

Aus der Medizinischen Klinik mit Schwerpunkt Nephrologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Datenintegration, Wissensrepräsentation und Datenanalyse –
Werkzeuge zur systematischen Untersuchung von
Einflussfaktoren auf das Langzeit-Outcome nephrologischer
Patienten

zur Erlangung des akademischen Grades

Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Danilo Schmidt

aus Berlin

Datum der Promotion: 04.06.2021

Inhaltsverzeichnis

Abstract (deutsch)	4
Abstract (english)	4
Liste der Abkürzungen	5
1. Einführung	6
1.1. Akutes Nierenversagen (ANV)	8
1.2. Nierentransplantationsmedizin	8
1.3. Webbasierte elektronische Patientenakte TBase	9
1.4. Plattform für die einfache Integration, Verwaltung und Ausleitung von Daten (i2b2)	10
1.5. Potentiale der facettierten Suche	10
1.6. Klinisches Text Mining	11
2. Studie A: “Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools”	12
2.1. Material und Methoden	12
2.1.1. Datenbasis	13
2.1.2. Integrationstool für i2b2	13
2.1.3. Vorverarbeitung und ETL	13
2.1.4. Auswertung	13
2.2. Ergebnisse	14
2.2.1. Analyse der Quelldaten	14
2.2.2. Analyse des Mappingverfahrens	14
2.3. Diskussion	15
2.3.1. Hardware Voraussetzungen	15
2.3.2. Mapping	16
2.3.3. Einschränkungen	16
3. Studie B: „A novel tool for the identification of correlations in medical data by faceted search“	17
3.1. Material und Methoden	17
3.1.1. Datenbasis	17
3.1.2. Systemarchitektur	17

3.1.2.1. Solr Suchplattform	18
3.1.2.2. Informationsextraktion aus unstrukturierten medizinischen Texten	18
3.2. Ergebnisse	19
3.2.1. Facettierte Suche	19
3.2.2. Darstellung der Informationsextraktion	19
3.3. Diskussion	20
4. Studie C: „The Incidence of Acute Kidney Injury and Associated Hospital Mortality - A Retrospective Cohort Study of Over 100 000 Patients at Berlin’s Charité Hospital”	21
4.1. Material und Methoden	21
4.1.1. Beschreibung der Kohorte	21
4.1.2. Definition der Stadien des akuten Nierenversagen	22
4.1.3. Statistische Methoden	22
4.2. Ergebnisse	22
4.3. Diskussion	24
4.3.1. ANV: ein unterschätzter Risikofaktor	24
4.3.2. Internationaler Vergleich	24
4.3.3. Unzureichende ANV-Codierung	24
4.3.4. Das Potenzial für die automatisierte Analyse klinischer Routinedaten	25
4.3.5. Stärken und Schwächen der Studie	26
5. Zusammenfassung	26
6. Literaturverzeichnis	27

Abstract (deutsch)

Das Gesundheitssystem wird sich durch die Digitalisierung in der Zukunft radikal ändern. Besserer Informationsaustausch durch vernetzte Fallakten, neue Versorgungsformen wie z.B. Telemedizinanwendungen können zukünftig das Problem des Fachärztemangels in strukturschwachen Gegenden vermindern. Neue Mobile Health Anwendungen werden die Patienten aktiver in Therapiemöglichkeiten einbinden und das Patient Empowerment verbessern. Zusätzlich werden durch die Digitalisierung immer mehr Daten produziert, die einen Beitrag zur medizinischen Forschung und Verbesserung von Therapien leisten können. Neben den Herausforderungen zum Datenschutz und zur Datensicherheit, müssen auch Fragen zur Interoperabilität, Nutzen und Transparenz geklärt werden. Diese Arbeit untersucht exemplarisch an drei konkreten Beispielen (zur Datenintegration, Wissensrepräsentation und Datenanalyse), welche Herausforderungen und Lösungen möglich sind, um medizinische Daten effektiv zu nutzen und die Forschung und Routineversorgung zu verbessern. In der Studie zur Datenintegration wurde untersucht, inwieweit sich eine auf einem relationalen Datenbankschema basierende medizinische Routinedatenbank mit Langzeitdaten von transplantierten Patienten, in eine Ontologie-basierte Forschungsdatenbank wie i2b2, ohne Informationsverlust überführen lässt. Des Weiteren wurde in der Studie zur Wissensrepräsentation untersucht, wie sich mit Hilfe von Open Source Entwicklungswerkzeugen eine Applikation zur Visualisierung von Informationen aus strukturierten und unstrukturierten medizinischen Daten implementieren lässt. Mit der entwickelten Applikation kann das medizinische Personal ohne Programmierkenntnisse Informationen aus dem medizinischen Datenpool extrahieren und systematisch analysieren. Das Thema Datenanalyse wurde durch die Studie zum akuten Nierenversagen näher beleuchtet. In dieser Studie wurde ein Algorithmus implementiert, der in einer großen Kohorte aus stationären Patientendaten, das Ereignis akutes Nierenversagen (ANV) detektieren kann. Nach der statistischen Auswertung der Ergebnisse dieses Algorithmus, konnte die Kohorte im Hinblick auf das Auftreten von akuten Nierenversagen und den damit verbundenen Krankheitscharakteristika und Risikoassoziationen umfassend beschrieben werden.

Abstract (english)

The digitalization will radically transform the healthcare system in the future. New forms of health care e.g. telemedicine or interconnected health records have the capability to reduce the problem of the shortage of medical experts in rural areas. New mobile health applications will involve patients more actively in their treatment options and will improve patient empowerment. Furthermore, the digitalization is producing more and more data, which should foster medical research and further improve of therapies. In addition to the challenges of data protection and data security, questions about interoperability, medical value and transparency must also be addressed. This thesis is based on three concrete examples (for data integration, knowledge representation and data analysis) and investigates which challenges and solutions are possible to use medical data

effectively and to improve research and routine medical care. The study on data integration examined the extent to which a relational database for routine medical care with long-term data from transplanted patients can be transferred to an ontology-based research database such as i2b2 without loss of information. The study on the representation of knowledge examined the implementation of an application for the visualization of information from structured and unstructured medical data by using open source development tools. With the fully developed application, medical personnel can now extract information from the medical data base and easily analyse data without programming knowledge. The study on acute kidney failure examined the topic of data analysis in more detail. In this study, an algorithm was implemented that can detect the event of acute kidney failure in a large cohort of inpatient hospital data. After the statistical analysis of the results of this algorithm, the cohort could be comprehensively described with regard to the occurrence of acute kidney failure and the associated disease characteristics and risk associations.

Liste der Abkürzungen

ABDAMED	Datenbank der Bundesvereinigung Deutscher Apothekerverbände e.V.
ANV	Akutes Nierenversagen
API	Application Programming Interface
BLOB	Binary Large Object (Datentyp)
CDISC	Clinical Data Interchange Standards Consortium
CKD	Chronic Kidney Disease
CSV	Comma-separated values (Dateiformat für den Austausch von Daten)
cTAKES	clinical Text Analysis and Knowledge Extraction System
EAV	Entity-Attribute-Value
eGFR	estimated Glomerular Filtration Rate
ESA	Erythropoiesis-Stimulating Agents
ETL	Extract, Transform and Load
FHIR	Fast Healthcare Interoperable Resources (Standard für den Austausch von medizinischen Daten)
GB	Gigabyte
GUI	Graphical User Interface
HCV	Hepatitis-C-Virus
HITEx	Health Information Text Extraction (Open Source Software für Natural Language Processing)
HL7	Health Level 7 (Standard für den Austausch von medizinischen Daten)
i2b2	Informatics for Integrating Biology and the Bedside
IBM	International Business Machines Corporation (IT-Unternehmen)
ICD-10-GM	International Classification of Diseases 10. Revision German Modification

IDRT	Integrated Data Repository Toolkit
JAVA	objektorientierte Programmiersprache
KDIGO	Kidney Disease Improval Global Outcome
LOINC	Logical Observation Identifiers Names and Codes
MedKAT	Medical Knowledge Analysis Tool (Software der Firma IBM)
NIH	National Institutes of Health
ODM	Operational Data Model
PHP	Hypertext Preprocessor (Open Source Script Sprache für die serverseitige Webentwicklung)
SAP	IT-Unternehmen
SNOMED	Systematisierte Nomenklatur der Medizin
Solr	Open-Source Suchmaschine der Apache Software Foundation
SQL	Structured Query Language
TBase	elektronische Patientenakte der Charité – Universitätsmedizin Berlin
UACR	Urine Albumin-to-Creatinine Ratio
XMI	XML Metadata Interchange
XML	Extensible Markup Language

1. Einführung

Die globale Digitalisierung aller Lebensbereiche des Menschen hat in der letzten Dekade auch das Gesundheitssystem erfasst. Dies bedeutet, dass sich das Gesundheitswesen auf radikale Veränderungen einstellen muss [1, 2]. Die technischen Errungenschaften bei der Herstellung von immer preiswerteren Speicher und leistungsfähigeren Prozessoren, sowie die Fortschritte in der Kommunikation- und Mobilfunktechnik (Glasfaser, bessere Bandbreiten im Mobilfunk, Smartphones etc.) haben dazu beigetragen, dass sich die Digitalisierung unserer Gesellschaft sehr schnell weiterentwickelt hat. Die datengetriebene Gerätetechnik und die Vernetzung der Geräte kommt ohne eine Vielzahl neuer Sensoren nicht aus [3, 4]. Hinzukommt, dass aufgrund der technischen Fortschritte bei der Herstellung von effizienteren Speichern und der steten Verbesserung von Rechenleistungen, die Bild-, Text- und Sprachverarbeitung deutlich besser geworden ist. Smartphones wären ohne diese Technologien nicht denkbar. Diese neuen technischen Möglichkeiten finden natürlich auch ihren Einsatz in der Medizin. Immer mehr Sensoren, erfassen in immer größeren Mengen Daten im medizinischen Routinealltag. Das bedeutet, dass nicht nur digitale Daten für den konkreten Versorgungsfall zur Verfügung stehen, sondern diese Daten auch für die Qualitätssicherung oder die Forschung verwendet werden können [2]. Die sekundäre Nutzung der medizinischen Daten gewinnt somit immer mehr an Bedeutung. Neben den speziellen Herausforderungen bei Datenschutz und Datensicherheit, sowie den allgemeinen Umgang mit Gesundheitsdaten, werden diese neuen technischen Möglichkeiten die Dienstleistungen, Geschäftsmodelle und Arbeitsprozesse im Gesundheitswesen grundsätzlich

verändern [1, 2, 5]. Es entstehen neue Versorgungsformen wie zum Beispiel Telemedizinanwendungen, die dabei helfen können Strukturprobleme (Fachärztemangel, fehlende flächendeckende Versorgung in ländlichen bzw. strukturschwachen Räumen etc.) zu beheben. Patienten werden nicht mehr nur passive Protagonisten ihrer Erkrankung sein. Sie werden zukünftig dank ihrer digitalen Gesundheitsdaten die Möglichkeit haben, ihre Erkrankung besser zu verstehen und zu managen. Zum Beispiel können Mobile-Health Anwendungen einen Beitrag leisten das Patient-Empowerment zu erhöhen [1, 2]. Die sekundäre Verwendung von medizinischen Daten (z.B. aus elektronischen Krankenakten, Gerätedaten aus dem Krankenhaus oder durch den Patienten selbst erhobene Daten via Wearables etc.) ist zu einem wichtigen Faktor geworden, um Krankheitsursachen und Zusammenhänge zu identifizieren, zu analysieren und so das Outcome der Patienten zu verbessern.

Diese Arbeit erklärt anhand von drei systematischen Untersuchungen, zum Thema Datenintegration, Wissensrepräsentation und Datenanalyse, die Möglichkeiten und Herausforderungen beim Verarbeiten von in der Routine erhobenen Gesundheitsdaten. Die Untersuchungen wurden in der Medizinischen Klinik mit Schwerpunkt Nephrologie der Charité - Universitätsmedizin Berlin durchgeführt und hatten zum Ziel einen Beitrag zur Outcome Forschung in diesem Fachbereich zu leisten.

Diese Arbeit ist eine gekürzte Zusammenfassung der folgenden drei ausgewählten Publikationen:

Studie A Datenintegration:

Maier C, Christoph J, **Schmidt D**, Ganslandt T, Prokosch HU, Kraus S, Sedlmayr M.; *Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools.*; J Healthc Eng. 2019 Jan 17;2019:5640685. doi: 10.1155/2019/5640685. eCollection 2019. (Impact Factor 1,261)

Studie B Wissensrepräsentation:

Schmidt D, Budde K, Sonntag D, Profitlich HJ, Ihle M, Staeck O.; *A novel tool for the identification of correlations in medical data by faceted search.*; Comput Biol Med. 2017 Jun 1;85:98-105. doi: 10.1016/j.compbiomed.2017.04.011. Epub 2017 Apr 14. (Impact Factor 2,115)

Studie C Datenanalyse:

Khadzhynov D*, **Schmidt D***, Hardt J, Rauch G, Gocke P, Eckardt KU, Schmidt-Ott KM.; *The Incidence of Acute Kidney Injury and Associated Hospital Mortality.*; Dtsch Arztebl Int. 2019 May 31;116(22):397-404. doi: 10.3238/arztebl.2019.0397. (Impact Factor 3,89)

* geteilte Erstautorenschaft

1.1. Akutes Nierenversagen (ANV)

Mit akutem Nierenversagen (ANV) bezeichnet man eine schnelle Verschlechterung der Nierenfunktion. Es ist ein häufiges relevantes klinisches Ereignis, welches bei 8 bis 22% aller Krankenhauspatienten vorkommt [6, 7] und mit einem 1,4- bis 15,4-fachen Anstieg des Mortalitätsrisikos verbunden ist [8-11]. Zudem haben Patienten mit akutem Nierenversagen auch ein erhebliches Risiko für die Entwicklung oder Verschlechterung einer chronischen Nierenerkrankung [12-15]. Derzeit wird das akute Nierenversagen in Abhängigkeit von Änderungen der Serumkreatinin- und Urinausscheidung gemäß den Kidney Disease Improval Global Outcome (KDIGO) - Richtlinien definiert und abgestuft [15, 16]. Gemäß internationaler Richtlinien sollen bei einer Diagnose von akutem Nierenversagen eine Reihe von Behandlungsschritten eingeleitet werden. Sie schreiben unter anderem die Identifizierung und Behandlung von auslösenden Faktoren, die Anpassung der Arzneimitteldosen, die Vermeidung von Nephrotoxinen, die Überwachung der Hämodynamik sowie des Flüssigkeits- und Elektrolythaushalts und die nephrologische Verlaufskontrolle vor. Dass die systematische Umsetzung dieser Maßnahmen mit einer absoluten Senkung des Mortalitätsrisikos um 16,6% einhergeht, konnte in einer kürzlich veröffentlichte Studie nachgewiesen werden [17]. Die frühe Erkennung des akuten Nierenversagens ist allerdings die Voraussetzung für die zeitnahe Umsetzung dieser Maßnahmen. Andererseits deuten eine Reihe von Studien darauf hin, dass ANV-Episoden in der Mehrzahl der Fälle (57% bis 99%) im klinischen Alltag unentdeckt bleiben [18-21]. Eine automatische Analyse von Serum-Kreatininwerten (Studie A), die in der klinischen Routine (siehe auch Studie C) erhoben wurden, soll helfen akutes Nierenversagen besser und schneller zu identifizieren.

