

**Nutzbarkeit von elektronischen Patientendaten für die Erstellung von  
retrospektiven Beobachtungsstudien**

Lehrstuhl für Medizinische Informatik

Der Medizinischen Fakultät  
der Friedrich-Alexander-Universität  
Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr. med.  
vorgelegt von  
Marcel von Lucadou

Als Dissertation genehmigt  
von der Medizinischen Fakultät  
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 13. April 2021

Vorsitzender des Promotionsorgans: Prof. Dr. Markus F. Neurath

Gutachter: Prof. Dr. Hans-Ulrich Prokosch

Gutachter: Prof. Dr. Rainer Fietkau

## **Inhaltsverzeichnis**

<b>I.</b>	<b>Zusammenfassung</b>	<b>1-2</b>
	Titel	
	Hintergrund und Ziel	
	Methoden	
	Ergebnisse und Beobachtungen	
	Schlussfolgerungen	
<b>II.</b>	<b>Einordnung in den wissenschaftlichen Kontext</b>	<b>3-11</b>
	Literatur	
<b>III.</b>	<b>Publikation</b>	<b>12</b>
	Titel	
	Autoren	
	Fachzeitschrift	
	Erscheinungsjahr	
	DOI	
	Anhang	
<b>IV.</b>	<b>Danksagungen</b>	<b>13</b>

# **I. Zusammenfassung**

## **Titel**

Nutzbarkeit von elektronischen Patientendaten für die Erstellung von retrospektiven Beobachtungsstudien

## **Hintergrund und Ziel**

Infolge der zunehmenden Digitalisierung im Gesundheitswesen besteht großes Interesse an Sekundärnutzungen elektronischer Patientendaten wie etwa für retrospektive Beobachtungsstudien. Die Eignung der im Rahmen der Patientenversorgung erhobenen Daten für retrospektive Beobachtungsstudien bleibt ungeklärt. In dieser Arbeit haben wir deshalb den allgemeinen Datenbedarf von retrospektiven Beobachtungsstudien und die Durchführbarkeit von drei Studien mit strukturierten elektronischen Patientendaten hinsichtlich Datenverfügbarkeit und -qualität untersucht.

## **Methoden**

Zur Erfassung des allgemeinen Datenbedarfs analysierten wir publizierte retrospektive Beobachtungsstudien der letzten 10 Jahre aus drei Fachabteilungen des Universitätsklinikums Erlangen auf verwendete Datenquellen bei der Studiererstellung. Anschließend wählten wir zufällig eine Studie aus jeder Fachabteilung der vorangegangenen Datenbedarfsanalyse aus, extrahierten Ein- und Ausschlusskriterien, untersuchten Expositionen und klinische Endpunkte sowie andere relevante Patientenmerkmale. Dann prüften wir die Verfügbarkeit und Eignung der korrespondierenden strukturierten elektronischen Patientendaten für die Durchführung der Studien. Die Datenqualität der verfügbaren elektronischen Daten wurde anhand vier Datenqualitätskriterien bewertet:

- (1) Vollständigkeit: Sind die strukturierten Daten für die Studie vollständig dokumentiert?
- (2) Granularität: Ist der Detailgrad der strukturierten Daten für das Patientenmerkmal ausreichend?
- (3) Korrektheit: Sind die strukturierten Daten richtig?

#### (4) Zeitliche Plausibilität: Ist die zeitliche Abfolge der strukturierten Daten plausibel?

Um die Vollständigkeit abzuschätzen, berechneten wir Häufigkeitsverteilungen der Patientenmerkmale und verglichen diese mit denen der publizierten Studie. Bei unzureichender Granularität wurde versucht, aus dem Zusammenschluss verschiedener verfügbarer strukturierter Daten indirekt auf das gesuchte Patientenmerkmal zu schließen, diese Vorgehensweise wird häufig im Englischen als „Phenotyping“ bezeichnet. Die Richtigkeit der strukturierten Daten wurde manuell durch den inhaltlichen Vergleich mit klinischen Freitexten bewertet. Durch Betrachtung der begleitenden Zeitstempel der strukturierten Daten wurde evaluiert, inwieweit hieraus ein plausibler klinischer Verlauf abgeleitet werden kann.

