

Ergebnisbericht

(gemäß Nr. 14.1 ANBest-IF)

Konsortialführung:	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS GmbH
Förderkennzeichen:	01VSF16020
Akronym:	PV-Monitor
Projekttitel:	Nutzung von Routinedaten zur Pharmakovigilanz in Deutschland: Methodenentwicklung und erste Anwendungen
AutorInnen:	Ronja Foraita, Roland Linder, Markus Schwaninger, Iris Pigeot
Förderzeitraum:	1. Januar 2017 – 30. Juni 2020

Tabellenverzeichnis.....	2
Abbildungsverzeichnis.....	2
Abkürzungsverzeichnis.....	3
1. Zusammenfassung	5
2. Beteiligte Projektpartner	6
3. Einleitung.....	6
4. Projektdurchführung	8
5. Methodik.....	10
6. Projektergebnisse und Schlussfolgerungen	19
7. Beitrag für die Weiterentwicklung der GKV-Versorgung und Fortführung nach Ende der Förderung.....	26
8. Erfolgte bzw. geplante Veröffentlichungen.....	27
9. Anlagen	28

Tabellenverzeichnis

Tabelle 1: Vierfeldertafel – Anzahlen der Versicherten mit einer bestimmten Kombination aus Ereignis (ja/nein) und Arzneimittel (ja/nein).	17
Tabelle 2: Dosierung von DOAKs in mg.	19
Tabelle 3: DOAK-Neuerschreibungen (Apixaban, Dabigatran, Edoxaban, Rivaroxaban) zwischen 2015-2016 gruppiert nach ICD-10-GM Aggregationsniveau (Verarbeitungskette „Methoden-Mix“).	21
Tabelle 4: Heatmap-Vergleich der gefundenen Signale mit bekannten UAWs für Rivaroxaban (Verarbeitungskette „Methoden-Mix“). Das stärkste positive Signal (Rang 1) könnte eine potenzielle UAW sein, hohe Ränge sprechen für eine mögliche Indikation (insgesamt 1.142 Ränge).	22
Tabelle 5: Häufigkeiten der Exposition am Indextag.	23
Tabelle 6: Unadjustierte und adjustierte Odds Ratios mit 95%-Konfidenzintervall (CI).	24
Tabelle 7: Erzielte Fehlklassifikationsraten im Benchmarktest.	1
Tabelle 8: Vorhersagewahrscheinlichkeit für die UAW nach 12 Monaten nach Erstverschreibung.	3
Tabelle 9: Sensitivität, Spezifität und False Discovery Rate (FDR) der Vorhersage.	20
Tabelle 10: Untersuchte Ereignisse und Signale entsprechend dem Aggregationsniveau der ICDs (Verarbeitungskette „Deep Learning“).	22
Tabelle 11: Ergebnisse der Signaldetektion, die auf Sepsis als mögliche UAW hinweisen (Verarbeitungskette „Methoden-Mix“). Im Studienzeitraum sind zu sieben Sepsis Codes (A207, A227, A267, A392, A393, A394, A427) keine Einträge vorhanden.	24
Tabelle 12: Unadjustierte und adjustierte Odds Ratios mit 95%-Konfidenzintervall (CI).	26

Abbildungsverzeichnis

Abbildung 1: Projektablauf mit Verantwortlichkeiten. Institutionen in fettgedruckter Schrift hatten für das Erreichen des Ergebnisses eine Schlüsselfunktion. Schwarze Kästen stellen Schlüsselergebnisse im Projektverlauf dar.	7
Abbildung 2. Schematische Darstellung des Deep Learning Ansatzes mit Verarbeitungskette.	13
Abbildung 3: Vereinfachtes Self-Controlled Case-Series (SCCS) Studiendesign.	15
Abbildung 4: Expositions- und Falldefinition. DOAK Nutzer 1 nimmt das Medikament zeitgleich zur UAW und wird zum aktuell exponierten („current use“) Fall. DOAK Nutzer 2 wird zum Fall nach Absetzen des DOAKs („recent use“, wenn die Zeitspanne zwischen UAW und Vorratsende weniger als ein halbes Jahr beträgt, ansonsten bezeichnet als „past use“). DOAK Nutzer 3 erleidet die UAW nicht und dient als potenzielle Kontrolle am Indextag für DOAK Nutzer 1 und 2.	19
Abbildung 5: Index of Prediction Accuracy als Ergebnis der temporalen Validierung.	4
Abbildung 6: Heatmap-Analyse der Signalstärken nach statistischer Methode basierend auf Kendall'schen Rangkorrelationskoeffizienten.	23