1.2. Nierentransplantationsmedizin

In den letzten Jahrzehnten sind Nierenerkrankungen zu einer globalen Gesundheitsbelastung geworden. Chronische Nierenerkrankungen betreffen 10% der Weltbevölkerung und 3-17% der Bevölkerung in der gesamten Europäischen Union, eine Prävalenz, die mit der von Diabetes vergleichbar ist [22]. Schätzungen zufolge, entfallen allein in Deutschland jährlich mehr als 3 Milliarden Euro auf die Behandlung von Nierenersatztherapien bei Patienten mit finalem Nierenversagen. Für Patienten mit Nierenerkrankungen im Endstadium, ist die Nierentransplantation gegenüber der Nierenersatztherapie, die beste Behandlungsform, da sie mit einer niedrigeren Morbidität und Mortalität, höheren Lebensqualität und letztlich auch niedrigeren Kosten im Gesundheitssystem assoziiert ist [23]. Im Jahr 2015 standen in Deutschland ca. 8000 Patienten auf der Warteliste. Nur ein Teil dieser Patienten konnte mit einem Transplantat versorgt werden. Aufgrund des eklatanten Organmangels in Deutschland kommt es zu langen Wartezeiten (mittlerweile ca. 8 Jahre). Die Mortalität auf der Warteliste ist dabei im Vergleich zur Dialyse deutlich erhöht. [24]

Die Nierentransplantation stellt demnach eine lebensverlängerte Maßnahme dar. Reduktion von Komplikationen und Erhöhung des langfristigen Transplantatüberlebens sind die Hauptziele der Therapie nach einer Nierentransplantation. Die Therapie ist meist sehr komplex, den nach Jahren der chronischen Nierenkrankheit und Dialysebehandlung die Transplantatempfänger weitgehend von komplexen Begleiterkrankungen wie Infektionen, Arzneimitteltoxizität, bösartigen Tumoren und Herz-Kreislauf-Erkrankungen betroffen sind. Darüber hinaus sind sie einem hohen Risiko für schwerwiegende Komplikationen wie Organabstoßungen ausgesetzt. Die Mehrzahl der Patienten muss nach einer Nierentransplantation lebenslang täglich bis zu 10 verschiedene Medikamente einnehmen [25]. Die Outcome-Forschung nach Nierentransplantation zielt darauf ab, alle relevante Risikofaktoren und Prädiktoren für Komplikationen zu identifizieren. Die Komplexität des Posttransplantationsmanagements erfordert eine lebenslange intensive Nachbetreuung der Nierentransplantationsempfänger. Daraus resultiert, dass eine Fülle von medizinischen Daten zur Verfügung stehen und die Entscheidungsfindung in der klinischen Praxis sehr komplex ist. Die erzeugten medizinischen Daten sind in der Regel schlecht strukturierte diskrete Daten oder unstrukturierte medizinische Berichte. [26]

Eine integrierte Facettensuchanwendung (siehe Studie B) soll helfen, Zusammenhänge in den Daten durch eine intelligente Wissensrepräsentation besser zu verstehen.

1.3. Webbasierte elektronische Patientenakte TBase

Die webbasierte elektronische Patientenakte TBase wurde im Jahre 1999 im Rahmen des Deutschen Nierentransplantationsprogramms in Kooperation von Charité – Universitätsmedizin Berlin und Humboldt Universität zu Berlin entwickelt. Es können in ihr allgemeine klinische und ambulante Routinedaten gespeichert, sowie fachspezifische Daten zur Nierentransplantation und Nachsorge erfasst werden. Das TBase System basiert auf einer einfachen relationalen Datenbankstruktur und ist in einer Client-Server Architektur eingebunden. Durch die webbasierte Lösung lässt sich das System sehr gut in bestehende IT Infrastrukturen integrieren. Dieser für damalige Verhältnisse zukunftsweisende Ansatz einer elektronischen Patientenakte ist für die Klinik mit Schwerpunkt Nephrologie ein hervorragendes System zur Erfassung von Langzeitdaten im Bereich Nierentransplantation und wird mittlerweile auch in acht weiteren deutschen Universitätskliniken eingesetzt. TBase erfasst alle relevanten medizinischen Daten, wie Anamnesen, Diagnosen, Laborbefunde, Arzneimitteltherapiedaten sowie Befunde von Radiologie, Pathologie, Mikrobiologie in strukturierter und unstrukturierter Form. Des Weiteren werden vollständig die Transplantations- und Posttransplantationsdaten der Empfänger und der Spenderdaten erfasst. Da transplantierte Patienten chronisch krank sind und eine lebenslange Nachsorgetherapie benötigen, fallen im Laufe ihres Lebens eine Vielzahl von Daten an [27]. Diese Daten sind die Basis zahlreicher Arbeiten zum Thema Outcome nach Nierentransplantation (siehe vollständiges Literaturverzeichnis).

1.4. Plattform für die einfache Integration, Verwaltung und Ausleitung von Daten (i2b2)

Eine Möglichkeit für die einfache Datenausleitung und Auswertung von medizinischen Daten bietet die mit Fördermitteln der National Institutes of Health (NIH) in den USA entwickelte Opensource Plattform i2b2 (Informatics for Integrating Biology and the Bedside). Mediziner können mit Hilfe dieser Plattform Daten ohne spezielle Programmierkenntnisse (z.B. SQL etc.) analysieren und ausleiten. Es verwendet ein proprietäres Datenbankschema, das auf einem Entity-Attribute-Value-Schema (EAV) basiert. Dies ist eine weit verbreitete Technik für biomedizinische Datenspeicher [28], da sie ein generisches Data Warehouse, wie das Sternschema (d.h. eine Faktentabelle und mehrere Dimensionstabellen) bereitstellt [29], welches eine einfache Integration heterogener Daten ermöglicht. Dadurch können verschiedene relationale Schemata leichter auf ein gemeinsames Datenmodell abgebildet, und die Daten verschiedener Quellen und Formate besser verglichen werden. Die Integration von Daten aus herkömmlichen Datenbankmodellen (z.B. relationale Datenbankstruktur) stellt eine besondere Herausforderung an den ETL-Prozess (Extract, Transform and Load) über den die Daten in das i2b2-Sternschema integriert werden. In der Literatur finden sich dazu verschiedene Ansätze [30, 31]. Diese erwarten jedoch, dass die Quelldaten in einem bestimmten Format vorliegen, sodass sie normalerweise nicht auf verschiedene Quelldatenschemata übertragbar sind. [27]

Eine Lösung, die in Studie A untersucht wurde, ist das Integrated Data Repository Toolkit (IDRT). Es verspricht die einfache Integration und Verwaltung von Daten aus heterogenen Quellen [32], indem es die Möglichkeit bietet, eine Konfiguration bereitzustellen, mit der das Schema der Quelldaten definiert werden kann. [27]

1.5. Potentiale der facettierten Suche

Im Bereich Medizin kommen für die facettierte Suche unterschiedliche Anwendungsszenarien in Betracht. Zum einen der Einsatz in der klinischen Forschung. Zum anderen für die Verwendung in der klinischen Routine als entscheidungsunterstützendes System. Im Rahmen der klinischen Forschung lassen sich mit der facettierten Suche Subgruppen mit Patienten mit bestimmten Eigenschaften charakterisieren. (z. B. für die Suche nach Patienten mit bestimmten medikamentösen Therapien, mit Laborwerten in einem definierten Bereich, speziellen Symptomen oder Übereinstimmungen von Diagnosen bzw. Patienten mit entsprechenden Einschlusskriterien für klinische Studien). Die facettierte Suche kann beispielsweise aber auch für die systematische Analyse nach Nebenwirkungen oder Arzneimittelwechselwirkungen genutzt werden, indem auffällige Laborwerte, Symptome oder Diagnosen identifiziert werden, die mit bestimmten Wirkstoffen oder Wirkstoffkombinationen verbunden sind. Andererseits ergeben sich Anwendungsfälle in der klinischen Routine und Behandlung des individuellen Patienten. So kann die facettierte Suche für die Identifizierung von Patienten mit besonderen Merkmalen bei neuen therapeutischen Optionen sowie der Identifizierung von Nebenwirkungen von Medikamenten oder

Kontraindikationen nützlich sein. Symptome oder Änderungen von Laborwerten könnten mit einer Rangliste potenzieller nicht identifizierter Diagnosen verknüpft werden. So ließe sich die Patientensicherheit erhöhen und die facettierte Suche kann dem Kliniker helfen eine optimale Therapie zu finden. Ebenso kann sie bei der Entwicklung eines Entscheidungsunterstützungssystems angewendet werden, das sowohl strukturierte als auch unstrukturierte medizinische Berichte großer Patientenkohorten analysiert, um Symptome mit einer Rangliste von Diagnosen zu verknüpfen [26]. Die Literatur beschreibt vorangegangene Projekte, die sich mit interaktiven Suchsystemen bzw. Volltextsuchen unter Einbeziehung von Suchfacetten in medizinischen Informationssystemen befassen. So wird zum Beispiel in Sacco et al. [33] ein interaktives Diagnosesystem mit satzorientierter Suchtechnik beschrieben, um den Benutzer beim freien (ungeordneten) Durchsuchen der Informationsdatenbank zu unterstützen. Biron et al. [34] beschreiben ein Datenabfragesystem für elektronische Patientenakten, das die Volltextsuche nach unstrukturierten medizinischen Informationen in verschiedenen Repositories über eine Solr-Suchmaschine ermöglicht. Sie verwendet Facetten aus administrativen Daten, um vorangegangene Krankenberichte der Patienten zu finden. [26]

1.6. Klinisches Text Mining

Klinisches Text Mining zielt darauf ab, relevante Informationen aus unstrukturierten klinischen Daten zu extrahieren und sie einem normalisierten Inhalt zuzuordnen [35-37]. Da sich klinische Texte erheblich von anderen Fachgebieten unterscheiden, werden hierfür spezielle Systeme entwickelt [38]. Bekannte englischsprachige Systeme sind cTAKES [39], HITEx [40] [41] und IBMs MedKAT [42]. Mit ausreichend großen Mengen an Trainingsdaten können basierend auf dem Vorhandensein einer großen Anzahl von Assoziationen zwischen Texten und dem normalisierten Inhalt, mit statistischen Methoden gute Ergebnisse erzielt werden [43] [44]. Ist die Anzahl der Assoziationen gering, sind wissensintensivere, trainingsunabhängige Methoden erforderlich [45]. Hier liefern regelbasierte Ansätze oft die besseren Ergebnisse [46] [47]. Es gibt bereits Text-Mining-Ansätze auf Basis webbasierter medizinischer Literatur. Zum Beispiel die Informationsextraktion zu unerwünschten Arzneimittelwirkungen [48] aus medizinischen Texten oder die Extraktion von Symptomen aus medizinischen Texten zu seltenen Krankheiten [49]. Die klinische Informationsextraktion aus Patientenakten ist jedoch im klinischen Umfeld immer noch unterrepräsentiert und noch nicht weit entwickelt. Frühere Arbeiten beschäftigten sich mit der Identifikation von semantischen Beziehungen wie „Substanz A behandelt Krankheit B“ [50]. Neuere Arbeiten beziehen auch die zeitliche Informationsextraktion mit ein [51]. Ebenso gibt es die Notwendigkeit der Entwicklung von Ontologiemodellen für die medizinische Terminologie mit der sich Arbeiten beschäftigen [52]. Aufgrund der enormen Annotationskosten werden überwiegend automatische Methoden eingesetzt [53]. Im klinischen Umfeld werden jedoch sehr detaillierte (und arbeitsintensive), überwachte regelbasierte Ansätze verwendet. [26]

2. Studie A: “Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools”

Das Umwandeln von Daten von einem Datenbankschema in ein anderes kann mit einem Kohärenzverlust zwischen den Entitäten des Quelldatenschemas einhergehen (d.h. einem Informationsverlust). Diese Studie sollte untersuchen, inwieweit sich Daten der Medizinischen Klinik für Nephrologie der Charité - Universitätsmedizin Berlin aus dem Dokumentationssystem TBase, automatisch in die Forschungsdatenbank i2b2 überführen lassen, ohne dass Beziehungen zwischen den Entitäten verloren gehen. Das relationale Datenbankschema von TBase und das Sternschema von i2b2 stellen grundlegend unterschiedliche Arten der Datenmodellierung dar. Dies kann dazu führen, dass die Beziehungen des Quelldatenschemas aufgrund der modellspezifischen Art der Erhaltung von Beziehungen im IDRT Tool, nicht ins Zielschema überführt werden können. Um die Qualität der Datentransformation zu bestimmen, haben wir evaluiert (a) ob die typischen Fragen der Kliniker zur Kohortenidentifikation überhaupt noch beantwortet werden können und wenn ja, ob sie (b) im grafischen Frontend von i2b2 oder (c) nur mithilfe von SQL in der i2b2-Datenbank selbst beantwortet werden konnten. Des Weiteren wollten wir herausfinden, welche Art von Fragen aufgrund des oben genannten möglichen Kohärenzverlustes aufgrund der Einschränkungen des Zieldatenschemas oder des Funktionsumfangs im IDRT nicht mehr beantwortet werden können. [27]

2.1. Material und Methoden

In einem ersten Schritt unserer Untersuchungen erfolgte eine Literaturrecherche bzgl. wissenschaftlicher Publikationen zum Thema i2b2 in der Nephrologie und Transplantationsmedizin. Trotz intensiver Literaturrecherche konnten wir diesbezüglich keine Veröffentlichungen finden. Anschließend identifizierten wir im TBase-Datenbankschema Entitäten, die häufig für den Zweck der Zusammenstellung von Patientenkohorten verwendet werden. Als Vorlage dienten Studienprotokolle der letzten zehn Jahre der Klinik für Nephrologie, in denen die Ein- und Ausschlusskriterien benannt waren. Der nächste Schritt bestand darin, die Datenstruktur dieser Entitäten zu analysieren und mögliche Unterschiede zwischen ihnen zu untersuchen. Gleichzeitig haben wir eine weitere Literaturrecherche durchgeführt und verschiedene Integrationstools für i2b2 verglichen. Anschließend implementierten wir die Vorverarbeitung des Quelldatenschemas, um sie danach an das Integrationstool weiterzuleiten. Nach dem Laden der Daten wurde die Datenintegration evaluiert indem wir, die durch die Studienvorgaben (Ein- und Ausschlusskriterien) erzeugten Kohorten, mit der Kohortenidentifikation in i2b2 verglichen haben. [27]

2.1.1. Datenbasis

Als Datenbasis wurde die in 1.4. beschriebene webbasierte elektronische Patientenakte TBase verwendet. Die Datenbank läuft seit dem Jahre 2000 im Routinebetrieb der Klinik mit Schwerpunkt Nephrologie Charité Campus Mitte und seit 2001 in der Klinik für Intensivmedizin und Nephrologie Charité Campus Virchow. Für diese Studie wurde der Datenpool der Nephrologie Campus Mitte betrachtet, Der Quelldatensatz umfasst 18 Jahre medizinische Langzeitdokumentation und hat eine Gesamtgröße von 1,2 GB strukturierter und unstrukturierter Daten. [27]

2.1.2. Integrationstool für i2b2

Ziel dieser Studie war es, bei der Suche nach einem geeigneten Integrationstool für i2b2, einen generischen Ansatz zu verfolgen, der es später erlaubt auch eine Integration für andere Quelldatenmodelle zu ermöglichen. Deshalb kamen nur Integrationsmodule in Betracht, die keine bestimmte Quelldatenstruktur erfordern. Nachdem verschiedene Module verglichen wurden [30, 31, 54, 55], fiel die Entscheidung auf das „Integrated Data Repository Toolkit“ (IDRT) [31], eine Open-Source-Software, die mit Talend Open Studio für die Datenintegration erstellt wurde [56]. Es führt eine Transformation und ein Laden von „verschiedenen Formaten wie CSV, SQL, CDISC, ODM“ [31] durch, die das i2b2-Schema adressieren und einen GUI-unterstützten Ontologie-Editor zum Neuordnen und Zuordnen von Patientendaten enthalten. [27]