#### **Ergebnisse und Beobachtungen**

Die Datenbedarfsanalyse in den ausgewählten Veröffentlichungen zeigte, dass neben der klinischen Routinedokumentation als primäre Datenquelle retrospektiver Beobachtungsstudien auch zusätzliche Patientendaten über Fragebögen, Interviews und Nachuntersuchungen nachträglich eingeholt werden.

Insgesamt wurden 847 Patienten in der Durchführbarkeitsanalyse der drei Studien eingeschlossen (Studie 1 n= 411, Studie 2 n=423, Studie 3 n=13). Für zwei Studien waren alle Einschlusskriterien in strukturierten Daten verfügbar. In einer Studie wurde versucht die gesuchte Studienpopulation anhand von „Phenotypes“ zu identifizieren. Sämtliche Patientenmerkmale konnten in einer Studie entweder in strukturierten oder unstrukturierten Daten wiedergefunden werden. Für zwei Studien waren alle Expositionen und klinische Endpunkte in strukturierten Daten verfügbar. Zum Zeitpunkt unserer Untersuchung lagen die strukturierten Daten hauptsächlich in Form von Kodierungen der International Classification of Diseases (ICD) und der Operationen- und Prozedurenschlüssel (OPS) aus Abrechnungsdaten vor. Die Analyse der Datenqualität zeigte, dass einige Patientenmerkmale eine wünschenswerte Granularität aufgrund der verwendeten Diagnose- und Prozedurenschlüssel nicht oder nur eingeschränkt erreichen. Es ergaben sich Schwierigkeiten, äquivalente Diagnose- oder Prozedurenschlüssel zu finden. Auch die Interpretation relevanter Kausalbeziehungen wie zum Beispiel pathophysiologischer Zusammenhänge gestaltete sich schwierig. Bei einigen kodierten

Diagnosen war nicht erkennbar, ob es sich um aktuell festgestellte oder alte Diagnosen handelte. Bei Betrachtung der Zeitstempel konnten wir eine plausible klinische Abfolge in kodierten Therapiemaßnahmen feststellen, bei Diagnosen hingegen nicht. Ein Vergleich der berechneten Patientenmerkmalsverteilungen mit denen der Publikationen zeigte für unterschiedliche Patientenmerkmale sowohl ähnliche als auch abweichende Häufigkeiten, die auf eine unvollständige Dokumentation hinweisen könnten. Das Datenqualitätskriterium Korrektheit wurde in einigen Datensätzen dadurch verletzt, dass Verdachtsdiagnosen nicht als solche gekennzeichnet waren und dass manche kodierte Diagnosen das eigentliche klinische Problem nicht richtig wiedergegeben haben.

### **Schlussfolgerungen**

Die ausgewerteten Beobachtungsstudien in der Datenbedarfsanalyse belegen studienspezifische Informationsbedarfe, welche nur durch erneuten Patientenkontakt gedeckt werden können. Die zur Sekundärnutzung verfügbaren strukturierten elektronischen Patientendaten bestehen derzeit primär aus Diagnose- und Prozedurenschlüssel von Abrechnungsdaten. Die Granularität von strukturierten Daten ist maßgebend für die Wiederverwertbarkeit in retrospektiven Beobachtungsstudien und derzeit nach unserer Auffassung der ausschlaggebende limitierende Faktor. Es bleibt deswegen eine Einzelfallentscheidung, inwieweit Diagnose- und Prozedurenschlüssel für retrospektive Beobachtungsstudien verwendet werden können. Freitexte sind weiterhin wichtige Informationsquellen für retrospektive Beobachtungsstudien.