Abkürzungsverzeichnis

AOK	Allgemeine Ortskrankenkasse
ARTP	Adaptive Runc Truncated Product
ATC	Anatomisch-therapeutisch-chemische Klassifikationssystem
BCPNN	Bayesian Confidence Propagation Neural Network
BfArM	Bundesinstituts für Arzneimittel und Medizinprodukte
BIPS	Leibniz-Institut für Präventionsforschung und Epidemiologie
BVA	Bundesversicherungsamt
CHA2DS2-VASC	Congestive heart failure, hypertension, age 75 years or older, diabetes mellitus, previous stroke or transient ischemic attack, vascular disease, age 65 to 74 years, female
DAK	Deutsche Angestellten-Krankenkasse
DDD	Defined Daily Dose
DL	Deep Learning
DOAK	Direkter oraler Gerinnungshemmer
DRG	Diag-nosis Related Groups
EBM	Einheitlicher Bewertungsmaßstab
EMA	European Medicines Agency
FDA	Food and Drug Administration
FDR	False Discovery Rate
FP	Falsch-positives Signal
FT	Functional Targets
GB	Gastrointestinaler Blutung
GePaRD	German Pharmacoepidemiological Research Database
GKV	Gesetzliche Krankenversicherung
GPGPU	General Purpose Computation on Graphics Processing Unit
HAS-BLED	Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (>65 years), Drugs/alcohol concomitantly
HDIA	Hauptentlassungsdiagnosen
HKK	Handelskrankenkasse
HMG	Hauptmorbiditygruppen
HPC	High-performance cluster
ICA	BCPNN, Information Component Alter-native
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
ICD-10-GM	ICD, 10. Revision, German modification
ICO	BCPNN, Information Component Original
IPT	Institut für Experimentelle und Klinische Pharmakologie und Toxikologie der Universität zu Lübeck
KNN	Künstliches Neuronales Netz
LASSO	Least Absolute Shrinkage and Selection Operator
LGPS	Longitudinal Gamma-Poisson Shrinker
LR	Logistische Regression
NREP	Number of Reports
OPS	Operationen- und Prozedurenschlüssel
POIS	Test of the Poisson Mean
PRR	Propor-tional Relative Risk
RF	Random Forest
RF_CORR	Random Forest Corrected Impurity Importance
RF_IMPU	Random Forest Impurity Importance

Akronym: PV-Monitor
Förderkennzeichen: 01VSF16020

RNN	<i>Recurrent Neural Network</i>
RSF	<i>Random Survival Forests</i>
SCCS	<i>Self-controlled Case-Series</i>
SD	<i>Standardabweichung</i>
SMR	<i>Spontanmelderegistern</i>
SSA	<i>Switched Sensitivity Analysis</i>
TA	<i>True association</i>
TK	<i>Techniker Krankenkasse</i>
TP	<i>Richtig-positives Signal</i>
TTD	<i>Therapeutic Targets Database</i>
UAW	<i>Unerwünschte Arzneimittelwirkung</i>
WINEG	<i>Wissenschaftlichem Institut der TK für Nutzen und Effizienz im Gesundheitswesen</i>

1. Zusammenfassung

Unerwünschte Arzneimittelwirkungen (UAWs) zählen zu den häufigen Todesursachen. Immer wieder kommt es zu Marktrücknahmen bei Arzneimitteln, da schwerwiegende Nebenwirkungen vor der Zulassung nicht erkannt wurden. Um diese Risiken aufzudecken, überwacht die sogenannte Pharmakovigilanz systematisch die Arzneimittelsicherheit nach der Zulassung. Medizinisches Personal und Betroffene melden unerwünschte Arzneimittelwirkungen an ein Melderegister. Diese sogenannten Spontanmeldedaten sind jedoch nur begrenzt geeignet, um potenzielle Sicherheitsrisiken zu identifizieren. Über den tatsächlichen Gesundheitszustand der betreffenden erkrankten Person ist zudem wenig bekannt.