2.1.3. Vorverarbeitung und ETL

Bevor der Transformations- und Ladeprozess mit dem IDRT Tool gestartet werden konnte, mussten wie in 2.1. erwähnt, die zur Identifizierung der ausgewählten Patientenkohorten (mit Ein- und Ausschlusskriterien) relevanten Entitäten und die jeweils unterschiedlichen Eigenschaften des TBase-Datenbankschemas identifiziert werden. Jede Variante musste anders behandelt werden, um die Daten für den Ladevorgang vorzubereiten: So fehlten einigen Entitäten Attribute, die für die Darstellung eines Fakts in i2b2 obligatorisch sind. Infolgedessen wurden Teile des Datenbankschemas denormalisiert, um die Anforderungen zu erfüllen. Zuletzt führte das IDRT den Transformations- und Ladeprozess durch. [27]

2.1.4. Auswertung

In einem ersten Schritt mussten die Fehlermeldungen des IDRT während des Ladevorgangs ermittelt und das IDRT Tool entsprechend angepasst werden. Danach wurde untersucht, ob die Art der Abbildung und Transformation der TBase-Quelldaten und deren Integration in das i2b2-Schema die Identifizierung von Kohorten nach den gegebenen Ein- und Ausschlusskriterien ermöglicht. Dafür wurden vier ausgewählte Studiendesigns herangezogen, und entsprechend der datenspezifischen Kriterien wurde untersucht, a) ob die Kohorten innerhalb des grafischen Frontends von i2b2, b) ob sie nur mithilfe von SQL in der i2b2-Datenbank selbst oder c) ob sie aufgrund einer fehlenden Beziehung überhaupt nicht erzeugt werden konnten. Letzteres könnte

darauf zurückzuführen sein, dass i2b2 die Beziehung aufgrund seines generischen Data-Warehouse-Formats nicht aufrechterhalten konnte oder dass IDRT keine Funktion zum Laden dieser Art von Daten hatte. [27]

2.2. Ergebnisse

2.2.1. Analyse der Quelldaten

Bei der Analyse der Quelldaten wurde zunächst geprüft, welche Arten von Tabellen im TBase-Schema verwendet werden und wie diese sich im i2b2 - EAV (Entity-Attribut-Value) Schema darstellen lassen. Wir identifizierten fünf verschiedene Tabellentypen im Quelldatenschema, welche vom ETL Prozess jeweils unterschiedlich behandelt werden mussten [27]:

1. Tabellen, die bereits in einem i2b2-kompatiblen EAV-Schema vorhanden sind.
2. Tabellen, die in einem i2b2-kompatiblen EAV-Schema nicht vorhanden sind und einen Fremdschlüssel enthalten, der auf die Transplantationstabelle verweist, oder eine andere Tabelle mit einem Fremdschlüssel für eine Tabelle, die auf die Transplantationstabelle verweist.
3. Tabellen, die in einem i2b2-kompatiblen EAV-Schema nicht vorhanden sind und keinen Fremdschlüssel enthalten, der auf die Transplantationstabelle verweist, oder einen anderen Schlüssel, der einen Verweis auf die Transplantationstabelle ermöglichen würde.
4. Tabellen, die Freitext-BLOBs enthalten
5. Tabellen, die Laborwerte mit zugehörigen Einheiten enthalten.

Bezüglich der Tabellentypen 1, 4 und 5 mussten geringe Änderungen am IDRT vorgenommen werden, um ein reibungslosen Import mittels IDRT Tool in das i2b2 EAV Schema zu ermöglichen. [27]

2.2.2. Analyse des Mappingverfahrens

Zur Auswertung des Mappingverfahrens erhielten wir von den medizinischen Experten vier Fragen zur Kohortenidentifikation. Diese Fragen wurden in der Vergangenheit von TBase-Experten durch die Programmierung von SQL-Skripten beantwortet. Jede Frage bestand aus mehreren Einschluss- und Ausschlusskriterien. Generell bestehen die meisten Abfragen aus einem oder mehreren der folgenden Eigenschaften [27]:

1. Alle Fakten die bei einem Ereignis aufgetreten sind (z.B. bestimmter Laborwert bei einer bestimmten Visite, etc.)
2. Parameter die größer, kleiner oder gleich zu einem gegebenen Ausgangswert sind
3. Datumswerte die größer, kleiner oder gleich zu einem gegebenen Ausgangswert sind
4. ein klinischer Fakt/Ereignis ist NULL (z.B. wurde nicht erfasst)
5. ein klinischer Fakt/Ereignis entspricht einem vorgegebenen String, oder nicht
6. Volltextsuche mittels regulärer Ausdrücke (Platzhalter)
7. klinische Fakten/ Ereignisse ereignen sich in bestimmter zeitlicher Reihenfolge
 - a. sind durch logische Operatoren UND / ODER verbunden

b. Fakt/ Ereignis A ereignete sich x Tage nach Fakt/ Ereignis B

Kohorte (I) beinhaltete Patienten, für die in einem Labortest drei verschiedene Laborwerte bestimmte Ausprägungen haben müssen. Dieses Mapping konnte in i2b2 dargestellt werden, da i2b2 die Gruppierung von Datensätzen durch sogenannte „Instanznummern“ erlaubt, so dass im Webclient ausgewählte Fakten mit derselben Instanznummer zurückgegeben werden können. IDRT erlaubte jedoch keine Angabe von benutzerdefinierten Instanznummern und verwendete lediglich einen inkrementierten Wert für jedes Attribut. Somit ist IDRT für diese Art der Kohortenabfrage nicht geeignet [27]. Eine relativ einfache Kohortenabfrage (II) hatte zum Ziel einen bestimmten Laborwert (HCV-Wert = positiv) bei Patienten mit erfolgter Nierentransplantation zu einem bestimmten Zeitpunkt (Datum x) abzufragen. Alle Anforderungen konnten sowohl von i2b2 als auch vom IDRT Tool erfüllt werden [27]. Kohorte (III) hatte folgende definierte Einschlusskriterien: Patient hat eine Nierentransplantation erhalten, die Niere wurde abgestoßen und zwischen Transplantation und dem Ereignis Nierenabstoßung lag mindestens ein Harnwegsinfekt vor. Ausgeschlossen wurden alle Patienten, die mit der Nierentransplantation eine kombinierte Pankreastransplantation erhalten haben. Da i2b2 v1.7.00 zeitliche Abfrageeinschränkungen unterstützt, war es möglich in dem Modell die Reihenfolge der Ereignisse zu definieren. Allerdings liegt in den Quelldaten das Ereignis „Harnwegsinfekt“ nicht als diskreter Wert vor, sondern wird in den Beispielabfragen mittels Freitextsuche mit regulären Ausdrücken (Platzhalter) gesucht. Im i2b2-Webclient konnte dies ebenfalls nachgebildet werden, da er eine Freitextsuche mit Platzhaltern unterstützt. Das IDRT-Tool war ursprünglich nicht in der Lage Freitext in i2b2 zu laden. Im Rahmen dieser Arbeit konnte es aber entsprechend angepasst werden, so dass die Ein- und Ausschlusskriterien dieses Abfragebeispiels modelliert werden konnten [27]. In Kohortenabfrage (IV) wurde Patienten mit einer chronischen Nierenerkrankung (CKD) gesucht, deren jüngster erfasster Hämoglobinwert, frühestens 183 Tage nach der Transplantation erfasst wurde und zwischen zwei statischen Werten liegt. Darüber hinaus durften sie mindestens sechs Monate nach der Transplantation, keine Behandlung mit Erythropoese-stimulierenden Mitteln (ESA) erhalten haben. Der erste gemessene Laborwert: „Verhältnis von Urinalbumin zu Kreatinin“ (UACR) nach frühestens 183 Tagen nach Nierentransplantation gemessen, durfte einen vorgegebenen statischen Wert nicht überschreiten. Seit i2b2 v1.7.00 unterstützt der Webclient die Suche nach einem Fakt/Ereignis, welches sich x Tage nach einem anderen Fakt/Ereignis stattgefunden hat. Daher konnten auch hier alle Anforderungen der Abfrage in i2b2 und im IDRT-Tool erfüllt werden. [27]

2.3. Diskussion

2.3.1. Hardware Voraussetzungen

Das IDRT Tool ist eine GUI-basierte Schnittstelle, welche generisch strukturierte Eingabedaten wie CSV- CDISC ODM- und SQL-Datenbanken akzeptiert. Es bietet eine einfache Konfiguration, um die erforderlichen Entitäten für die Transformationsprozedur zu definieren. Die Verarbeitung von größeren Quelldatenbanken (> 1 GB) verbraucht das IDRT Tool eine Menge Ressourcen.

Grund ist zum einen, dass die Quelldateien beim Umwandeln in das EAV Schema vollständig in den Arbeitsspeicher eingelesen und in verschiedene Zwischenformate umgewandelt werden. Dies führt dazu, dass der Arbeitsspeicher schnell ausgelastet ist. Um dieses Problem zu umgehen muss man entsprechende ausgestattete Hardware vorhalten. Falls entsprechende Hardware Lösungen nur eingeschränkt zur Verfügung stehen, könnten auch andere Lösungen (z.B. Aufteilung großer Quelldateien in viele kleinere Dateien oder das Weglassen von nicht benötigten Tabellen im i2b2 Schema) den Ressourcenverbrauch erheblich reduzieren. [27]

2.3.2. Mapping

Wir konnten zeigen, dass sich auch komplexe relationale Datenbanken auf das i2b2 Sternschema abbilden lassen. Es ließen sich mit Hilfe des Webclients alle Beispielabfragen aus 2.2.2. zur Kohortengenerierung verarbeiten. Allerdings können erst ab i2b2 Version 1.7.00 Abfragen mit Berücksichtigung des zeitlichen Verlaufes verarbeitet werden. Wir konnten zeigen, dass alle Beziehungen in den Entitäten von TBase auch nach der Integration ins i2b2 EAV Schema erhalten geblieben sind [27]. Für die Analyse eines generischen Importwerkzeugs für i2b2 haben wir uns nach intensiver Literaturrecherche für das IDRT Tool entschieden, da es im Gegensatz zu anderen beschriebenen Importwerkzeugen kein vordefiniertes Modell erfordert. In unserer Untersuchung stellte sich allerdings heraus, dass nicht alle Datenmodelle unterstützt wurden, die im TBase Datenbankschema verwendet werden. Das IDRT musste deshalb im Rahmen dieser Studie nachträglich verändert werden. Aufgrund der Komplexität von IDRT könnten die Änderungen aber nicht ohne weiteres in die GUI Oberfläche überführt werden. Generell ist die Dokumentation von IDRT relativ spärlich. Bei Fehlern während des Ladevorganges werden oft nur JAVA Stack Meldungen angezeigt, weshalb das Debugging für weniger technisch versierte Anwender wahrscheinlich ohne zusätzliches Expertenwissen nicht leicht zu lösen sind. [27]

2.3.3. Einschränkungen

Vier Abfragen für eine beispielhafte Kohortenzusammenstellung waren relativ wenig. Deshalb zerlegten wir die gegebenen Beispiele in acht Hauptkriterien, die sehr oft für Abfragen zum Outcome nach Nierentransplantation verwendet werden. Bei unseren Untersuchungen konzentrierten wir uns ausschließlich auf die Durchführung der Mapping- und Ladeverfahren in i2b2. Zur Kontrolle, ob alle Daten vollständig geladen wurden, verglichen wir die Anzahl der Fakten in beiden Schemata vor und nach der Transformation. [27]

3. Studie B: „A novel tool for the identification of correlations in medical data by faceted search“

Diese Studie beschäftigt sich mit dem Aufbau und der Architektur einer facettierten Suchanwendung, welche für den Fachbereich Nephrologie und Nierentransplantation entwickelt wurde. Sie zeigt auf, wie ein System für die einfache Informationsextraktion und Datenanalyse von umfangreichen Datensätzen (z. B. Laborwerte, Vitaldaten, Medikamente) und teilweise unstrukturierten medizinischen Daten, wie Entlassungsbriefen, Diagnoseberichten und klinischen Notizen, implementiert werden kann. Diese Studie beschreibt die Details der Anwendung für die Nierentransplantationsmedizin und die daraus resultierende technische Architektur der facettierten Suchanwendung. [26]

3.1. Material und Methoden

3.1.1. Datenbasis

Als Datenbasis wurde die in 1.4. beschriebene webbasierte elektronische Patientenakte TBase verwendet. Die Datenbank läuft seit dem Jahr 2000 im Routinebetrieb der Klinik mit Schwerpunkt Nephrologie Charité Campus Mitte und seit 2001 in der Klinik für Intensivmedizin und Nephrologie Charité Campus Virchow. Für diese Studie wurde der Datenpool der Nephrologie Campus Mitte betrachtet, welche sich auch folgenden Daten zusammensetzte: 6.500 Transplantatempfänger und Patienten auf der Warteliste, 9,9 Millionen Laborbefunde, 237.821 dokumentierte Medikationen, 146.851 diagnostische Befunde aus der Radiologie, Pathologie und Virologie (Befunde) sowie administrative Daten aus dem SAP-System der Charité (70.591 Entlassungsdiagnosen, 25.520 Krankenhauseinweisungen. Darüber hinaus wurden zu dieser Zeit 131.584 ambulante Verlaufsvsiten in TBase dokumentiert. [26]

3.1.2. Systemarchitektur

Das Datenrepository für die im Rahmen dieser Studie entwickelte facettierte Suche, basiert auf einer i2b2-Datenbankstruktur (Informatics for Integrating Biology & the Bedside) [57]. Es bietet wie unter 1.5. bereits beschrieben, die notwendigen Tools, um medizinische Berichte und medizinische Forschungsdaten in einer einzigen Software-Suite zu integrieren. Für die Informationsextraktion aus unstrukturierten Texten wurden Textdaten aus der TBase Datenbank mittels einer kommerziellen Text-Mining-Pipeline, die von der Firma Averbis entwickelt wurde, separat verarbeitet und die resultierenden strukturierten Annotationsergebnisse im sogenannten XMI (XML Metadata Interchange) Format gespeichert [26]. Anschließend wurden diese Daten in die i2b2-Struktur integriert. Die wichtigsten Komponenten der facettierten Suche sind die Solr-Suchplattform, das Informationsextraktionsmodul und das Benutzeroberflächenmodul der facettierten Suche. [26]

3.1.2.1. Solr Suchplattform

Solr ist eine weitverbreitete Open Source Suchmaschine, welche in webbasierten Anwendungen für die Volltextsuche eingesetzt werden kann. Solr enthält auch eine Plugin-Architektur, mit der erweiterbare Anpassungsmöglichkeiten unterstützt werden. Es kann auch auf viele Arten von Anwendungen zugeschnitten werden, ohne dass eine JAVA-Codierung erforderlich ist. Die verfügbaren Funktionen wie facettierte Navigation für die strukturierte und textuelle Suche, die Möglichkeit zur automatischen Zusammenfassung von Ergebnissen (basierend auf der Open Source Software *carrot²*), sowie die Skalier- und Erweiterbarkeit durch Plug-Ins waren ausschlaggebende Faktoren für die Verwendung des Solr Systems für diese Studie. [26]

3.1.2.2. Informationsextraktion aus unstrukturierten medizinischen Texten

Um das Ziel der Wissensrepräsentation nicht nur auf strukturiert vorliegende Daten zu beschränken, war ein erster Schritt dieser Studie, die Informationsextraktion aus unstrukturierten Texten umzusetzen. Dafür wurden mit Hilfe einer Textanalyse-Pipeline der Firma Averbis, relevante Informationseinheiten aus 5000 unstrukturierten medizinischen Texten (Befunde, Untersuchungen, klinische Entlassungsberichte und Follow-up-Berichte) der TBase Datenbank extrahiert. Das verwendete Textanalysetool bestand aus vier Stufen: das Dateneinlesen, die Vorverarbeitung, die Informationsextraktion und die Endausgabe. [26]