## **II. Einordnung in den fachwissenschaftlichen Kontext**

Im Sinne der evidenzbasierten Medizin soll die ärztliche Entscheidungsfindung auf der eigenen klinischen Erfahrung und der besten verfügbaren wissenschaftlichen Evidenz aus Studien beruhen [1]. Beobachtungs- und Interventionsstudien können als zwei grundlegende Studientypen unterschieden werden, die sich durch die Art der Einflussnahme charakterisieren lassen. In Beobachtungsstudien werden Patienten lediglich beobachtet, ohne dass studienspezifische Expositionen wie in Interventionsstudien experimentell zugewiesen werden. Im Allgemeinen untersuchen beide Studientypen fragliche kausale Zusammenhänge zwischen ausgewählten Expositionen und Endpunkten. Expositionen können beispielsweise Interventionen oder Umweltfaktoren sein, Endpunkte hingegen umfassen alle potentiellen Auswirkungen einer Exposition, wie etwa das Auftreten von Erkrankungen.

Durch sogenannte Störfaktoren, die mit der Exposition assoziiert sind und zugleich Auswirkung auf den Endpunkt haben, können Studienergebnisse beeinflusst werden. In der Studierstellung sollten demnach möglichst Störfaktoren methodisch berücksichtigt werden. Ziel einer Studie ist eine hohe Verallgemeinerbarkeit und Gültigkeit der Studienergebnisse, beides findet sich je nach verwendetem Studientyp unterschiedliche Gewichtung [2]. Methodisch gut gemachte Interventionsstudien erfahren in Bezug auf ihre Evidenz oft mehr Beachtung im klinischen Alltag, jedoch wird auch eine Gleichwertigkeit mit Beobachtungsstudien diskutiert [3-5].

Nach dem zeitlichen Studienverlauf kann weiter zwischen retrospektiven Studien, die vergangene Ereignisse untersuchen, und prospektive Studien, die Ereignisse in der Zukunft untersuchen, differenziert werden.

Manche Fragestellungen lassen sich aus ethischen oder praktischen Gründen nur mit bereits gesammelten Daten aus der Vergangenheit beantworten, sodass grundsätzlich die Fragestellung den Studientyp bestimmt; retrospektive Beobachtungsstudien sind hier die Studien der Wahl. Der ethische Aspekt lässt sich mit einem hypothetischen Beispiel, erschienen in der Weihnachtsausgabe des British Medical Journal im Jahr 2003, veranschaulichen: Eine Übersichtsarbeit über die Effektivität von Fallschirmen fand

erwartungsgemäß keine Interventionsstudien, die beschirmte und unbeschirmte Flugzeugausstiege hinsichtlich Mortalität und Verletzungsprävention untersuchten [6]. Dieses überspitzte Beispiel lässt sich auf viele medizinische Fragestellungen übertragen – wie würde man Medikamente auf Embryotoxizität während der Schwangerschaft prüfen? Die Beantwortung etwaiger Fragestellungen mit Interventionsstudien wäre ethisch nicht vertretbar und könnte erst rückblickend festgestellt werden, wie beispielsweise die fruchtschädigende Wirkung des Arzneimittels Thalidomid, das in den 50er Jahren Missbildungen bei Neugeborenen hervorgerufen hat [7].

Die klinische Routedokumentation wird häufig als Datenquelle retrospektiver Beobachtungsstudien verwendet, klassischerweise müssen hier Informationen aus Patientenakten händisch extrahiert und in eine strukturierte Form für die spätere Analyse gebracht werden. Der verbundene Arbeitsaufwand für die Erstellung ist groß. Grundsätzlich kann nur das untersucht werden, was auch dokumentiert wurde, andernfalls müssen die Informationen durch erneuten Patientenkontakt nachträglich eingeholt werden. Ein bekanntes historisches Beispiel für eine retrospektive Beobachtungsstudie ist die Studie des britischen Arztes Richard Doll und dem britischen Epidemiologen und Statistiker Bradford Hill über die Assoziation von Rauchen und Lungenkarzinomen [8].

