	100.0
L01AX04	80	0	80	0.0	100.0
N02CC01	0	80	80	100.0	0.0
R06AX27	0	77	77	100.0	0.0
N07BB01	77	0	77	0.0	100.0
S01XC07	0	77	77	100.0	0.0
M01AE02	0	77	77	100.0	0.0
A03FP30	0	74	74	100.0	0.0
N07AA01	0	73	73	100.0	0.0
J01DD01	0	71	71	100.0	0.0
C02AB01	0	69	69	100.0	0.0
C09XA02	0	68	68	100.0	0.0
A11HA02	0	67	67	100.0	0.0
J02AX04	0	66	66	100.0	0.0
N03AX15	0	66	66	100.0	0.0
S01ED69	0	64	64	100.0	0.0
L02BG03	0	63	63	100.0	0.0
N06DP01	0	62	62	100.0	0.0
S01EC01	0	61	61	100.0	0.0
D01AE14	0	58	58	100.0	0.0
N06AG02	0	58	58	100.0	0.0
S01EA05	0	58	58	100.0	0.0
N05BA12	0	57	57	100.0	0.0
C02KX08	0	57	57	100.0	0.0
C03EA21	0	56	56	100.0	0.0
N05AF03	0	55	55	100.0	0.0
L01AX03	0	54	54	100.0	0.0
D07CC01	0	54	54	100.0	0.0
D07AB02	0	54	54	100.0	0.0
A03AB02	0	53	53	100.0	0.0
C01CA01	0	53	53	100.0	0.0
R01AX28	0	52	52	100.0	0.0
V08EA01	0	51	51	100.0	0.0
S01XA02	0	51	51	100.0	0.0
S01BA04	0	50	50	100.0	0.0
N04AA02	0	50	50	100.0	0.0
R03DA04	0	50	50	100.0	0.0
V03AF07	0	49	49	100.0	0.0

J01GB01	0	48	48	100.0	0.0
C01BC04	0	47	47	100.0	0.0
J04BA02	0	47	47	100.0	0.0
G02CB03	0	46	46	100.0	0.0
C01CA17	0	46	46	100.0	0.0
D07AC13	0	46	46	100.0	0.0
B05XA31	0	43	43	100.0	0.0
C03BA04	0	43	43	100.0	0.0
N05AE04	0	43	43	100.0	0.0
N03AX22	0	43	43	100.0	0.0
D01AC20	0	43	43	100.0	0.0
G04CA03	0	42	42	100.0	0.0
N07AB02	0	41	41	100.0	0.0
G03DA04	0	40	40	100.0	0.0
R05CB02	0	40	40	100.0	0.0
B05BA02	0	40	40	100.0	0.0
H05BA01	0	39	39	100.0	0.0
J05AD01	0	39	39	100.0	0.0
J01XE01	0	39	39	100.0	0.0
D06AX01	0	39	39	100.0	0.0
V04CZ10	0	38	38	100.0	0.0
D08AG02	0	38	38	100.0	0.0
V03AE17	0	37	37	100.0	0.0
J02AA01	0	36	36	100.0	0.0
L04AB01	0	36	36	100.0	0.0
A01AD69	0	36	36	100.0	0.0
D05AX02	0	36	36	100.0	0.0
D01AA91	0	35	35	100.0	0.0
R03AA01	0	35	35	100.0	0.0
N04BC08	0	35	35	100.0	0.0
J06BB01	0	34	34	100.0	0.0
D04AX07	0	34	34	100.0	0.0
P03AX10	0	34	34	100.0	0.0
D08AJ57	0	33	33	100.0	0.0
M01AB01	0	32	32	100.0	0.0
V03AB15	0	32	32	100.0	0.0
C01CA30	0	31	31	100.0	0.0
N06BA04	0	30	30	100.0	0.0
N05BA05	0	30	30	100.0	0.0
B05AA56	0	30	30	100.0	0.0
J04AC01	0	30	30	100.0	0.0
N04BC07	0	30	30	100.0	0.0
L04AB02	0	30	30	100.0	0.0
R01AD09	0	29	29	100.0	0.0
C04AG01	0	29	29	100.0	0.0
S01BC05	0	29	29	100.0	0.0
S01AA11	0	28	28	100.0	0.0
N05AL01	0	28	28	100.0	0.0
N05CP01	0	27	27	100.0	0.0