Die Vorverarbeitungsstufe besteht aus mehreren linguistischen Analysekomponenten (z. B. Satzerkennung, Tokenisierung (Erkennung von einzelnen Wörtern), Part-of-Speech (Erkennung von Wortarten), Stemming (Zuordnung auf gemeinsamen Wortstamm) und Segmentierung (Zerlegung in kleinste Bedeutungseinheiten). [26]

Eine sogenannter Concept Mapper (ein konfigurierbarer und lexikonbasierter Annotator) kann nun in der Informationsextraktionsstufe die erzeugten Annotationen mit den Konzepten einer vorhandenen Terminologie und den Freitextphrasen mit den unterschiedlichen morphologischen Ebenen (kleinste Bedeutungseinheiten), wie Wort-, Stamm- und Unterworbenebene, zusammenführen. Der Concept Mapper enthält eine Reihe von Funktionen, z.B. Begriffsklärung und Filterung, um die Mapping-Prozesses zu verbessern. Die Hintergrundterminologien bestimmen jedoch im Wesentlichen die Gesamtqualität des Mappings. In dieser Studie wird das Concept Mapping zur Identifizierung z.B. von Diagnosen (via ICD-10-GM), Laborparametern (via LOINC) und Arzneimitteln (via ABDAMED) verwendet. In der letzten Stufe wurden die relevanten medizinischen Informationseinheiten in der i2b2 Datenstruktur gespeichert. [26]

3.2. Ergebnisse

3.2.1. Facettierte Suche

Die Benutzeroberfläche des Systems besteht aus einer Webseite für das Durchsuchen des Datenpools und aus einer zweiten Webseite zum Anzeigen der Ergebnisse aus dem Informationsextraktionstool für Freitexte. Die Benutzeroberfläche wurde mit der Open-Source-Skriptsprache für die Webprogrammierung PHP entwickelt, mit der serverseitige Abfragen an die Solr-API gestellt werden können. Die Weboberfläche erlaubt eine Suche über vordefinierte Menüs und Texteingabefelder, welche als Facettenvariablen zum Abfragen von Solr verwendet werden. Die Suchfacetten werden in vier Spalten dargestellt: die Definition („dictCanon“), das übereinstimmende Synonym („matchedTerm“), den Textteil („coveredText“) und die Referenz auf das entsprechende („uniqueId“). Darüber hinaus kann die Suche auf einen bestimmten Texttyp (z.B. Entlassungsbrief etc.) beschränkt werden. Bei den Laborwerten verwendeten wir anstelle der numerischen Werte eine vorhandene Kategorisierung des Krankenhauslabors (niedrig, normal, hoch), da dies eine bessere Gruppierung der Werte ermöglicht. Negierte Diagnosen oder Medikamente wurden als separate Facettenwerte integriert, um eine Unterscheidung zwischen dem Nichtvorkommen eines Begriffs und seiner expliziten Negation zu ermöglichen. Für jede Facette wird dem Benutzer eine kurze Liste der häufigsten Elemente zusammen mit der jeweiligen Anzahl angezeigt. Zusätzlich kann er oder sie aus einer alphabetischen Liste aller Elemente auswählen. Nach einer Suche werden ausgewählte Elemente hervorgehoben und können in beliebiger Reihenfolge aus den Sucheinschränkungen entfernt werden. Zusätzlich sind alle Elemente, die bei allen verbliebenen Patienten ebenfalls vorhanden sind, mit einem OK-Zeichen gekennzeichnet und unterstützen den Benutzer beim Durchsuchen des Suchbereichs und beim Finden von Korrelationen. [26]

3.2.2. Darstellung der Informationsextraktion

Die Ergebnisanzeige des Informationsextraktionstools für Freitexte besteht aus zwei Teilen: Der obere Teil zeigt den Originaltext mit markierten annotierten Textstellen. Der untere Teil enthält Registerkarten mit den verschiedenen relevanten Annotationen. Die Elemente in diesen Listen zeigen den „coveredText“, das „matchedTerm“, die „Definition“ und die Quelle der Annotation und der möglichen Negationen. Der Benutzer kann die Richtigkeit der Annotationen überprüfen und gegebenenfalls kommentieren. Der ursprüngliche XMI-Inhalt (siehe 3.1.2), der die vollständigen ursprünglichen Annotationsergebnisse darstellt, wird in einem Popup-Fenster angezeigt, wenn auf eine im Text hervorgehobene Annotationsmarkierung geklickt wird. Dementsprechend dient diese Seite zwei unterschiedlichen Zwecken: (a) der Darstellung des Originaltextes und (b) der Validierung der Annotationen. [26]

3.3. Diskussion

Ein Lösungsansatz, wie zum Beispiel der von Sacco et al. [33], entwickelte ein interaktives Suchsystems mit satzorientierter Suchtechnik. Diesen Ansatz haben wir aufgegriffen und um eine spezielle Facettenfunktionalität erweitert. Ebenso verfolgten wir die Idee von Biron et al. [34], die interaktive Suche auf Basis von Solr zu entwickeln. Im Gegensatz zu Biron et al. implementierten wir unsere Facetten ausschließlich aus medizinischen Informationen. Dies ermöglicht eine Extraktion von deskriptiven Daten aus unstrukturierten medizinischen Texten, während Biron et al. [34] lediglich vorangegangene Krankenberichte ohne weitere Informationsextraktion identifizieren konnte. Die Berücksichtigung expliziter Negationen als separate Konzepte ist ein grundlegendes Erfordernis im Umgang mit medizinischen Texten [34]. Die Feststellung von Biron et al. [34], dass die Erkennung von Negationen unerlässlich ist, um die zahlreichen falsch-positiven Fehler zu vermeiden, und dass die Berücksichtigung von Synonymen und Zeitangaben zu wertvollen zusätzlichen Ergebnissen führen wird, griffen wir ebenfalls auf. In unserer Anwendung werden sowohl das Negationshandling (über Advanced Text Mining) als auch Synonyme (unter Verwendung medizinischer Ontologien) behandelt und frühere Arbeiten erweitert. So führt unsere (Offline-) Trennung von Informationsextraktion (Semantische Analyse aus medizinischen Texten) und der eigentlichen Suchoperation zu einer sehr performanten Suche. Durch die Angabe der Kardinalitäten von Facettensätzen unterstützt unsere Anwendung die Ärzte bei der Suche nach möglichen medizinischen Zusammenhängen bei der Hypothesengenerierung bzw. bei der Kohortenbildung von Patienten. Die Rangfolge der Facettenwerte nach Kardinalität unterstützt die Erfassung der verbleibenden Teilmengen. Die interaktive Suche wird von unserer Anwendung zum Beispiel dadurch unterstützt, dass der Benutzer alle Einschränkungen entfernen kann, die er in den vorherigen Schritten vorgenommen hat. Dies ermöglicht eine viel bessere Navigation durch den Suchraum, wo verwandte Systeme nur das anschließende Ausdünnen erlauben [58] [59]. Die verwendeten Modelle für Negation, Multi-Term-Extraktion und Relationsextraktion (z.B. Krankheit/Symptom-Beziehung) basieren auf hochpräzisen Text-Mining- und Informationsextraktionsmodulen nach dem neuesten Stand der Technik. [26]

4. Studie C: „The Incidence of Acute Kidney Injury and Associated Hospital Mortality - A Retrospective Cohort Study of Over 100 000 Patients at Berlin’s Charité Hospital”

Epidemiologische Studien identifizieren häufig ANV-Episoden aus administrativen Krankenhausdaten [60-62]. Inwieweit administrative Daten die tatsächliche Inzidenz von ANV widerspiegeln, ist jedoch umstritten. In den USA durchgeführte Studien legen nahe, dass administrative Daten nur einen Teil der klinischen ANV-Episoden identifizieren [18, 19, 63]. Der Nutzen von Verwaltungsdaten zur Identifizierung von ANV-Episoden im Krankenhaussektor in Deutschland ist derzeit unklar. Seit 2011 empfiehlt die Deutsche Gesellschaft für Nephrologie eine akute Nierenschädigung mit N17 zu codieren, wenn ein Anstieg des Kreatinins um 50% innerhalb von 7 Tagen oder ein Anstieg des Kreatinins um 0,3 mg / dl innerhalb von 48 Stunden beobachtet wird [64]. Inwieweit diese Empfehlungen in der Codierungspraxis umgesetzt werden, wurde bisher nicht untersucht. [21]

Für diese Studie entwickelten wir einen Algorithmus auf Basis der (KDIGO)-ANV-Kriterien, mit dem wir retrospektiv das Vorhandensein oder Fehlen von ANV-Episoden über einen Zeitraum von 3,5 Jahren (2014–2017) auf der Grundlage von seriellen Kreatinin-konzentrationsmessungen im klinischen Alltag (während aller Krankenhausaufenthalte in der Charité – Universitätsmedizin Berlin) untersuchten. Zusätzlich wurden Patienteneigenschaften, Komorbiditäten, ANV-Kodierungen (d.h. mit dem N17-Code) sowie Todesfälle auf der Grundlage von Verwaltungsdaten für alle Krankenhausaufenthalte einbezogen. Dies ermöglichte es, die Inzidenz und die Ergebnisse von Kreatinin-basierten ANV-Episoden zu bestimmen und den Anteil codierter ANV-Episoden im Verhältnis zur geschätzten tatsächlichen Inzidenz von ANV in der definierten Studienpopulation zu berechnen. [21]

4.1. Material und Methoden

4.1.1. Beschreibung der Kohorte

Diese explorative retrospektive Studie umfasste vom 1. Januar 2014 bis zum 30. Juni 2017 stationär behandelte erwachsene Patienten an der Charité Universitätsmedizin Berlin. Wir analysierten retrospektiv alle ANV-Episoden über einen Zeitraum von 3,5 Jahren (2014–2017) auf der Grundlage routinemäßig erhaltener Serumkreatininmessungen bei 103.161 Patienten, deren Kreatinin mindestens zweimal gemessen worden war und die im Krankenhaus für mindestens zwei aufeinanderfolgenden Tagen in der Charité Universitätsmedizin behandelt wurden. Der durchschnittliche Beobachtungszeitraum lag unter Berücksichtigung aller verfügbaren Wiedervorstellungen pro Patient bei 248 Tagen. [21]

4.1.2. Definition der Stadien des akuten Nierenversagen

Die Definition und Einstufung eines akuten Nierenschadens (ANV) erfolgte nach den KDIGO-Kreatininkriterien [16]. Die verschiedenen Szenarien von ANV-Episoden und die damit verbundene Bestimmung des Basiskreatinins [65, 66] haben wir wie folgt definiert [21]:

- a) Die stationäre Aufnahme erfolgt mit bereits erhöhtem Kreatinin, ohne dass frühere Kreatinin-Werte bekannt sind. Im weiteren Verlauf kommt es zur Erholung der Nierenfunktion und Abfall des Kreatinins auf den angenommenen Baseline-Wert.
- b) Die stationäre Aufnahme erfolgt mit bereits erhöhtem Kreatinin. Der Baseline-Kreatininwert ist aus Voraufenthalten bekannt.
- c) Das Kreatinin ist zum Zeitpunkt der Aufnahme noch nicht erhöht, steigt jedoch im weiteren stationären Verlauf an.

Parallel dazu untersuchten wir die administrative Kodierung von Entlassungsdiagnosen derselben Patienten mit Codes aus der Internationalen Klassifikation von Krankheiten (ICD-10-GM). [21]

4.1.3. Statistische Methoden

Mortalitätsdaten haben wir aus den verfügbaren Krankenhausdaten ermittelt. Somit wurden aber nur Todesfälle an der Charité-Universitätsklinik Berlin erfasst. Die deskriptiven Analysen und Kaplan-Meier-Kurven wurden nach ANV-Stadium dargestellt. Es wurden uni- und multivariate Cox-Regressionsanalysen durchgeführt, um Prädiktoren für die Mortalität zu identifizieren. [21]

4.2. Ergebnisse

Im Rahmen dieser Studie implementierten wir einen Algorithmus mit definierten Ein- und Ausschlusskriterien (erwachsene Patienten, Behandlung in der Charité an mindestens zwei aufeinanderfolgenden Tagen, Zeitraum 2014-2017) auf Basis der KDIGO-Kriterien zur Identifizierung von ANV Episoden anhand von seriellen Kreatininmessungen. Der Algorithmus erzeugte eine Gesamtkohorte von 103.161 Patienten, welche die Einschlusskriterien (erwachsene Patienten, Behandlung in der Charité an mindestens zwei aufeinanderfolgenden Tagen, im Zeitraum 2014-2017) erfüllten. Der durchschnittliche Beobachtungszeitraum lag inklusive aller verfügbaren Kontrolluntersuchungen bei 248 Tagen pro Patient. [21]

Im Beobachtungszeitraum 2014-2017 erfüllten insgesamt 32.238 Patienten die Kriterien für ein ANV (31,3% der Gesamtpopulation). Bei 19.009 Patienten (18,4%) wurde ein ANV Stadium 1 (= 0,3 mg / dl oder 1,5-facher Kreatininanstieg) identifiziert. Ein ANV Stadium 2 (= zweifacher Kreatininanstieg) wurde bei 7499 Patienten (7,3%) und ANV Stadium 3 (= dreifacher Anstieg des Kreatins oder ein Anstieg auf = 4,0 mg / dl) wurde bei 5730 Patienten (5,6%) beobachtet. Unter den ANV-Patienten gab es mehr Männer als Frauen, während bei Patienten ohne ANV das Geschlecht gleichmäßiger verteilt war. Bei Patienten mit ANV waren Komorbiditäten wie Diabetes mellitus, Bluthochdruck, Herzinsuffizienz und koronare Herzkrankheit häufiger assoziiert als bei Patienten ohne ANV. Außerdem hatten Patienten mit ANV ein höheres Durchschnittsalter als Patienten ohne ANV. Bei Patienten mit ANV war die Baseline der

geschätzten glomerulären Filtrationsrate (eGFR) etwas niedriger als bei Patienten ohne ANV. Rekurrente ANV-Episoden traten signifikant häufiger auf, wenn Patienten mindestens zu einem Zeitpunkt ein ANV-Stadium 3 aufwiesen. [21]

Insgesamt lagen bei der oben genannten Patientenkohorte 185.760 Krankenhausfälle im Beobachtungszeitraum vor. Diese wurden in einer anschließenden fallbezogenen Analyse untersucht, um kodierte Haupt- und Nebendiagnosen, sowie Prozeduren zu identifizieren, welche im Zusammenhang mit ANV kodiert wurden. Dabei wurden in 21,4% aller Krankenhausfälle ANV-Episoden beobachtet. Bei diesen Fällen wurden häufiger ein akutes Koronarsyndrom, eine akute Atemwegserkrankung oder Krebs als Hauptdiagnose bzw. Sepsis als Haupt- oder Nebendiagnose kodiert. Mit einer klaren Zuordnung zum KDIGO-Stadium wurden bei ANV-Fällen vermehrt chirurgische Eingriffe und mechanische Beatmung kodiert. Bei Fällen mit dem ANV-Stadium 3 wurden 29,9% als intermittierende Hämodialyse und 21,4% als kontinuierliche Nierenersatzverfahren kodiert. In Bezug auf die Verweildauer der Krankenhausfälle hatten Patienten mit ANV signifikant längere Aufenthalte und sowie eine erhöhte Krankenhaussterblichkeit. Zudem benötigten Patienten mit ANV zum Zeitpunkt der Entlassung aus dem Krankenhaus auch häufiger eine Nierenersatztherapie. Um Langzeitergebnisse zu analysieren, wurden für den verfügbaren Nachbeobachtungszeitraum fallbezogen die Mortalitätsdaten (beginnend mit dem ersten dokumentierten Krankenhausaufenthalt während des Beobachtungszeitraums) untersucht. Es zeigte sich ein zunehmend schlechteres Langzeitüberleben von Stadium 1 bis Stadium 3. Ein Zusammenhang zwischen ANV und Mortalität, konnte mit der univariaten Cox-Regressionsanalyse bestätigt werden. Ebenso blieb ANV im multivariaten Cox-Regressionsmodell adjustiert für Alter, männliches Geschlecht, Begleiterkrankungen, eGFR, Sepsis und mechanische Beatmung, mit Mortalität assoziiert (Hazard Ratio [HR] = 4,71; 95% Konfidenzintervall [KI]: [4,42; 5,00]). [21]