In diesem Projekt wurden neue Methoden zur Anwendung in der Pharmakovigilanz erforscht, die auf Abrechnungsdaten von gesetzlichen Krankenversicherungen (GKV) zurückgreifen. Aus den Abrechnungsdaten lassen sich wesentlich mehr Informationen gewinnen als aus Spontanmelderegistern, so z.B. zur Art der abgegebenen Arzneimittel und zum Auftreten von Diagnosen. Zudem lässt sich die Gesamtzahl der Exponierten bestimmen. Dementsprechend begründen sich die Ziele in der Methodenentwicklung u.a. auf die Ausnutzung von Analogien zur genetischen Epidemiologie sowie auf Methoden der künstlichen Intelligenz und der Erschließung dieser Methoden für die Pharmakovigilanz, und zwar (1) zur Vermeidung falsch-positiver Signale für Arzneimittelrisiken, (2) zur Erkennung seltener Risiken und (3) zur Ermittlung von Risikoprofilen von Patienten und Patientinnen. Im Rahmen einer Machbarkeitsstudie wurden zwei von uns entwickelte Verarbeitungsketten zur Anwendung auf GKV-Daten eingesetzt, um potenzielle UAWs direkter oraler Gerinnungshemmer (DOAKs) zu identifizieren, die von Patienten und Patientinnen mit u.a. Vorhofflimmern eingenommen werden. Zur Validierung wurden sieben Signale in einer anschließenden pharmakoepidemiologischen Studie basierend auf der pharmakoepidemiologischen Forschungsdatenbank (*GePaRD*) genauer analysiert. Unter Verwendung eines Active-Comparator Designs wurde das Risiko des Events unter DOAK Exposition mit dem unter Exposition durch die Einnahme von Vitamin-K-Antagonisten Phenprocoumon verglichen. Ein Signal führte zu einer interessanten Neuentdeckung. Bevor diese Entdeckung allerdings publiziert werden kann, ist der Einbezug von weiterem Expertenwissen notwendig, um eine unnötige Verunsicherung der Patienten und Patientinnen zu vermeiden.

Die auf Abrechnungsdaten basierenden Verarbeitungsketten stellen eine sinnvolle Ergänzung zu dem bestehenden Spontanmelderegister dar und können zur Verbesserung der Patientensicherheit beitragen. Insbesondere die hohe Anzahl von GKV-Versicherten sowie die longitudinale Erfassung von Verschreibungen und Erkrankungen ermöglichen es zeitnahe Analysen zu möglichen seltenen, schweren und späten Risiken von neuen Arzneimitteln nach deren Zulassung durchzuführen.

2. Beteiligte Projektpartner

Institut	Name, E-Mail	Verantwortlichkeiten
Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS	Prof. Dr. rer. nat. Iris Pigeot-Kübler Tel: 0421 / 218-56942 <i>pigeot@leibniz-bips.de</i>	Projektleitung/Konsortialführung
Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS	Dr. rer. nat. Ronja Foraita <i>foraita@leibniz-bips.de</i>	Methodenentwicklung, Biometrie Signalgenerierung mit Methoden-Mix.
Techniker Krankenkasse	Prof. Dr. med. Roland Linder <i>prof.dr.roland.linder@tk.de</i>	Projektverantwortlicher der TK / Signalgenerierung mit Deep Learning
Institut für Experimentelle und Klinische Pharmakologie und Toxikologie der Universität zu Lübeck	Prof. Dr. med. Markus Schwaninger <i>markus.schwaninger@uni-luebeck.de</i>	Projektverantwortlicher Pharmakologe