Im Rahmen der zunehmenden Digitalisierung im Gesundheitswesen ist ein großer elektronischer Datenschatz entstanden, der für Forschungszwecke genutzt werden kann und in der Sekundärnutzung viele neue Forschungsmöglichkeiten eröffnet [9]. Für klinische Studien finden bereits automatische elektronische Rekrutierungen von Studienteilnehmern weitläufig und erfolgreich Anwendung [10]. Auch die Nutzung elektronischer Patientendaten für retrospektive Beobachtungsstudien ist von großem Interesse, insbesondere aufgrund der Möglichkeit große Studienpopulation untersuchen zu können [11]. Durch die bessere Zugänglichkeit elektronischer Daten werden Zeit und Kosten gespart [12, 13].

Trotz der vielversprechenden Nutzungsmöglichkeiten elektronischer Patientendaten haben sich mehrere Studien zu Datenqualitätsproblemen in der Vollständigkeit, Plausibilität oder Korrektheit der Daten geäußert [14, 15]. Ein einheitliches Rezept für

den Umgang mit diesen Datenqualitätsproblemen scheint durch die heterogenen Organisation- und Dokumentationsstrukturen in Krankenhäusern schwierig [16]. Unvermeidlich ergeben sich Datenqualitätsunterschiede durch die unterschiedlichen Datenanforderungen in der Krankenhausumgebung wie etwa bei der Abrechnung, Patientenversorgung oder Studiendatenerhebung; weiter wird die Datenqualität auch durch individuelle Dokumentationsgewohnheiten, -sorgfalt und -vorgaben beeinflusst. Zum Beispiel können bei der Abrechnungskodierung von therapeutischen und diagnostischen Maßnahmen mehrere Fehler auftreten [17], zusätzlich ist eine Beeinflussung durch finanzielle Anreize möglich [18-20]. Nach der Einführung des Diagnosis Related Group (DRG) basierenden Abrechnungssystems haben Studien tendenziell eine veränderte Kodierungsgewohnheit festgestellt [21, 22]. In Berücksichtigung der Komplexität und Heterogenität elektronischer Patientendaten kann eine unkritische Verwendung, ungeachtet des Datenursprungs, Studienergebnisse verfälschen [14, 15].

Die Erschließung und Bereitstellung elektronischer Patientendaten in einheitliche Forschungsinfrastrukturen zur Sekundärnutzung ist eine Herausforderung; am Universitätsklinikum Erlangen wurde begonnen, die Forschungsdatenbank „Informatics for Integrating Biology and the Bedside“ (i2b2) zu etablieren. Wir haben in dieser Arbeit untersucht, inwieweit der derzeitige elektronische Datenbestand eine Durchführung von retrospektiven Beobachtungsstudien zulässt. Hierzu haben wir zwei Analysen durchgeführt:

1. Datenbedarfsanalyse einer Auswahl von retrospektiven Beobachtungsstudien
2. Durchführbarkeitsanalyse von drei retrospektiven Beobachtungsstudien mit strukturierten elektronischen Patientendaten

Methodisch sind wir dabei folgenderweise vorgegangen:

Zur Bestimmung des allgemeinen Datenbedarfs bei der Erstellung von retrospektiven Beobachtungsstudien in Bezug auf verwendete Datenquellen analysierten wir auf

MEDLINE gelistete retrospektive Beobachtungsstudien von drei Fachabteilungen des Universitätsklinikums Erlangen der letzten 10 Jahre.