Es lässt sich zusammenfassend feststellen, dass bei ANV-Episoden in Abhängigkeit vom KDIGO-Stadium, Assoziationen mit kardiovaskulären Komorbiditäten, schweren Krankheitsverläufen, ungünstigen renalen Outcomes sowie Langzeit- und Kurzzeitmortalität zu sehen sind. Schon ANV-Stadium 1 ist mit einer Zunahme von ungünstigem Outcome assoziiert. Abschließend haben wir untersucht, wie oft in den betrachteten Krankenhausaufenthalten (N=185.760), Fälle mit ANV-Episoden mit der administrativen Kodierung N17 (ICD-10-GM) tatsächlich kodiert wurden. In der Gesamtanalyse stellte sich heraus, dass ANV-Episoden mit Stadium 1 nur selten 18,4% (N=4.670) codiert waren. Der Anteil der entsprechend kodierten Fälle stieg mit zunehmenden ANV-Stadium auf 35,9% (ANV-Stadium 2; N=3.054) bzw. 63,9% (ANV-Stadium 3; N=3.757). Auf die einzelnen Jahre getrennt analysiert, lässt sich ein schrittweiser Anstieg der Kodierungen feststellen (z.B. Kodierung ANV-Stadium 1 13,4% im Jahr 2014 und 23,8% im Jahr 2017), was vermutlich mit der verstärkten Umsetzung der Kodierungsempfehlungen zusammenhängt. [21]

Dennoch zeigen unsere Untersuchungen eindeutig eine administrative Untercodierung von ANV, die auch in höheren ANV-Stadien erkennbar ist. Wir konnten in einer weiteren Analyse zeigen, dass auch die nicht kodierten ANV-Episoden (N=97.126) mit erhöhter Langzeitmortalität

assoziiert sind und ein eindeutiger Zusammenhang von Langzeitmortalität und ANV Stadium besteht. Diese Beobachtung legt nahe, dass sogar klinisch relevante ANV-Episoden nicht administrativ erfasst werden. [21]

4.3. Diskussion

4.3.1. ANV: ein unterschätzter Risikofaktor

Akutes Nierenversagen wurde lange Zeit als reversibel angesehen, solange der Patient die Bedingungen überlebte, unter denen es sich entwickelte. In den letzten 15 Jahren durchgeführte Studien haben diese Ansicht widerlegt und gezeigt, dass selbst ein vergleichsweise leichter vorübergehender Rückgang der Nierenfunktion mit einer ungünstigen Prognose verbunden ist. Beispielsweise war ein Anstieg des Kreatinins um 0,3–0,4 mg / dl mit einem 1,7-fachen Anstieg des Mortalitätsrisikos verbunden [9]. Unsere retrospektive Analyse eines Zeitraums von 3,5 Jahren an der Charité in Berlin unterstreicht die Bedeutung von ANV als Risikofaktor bei hospitalisierten Patienten. Fast ein Drittel aller stationären Patienten, die während mindestens zwei Tagen behandelt wurden, und mehr als ein Fünftel aller Krankenhausaufenthalte, waren in diesem Zeitraum mit ANV-Episoden verbunden. Die Tatsache, dass insgesamt weniger als 30% aller ANV-Fälle administrativ kodiert wurden (18% – 64% je nach ANV-Stadium), legt nahe, dass die klinische Erkennung von ANV immer noch unzureichend ist. [21]

4.3.2. Internationaler Vergleich

Wir beobachteten einen stadienabhängigen Zusammenhang zwischen ANV-Episoden und kurz- und langfristiger Krankenhaussterblichkeit, sowie ausgeprägte stadienabhängige Assoziation zwischen ANV und zunehmender Verweildauer und Dialyseabhängigkeit. Dies bestätigen auch internationale Studien [9, 10, 67-69]. Darüber hinaus bestätigte unsere detaillierte Analyse von Primär- und Sekundärdiagnosen und -verfahren die zuvor beschriebenen Zusammenhänge zwischen ANV-Episoden und Sepsis [70, 71], Mechanische Lüftung [70, 71], Herz-Kreislauf-Erkrankungen [72], Krebs [73], Lebererkrankung [74, 75], Chirurgische Eingriffe [76]. Diese Vergleiche zeigen, dass ANV-Episoden, ungeachtet der teilweise signifikanten Unterschiede in der stationären Krankenhausversorgung in den verschiedenen Gesundheitssystemen, einen bemerkenswert konsistenten Risikofaktor in allen Disziplinen darstellen. [21]

4.3.3. Unzureichende ANV Codierung

In fast 40.000 Fällen erkannten wir deutliche Hinweise auf eine Unterkodierung insbesondere bei ANV-Episoden der Stadien 1 und 2. Wir vermuten, dass nicht alle ANV-Episoden im klinischen Alltag erkannt werden. Da unsere Analyse auf fehlende ANV-Episoden aber ausschließlich administrativen Daten untersuchte, lässt sich nicht mit Sicherheit darauf schließen, dass der Arzt die Diagnose nicht gestellt hat. Da die Kodierung aber in der Regel von qualifiziertem Dokumentationspersonal mit Hilfe von medizinischen Berichten durchgeführt wird, kann man davon ausgehen, dass ein erheblicher Teil der ANV-Episoden tatsächlich nicht erkannt wird.

Besonders relevant dabei ist, dass wir zeigen konnten, dass auch in der Untergruppe der nicht administrativ dokumentierten ANV-Fälle ein eindeutiger Zusammenhang zwischen ANV-Episoden und Mortalität bestand. Unsere Beobachtungen decken sich auch mit internationalen Veröffentlichungen. So berichten Grams et al. [63] von einer eingeschränkten Sensitivität (<20%) von administrativen Abrechnungs-codes für die Identifikation von Patienten mit kreatininbasierten ANV in einer US-amerikanischen Population. Ebenso berichtet eine andere retrospektive Kohortenstudie aus einem Universitätskrankenhaus in den USA (Wilson et al. [19]), dass ANV-Episoden nur bei 43% der Patienten mit ANV (definiert als Verdoppelung von Kreatinin) in den medizinischen Unterlagen vermerkt wurde. Interessanterweise berichtet diese Forschergruppe auch, dass die unadjustierte Analyse zeigte, dass die Dokumentation von ANV mit einer höheren Sterblichkeit assoziiert war. Ein Phänomen, welches auch in unserer Kohorte beobachtet wurde. Eine wahrscheinliche Erklärung könnte sein, dass nicht codierte ANV-Episoden mit einer niedrigeren Fallschwere und einer niedrigeren Mortalität assoziiert sind. Dies könnte dadurch belegt werden, dass Wilson et al. [19] berichten, dass sich die Assoziation nach Adjustierung nach dem Schweregrad der Erkrankung umkehrte und eine nicht dokumentierte ANV-Episode mit einer erhöhten Mortalität assoziiert war. Studien in anderen Ländern ergaben ebenfalls eine Nichtanerkennungsrate von ANV von über 70% [20], was belegt, dass der Zeitpunkt der ANV-Erkennung ebenfalls mit der Krankenhaussterblichkeit zusammenhängt. [21]

4.3.4. Potenzial einer automatischen Analyse klinischer Routinedaten

Unsere Untersuchung basierte auf der Implementierung eines Algorithmus zur Identifizierung von ANV-Episoden auf der Basis von Kreatininbestimmungen, die im Rahmen der klinischen Routine erfolgten. Ähnliche ANV-Algorithmen wurden zuvor in anderen Studien angewandt [14, 77-80]. Die meisten dieser Algorithmen sind darauf ausgelegt, ein sich entwickelndes ANV auf Grundlage steigender Kreatininspiegel zu identifizieren. Um eine umfassende Analyse der ANV-Episoden durchführen zu können, haben wir einen erweiterten Algorithmus entwickelt, der zusätzlich einen Abfall des Serumkreatininspiegels während eines Krankenhausaufenthalts identifiziert und so die Auflösung einer ANV-Episode anzeigt.

Unsere Studie zeigt, dass eine relativ einfache automatisierte „elektronische“ Auswertung von gespeicherten Gesundheitsdaten wichtige Zusatzinformationen liefern kann und ein hohes Potenzial hat, um Krankheitscharakteristika und Risikoassoziationen zu erfassen. Die einfache computergestützte Analyse eines einzelnen Laborparameters (Kreatininkonzentration) aus einer Datenbank und die Kombination mit klinischen Daten und Codierungsdaten (Therapien, Haupt- und Nebendiagnosen) aus einer anderen Verwaltungsdatenbank erlaubte es uns, eine umfassende Statusbeschreibung zu erstellen. [21]

4.3.5. Stärke und Schwächen der Studie

Zu den Stärken der Studie gehört die Größe der Kohorte sowie die Tatsache, dass alle Kreatininwerte in einem Zentrallabor bestimmt wurden und somit eine mögliche Methodenvarianz ausgeschlossen werden kann. Eine weitere Stärke ist der implementierte Algorithmus, der sich entwickelnden und abklingenden ANV-Episoden automatisch detektieren kann. Zu den Schwächen der Studie gehört der retrospektive Charakter und damit mögliche nichtberücksichtigte Störgrößen (Confounder), welche einen Einfluss auf die Ergebnisse haben könnten. Außerdem wurden keine anderen ANV-Kriterien (wie zum Beispiel Urinausfuhr, Dialysetherapie etc.) berücksichtigt. Da der implementierte Algorithmus mindestens zwei Kreatininmessungen für die Detektion der ANV-Episoden benötigt, ist nicht auszuschließen, dass dieser Ansatz zu einem Selektionsbias zugunsten schwerkranker Patienten geführt hat. Eventuell wurde die ANV-Inzidenz insgesamt überschätzt. Andererseits blieb unklar, ob es ANV-Episoden gab, die nicht identifiziert wurden, weil es keine Kreatininmessungen gab. Eine weitere Schwäche der Studie ist die Tatsache, dass für die Analysen zum Patientenüberleben nur Krankenhausmortalitätsdaten in der Charité vorlagen. Des Weiteren handelt es sich um eine Monozentrische Studie, sodass die Ergebnisse möglicherweise nicht auf andere Zentren und Kliniken übertragbar sind. [21]

5. Zusammenfassung

In dieser Arbeit wurden die Themen Datenintegration, Wissensrepräsentation und Datenanalyse zur Unterstützung von Outcome Forschung im Fachbereich der Medizinischen Klinik mit Schwerpunkt Nephrologie bearbeitet. In der Studie zur Datenintegration wurde untersucht, inwieweit sich eine auf einem relationalen Datenbankschema basierende medizinische Routinedatenbank mit Langzeitdaten von transplantierten Patienten, in eine Ontologie-basierte Forschungsdatenbank wie i2b2, ohne Informationsverlust überführen lässt. Des Weiteren wurde in der Studie zur Wissensrepräsentation untersucht, wie sich mit Hilfe von Open Source Entwicklungswerkzeugen eine Applikation zur Visualisierung von Informationen aus strukturierten und unstrukturierten medizinischen Daten implementieren lässt. Mit der entwickelten Applikation kann das medizinische Personal ohne Programmierkenntnisse Informationen aus dem medizinischen Datenpool extrahieren und systematisch analysieren. Das Thema Datenanalyse wurde durch die Studie zum akuten Nierenversagen näher beleuchtet. In dieser Studie wurde ein Algorithmus implementiert, der in einer großen Kohorte aus stationären Patientendaten, das Ereignis akutes Nierenversagen (ANV) detektieren kann. Nach der statistischen Auswertung der Ergebnisse dieses Algorithmus, konnte die Kohorte im Hinblick auf das Auftreten von akuten Nierenversagen und den damit verbundenen Krankheitscharakteristika und Risikoassoziationen umfassend beschrieben werden.

In zukünftigen Arbeiten müsste untersucht werden in wieweit Interoperabilität die Probleme bei der Datenintegration lösen kann. Offen ist bisher, wie breit bestehende Standards zum Beispiel SNOMED, LOINC, HL7-FHIR etc. in Deutschland in der Routineversorgung bereits verwendet

werden, ob diese ausreichen oder zusätzliche Standards geschaffen werden müssen, um die Vernetzung und den Austausch von Daten der Akteure im Gesundheitswesen untereinander weiter auszubauen. Beim Thema Wissensrepräsentation ist die weitere Untersuchung der Fragestellung interessant, wie zum Beispiel eine verbesserte Informationsextraktion aus Daten die Weiterentwicklung KI-Anwendungen unterstützen kann. Darüber hinaus wird es zukünftig wichtig sein, durch eine bessere Wissensrepräsentation in den KI-Modellen selbst, eine Transparenz zu schaffen, die es den Anwendern erlaubt zu verstehen, auf welcher Basis KI - Systeme und Algorithmen bestimmte Entscheidungen getroffen haben. Weitere Studien könnten untersuchen, wie neuartige KI-Systeme den Arbeitsalltag (z.B. Entscheidungsunterstützung für Ärzte) im Gesundheitswesen verändern und welche Herausforderungen und Möglichkeiten diese für Ärzte und Patienten mit sich bringen (z.B. bezüglich der Akzeptanz, Patient-Empowerment etc.). Durch valide klinische Studien sollten wie auch in der evidenzbasierten Medizin überprüft werden, inwieweit die im Routinebetrieb genutzten computergestützten Applikationen einen Einfluss auf den Patienten-Outcome haben. Nur so kann ein valider Nachweis über ihren Nutzen für die Patienten erbracht werden.

6. Literaturverzeichnis

1. Veit K., Wessels M., and Deiters W., *Gesundheitsdaten und Digitalisierung – Neue Anforderungen an den Umgang mit Daten im Gesundheitswesen*, in *Digitale Transformation von Dienstleistungen im Gesundheitswesen VI: Impulse für die Forschung*, M.A. Pfannstiel, P. Da-Cruz, and H. Mehlich, Editors. 2019, Springer Fachmedien Wiesbaden: Wiesbaden. p. 19-33.
2. Kaiser M.J. and Fränken J., *Präferenzanalytische Untersuchung von Chancen durch Digitalisierung für eine patientengesteuerte Gesundheitsversorgung mittels elektronischer Patientenakte*, in *Digitale Transformation von Dienstleistungen im Gesundheitswesen VI: Impulse für die Forschung*, M.A. Pfannstiel, P. Da-Cruz, and H. Mehlich, Editors. 2019, Springer Fachmedien Wiesbaden: Wiesbaden. p. 117-137.
3. Topol E.J., Steinhubl S.R., and Torkamani A., *Digital medical tools and sensors*. JAMA, 2015. 313(4): p. 353-354.
4. Bhavnani S.P., Narula J., and Sengupta P.P., *Mobile technology and the digitization of healthcare*. European heart journal, 2016. 37(18): p. 1428-1438.
5. Hanika H., *Digitalisierung, Big Data und Big To-dos aus Sicht der Rechtswissenschaft*. 2019. p. 67-87.
6. Sawhney S., Marks A., Fluck N., Levin A., Prescott G., and Black C., *Intermediate and Long-term Outcomes of Survivors of Acute Kidney Injury Episodes: A Large Population-Based Cohort Study*. Am J Kidney Dis, 2017. 69(1): p. 18-28.
7. Wang H.E., Muntner P., Chertow G.M., and Warnock D.G., *Acute kidney injury and mortality in hospitalized patients*. Am J Nephrol, 2012. 35(4): p. 349-55.