3. Einleitung

Ausgangslage des Projekts

Unerwünschte Arzneimittelwirkungen (UAW) verursachen europaweit jährlich ca. 200.000 Todesfälle und sind für 5% der Krankenhauseinweisungen verantwortlich (European Medicines Agency, 2014). Ziel dieses methodischen Projekts war es, durch die Entwicklung innovativer Methoden basierend auf Abrechnungsdaten gesetzlicher Krankenversicherungen (GKV) zur Verbesserung der Arzneimittelsicherheit nach Marktzulassung beizutragen. Üblicherweise wird zur Monitorierung der Arzneimittelsicherheit eine systematische Überwachung von Spontanmelderegistern (SMR, Sammlung von Meldungen über UAW) durchgeführt, um die zum Zeitpunkt der Zulassung noch unbekannt Sicherheitsrisiken („Signale“) aufzudecken. Jedoch haben SMR viele Limitationen (Goldman 1998): U.a. ist die Anzahl derjenigen, die das Medikament eingenommen haben (Exponierte), nicht bekannt, es können Duplikate auftreten und die Datenqualität leidet unter einer möglichen Unter- oder auch Übererfassung (Suling & Pigeot 2012). In diesem Projekt wurde daher der Nutzen von Abrechnungsdaten der GKV als Ergänzung zu SMR untersucht.

Ziele und Fragestellungen/Hypothesen des Projekts

Der Forschungsschwerpunkt dieses Projekts lag auf der methodischen Weiterentwicklung von innovativen statistischen Verfahren zur Identifizierung von Arzneimittelrisiken und deren Überprüfung im Rahmen von Simulationsstudien und einer Machbarkeitsstudie.

Das Ziel der Methodenentwicklung war es, unter Ausnutzung von Analogien zur genetischen Epidemiologie sowie von Ansätzen der künstlichen Intelligenz entsprechende Methoden für die Pharmakovigilanz zu erarbeiten, und zwar speziell (1) zur Vermeidung falsch-positiver Signale für Arzneimittelrisiken, (2) zur Erkennung seltener Risiken und (3) zur Ermittlung von Risikoprofilen von Patienten und Patientinnen. Die Eignung der entwickelten Verfahren wurde anhand von Simulationsstudien, Benchmarkdatensätzen

und anhand von GKV-Daten für ein konkretes Anwendungsbeispiel, und zwar zu direkten oralen Antikoagulanzen (DOAK) untersucht.

Es wurden zudem zwei unterschiedliche Verarbeitungsketten aus automatisierten Analysen und einer interdisziplinären Expertise-Diskussionen entwickelt, die zukünftig leicht auf neue Arzneimittelgruppen angepasst werden können.

Die langfristigen Ziele bestanden in der Etablierung eines interdisziplinären Forschungsverbunds und der Bereitstellung von Methoden zur nachhaltigen Nutzung von Routinedaten zur Pharmakovigilanz in Deutschland (als Ergänzung zum Spontanmeldesystem des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM)).

Projektstruktur (Projektaufbau, Strukturen, Verantwortlichkeiten, ggf. Besonderheiten in der Projektstruktur).