G03CA03	0	27	27	100.0	0.0
J06BB09	0	26	26	100.0	0.0
S01EB01	0	26	26	100.0	0.0
J04AK01	0	26	26	100.0	0.0
A03BA01	0	26	26	100.0	0.0
N04BX01	0	26	26	100.0	0.0
A06AG11	0	26	26	100.0	0.0
A01AB03	0	25	25	100.0	0.0
C01BD07	0	24	24	100.0	0.0
G03CD01	0	24	24	100.0	0.0
D02AE01	0	24	24	100.0	0.0
G01AF02	0	24	24	100.0	0.0
A10BX02	0	23	23	100.0	0.0
A07BA01	0	23	23	100.0	0.0
A07AA02	0	23	23	100.0	0.0
L01BB02	0	22	22	100.0	0.0
C01CA03	0	22	22	100.0	0.0
N04AA01	0	21	21	100.0	0.0
N04BX02	0	21	21	100.0	0.0
L04AC02	0	21	21	100.0	0.0
N03AA02	0	20	20	100.0	0.0
D07AD01	0	20	20	100.0	0.0
J04AK02	0	19	19	100.0	0.0
N07BB03	0	19	19	100.0	0.0
N06AA04	0	19	19	100.0	0.0
A02BB01	0	18	18	100.0	0.0
D10AF54	0	18	18	100.0	0.0
N02BA51	0	18	18	100.0	0.0
A10AD04	0	17	17	100.0	0.0
S01FA04	0	17	17	100.0	0.0
R04AP30	0	16	16	100.0	0.0
N06BA07	0	16	16	100.0	0.0
N07BB06	0	16	16	100.0	0.0
H02AB01	0	16	16	100.0	0.0
S01AA04	0	16	16	100.0	0.0
B05XA03	0	16	16	100.0	0.0
L01AA07	0	15	15	100.0	0.0
A01AE51	0	15	15	100.0	0.0
G04BP01	0	15	15	100.0	0.0
P01BD01	0	14	14	100.0	0.0
J01AA12	0	14	14	100.0	0.0
J05AP01	0	14	14	100.0	0.0
A11HA30	0	14	14	100.0	0.0
C08DB01	0	14	14	100.0	0.0
J02AC02	0	13	13	100.0	0.0
C01DA14	0	13	13	100.0	0.0
M01CC01	0	12	12	100.0	0.0
C05BA03	0	12	12	100.0	0.0
S01FA01	0	12	12	100.0	0.0