8. Chertow G.M., Soroko S.H., Paganini E.P., Cho K.C., Himmelfarb J., Ikizler T.A., and Mehta R.L., *Mortality after acute renal failure: models for prognostic stratification and risk adjustment*. *Kidney Int*, 2006. 70(6): p. 1120-6.
9. Chertow G.M., Burdick E., Honour M., Bonventre J.V., and Bates D.W., *Acute kidney injury, mortality, length of stay, and costs in hospitalized patients*. *J Am Soc Nephrol*, 2005. 16(11): p. 3365-70.
10. Liangos O., Wald R., O'Bell J.W., Price L., Pereira B.J., and Jaber B.L., *Epidemiology and outcomes of acute renal failure in hospitalized patients: a national survey*. *Clin J Am Soc Nephrol*, 2006. 1(1): p. 43-51.
11. Uchino S., Bellomo R., Goldsmith D., Bates S., and Ronco C., *An assessment of the RIFLE criteria for acute renal failure in hospitalized patients*. *Crit Care Med*, 2006. 34(7): p. 1913-7.
12. Chawla L.S., Amdur R.L., Amodeo S., Kimmel P.L., and Palant C.E., *The severity of acute kidney injury predicts progression to chronic kidney disease*. *Kidney Int*, 2011. 79(12): p. 1361-9.
13. Coca S.G., Singanamala S., and Parikh C.R., *Chronic kidney disease after acute kidney injury: a systematic review and meta-analysis*. *Kidney Int*, 2012. 81(5): p. 442-8.
14. Heung M., Steffick D.E., Zivin K., Gillespie B.W., Banerjee T., Hsu C.Y., Powe N.R., Pavkov M.E., Williams D.E., Saran R., Shahinian V.B., Centers for Disease C., and Prevention C.K.D.S.T., *Acute Kidney Injury Recovery Pattern and Subsequent Risk of CKD: An Analysis of Veterans Health Administration Data*. *Am J Kidney Dis*, 2016. 67(5): p. 742-52.
15. Levey A.S., Levin A., and Kellum J.A., *Definition and classification of kidney diseases*. *Am J Kidney Dis*, 2013. 61(5): p. 686-8.
16. *Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury*. *Kidney Int*, 2012. Suppl(2): p. 1-138.
17. Meersch M., Schmidt C., Hoffmeier A., Van Aken H., Wempe C., Gerss J., and Zarbock A., *Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial*. *Intensive Care Med*, 2017. 43(11): p. 1551-1561.
18. Bihorac A., Yavas S., Subbiah S., Hobson C.E., Schold J.D., Gabrielli A., Layon A.J., and Segal M.S., *Long-term risk of mortality and acute kidney injury during hospitalization after major surgery*. *Ann Surg*, 2009. 249(5): p. 851-8.
19. Wilson F.P., Bansal A.D., Jasti S.K., Lin J.J., Shashaty M.G., Berns J.S., Feldman H.I., and Fuchs B.D., *The impact of documentation of severe acute kidney injury on mortality*. *Clin Nephrol*, 2013. 80(6): p. 417-25.
20. Yang L., Xing G., Wang L., Wu Y., Li S., Xu G., He Q., Chen J., Chen M., Liu X., Zhu Z., Yang L., Lian X., Ding F., Li Y., Wang H., Wang J., Wang R., Mei C., Xu J., Li R.,

- Cao J., Zhang L., Wang Y., Xu J., Bao B., Liu B., Chen H., Li S., Zha Y., Luo Q., Chen D., Shen Y., Liao Y., Zhang Z., Wang X., Zhang K., Liu L., Mao P., Guo C., Li J., Wang Z., Bai S., Shi S., Wang Y., Wang J., Liu Z., Wang F., Huang D., Wang S., Ge S., Shen Q., Zhang P., Wu L., Pan M., Zou X., Zhu P., Zhao J., Zhou M., Yang L., Hu W., Wang J., Liu B., Zhang T., Han J., Wen T., Zhao M., Wang H., and Consortiums I.A.b.C., *Acute kidney injury in China: a cross-sectional survey*. Lancet, 2015. 386(10002): p. 1465-71.
21. Khadzhynov D., Schmidt D., Hardt J., Rauch G., Gocke P., Eckardt K.U., and Schmidt-Ott K.M., *The Incidence of Acute Kidney Injury and Associated Hospital Mortality*. Dtsch Arztebl Int, 2019. 116(22): p. 397-404.
 22. Levin A., Tonelli M., Bonventre J., Coresh J., Donner J.A., Fogo A.B., Fox C.S., Gansevoort R.T., Heerspink H.J.L., Jardine M., Kasiske B., Kottgen A., Kretzler M., Levey A.S., Luyckx V.A., Mehta R., Moe O., Obrador G., Pannu N., Parikh C.R., Perkovic V., Pollock C., Stenvinkel P., Tuttle K.R., Wheeler D.C., Eckardt K.U., and participants I.S.N.G.K.H.S., *Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy*. Lancet, 2017. 390(10105): p. 1888-1917.
 23. Wolfe R., Ashby V., Milford E., Ojo A., Ettenger R., Agodoa L., Held P., and Port F., *Comparison of Mortality in All Patients on Dialysis, Patients on Dialysis Awaiting Transplantation, and Recipients of a First Cadaveric Transplant*. The New England journal of medicine, 1999. 341: p. 1725-30.
 24. Organspende D.S. 2017 [cited 2017 4. April 2017]; Available from: <http://www.dso.de/organspende-und-transplantation/transplantation/nierentransplantation.html>.
 25. Glander P. *Tablettenlast nach Transplantation*. Lebenslinien 2/2010 2010 2010 [cited 2 2/2010]; 10-11].
 26. Schmidt D., Budde K., Sonntag D., Profitlich H.J., Ihle M., and Staeck O., *A novel tool for the identification of correlations in medical data by faceted search*. Comput Biol Med, 2017. 85: p. 98-105.
 27. Maier C., Christoph J., Schmidt D., Ganslandt T., Prokosch H.U., Kraus S., and Sedlmayr M., *Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools*. J Healthc Eng, 2019. 2019: p. 5640685.
 28. Dinu V. and Nadkarni P., *Guidelines for the Effective Use of Entity-Attribute-Value Modeling for Biomedical Databases*. International journal of medical informatics, 2007. 76: p. 769-79.
 29. Bellazzi R., Masseroli M., Murphy S., Shabo A., and Romano P., *Clinical Bioinformatics: Challenges and opportunities*. BMC bioinformatics, 2012. 13 Suppl 14: p. S1.
 30. Post A., Kurc T., Rathod H., Agravat S., Mansour M., Torian W., and Saltz J., *Semantic ETL into i2b2 with Eureka!* AMIA Summits Transl Sci Proc, 2013. accepted.
 31. Haarbrandt B., Tute E., and Marschollek M., *Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository*. J Biomed Inform, 2016. 63: p. 277-294.

32. Bauer C., Ganslandt T., Baum B., Christoph J., Engel I., Löbe M., Mate S., Stäubert S., Drepper J., Prokosch H.-U., Winter A., and Sax U., *Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data*. *Methods of information in medicine*, 2015. 54.
33. Sacco G., *Guided interactive diagnostic systems*. 2005. 117-122.
34. Biron P., Metzger M.H., Pezet C., Sebban C., Barthuet E., and Durand T., *An Information Retrieval System for Computerized Patient Records in the Context of a Daily Hospital Practice: the Example of the Léon Bérard Cancer Center (France)*. *Applied clinical informatics*, 2014. 5: p. 191-205.
35. Jensen P.B., Jensen L.J., and Brunak S., *Mining electronic health records: towards better research applications and clinical care*. *Nat Rev Genet*, 2012. 13(6): p. 395-405.
36. Meystre S.M., Savova G.K., Kipper-Schuler K.C., and Hurdle J.F., *Extracting information from textual documents in the electronic health record: a review of recent research*. *Yearb Med Inform*, 2008: p. 128-44.
37. Krauthammer M. and Nenadic G., *Term identification in the biomedical literature*. *J Biomed Inform*, 2004. 37(6): p. 512-26.
38. Friedman C., *A broad-coverage natural language processing system*. *Proc AMIA Symp*, 2000: p. 270-4.
39. Savova G.K., Masanz J.J., Ogren P.V., Zheng J., Sohn S., Kipper-Schuler K.C., and Chute C.G., *Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications*. *J Am Med Inform Assoc*, 2010. 17(5): p. 507-13.
40. *Health Information Text Extraction (HITEx)*. 2017 [cited 2017 4. April 2017]; Available from: https://www.i2b2.org/software/projects/hitex/hitex_manual.html.
41. Zeng Q.T., Goryachev S., Weiss S., Sordo M., Murphy S.N., and Lazarus R., *Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system*. *BMC Med Inform Decis Mak*, 2006. 6: p. 30.
42. Coden A., Savova G., Sominsky I., Tanenblatt M., Masanz J., Schuler K., Cooper J., Guan W., and de Groen P.C., *Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model*. *J Biomed Inform*, 2009. 42(5): p. 937-49.
43. Slee V.N., Slee D., and Schmidt H.J., *The tyranny of the diagnosis code*. *N C Med J*, 2005. 66(5): p. 331-7.
44. Pakhomov S.V., Buntrock J.D., and Chute C.G., *Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques*. *J Am Med Inform Assoc*, 2006. 13(5): p. 516-25.
45. Ehrler F., Geissbuhler A., Jimeno A., and Ruch P., *Data-poor categorization and passage retrieval for gene ontology annotation in Swiss-Prot*. *BMC Bioinformatics*, 2005. 6 Suppl 1: p. S23.

46. Markó K., Daumke P., Schulz S., and Hahn U., *Cross-language MeSH Indexing using Morpho-Semantic Normalization*. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2003. 2003: p. 425-9.
47. Sebastiani F., *Machine Learning in Automated Text Categorization*. ACM Computing Surveys, 2001. 34: p. 1-47.
48. Odom P., Bangera V., Khot T., Page D., and Natarajan S., *Extracting Adverse Drug Events from Text using Human Advice*. Artif Intell Med Conf Artif Intell Med (2005-), 2015. 2015: p. 195-204.
49. Métivier J.-P., Serrano L., Charnois T., Cuissart B., and Widlöcher A. *Automatic Symptom Extraction from Texts to Enhance Knowledge Discovery on Rare Diseases*. in *Artificial Intelligence in Medicine*. 2015. Cham: Springer International Publishing.
50. Vintar Š., Todorovski L., Sonntag D., and Buitelaar P., *P.: Evaluating context features for medical relation mining*. 2003.
51. Mkrtchyan T. and Sonntag D. *Deep Parsing at the CLEF2014 IE Task (DFKI-Medical)*. 2014.
52. Daniel S., Pinar W., Paul B., and Sonja Z., *Pillars of Ontology Treatment in the Medical Domain*. Journal of Cases on Information Technology (JCIT), 2009. 11(4): p. 47-73.
53. Alicante A., Corazza A., Isgro F., and Silvestri S., *Unsupervised entity and relation extraction from clinical records in Italian*. Comput Biol Med, 2016. 72: p. 263-75.
54. Segagni D., Tibollo V., Dagliati A., Zambelli A., Priori S., and Bellazzi R., *An ICT infrastructure to integrate clinical and molecular data in oncology research*. BMC bioinformatics, 2012. 13 Suppl 4: p. S5.
55. Mate S., Köpcke F., Toddenroth D., Martin M., Prokosch H.-U., Bürkle T., and Ganslandt T., *Ontology-Based Data Integration between Clinical and Research Systems*. PLoS ONE, 2015.
56. Inc. T. *Open studio for data integration*. 2016 [cited 2016; Available from: <http://www.talend.com/products/data-integration>].
57. Murphy S.N., Weber G., Mendis M., Gainer V., Chueh H.C., Churchill S., and Kohane I., *Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)*. J Am Med Inform Assoc, 2010. 17(2): p. 124-30.
58. Sacco G., *Dynamic taxonomies and guided searches*. Journal of the American Society for Information Science and Technology, 2006. 57: p. 792-796.
59. Giovanni M.S., *Dynamic Taxonomies for Intelligent Information Access*, in *Encyclopedia of Information Science and Technology, Second Edition*. 2009, IGI Global: Hershey, PA, USA. p. 1209-1215.
60. Lenihan C.R., Montez-Rath M.E., Mora Mangano C.T., Chertow G.M., and Winkelmayr W.C., *Trends in acute kidney injury, associated use of dialysis, and mortality after cardiac surgery, 1999 to 2008*. Ann Thorac Surg, 2013. 95(1): p. 20-8.

61. Waikar S.S., Curhan G.C., Wald R., McCarthy E.P., and Chertow G.M., *Declining mortality in patients with acute renal failure, 1988 to 2002*. J Am Soc Nephrol, 2006. 17(4): p. 1143-50.
62. Xue J.L., Daniels F., Star R.A., Kimmel P.L., Eggers P.W., Molitoris B.A., Himmelfarb J., and Collins A.J., *Incidence and mortality of acute renal failure in Medicare beneficiaries, 1992 to 2001*. J Am Soc Nephrol, 2006. 17(4): p. 1135-42.
63. Grams M.E., Waikar S.S., MacMahon B., Whelton S., Ballew S.H., and Coresh J., *Performance and limitations of administrative data in the identification of AKI*. Clin J Am Soc Nephrol, 2014. 9(4): p. 682-9.
64. *Deutsche Gesellschaft für Nephrologie: Aktualisierte Stellungnahme zur Kodierung von Nierenerkrankungen (AKI, CKD)*. 2019 [cited 2019 15.03.2019]; Available from: <https://www.dgfn.eu/kommission-drg-details/dgfn-aktualisierte-stellungnahme-zur-kodierung-von-nierenerkrankungen-aki-ckd-22.html>.
65. Levey A.S., Stevens L.A., Schmid C.H., Zhang Y.L., Castro A.F., 3rd, Feldman H.I., Kusek J.W., Eggers P., Van Lente F., Greene T., Coresh J., and Ckd E.P.I., *A new equation to estimate glomerular filtration rate*. Ann Intern Med, 2009. 150(9): p. 604-12.
66. Sawhney S., *Automated alerts for acute kidney injury warrant caution*. BMJ, 2015. 350: p. h19.
67. See E.J., Jayasinghe K., Glassford N., Bailey M., Johnson D.W., Polkinghorne K.R., Toussaint N.D., and Bellomo R., *Long-term risk of adverse outcomes after acute kidney injury: a systematic review and meta-analysis of cohort studies using consensus definitions of exposure*. Kidney Int, 2018.
68. Zeng X., McMahon G.M., Brunelli S.M., Bates D.W., and Waikar S.S., *Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals*. Clin J Am Soc Nephrol, 2014. 9(1): p. 12-20.
69. Liang K.V., Sileanu F.E., Clermont G., Murugan R., Pike F., Palevsky P.M., and Kellum J.A., *Modality of RRT and Recovery of Kidney Function after AKI in Patients Surviving to Hospital Discharge*. Clin J Am Soc Nephrol, 2016. 11(1): p. 30-8.
70. Liborio A.B., Leite T.T., Neves F.M., Teles F., and Bezerra C.T., *AKI complications in critically ill patients: association with mortality rates and RRT*. Clin J Am Soc Nephrol, 2015. 10(1): p. 21-8.
71. Wald R., McArthur E., Adhikari N.K., Bagshaw S.M., Burns K.E., Garg A.X., Harel Z., Kitchlu A., Mazer C.D., Nash D.M., Scales D.C., Silver S.A., Ray J.G., and Friedrich J.O., *Changing incidence and outcomes following dialysis-requiring acute kidney injury among critically ill adults: a population-based cohort study*. Am J Kidney Dis, 2015. 65(6): p. 870-7.
72. Chawla L.S., Amdur R.L., Shaw A.D., Faselis C., Palant C.E., and Kimmel P.L., *Association between AKI and long-term renal and cardiovascular outcomes in United States veterans*. Clin J Am Soc Nephrol, 2014. 9(3): p. 448-56.