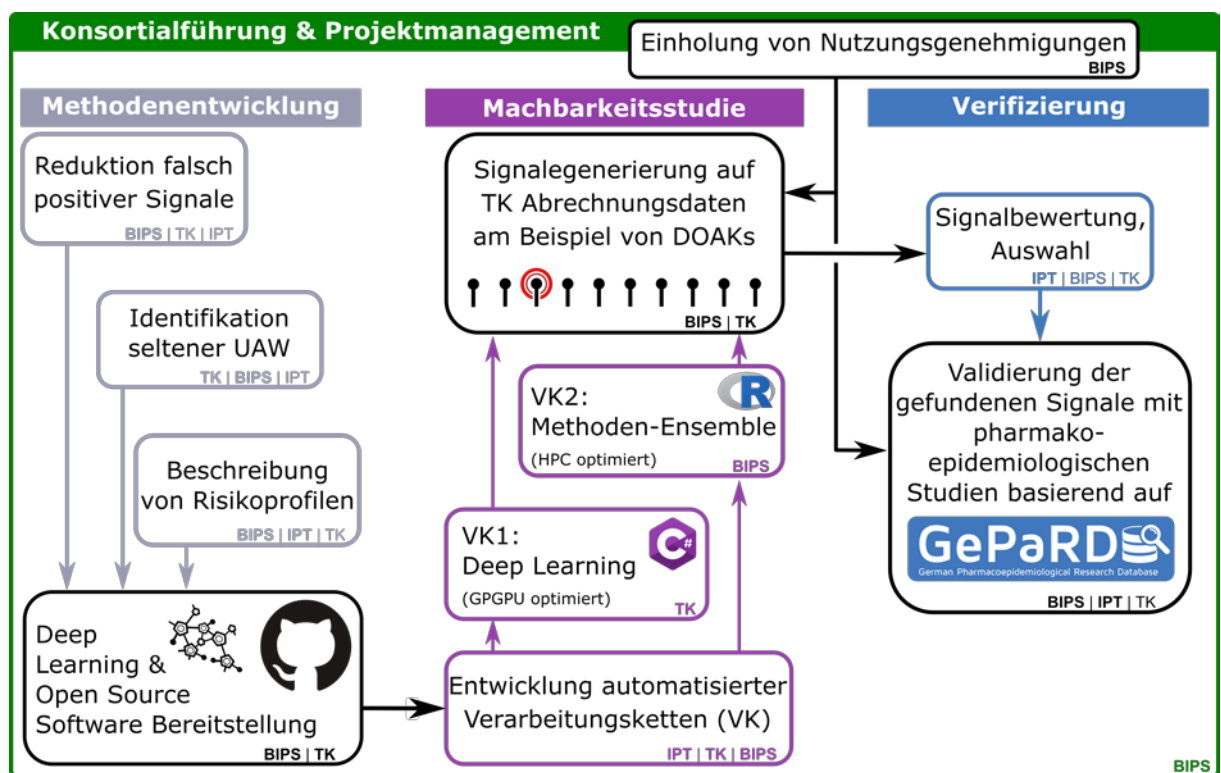


Abbildung 1: Projektablauf mit Verantwortlichkeiten. Institutionen in fettgedruckter Schrift hatten für das Erreichen des Ergebnisses eine Schlüsselfunktion. Schwarze Kästen stellen Schlüsselergbnisse im Projektverlauf dar.

GePaRD: German Pharmacoepidemiological Research Database ; GPGPU: General Purpose Computation on Graphics Processing Unit; HPC: High-performance Cluster; BIPS, IPT, TK: die beteiligten Projektpartner (siehe S. 6).

Zusammenarbeit mit den Konsortialpartnern sowie den ggf. beteiligten Kooperationspartnern

Die Zusammenarbeit zwischen BIPS, TK und IPT war konstruktiv, lösungsorientiert und durch kritische Diskussionen geprägt. Die Ergebnisse wurden stets zwischen allen Projektpartnern geteilt und alle Meinungen, Ideen und Zweifel wurden ernst genommen. Der Austausch wurde durch eine gemeinsame Plattform erleichtert. Regelmäßige Telefonkonferenzen und persönliche Treffen erlaubten eine zielgerichtete und direkte Kommunikation zwischen allen Partnern.

Im Projektverlauf traten allerdings zwei unvorhersehbare Probleme auf, die zwar einen entscheidenden Einfluss auf die Projektabwicklung hatten, aber trotzdem keine Anpassung des angestrebten Projektziels zur Folge hatten. So wurde zum einen zum 31. Mai 2018 das WINEG aufgelöst, was eine Umstrukturierung der zu übernehmenden Aufgaben zwischen den Partnern zur Folge hatte. Zum anderen behinderte die unterschiedliche IT-Ausstattung am BIPS und an der TK die Erstellung gemeinsamer Software. Ursprünglich war geplant, dass die bei der TK und am BIPS entwickelten Algorithmen zwischen den Institutionen ausgetauscht werden. Schnell zeigte sich jedoch, dass die Implementierung der im BIPS verwendeten R Packages dem Sicherheitskonzept der TK entgegensteht: Den Angestellten in der TK ist es nicht gestattet und durch die Firewall-Einstellungen auch technisch unmöglich, selbständig Softwarepakete aus dem Internet herunterzuladen und ungeprüft zu nutzen. Umgekehrt verfügte das BIPS über keinen leistungsstarken GPGPU-Rechner mit Windows Server als Betriebssystem, um den von der TK in C#.NET geschriebenen Deep Learning Ansatz ausführen zu können. Hierzu wurde eine intelligente Lösung gefunden, die weiter unten beschrieben ist (siehe Kapitel 5).