S01GX01	0	12	12	100.0	0.0
A01AP03	0	11	11	100.0	0.0
D07AC01	0	11	11	100.0	0.0
B01AB09	0	11	11	100.0	0.0
A02AD05	0	11	11	100.0	0.0
R07AA03	0	11	11	100.0	0.0
D03AX03	0	11	11	100.0	0.0
L01XX14	0	10	10	100.0	0.0
P03AC04	0	10	10	100.0	0.0
A07CA50	0	10	10	100.0	0.0
N01BB20	0	10	10	100.0	0.0
D02AB01	0	10	10	100.0	0.0
C01CA24	0	10	10	100.0	0.0
G03CA04	0	9	9	100.0	0.0
A03AB20	0	9	9	100.0	0.0
N05AB10	0	9	9	100.0	0.0
A12CC14	0	9	9	100.0	0.0
N03AX03	0	9	9	100.0	0.0
N01AH01	0	9	9	100.0	0.0
D11AH01	0	9	9	100.0	0.0
V03AB48	0	8	8	100.0	0.0
N05CC01	0	8	8	100.0	0.0
A07XP03	0	8	8	100.0	0.0
J01EC02	0	8	8	100.0	0.0
L01BC06	0	8	8	100.0	0.0
B05CB10	0	8	8	100.0	0.0
C03DA02	0	8	8	100.0	0.0
V08AB02	0	8	8	100.0	0.0
A11HA03	0	8	8	100.0	0.0
D05BB02	0	8	8	100.0	0.0
R02AP52	0	8	8	100.0	0.0
C01AA05	0	7	7	100.0	0.0
B01AD04	0	7	7	100.0	0.0
V07AB02	0	7	7	100.0	0.0
A11CC80	0	7	7	100.0	0.0
R02AD01	0	7	7	100.0	0.0
N05AF01	0	7	7	100.0	0.0
P02CA01	0	7	7	100.0	0.0
R03AC12	0	7	7	100.0	0.0
V03AB23	0	7	7	100.0	0.0
A02BA01	0	7	7	100.0	0.0
S01CA53	0	7	7	100.0	0.0
J01XB01	0	6	6	100.0	0.0
C01CE03	0	6	6	100.0	0.0
J07AJ52	0	6	6	100.0	0.0
D11AH02	0	6	6	100.0	0.0
B05AA57	0	6	6	100.0	0.0
B01AC23	0	6	6	100.0	0.0
J07AM51	0	6	6	100.0	0.0

H01BB02	0	6	6	100.0	0.0
D08AX06	0	6	6	100.0	0.0
N06AA21	0	6	6	100.0	0.0
L01DB07	0	6	6	100.0	0.0
A06AX02	0	6	6	100.0	0.0
A16AA01	0	6	6	100.0	0.0
M03CA01	0	6	6	100.0	0.0
M02AH20	0	6	6	100.0	0.0
L04AA24	0	6	6	100.0	0.0
C03BA11	0	6	6	100.0	0.0
B05BB14	0	5	5	100.0	0.0
G03BA03	0	5	5	100.0	0.0
R01AA05	0	5	5	100.0	0.0
H02AB58	0	5	5	100.0	0.0
C05AA61	0	5	5	100.0	0.0
V03AB14	0	5	5	100.0	0.0
S01XA28	0	5	5	100.0	0.0
S01EA03	0	5	5	100.0	0.0
S01BC03	0	5	5	100.0	0.0
J01DC04	0	5	5	100.0	0.0
D08AC52	0	5	5	100.0	0.0
C01BC03	0	5	5	100.0	0.0
D06BA01	0	4	4	100.0	0.0
V03AB25	0	4	4	100.0	0.0
L01BC05	0	4	4	100.0	0.0
D01AC52	0	4	4	100.0	0.0
D02AB51	0	4	4	100.0	0.0
C04AX21	0	4	4	100.0	0.0
V06DE01	0	4	4	100.0	0.0
G02CB01	0	4	4	100.0	0.0
L01DB02	0	4	4	100.0	0.0
J01GB06	0	4	4	100.0	0.0
B02BD31	0	4	4	100.0	0.0
J01CG01	0	4	4	100.0	0.0
D09AA02	0	4	4	100.0	0.0
N06DA04	0	4	4	100.0	0.0
L02AE03	0	4	4	100.0	0.0
D03AX79	0	4	4	100.0	0.0
B01AE03	0	4	4	100.0	0.0
R06AD02	0	4	4	100.0	0.0
S01GA02	0	3	3	100.0	0.0
M05BA03	0	3	3	100.0	0.0
N05AB02	0	3	3	100.0	0.0
L04AB04	0	3	3	100.0	0.0
B05XA18	0	3	3	100.0	0.0
R03AK05	0	3	3	100.0	0.0
N06AX03	0	3	3	100.0	0.0
S01AE05	0	3	3	100.0	0.0
C08CA06	0	3	3	100.0	0.0