73. Perazella M.A. and Rosner M.H., *Acute Kidney Injury in Patients With Cancer*. Oncology (Williston Park), 2018. 32(7): p. 351-9.
74. Chen N., Chen X., Ding X., and Teng J., *Analysis of the high incidence of acute kidney injury associated with acute-on-chronic liver failure*. Hepatol Int, 2018. 12(3): p. 262-268.
75. Gessolo Lins P.R., Carvalho Padilha W.S., Magalhaes Giradin Pimentel C.F., Costa Batista M., and Teixeira de Gois A.F., *Risk factors, mortality and acute kidney injury outcomes in cirrhotic patients in the emergency department*. BMC Nephrol, 2018. 19(1): p. 277.
76. O'Connor M.E., Kirwan C.J., Pearse R.M., and Prowle J.R., *Incidence and associations of acute kidney injury after major abdominal surgery*. Intensive Care Med, 2016. 42(4): p. 521-530.
77. Broce J.C., Price L.L., Liangos O., Uhlig K., and Jaber B.L., *Hospital-acquired acute kidney injury: an analysis of nadir-to-peak serum creatinine increments stratified by baseline estimated GFR*. Clin J Am Soc Nephrol, 2011. 6(7): p. 1556-65.
78. Park S., Baek S.H., Ahn S., Lee K.H., Hwang H., Ryu J., Ahn S.Y., Chin H.J., Na K.Y., Chae D.W., and Kim S., *Impact of Electronic Acute Kidney Injury (AKI) Alerts With Automated Nephrologist Consultation on Detection and Severity of AKI: A Quality Improvement Study*. Am J Kidney Dis, 2018. 71(1): p. 9-19.
79. Porter C.J., Juurlink I., Bisset L.H., Bavakunji R., Mehta R.L., and Devonald M.A., *A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital*. Nephrol Dial Transplant, 2014. 29(10): p. 1888-93.
80. Siew E.D., Parr S.K., Abdel-Kader K., Eden S.K., Peterson J.F., Bansal N., Hung A.M., Fly J., Speroff T., Ikizler T.A., and Matheny M.E., *Predictors of Recurrent AKI*. J Am Soc Nephrol, 2016. 27(4): p. 1190-200.

Eidesstattliche Versicherung

„Ich, Danilo Schmidt, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „**Datenintegration, Wissensrepräsentation und Datenanalyse – Werkzeuge zur systematischen Untersuchung von Einflussfaktoren auf das Langzeit-Outcome nephrologischer Patienten**“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilerklärung an den erfolgten Publikationen

Danilo Schmidt hatte folgenden Anteil an den ausgewählten Publikationen

Publikation 1:

Maier C, Christoph J, **Schmidt D**, Ganslandt T, Prokosch HU, Kraus S, Sedlmayr M., *Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools*, J Healthc Eng, 2019

Beitrag von **Danilo Schmidt** im Einzelnen:

Bereitstellung und Vorverarbeitung der Datenbasis, Aufbereitung der Datenstruktur und Pseudonymisierung der Daten. Aufbau der i2b2 Datenbank in der Klinik für Nephrologie. Datentransformation aus TBase in i2b2. Suche und Analyse von geeigneten Studienmodellen aus dem Fachbereich Nephrologie für die Bereitstellung eines geeigneten Vergleichsmodells für die Auswertung der Ergebnisse. Harmonisierung der Daten (Laborbefunde und Transplantationsdaten). Bearbeitung der entsprechenden Publikationsabschnitte.

Publikation 2:

Schmidt D, Budde K, Sonntag D, Profitlich HJ, Ihle M, Staeck O.; *A novel tool for the identification of correlations in medical data by faceted search.*; Comput Biol Med. 2017 Jun 1;85:98-105. doi: 10.1016/j.combiomed.2017.04.011. Epub 2017 Apr 14.

Beitrag von **Danilo Schmidt** im Einzelnen:

Erzeugung eines Demodatensatzes für die Entwicklung der SOLr Indexierung durch das DFKI sowie für das Textminingmodul durch Averbis. Bereitstellung und Vorverarbeitung der Datenbasis für die Applikationsentwicklung innerhalb der Charité (Fachabteilung Nephrologie): Aufbereitung der Datenstruktur und Pseudonymisierung der Daten. Anteil am Design und Festlegung der Funktionalitäten des Frontends der facettierten Suche. Harmonisierung der Daten. Entwicklung der SQL Abfragen für die Datenrepräsentation. Annotationen von Arztbriefen für die Erstellung des Modells für die Informationsextraktion aus Texten. Durchführung der erforderlichen Literaturrecherche. Entwurf und Bearbeitung des Manuskripttextes durch den Review Prozess, Einreichung der Publikation.

Publikation 3:

Khadzhynov D*, Schmidt D*, Hardt J, Rauch G, Gocke P, Eckardt KU, Schmidt-Ott KM.; *The Incidence of Acute Kidney Injury and Associated Hospital Mortality.*; Dtsch Arztebl Int. 2019 May 31;116(22):397-404. doi: 10.3238/arztebl.2019.0397.

***geteilte Erstautorenschaft**

Beitrag von **Danilo Schmidt** im Einzelnen:

Durchführung der erforderlichen Literaturrecherche. Implementierung der Datenanbindung und Auswahl von geeigneten Parametern. Programmierung des Algorithmus für die ANV Detektion. Diskussion und Anpassung des Algorithmus. Beschreibung der Kohorte und Analysen (Table 2, Table 3, Table 4, Figure 2 und eFigure - study flow chart). Bearbeitung der entsprechenden Publikationsabschnitte.

Beitrag von **Dmytro Khadzhynov** im Einzelnen:

Durchführung der erforderlichen Literaturrecherche. Einführung in das Thema und Definition der AKI Episoden Szenarien (Tabelle 1, Grafik 1). Diskussion und internationaler Vergleich. Bearbeitung der entsprechenden Publikationsabschnitte. Entwurf und Bearbeitung des Manuskripttextes durch den Review Prozess, Einreichung der Publikation.

Unterschrift Dmytro Khadzhynov

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

Unterschrift des Doktoranden/der Doktorandin

Druckexemplar Publikation 1:

Maier C, Christoph J, **Schmidt D**, Ganslandt T, Prokosch HU, Kraus S, Sedlmayr M., *Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools*, J Healthc Eng, 2019

<https://doi.org/10.1155/2019/5640685>

Druckexemplar Publikation 2:

Schmidt D, Budde K, Sonntag D, Profitlich HJ, Ihle M, Staeck O.; *A novel tool for the identification of correlations in medical data by faceted search.*; Comput Biol Med. 2017 Jun 1;85:98-105. doi: 10.1016/j.combiomed.2017.04.011. Epub 2017 Apr 14.

<https://doi.org/10.1016/j.combiomed.2017.04.011>

Druckexemplar Publikation 3:

Khadzhynov D*, Schmidt D*, Hardt J, Rauch G, Gocke P, Eckardt KU, Schmidt-Ott KM.; *The Incidence of Acute Kidney Injury and Associated Hospital Mortality.*; Dtsch Arztebl Int. 2019 May 31;116(22):397-404. doi: 10.3238/arztebl.2019.0397.

***geteilte Erstautorenschaft**

<https://doi.org/10.3238/arztebl.2019.0397>

CURRICULUM VITAE

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Vollständige Publikationsliste

Khadzhynov D*, **Schmidt D***, Hardt J, Rauch G, Gocke P, Eckardt KU, Schmidt-Ott KM., *The Incidence of Acute Kidney Injury and Associated Hospital Mortality.*, Dtsch Arztebl Int. 2019 May 31;116(22):397-404. doi: 10.3238/arztebl.2019.0397., PMID: 31366430

*geteilte Erstautorenschaft

Maier C, Christoph J, **Schmidt D**, Ganslandt T, Prokosch HU, Kraus S, Sedlmayr M., *Experiences of Transforming a Complex Nephrologic Care and Research Database into i2b2 Using the IDRT Tools.*, J Healthc Eng. 2019 Jan 17;2019:5640685. doi: 10.1155/2019/5640685. eCollection 2019., PMID: 30800257

Kleinsteuber A, Halleck F, Khadzhynov D, Staeck A, Lehner L, Duerr M, Glander P, **Schmidt D**, Budde K, Staeck O., *Impact of Pre-existing Comorbidities on Long-term Outcomes in Kidney Transplant Recipients.*, Transplant Proc. 2018 Dec;50(10):3232-3241. doi: 10.1016/j.transproceed.2018.08.028. Epub 2018 Sep 7., PMID: 30577191

Rissling O, Naik M, Brakemeier S, **Schmidt D**, Staeck O, Hohberger A, Neumayer HH, Budde K., *High frequency of valganciclovir underdosing for cytomegalovirus prophylaxis after renal transplantation.* Clin Kidney J. 2018 Aug;11(4):564-573. doi: 10.1093/ckj/sfx145. Epub 2018 Jan 12., PMID: 30094022

Kara E, Zeen T, Gabryszak A, Budde K, **Schmidt D**, Roller R, *A Domain-adapted Dependency Parser for German Clinical Text*, In Proceedings of KONVENS 2018 "The Conference on Natural Language Processing", Vienna, Austria

Lachmann N, Niemann M, Reinke P, Budde K, **Schmidt D**, Halleck F, Pruß A, Schönemann C, Spierings E, Staeck O., *Invited letter in response to "Predicted indirectly recognizable HLA epitopes (PIRCHE): Only the tip of the iceberg?"* Am J Transplant. 2018 Feb;18(2):523-524. doi: 10.1111/ajt.14535. Epub 2017 Nov 4., PMID: 28992384

Khadzhynov D, Halleck F, Lehner L, **Schmidt D**, Schrezenmeier E, Budde K, Staeck O. *Immunologic Long-term Outcomes of Living-Related Kidney Transplantations Depending on the Donor-Recipient Relationship.* Transplant Proc. 2017 Dec;49(10):2265-2268.

Halleck F, **Schmidt D**, Georgi S, Graf V, Gaede S, Roller R, Budde K, Staeck O. *mHealth and digital management after kidney transplantation.*, Nieren- und Hochdruckkrankheiten. Volume 46 11/2017, Pages 474-480

Dürr M, Lachmann N, Zukunft B, **Schmidt D**, Budde K, Brakemeier S., *Late Conversion to Belatacept After Kidney Transplantation: Outcome and Prognostic Factors.*, Transplant Proc. 2017 Oct;49(8):1747-1756.e1. doi: 10.1016/j.transproceed.2017.05.013., PMID: 28923620

Brakemeier S, Taxeidi SI, Zukunft B, **Schmidt D**, Gaedeke J, Dürr M, Hansen S, Budde K., *Extended-Spectrum Beta-Lactamase-Producing Enterobacteriaceae-Related Urinary Tract Infection in Kidney Transplant Recipients: Risk Factors, Treatment, and Long-Term Outcome.*, Transplant Proc. 2017 Oct;49(8):1757-1765. doi: 10.1016/j.transproceed.2017.06.033., PMID: 28923621

Roller R, Rethmeier N, Thomas P, Hübner M, Uszkoreit H, Halleck F, Staeck O, Budde K, **Schmidt D** (2017): *Detecting Named Entities and Relations in German Clinical Reports. In: Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology, GSCL 2017, Berlin, Germany.*

Lachmann N, Niemann M, Reinke P, Budde K, **Schmidt D**, Halleck F, Pruss A, Schonemann C, Spierings E, Staeck O. *Donor-Recipient Matching Based on Predicted Recognizable HLA Epitopes Predicts the Incidence of De Novo Donor-Specific HLA Antibodies Following Renal Transplantation.* American Journal of Transplantation. 2017 Jun 14.

Schmidt D, Budde K, Sonntag D, Profitlich HJ, Ihle M, Staeck O. *A novel tool for the identification of correlations in medical data by faceted search.*, Comput Biol Med. 2017 Jun 1;85:98-105. doi: 10.1016/j.combiomed.2017.04.011. Epub 2017 Apr 14. PubMed PMID: 28499136.

Brakemeier S, Dürr M, Bachmann F, **Schmidt D**, Gaedeke J, Budde K., *Risk Evaluation and Outcome of Pneumocystis jirovecii Pneumonia in Kidney Transplant Patients.*, Transplant Proc. 2016 Nov;48(9):2924-2930. doi: 10.1016/j.transproceed.2016.05.017., PMID: 27932109

Halleck F, Khadzhynov D, Liefeldt L, Schrezenmeier E, Lehner L, Duerr M, **Schmidt D**, Bamoulid J, Lachmann N, Waiser J, Budde K, Staeck O. *Immunologic outcome in elderly kidney transplant recipients: is it time for HLA-DR matching?* Nephrol Dial Transplant. 2016 Dec;31(12):2143-2149. Epub 2016 Jul 1. PubMed PMID: 27369852.

Halleck F, **Schmidt D**, Staeck O, Schaaf T, Tolxdorff T, Löser A, Yu Xu F, Uszkoreit H, Legge P, Sachs K, Oechsler S, Budde K. *Integrated care of renal transplant patients – Development of an electronic health care service platform.* Dialyse aktuell 2016; 20(06): 285-290, Georg Thieme Verlag Stuttgart, New York 7/2016.

Brakemeier S, Kannenkeril D, Dürr M, Braun T, Bachmann F, **Schmidt D**, Wiesener M, Budde K. *Experience with belatacept rescue therapy in kidney transplant recipients*. *Transpl Int*. 2016 Nov;29(11):1184-1195. doi: 10.1111/tri.12822. Epub 2016 Sep 14. PubMed PMID: 27514317.

Brakemeier S, Taxeidi SI, Dürr M, Hofmann J, **Schmidt D**, Bachmann F, Gaedeke J, Budde K. *Clinical outcome of norovirus infection in renal transplant patients*. *Clin Transplant*. 2016 Oct;30(10):1283-1293. doi: 10.1111/ctr.12820. Epub 2016 Aug 23. PubMed PMID: 27447127.

Matz M, Fabritius K, Lorkowski C, Dürr M, Gaedeke J, Durek P, Grün JR, Goestemeyer A, Bachmann F, Wu K, Rudolph B, **Schmidt D**, Weber U, Haftmann C, Unterwalder N, Lachmann N, Radbruch A, Neumayer HH, Mashreghi MF, Budde K. *Identification of T Cell-Mediated Vascular Rejection After Kidney Transplantation by the Combined Measurement of 5 Specific MicroRNAs in Blood*. *Transplantation*. 2016 Apr;100(4):898-907. doi: 10.1097/TP.0000000000000873. PubMed PMID: 26444957.

Wu K, Budde K, **Schmidt D**, Neumayer HH, Lehner L, Bamoulid J, Rudolph B. *The inferior impact of antibody-mediated rejection on the clinical outcome of kidney allografts that develop de novo thrombotic microangiopathy*. *Clin Transplant*. 2016 Feb;30(2):105-17. doi: 10.1111/ctr.12645. Epub 2016 Jan 5. PubMed PMID: 26448478.