Im Januar 2019 fand ein Arbeitstreffen der Projektgruppe mit der Forschungsgruppe „Arzneimittelallergien, Nebenwirkungen an der Haut und spezielle Fragen der Arzneimittelsicherheit“ des BfArM unter der Leitung von Herrn Prof. Dr. Sachs statt. In der Folge entstand insbesondere zwischen dem IPT und dem BfArM eine engere Kooperation. Dies führte zu einer Auswertung von Spontanberichten zu Nebenwirkungen von Dimethylfumarat, aus der ein Publikationsmanuskript entstand, das 2020 beim British Journal of Clinical Pharmacology eingereicht wurde (Dubrall et al. 2020).

4. Projektdurchführung

Der Projektablauf sah vor, dass in den ersten zwei Jahren die statistischen Methoden entwickelt werden, mit denen eine Signalgenerierung auf GKV-Routinedaten durchgeführt werden soll. Die Leistungsfähigkeit der statistischen Methoden sollten mittels Simulationsstudien untersucht werden. Im Projektverlauf wurde erkannt, dass eine Simulationsstudie nicht bei allen Fragestellungen zum Ziel führte. So sollte die entwickelten maschinellen Lernverfahren hinsichtlich ihrer Fähigkeit, unbekannte seltene Ereignisse zu finden, anhand einer Simulationsstudie miteinander verglichen werden. Hierzu wurde ein Datensimulator entwickelt, der die Abhängigkeitsstruktur longitudinaler GKV-Daten generiert. Um die hohe Komplexität von GKV-Daten zu erfassen, musste für jedes Arzneimittel und für jede potenzielle UAW eine Zeitreihe simuliert werden. Um dann zu untersuchen, wie die Arzneimittelexposition das Risiko einer UAW beeinflusst, musste eine Vielzahl an Parametern variiert werden. Dabei ist die Wahl der entsprechenden Parameterwerte nicht trivial, da häufig keine oder nur geringe Kenntnisse darüber vorliegen, welche Werte in der Praxis auftreten. Für einen fairen Methodenvergleich hätten wir demnach eine zu große Anzahl verschiedener Parameterkonstellationen untersuchen müssen. Zudem bedarf die Untersuchung von seltenen Ereignissen sehr großer Datensätze, die sowohl bei der Generierung als auch bei der Analyse viel Rechenzeit benötigen. Obwohl das BIPS über einen HP-Cluster (HPC) verfügt, war dies aufgrund der beschränkten Rechnerkapazität nicht möglich. Wir haben uns daher entschlossen, alternativ die Performanz der Methoden hinsichtlich ihrer Vorhersagekraft zu bewerten. Der Deep Learning Ansatz wurde statt mit simulierten Daten mit Hilfe von existierenden Benchmarkdatensätzen evaluiert.

Im zweiten Teil des Projekts sollte in einer Machbarkeitsstudie untersucht werden, ob sich GKV-Routinedaten für die Identifizierung von UAWs eignen und damit die Voraussetzung

gegeben ist, das etablierte Spontanmeldesystem von BfArM/EMA auf diesem Weg sinnvoll zu ergänzen. Es wurde erwartet, dass sich zum einen die in der Arzneimittelinformation beschriebene und damit bekannten UAWs in den Analysen wiederfinden lassen und zum anderen idealerweise bisher nicht beschriebene UAWs entdeckt werden können. Gefundene Signale von bisher unbekanntem UAWs sollten mit der pharmakoepidemiologischen Forschungsdatenbank (German Pharmacoepidemiological Research Database, abgekürzt GePaRD) am BIPS reproduziert und vom IPT plausibilisiert werden. Die pharmakoepidemiologische Forschungsdatenbank GePaRD enthält Abrechnungsdaten von vier gesetzlichen Krankenversicherungen und umfasst Informationen von derzeit ungefähr 25 Millionen Personen, die seit 2004 bei einer der teilnehmenden Krankenkassen versichert waren. GePaRD erlaubt u.a. die Identifizierung seltener Risiken von Arzneimitteln über längere Zeiträume. Der Zugriff auf die Datenbank ist nur für Beschäftigte des BIPS im Rahmen von behördlich genehmigten Forschungsprojekten möglich. Für PV-Monitor war es daher notwendig, ein spezifisches Datenschutzkonzept auszuarbeiten sowie bei den an GePaRD beteiligten Krankenkassen und den jeweils zuständigen Behörden einen Genehmigungsantrag auf Überlassung dieser Sozialdaten zu Forschungszwecken zu stellen. Aufgrund der Revision von § 75 Sozialgesetzbuch (SGB) X im Jahr 2018 hat das damalige Bundesversicherungsamt (BVA; jetzt Bundesamt für Soziale Sicherung) sein Antragsverfahren zur Nutzung der Krankenversicherungsdaten geändert. Obwohl dies zu Verzögerungen bei der Bearbeitung der Nutzungsanträge führte, hat unser Projekt die erforderlichen Genehmigungen vom BVA für die Datennutzung der Krankenkassen TK, HKK und DAK als auch von der damaligen Senatorin für Wissenschaft, Gesundheit und Verbraucherschutz für die AOK Bremen/Bremerhaven rechtzeitig erhalten.