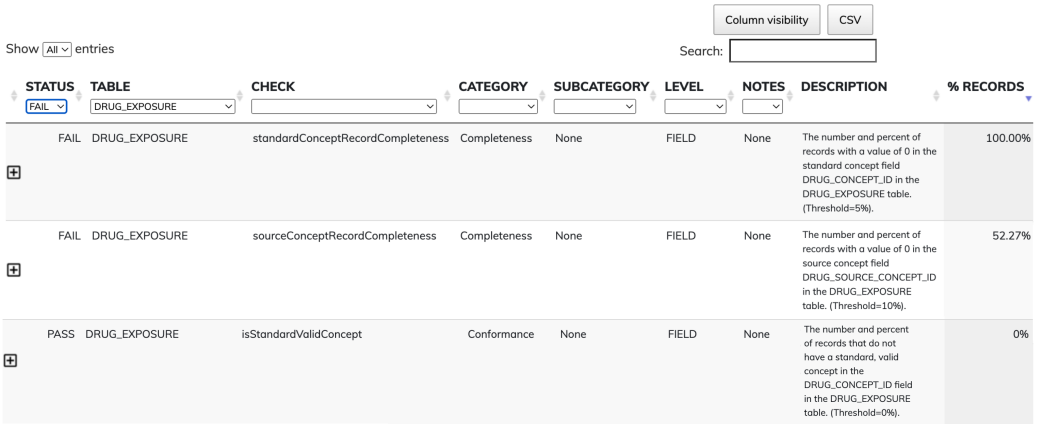
D08AF01	0	3	3	100.0	0.0
L01BB05	0	3	3	100.0	0.0
G01AA10	0	3	3	100.0	0.0
B05BB12	0	3	3	100.0	0.0
L01XX02	0	2	2	100.0	0.0
S01FA06	0	2	2	100.0	0.0
J01DE01	0	2	2	100.0	0.0
N05CM18	0	2	2	100.0	0.0
H01AA02	0	2	2	100.0	0.0
L02BB01	0	2	2	100.0	0.0
J02AC05	0	2	2	100.0	0.0
L01DC03	0	2	2	100.0	0.0
C07AA07	0	2	2	100.0	0.0
N01AX03	0	2	2	100.0	0.0
J06BB16	0	2	2	100.0	0.0
G03CC06	0	2	2	100.0	0.0
L01XB01	0	2	2	100.0	0.0
N01AH03	0	2	2	100.0	0.0
C01EB10	0	2	2	100.0	0.0
R04AP03	0	2	2	100.0	0.0
D03AX11	0	2	2	100.0	0.0
B02BD02	0	2	2	100.0	0.0
B05XB02	0	1	1	100.0	0.0
L01XX19	0	1	1	100.0	0.0
L01AA06	0	1	1	100.0	0.0
B05CA05	0	1	1	100.0	0.0
L03AA02	0	1	1	100.0	0.0
N04BC06	0	1	1	100.0	0.0
L01XX01	0	1	1	100.0	0.0
L01CD01	0	1	1	100.0	0.0
C01BA05	0	1	1	100.0	0.0
R03CC14	0	1	1	100.0	0.0
L01XA03	0	1	1	100.0	0.0
B02BD06	0	1	1	100.0	0.0
R07AA05	0	1	1	100.0	0.0
D05BX02	0	1	1	100.0	0.0
S01AX02	0	1	1	100.0	0.0
G04BP51	0	1	1	100.0	0.0
C01CA04	0	1	1	100.0	0.0
L01AC01	0	1	1	100.0	0.0
C01CA07	0	1	1	100.0	0.0
N01BB51	0	1	1	100.0	0.0
B02BC07	0	1	1	100.0	0.0
J06BB03	0	1	1	100.0	0.0
M02AB53	0	1	1	100.0	0.0
C05BB02	0	1	1	100.0	0.0
B02BC01	0	1	1	100.0	0.0
V03AB46	0	1	1	100.0	0.0
D03BA20	0	1	1	100.0	0.0