Cotik V, Roller R, Xu F, Uszkoreit H, Budde K, **Schmidt D**, *Negation Detection in Clinical Reports Written in German, In Proceedings of the 5th Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM)*, 2016 Osaka, Japan

Roller R, Uszkoreit H, Xu F, Seiffe L, Mikhailov M, Staek O, Budde K, Halleck F, **Schmidt D**, *A fine-grained corpus annotation schema of German nephrology records, In Proceedings of the Clinical Natural Language Processing Workshop*, 2016 Osaka, Japan

Sonntag D, Tresp V, Zillner S, Cavallaro A, Hammon M, Reis A, Fasching PA, Sedlmayr M, Ganslandt T, Prokosch HU, Budde K, **Schmidt D**, Hinrichs C, Wittenberg T, Daumke P, Oppelt PG, *The Clinical Data Intelligence Project - A Smart Data Initiative*, *Informatik Spektrum*, 39(4): 290-300 (2016), Official Link: <https://doi.org/10.1007/s00287-015-0913-x>

Staek O, Slowinski T, Lieker I, Wu K, Rudolph B, **Schmidt D**, Brakemeier S, Neumayer HH, Wei C, Reiser J, Budde K, Halleck F, Khadzhynov D. *Recurrent Primary Focal Segmental Glomerulosclerosis Managed With Intensified Plasma Exchange and Concomitant Monitoring of Soluble Urokinase-Type Plasminogen Activator Receptor-Mediated Podocyte β 3-integrin Activation*. *Transplantation*. 2015 Dec;99(12):2593-7. doi: 10.1097/TP.0000000000000914. PubMed PMID: 26371597; PubMed Central PMCID: PMC4900174.

Esteban C, **Schmidt D**, Krompaß D, Tresp V, *Predicting Sequences of Clinical Events by Using a Personalized Temporal Latent Embedding Model*, Proceedings IEEE International Conference on Healthcare Informatics(ICH), Dallas, TX, USA, 2015: 130-139, Official Link: <https://doi.org/10.1109/ICHI.2015.23>

Wu K, Budde K, **Schmidt D**, Neumayer HH, Rudolph B. *The Relationship of the Severity and Category of Acute Rejection With Intimal Arteritis Defined in Banff Classification to Clinical Outcomes. Transplantation.* 2015 Aug;99(8):e105-14. doi: 10.1097/TP.0000000000000640. PubMed PMID: 25719260.

Huber L, Lachmann N, Niemann M, Naik M, Liefeldt L, Glander P, **Schmidt D**, Halleck F, Waiser J, Brakemeier S, Neumayer HH, Schönemann C, Budde K. *Pretransplant virtual PRA and long-term outcomes of kidney transplant recipients.* Transpl Int. 2015 Jun;28(6):710-9. doi: 10.1111/tri.12533. PubMed PMID: 25639331.

Wu K, Budde K, **Schmidt D**, Neumayer HH, Rudolph B. *Designation of T-cell-mediated rejection type Ib and IIa reflects the type of rejection rather than the severity in the Banff classification.* Exp Clin Transplant. 2014 Oct;12(5):415-23. PubMed PMID: 25299369.

Wu K, Budde K, **Schmidt D**, Neumayer HH, Rudolph B. *Pathologic characteristics of early or late acute cellular rejection and outcome after kidney transplant.* Exp Clin Transplant. 2014 Aug;12(4):314-22. PubMed PMID: 25095709.

Kurbalija V, Radovanović M, Ivanović M, **Schmidt D**, von Trzebiatowski GL, Burkhard HD, Hinrichs C. *Time-series analysis in the medical domain: a study of Tacrolimus administration and influence on kidney graft function.* Comput Biol Med. 2014 Jul;50:19-31. doi: 10.1016/j.compbiomed.2014.04.007. Epub 2014 Apr 18. PubMed PMID: 24813681.

Wu K, Budde K, Lu H, **Schmidt D**, Liefeldt L, Glander P, Neumayer HH, Rudolph B. *The severity of acute cellular rejection defined by Banff classification is associated with kidney allograft outcomes.* Transplantation. 2014 Jun 15;97(11):1146-54. doi: 10.1097/01.TP.0000441094.32217.05. PubMed PMID: 24892962.

Wu KY, Budde K, **Schmidt D**, Neumayer HH, Rudolph B., *Acute cellular rejection with isolated v-lesions is not associated with more favorable outcomes than vascular rejection with more tubulointerstitial inflammations.*, Clin Transplant. 2014 Apr;28(4):410-8. doi: 10.1111/ctr.12333. Epub 2014 Mar 11., PMID: 24617534

Friedersdorff F, Roller C, Manus P, Cash H, Stier K, **Schmidt D**, Budde K, Kemkensteffen C, Busch J, Fuller TF, Giessing M. *Fate of finally transplanted deceased donor kidneys initially rejected at other kidney transplantation centers.* Urol Int. 2014;93(4):474-81. doi: 10.1159/000360302. Epub 2014 Aug 22. PubMed PMID: 25171397.

Schürmann M, Schürmann D, Schindler R, Meisel C, Liman P, Kruse J, Enghard P, König J, **Schmidt D**, Reinke P, Nickel P. *Impaired thymic function and CD4+ T lymphopenia, but not mannose-binding lectin deficiency, are risk factors for Pneumocystis jirovecii pneumonia in kidney transplant recipients.* Transpl Immunol. 2013 Jun;28(4):159-63. doi: 10.1016/j.trim.2013.05.003. Epub 2013 May 15. PubMed PMID: 23684946.

Tresp V, Zillner S, Costa MJ, Huang Y, Cavallaro A, Fasching PA, Reis A, Sedlmayr M, Ganslandt T, Budde K, Hinrichs C, **Schmidt D**, Daumke P, Sonntag D, Wittenberg T, Oppelt PG, Krompaß D, *Towards a New Science of a Clinical Data Intelligence*, Proceedings NIPS Workshop on Machine Learning for Clinical Data Analysis and Healthcare(NIPS Workshop), Lake Tahoe, NV, USA, 2013: null.

Cash H, Slowinski T, Buechler A, Grimm A, Friedersdorff F, **Schmidt D**, Miller K, Giessing M, Fuller TF. *Impact of surgeon experience on complication rates and functional outcomes of 484 deceased donor renal transplants: a single-centre retrospective study.* BJU Int. 2012 Oct;110(8 Pt B):E368-73. doi: 10.1111/j.1464-410X.2012.011024.x. Epub 2012 Mar 9. PubMed PMID: 22404898.

Liefeldt L, Brakemeier S, Glander P, Waiser J, Lachmann N, Schönemann C, Zukunft B, Illigens P, **Schmidt D**, Wu K, Rudolph B, Neumayer HH, Budde K., *Donor-specific HLA antibodies in a cohort comparing everolimus with cyclosporine after kidney transplantation.*, Am J Transplant. 2012 May;12(5):1192-8. doi: 10.1111/j.1600-6143.2011.03961.x. Epub 2012 Feb 2., PMID: 22300538

Pagonas N, Hörstrup J, **Schmidt D**, Benz P, Schindler R, Reinke P, van der Giet M, Zidek W, Westhoff TH., *Prophylaxis of recurrent urinary tract infection after renal transplantation by cranberry juice and L-methionine.*, Transplant Proc. 2012 Dec;44(10):3017-21. doi: 10.1016/j.transproceed.2012.06.071., PMID: 23195017

Giessing M, Fuller TF, Friedersdorff F, Deger S, **Schmidt D**, Neumayer HH, Budde K, Liefeldt L. *Comparison of first and second kidney transplants from the same deceased donor.* Nephrol Dial Transplant. 2010 Dec;25(12):4055-61. doi: 10.1093/ndt/gfq338. Epub 2010 Jun 17. PubMed PMID: 20558663.

Glander P, Budde K, **Schmidt D**, Fuller TF, Giessing M, Neumayer HH, Liefeldt L. *The 'blood group O problem' in kidney transplantation--time to change?* Nephrol Dial Transplant. 2010 Jun;25(6):1998-2004. doi: 10.1093/ndt/gfp779. Epub 2010 Jan 25. PubMed PMID: 20100733.

Lachmann N, Terasaki PI, Budde K, Liefeldt L, Kahl A, Reinke P, Pratschke J, Rudolph B, **Schmidt D**, Salama A, Schönemann C. *Anti-human leukocyte antigen and donor-specific antibodies detected by luminex posttransplant serve as biomarkers for chronic rejection of renal allografts.* Transplantation. 2009 May 27;87(10):1505-13. doi: 10.1097/TP.0b013e3181a44206. PubMed PMID: 19461487.

Giessing M, Fuller TF, Friedersdorff F, Deger S, Wille A, Neumayer HH, **Schmidt D**, Budde K, Liefeldt L. *Outcomes of transplanting deceased-donor kidneys between elderly donors and recipients.* J Am Soc Nephrol. 2009 Jan;20(1):37-40. doi: 10.1681/ASN.2008040423. Epub 2008 Dec 10. PubMed PMID: 19073824; PubMed Central PMCID: PMC2615731.

Hildebrand F, Lindemann G, **Schmidt D**: *Application of Case Retrieval Nets in the OpEN.SC Project*, Informatikberichte, Concurrency, Specification and Programming CS&P'2008, pp. 218-228, ISSN: 0863-095X, Groß Väter See, 29. September – 01. Oktober, Germany, 2008.

Gestewitz MD, Lindemann G, **Schmidt D**: *Ontologies in OpEN.SC – Building an Ontology for the domains Nephrology and Pathology*, Informatikberichte, Concurrency, Specification and Programming CS&P'2008, pp. 138-144, ISSN: 0863-095X, Groß Väter See, 29. September – 01. Oktober, Germany, 2008.

Lindemann G, **Schmidt D**, Schrader T, Beil M, Schaaf T, Burkhard H-D: *The System Architecture of the Open European Nephrology Science Centre*, Proceedings of World Academy of Science, Engineering and Technology, CESSE 2007, pp. 422-427, ISSN 1307-6884, Venice, 23.-25. November, Italy, 2007.

Lindemann G, **Schmidt D**, Schrader T, Keune D: *The Resource Description Framework (RDF) as a modern Structure for Medical Data*, Proceedings of World Academy of Science, Engineering and Technology, CESSE 2007, pp. 422-427, ISSN 1307-6884, Venice, 23.-25. November, Italy, 2007.

Schrader T, **Schmidt D**, Lindemann G, Niepage S, Hahn C, Keune D, Wetzel T, Krieger S, Weckend T, Peters H, Budde K, Rudolf B: *The Open European Nephrology Science Center: An International Research Resource Centre for Kidney Diseases*, Iran Journal of Kidney Diseases, 11th International Congress of Nephrology, Dialysis and Transplantation, pp. 4-5, ISSN 1735-8582, 21.-23. November, Iran, Tehran 2007.

Schmidt D, Lindemann G, Budde K, Schrader T: *TBase – A web-based Electronic Patient Record in clinical practice*, Iran Journal of Kidney Diseases, 11th International Congress of Nephrology, Dialysis and Transplantation, pp. 4-5, ISSN 1735-8582, , 21.-23. November, Iran, Tehran 2007.

Schmidt D, Lindemann G, Schrader T: *First Steps towards an Intelligent Catalogue within the Open European Nephrology Science Center – OpEN.SC*, 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), pp. 39-44, 29.-31. October, Greece, Patras 2007.

Keune D, Lindemann G, **Schmidt D**, Schrader T: *A modern Database Structure for Patient Data within the OpEN.SC project*, Proceedings of the CS&P2007, Concurrency, Specification & Programming 2007, pp. 357-364, ISBN: 978-83-88374-28-9, Lagow, 27.-29. September, Poland, 2007.

Schrader T, Lindemann G, **Schmidt D**: *Open European Nephrology Science Centre - webservices based knowledge management portal*, International Journal of Computer Assisted radiology and Surgery, Proceedings of the 21th International Congress and Exhibition Berlin, Volume 2, pp. S508, DOI: 10.1007/s11548-007-0116-2, Springer, ISSN 1861-6410, 27.-30. June, Germany, Berlin 2007.

Schrader T, Beil M, Schaaf T, **Schmidt D**, Lindemann G: *Open European Nephrology Science Center – OpEN.SC Geschäftsprozessmodellierung im Informationszentrum für Nierenerkrankungen und Nephrologischer Transplantationsmedizin*, 91. Jahrestagung der Deutschen Gesellschaft für Pathologie, 30. May – 02. June, Germany, Magdeburg 2007

Lindemann G, **Schmidt D**: *Case Based Reasoning in Medical Domains*, 2st Workshop of the Open European Nephrology Science Centre 2007, Germany, Berlin 2007.

Lindemann G, **Schmidt D**, Schrader T: *The Intelligent Catalogue as part of the Open European Nephrology Science Center - a case study -*, ICICIS 2007, Cairo 2007.

Schrader T, Beil M, **Schmidt D**, Lindemann G: *Open European Nephrology Science Centre (OpEN.SC) & Integrated Virtual Microscopy*, Medical Imaging 2007: PACS and Imaging Informatics. Edited by Horii, Steven C.; Andriole, Katherine P.. Proceedings of the SPIE, Volume 6516, pp. 651606 (2007). (SPIE Homepage); SPIE--The International Society for Optical Engineering; DOI: 10.1117/12.709623; March, USA, Atlanta 2007.

Lindemann G, Schrader T, **Schmidt D**: *The Aim of the Open European Nephrology Science Centre - OpEN.SC*, Proceedings of the CS&P2006, Concurrency, Specification & Programming 2006, pp. 480-486, ISSN: 0863-095X, Wandlitz, 27.-29. September, Germany, 2006.

Buchbeiträge

Schmidt D, Graf V, Roller R, Staeck O, Tolxdorff T, Schaaf T, Löser A, Xu F, Uszkoreit H, Legge P, Lurz G, Wedekind M, Sachs K, Oechsler S, Budde K, Gaede S, Halleck F: *Integrierte Versorgung chronisch kranker Patienten am Beispiel von MACSS*; Workshop “IT-Anwendungen für soziale Inklusion”; De Gruyter Verlag, ISBN 3110558971, 9783110558975, Berlin, 2018.

Düttmann-Rehnolt W, **Schmidt D**, Halleck F, Staeck O, Roller R, Högl M, Neuartige *Kommunikationswege und Strukturen zur Optimierung der häuslichen Versorgung am Beispiel von nierentransplantierten Patienten.*, in: Pfannstiel M., Da-Cruz P., Mehlich H. (eds) *Digitale Transformation von Dienstleistungen im Gesundheitswesen VI*. Springer Gabler, Wiesbaden, ISBN 978-3-658-25460-5, DOI https://doi.org/10.1007/978-3-658-25461-2_21

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei Herrn Professor Dr. Ulrich Frei und Herrn Professor Dr. Jan Steffen Jürgensen für die Betreuung und Überlassung des Dissertationsthemas bedanken. Ganz besonderer Dank gilt Professor Dr. Klemens Budde, Dr. Petra Glander, PD Dr. Oliver Staeck, Dr. Fabian Halleck, Dr. Dmytro Khadzhynov sowie Dr. Lutz Liefeldt. Ihre inspirierenden und innovativen Ideen die medizinische Forschung durch neue Lösungen voranzutreiben haben mich immer begeistert und mitgetragen. Sie haben mich mit ihren visionären Ansichten bezüglich dem Einsatz von digitalen Technologien in Forschung und Routineversorgung, in meiner Arbeit stets unterstützt und somit zum Gelingen dieser Doktorarbeit beigetragen. Großer Dank gilt ebenfalls Frau Dr. Gabriela Lindemann-von Trzebiatowski von der Humboldt Universität zu Berlin. Sie hat mich als Mentorin und gute Freundin auf meinem wissenschaftlichen Werdegang begleitet. Ich habe viel von ihr gelernt.

Ebenso danke ich dem gesamten Ärzteteam der Medizinischen Klinik für Nephrologie für die inspirierende Zusammenarbeit. Sowie dem immer größer gewordenen TBase-Team, das fleißig dafür sorgt, dass die gewonnenen Erkenntnisse für die wissenschaftliche Arbeit der Ärztinnen und Ärzte und zum Wohle der Patienten eingesetzt werden können.

Vielen Dank auch an meine Frau Ann-Kathrin, die mich für den Abschluss dieser Arbeit auf den letzten Metern unterstützt und mit der Geburt unserer Tochter Ellen reich beschenkt hat.

Nicht zuletzt gilt mein Dank auch meinen Eltern und meinem Bruder und Familie, die meinen Werdegang stets mit viel Anteilnahme und sehr interessiert verfolgt haben.