Anschließend prüften wir die Durchführbarkeit von drei Studien mit elektronischen Patientendaten. Jeweils eine Studie der vorangegangenen Datenbedarfsanalyse wurde zufällig aus jeder Fachabteilungen ausgewählt [23-25] und alle studienrelevanten Patientenmerkmale (Ein- und Ausschlusskriterien, Expositionen und klinische Endpunkte, andere Patientenmerkmale) extrahiert. Für die Identifizierung der jeweiligen Studienpopulationen mussten die extrahierten Ein- und Ausschlusskriterien an die Struktur der Forschungsdatenbank angepasst werden, welche zum Zeitpunkt unserer Untersuchung hauptsächlich in Form von Kodierungen der International Classification of Diseases (ICD) und der Operationen- und Prozedurenschlüssel (OPS) aus Abrechnungsdaten vorlagen. Durch den unterschiedlichen Detailgrad einiger Diagnose- und Prozedurenschlüssel lässt sich in manchen Fällen ein klinisches Problem nicht eindeutig wiedergeben [26]; zusätzlich gibt es mehrere Kodierungsmöglichkeiten für ein klinisches Problem [27]. Aus dieser Komplexität heraus würde eine Vorabfestlegung auf einzelne bestimmte Kodierungen bei unklaren Kodierungskonstellationen dazu führen, dass eigentlich studienzulässige Patienten fälschlicherweise ausgeschlossen werden. In unklaren Fällen haben wir deswegen zunächst übergeordnete Kodierungen aus der hierarchischen Struktur der ICD- und OPS-Kodierungen gewählt, die mit großer Wahrscheinlichkeit das gesuchte Patientenmerkmal enthalten. In der späteren händischen Analyse wurde innerhalb der gewählten übergeordneten Kategorie nach einer spezifischeren bzw. „besser passenden“ Kodierung gesucht. Konnte hiermit keine entsprechende Kodierung gefunden werden, wurde versucht durch sog. „Phenotypes“ indirekt auf das Auswahlkriterium zu schließen. „Phenotypes“ sind aus einem Zusammenschluss verschiedener Daten (z.B. Labor- und Medikationsdaten) mit logischen Ausdrücken generierte Merkmale [28]. Ein einfaches Beispiel für ein „Phenotype“ wäre, wenn man das Merkmal „Diabetes mellitus“ jedem Patienten mit einem HbA1c größer 6.5% zuschreibt.

Die händische Analyse der elektronischen Patientendaten erfolgte über Häufigkeitstabellen aller verfügbaren Patientenmerkmale und einer individuellen

zeitlichen Darstellung des diagnostischen und therapeutischen Verlaufs. Die detaillierte Vorgehensweise der Studienpopulationsauswahl kann im Anhang der Publikation nachvollzogen werden.

Nach Auswahl der Studienpopulation wurde die Verfügbarkeit aller anderen Patientenmerkmale geprüft. Ergaben sich keine äquivalenten Kodierungen, wurden auch hier nach Möglichkeit „Phenotypes“ generiert, andernfalls wurde das Patientenmerkmal als nicht verfügbar gewertet. Die Datenqualität wurde anhand vier Datenqualitätskriterien bewertet:

- (1) Vollständigkeit: Sind die strukturierten Daten für die Studie vollständig dokumentiert?
- (2) Granularität: Ist der Detailgrad der strukturierten Daten für das Patientenmerkmal ausreichend?
- (3) Korrektheit: Sind die strukturierten Daten richtig?
- (4) Zeitliche Plausibilität: Ist die zeitliche Abfolge der strukturierten Daten plausibel?

Um die Vollständigkeit der strukturierten Daten in unserer Forschungsdatenbank abzuschätzen, berechneten wir im Falle zureichender Granularität bzw. der Anwendungsmöglichkeit von „Phenotypes“ eine Häufigkeitsverteilung der Patientenmerkmale und verglichen diese mit der publizierten Studie. Eine große Abweichung interpretieren wir als eine mögliche unvollständige Dokumentation oder Variation in unserer Studienpopulation.

Der Vergleich der Häufigkeitsverteilungen wird ermöglicht durch die Beschränkung der Studienausswahl auf Erlangen und der letzten 10 Jahre, da wir damit eine ähnlich exponierte Studienpopulation mit ähnlich aktueller Routinedokumentation untersuchen [29].