L01DB01	0	1	1	100.0	0.0
N05CP30	0	1	1	100.0	0.0
N01BB01	0	1	1	100.0	0.0
S01KA02	0	1	1	100.0	0.0
N01AX10	0	1	1	100.0	0.0
L01CA03	0	1	1	100.0	0.0

E Anhang: ATC-GM Vokabular

concept_id	concept_name
400000008	Verschiedene, Andere Mittel gegen Ektoparasiten, inkl. Antiscabiosa
400000002	unspec
400000003	nomed
400000004	Nahrungsergaenzungsmittel
400000005	Polidocanol (Lauromacrogol 400), Kombinationen
400000011	Salbeiblätter
400000013	Kombinationen, Homöopathische und anthroposophische Zubereitungen gegen Gelenk- und Muskelschmerzen zur topischen Anwendung
400000012	Trospium
400000006	Rituximab
400000007	Doxylamin
400000009	Kombinationen, Pflanzliche Brusteinreibungen und Inhalate, inkl. Bäder
400000010	Ambroxol

F Screenshots DQD Dashboard



Column visibility CSV

Show All entries Search:

STATUS	TABLE	CHECK	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	DRUG_EXPOSURE	standardConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the standard concept field DRUG_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=5%).	100.00%
FAIL	DRUG_EXPOSURE	sourceConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the source concept field DRUG_SOURCE_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=10%).	52.27%
PASS	DRUG_EXPOSURE	isStandardValidConcept	Conformance	None	FIELD	None	The number and percent of records that do not have a standard, valid concept in the DRUG_CONCEPT_ID field in the DRUG_EXPOSURE table. (Threshold=0%).	0%

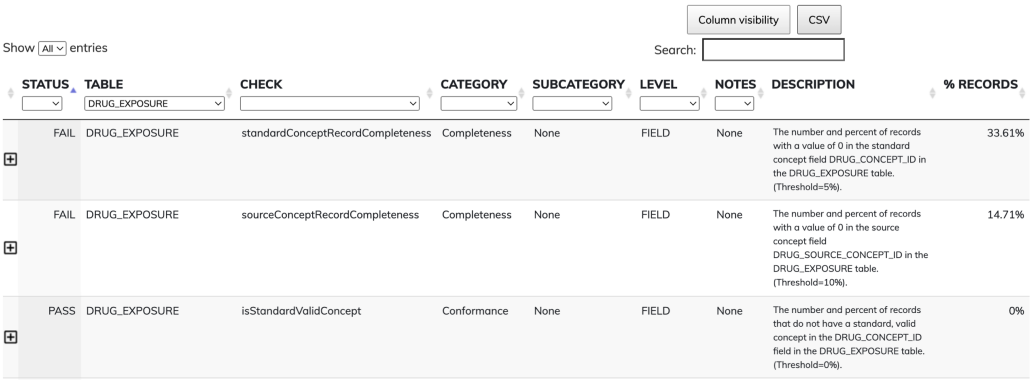
Abbildung F.1: DQD Dashboard Schritt 1: initiale Analyse der Medikationsverordnungen

Column visibility CSV

Show 5 entries Search:

STATUS	TABLE	CHECK	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	DRUG_EXPOSURE	standardConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the standard concept field DRUG_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=5%).	100.00%
FAIL	DRUG_EXPOSURE	sourceConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the source concept field DRUG_SOURCE_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=10%).	14.82%
PASS	DRUG_EXPOSURE	isStandardValidConcept	Conformance	None	FIELD	None	The number and percent of records that do not have a standard, valid concept in the DRUG_CONCEPT_ID field in the DRUG_EXPOSURE table. (Threshold=0%).	0%

Abbildung F.2: DQD Dashboard Schritt 2: nach den Maßnahmen zur Verbesserung der Datenstruktur



Column visibility CSV

Show All entries Search:

STATUS	TABLE	CHECK	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	DRUG_EXPOSURE	standardConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the standard concept field DRUG_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=5%).	33.61%
FAIL	DRUG_EXPOSURE	sourceConceptRecordCompleteness	Completeness	None	FIELD	None	The number and percent of records with a value of 0 in the source concept field DRUG_SOURCE_CONCEPT_ID in the DRUG_EXPOSURE table. (Threshold=10%).	14.71%
PASS	DRUG_EXPOSURE	isStandardValidConcept	Conformance	None	FIELD	None	The number and percent of records that do not have a standard, valid concept in the DRUG_CONCEPT_ID field in the DRUG_EXPOSURE table. (Threshold=0%).	0%

Abbildung F.3: DQD Dashboard Schritt 3: nach den Maßnahmen zur Überführung nach RxNorm

Erklärung zur Eröffnung des Promotionsverfahrens

Technische Universität Dresden
Medizinische Fakultät Carl Gustav Carus

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
2. Bei der Auswahl und Auswertung des Materials sowie bei der Erstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:
   Beratende Unterstützung durch die an den Journal Publikationen beteiligten Personen
3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters bzw. einer kommerziellen Promotionsberaterin in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht:
   Reinecke, I., Henke, E., Peng, Y., Sedlmayr, M., & Bathelt, F. (2023). Fitness for Use of Anatomical Therapeutic Chemical Classification for Real World Data Research. In M. Hägglund, M. Blusi, S. Bonacina, L. Nilsson, I. Cort Madsen, S. Pelayo, A. Moen, A. Benis, L. Lindsköld & P. Gallos (Hrsg.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI230245

   Reinecke, I., Siebel, J., Fuhrmann, S., Fischer, A., Sedlmayr, M., Weidner, J., & Bathelt, F. (2023). Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation. *JMIR Medical Informatics*, *11*, e40312. https://doi.org/10.2196/40312

   Reinecke, I., Bathelt, F., Sedlmayr, M., & Kühn, A. (2022). Pharmaceutical Feedback Loop – A Concept to Improve Prescription Safety and Data Quality. *Studies in Health Technology and Informatics*. https://doi.org/10.3233/SHTI220910

   Reinecke, I., Zoch, M., Wilhelm, M., Sedlmayr, M., & Bathelt, F. (2021). Transfer of Clinical Drug Data to a Research Infrastructure on OMOP - A FAIR Concept. *Studies in Health Technology and Informatics*, *287*, 63–67. https://doi.org/10.3233/SHTI210815

   Vass, A., Reinecke, I., Boeker, M., Prokosch, H.-U., & Gulden, C. (2022). Availability of Structured Data Elements in Electronic Health Records for Supporting Patient Recruitment in Clinical Trials. In P. Otero, P. Scott, S. Z. Martin & E. Huesing (Hrsg.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI220046

   Reinecke, I., Zoch, M., Reich, C., Sedlmayr, M., & Bathelt, F. (2021). The Usage of OHDSI OMOP - A Scoping Review. [Place: Netherlands]. *Studies in health technology and informatics*, *283*, 95–103. https://doi.org/10.3233/SHTI210546

6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.
7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.
8. Ich habe die Zitierrichtlinien fur Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.
9. Ich bin mit den an der Technischen Universität Dresden geltenden „Richtlinien zur Sicherung guter wissenschaftlicher Praxis, zur Vermeidung wissenschaftlichen Fehlverhaltens und für den Umgang mit Verstößen“ einverstanden.

Freiberg den 11. Juli 2023

Unterschrift der Promovierenden

Bestätigung über Einhaltung der aktuellen gesetzlichen Vorgaben

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation:

☐ die Einhaltung der Bestimmungen des Tierschutzgesetzes Aktenzeichen der Genehmigungsbehörde:
die Einhaltung des Gentechnikgesetzes Projektnummer:

☐ das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen
Aktenzeichen der zuständigen Ethikkommission:

☐ die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät Carl Gustav Carus und des Universitätsklinikums Carl Gustav Carus.

Freiberg den 11. Juli 2023

Unterschrift der Promovierenden