Die Korrektheit der strukturierten Daten wurde manuell durch den inhaltlichen Vergleich mit klinischen Freitexten bewertet. Durch Betrachtung der begleitenden Zeitstempel der

strukturierten Daten wurde evaluiert, inwieweit hieraus ein plausibler klinischer Verlauf abgeleitet werden kann.

## Literatur

1. Sackett, D.L., et al., *Evidence based medicine: what it is and what it isn't*. BMJ, 1996. **312**(7023): p. 71-2.
2. Grimes, D.A. and K.F. Schulz, *An overview of clinical research: the lay of the land*. The lancet, 2002. **359**(9300): p. 57-61.
3. Frieden, T.R., *Evidence for Health Decision Making — Beyond Randomized, Controlled Trials*. New England Journal of Medicine, 2017. **377**(5): p. 465-475.
4. Vandembroucke, J.P., *When are observational studies as credible as randomised trials?* The Lancet. **363**(9422): p. 1728-1731.
5. Concato, J., N. Shah, and R.I. Horwitz, *Randomized, controlled trials, observational studies, and the hierarchy of research designs*. N Engl J Med, 2000. **342**(25): p. 1887-92.
6. Smith, G.C. and J.P. Pell, *Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials*. Bmj, 2003. **327**(7429): p. 1459-61.
7. Thomann, K.-D., *Die Contergan-Katastrophe: Die trügerische Sicherheit der "harten" Daten*. Dtsch Arztebl International, 2007. **104**(41): p. 2778-.
8. Doll, R. and A.B. Hill, *Smoking and carcinoma of the lung; preliminary report*. Br Med J, 1950. **2**(4682): p. 739-48.
9. Prokosch, H.U. and T. Ganslandt, *Perspectives for medical informatics. Reusing the electronic medical record for clinical research*. Methods Inf Med, 2009. **48**(1): p. 38-44.
10. Köpcke, F. and H.U. Prokosch, *Employing Computers for the Recruitment into Clinical Trials: A Comprehensive Systematic Review*. J Med Internet Res, 2014. **16**(7).
11. Nair, S., D. Hsu, and L.A. Celi, *Challenges and Opportunities in Secondary Analyses of Electronic Health Record Data*, in *Secondary Analysis of Electronic Health Records*. 2016, Springer International Publishing: Cham. p. 17-26.

12. Frankovich , J., C.A. Longhurst , and S.M. Sutherland *Evidence-Based Medicine in the EMR Era*. New England Journal of Medicine, 2011. **365**(19): p. 1758-1759.
13. Xu, H., et al., *Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality*. Journal of the American Medical Informatics Association, 2014.
14. Botsis, T., et al., *Secondary Use of EHR: Data Quality Issues and Informatics Opportunities*. Summit on Translat Bioinforma, 2010. **2010**: p. 1-5.
15. Hersh, W.R., et al., *Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research*. Med Care, 2013. **51**(8 0 3): p. S30-7.
16. Hripcsak, G. and D.J. Albers, *Next-generation phenotyping of electronic health records*. J Am Med Inform Assoc, 2013. **20**(1): p. 117-21.
17. O'Malley, K.J., et al., *Measuring diagnoses: ICD code accuracy*. Health Serv Res, 2005. **40**(5 Pt 2): p. 1620-39.
18. Simborg , D.W., *DRG Creep*. New England Journal of Medicine, 1981. **304**(26): p. 1602-1604.
19. Steinwald, B. and L.A. Dummit, *Hospital case-mix change: sicker patients or DRG creep?* Health Aff (Millwood), 1989. **8**(2): p. 35-47.
20. Serden, L., R. Lindqvist, and M. Rosen, *Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data?* Health Policy, 2003. **65**(2): p. 101-7.
21. Derby, C.A., et al., *Possible effect of DRGs on the classification of stroke: implications for epidemiological surveillance*. Stroke, 2001. **32**(7): p. 1487-91.
22. Assaf, A.R., et al., *Possible influence of the prospective payment system on the assignment of discharge diagnoses for coronary heart disease*. N Engl J Med, 1993. **329**(13): p. 931-5.
23. Brunner, T.B., R. Sauer, and R. Fietkau, *Gemcitabine/cisplatin versus 5-fluorouracil/mitomycin C chemoradiotherapy in locally advanced pancreatic cancer: a retrospective analysis of 93 patients*. Radiat Oncol, 2011. **6**: p. 88.
24. Hagel, A.F., et al., *Perforation during esophageal dilatation: a 10-year experience*. J Gastrointestin Liver Dis, 2013. **22**(4): p. 385-9.

25. Loos, B., V. Puschkin, and R.E. Horch, *50 years experience with Dupuytren's contracture in the Erlangen University Hospital--a retrospective analysis of 2919 operated hands from 1956 to 2006*. BMC Musculoskelet Disord, 2007. **8**: p. 60.
26. Gologorsky, Y., et al., *Improving discharge data fidelity for use in large administrative databases*. Neurosurg Focus, 2014. **36**(6): p. E2.
27. Dixon, J., et al., *Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals*. J Public Health Med, 1998. **20**(1): p. 63-9.
28. *Electronic Health Records-Based Phenotyping*. Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials. 2014 [cited 2019; Available from: <https://rethinkingclinicaltrials.org/resources/ehr-phenotyping/>].
29. Madigan, D., et al., *Evaluating the Impact of Database Heterogeneity on Observational Study Results*. Am J Epidemiol, 2013. **178**(4): p. 645-51.

### **III. Publikation**

#### **Titel**

Feasibility analysis of conducting observational studies with the electronic health record

#### **Autoren**

Marcel von Lucadou, Thomas Ganslandt, Hans-Ulrich Prokosch, Dennis Toddenroth

#### **Fachzeitschrift**

BMC Medical Informatics and Decision Making

#### **Erscheinungsjahr**

2019

#### **DOI**

10.1186/s12911-019-0939-0

#### **Anhang**

1. Screenshot: frequency tables, patient files, pp. 2–3. Overview of temporal data requirements for DPS/DG/DRO studies, pp. 4–6. Mapping of eligibility criteria for DPS/DG/DRO studies, pp. 7–9. Overview of available data sources, p. 10. Overview of available data in accordance with eligibility criteria applied, for DPS/DG/DRO studies, pp. 11–13. Fluctuating numbers of diagnoses per encounter, p. 14. Coding inconsistencies in chronic diseases and lifestyle factors, p. 15. Overview of eligible radiochemotherapy protocols, p. 16. Overview of assigned ICD:C25 throughout the encounters per patient ID, p. 17.

## **IV. Danksagungen**

Ganz besonderen Dank gilt Herrn Dr. Toddenroth, der mich über die Jahre mit außerordentlichem Engagement und Ausdauer betreut hat. Die zahlreichen kritischen Anmerkungen haben die Arbeit inhaltlich reifen lassen. Über den regen Austausch konnte ich das wissenschaftliche Arbeiten erlernen. Für die lange, enge und unkomplizierte Zusammenarbeit möchte mich ganz herzlich bedanken.

Meinem Doktorvater Herrn Professor Prokosch, Lehrstuhl für medizinische Informatik an der Universität Erlangen-Nürnberg, danke ich für die Überlassung des Dissertationsthemas, Betreuung und Unterstützung.

Herrn Professor Ganslandt, Abteilung für Biomedizinische Informatik am Heinrich-Lanz-Zentrum, Universitätsmedizin Mannheim, danke ich für die Bereitstellung der Patientendaten, Betreuung und Unterstützung während der Dissertation.

Außerdem möchte ich mich bei allen Mitarbeitern der medizinischen Informatik, die bei der dieser Arbeit mitgewirkt haben, bedanken.

Anschließend bedanke ich mich bei meiner Familie, die mich während des Studiums und dieser Arbeit immer unterstützt hat.