Für die Signalgenerierung wurden zwei Verarbeitungsketten im Projekt entwickelt: Der Deep Learning Ansatz wurde für GPGPU-Rechner mit Windows Server als Betriebssystem optimiert. Der Ansatz des „Methoden-Mix“ basiert auf der Sprache R und ist HPC optimiert. Ein Vergleich beider Verarbeitungsketten konnte dennoch realisiert werden, weil GePaRD u.a. mit TK-Routinedaten gespeist wird. So konnten die entwickelten Methoden auf einer gemeinsamen Schnittmenge der jeweiligen Datenbanken laufen, die beiden Einrichtungen zur Verfügung stand. Anders als das BIPS ist die TK an kürzere Löschfristen gebunden, im Fall der ambulanten Daten waren das zum Zeitpunkt der Untersuchung vier Jahre plus das laufende Jahr (§304 SGB V). Berücksichtigt man den 6-9 monatigen Lieferverzug dieser Daten an die TK, den zeitlichen Verzug aufgrund der nur jährlichen Datenübermittlung von der TK an die Vertrauensstelle von GePaRD sowie die Zeit für Qualitätssicherung und Aufbereitung innerhalb der Vertrauensstelle, überlappten sich die Abrechnungsdaten für die Jahre 2015 und 2016 und standen so de facto für einen Methodenvergleich zur Verfügung.

Nach der automatisierten Signalgenerierung haben die Kooperationspartner gemeinsam interessante Signale ausgewählt. Diese Signale wurden anschließend in spezifischen pharmakoepidemiologischen Studien auf Basis von GePaRD überprüft. Unter Verwendung eines Active-Comparator Designs wurde das Risiko des Events unter DOAK-Exposition mit dem unter Exposition durch den Vitamin-K-Antagonisten Phenprocoumon verglichen. Ein Signal führte zu einer interessanten Neuentdeckung. Bevor diese Entdeckung allerdings publiziert werden kann, ist der Einbezug von weiterem Expertenwissen notwendig. Wenn sich dieser Kandidat bestätigen sollte, wird dies dem BfArM mitgeteilt, das eine erneute genaue Prüfung seiner Spontanmeldedaten zu dem jeweiligen Arzneimittel zugesagt hat.

5. Methodik

Im Rahmen dieses Projekts wurden statistische Methoden aus der genetischen Epidemiologie und der künstlichen Intelligenz adaptiert und genutzt, um Arzneimittelrisiken anhand von Routinedaten zu identifizieren (Signalgenerierungsansätze). Zudem sollte ihr Anwendungspotential für die Pharmakovigilanz untersucht werden, und zwar hinsichtlich (1) der Vermeidung falsch-positiver Signale für Arzneimittelrisiken, (2) der Erkennung seltener Risiken mit maschinellen Lernverfahren und (3) zur Ermittlung von Risikoprofilen von Patienten und Patientinnen.

A: Statistische Methodenentwicklung

(1) Vermeidung falsch-positiver Signale für Arzneimittelrisiken

Das erste Ziel war es, statistische Methoden zu identifizieren, mit denen im Rahmen einer Signalerkennung die Anzahl falsch-positiver Signale reduziert werden kann. Dafür wurden insbesondere Methoden zusammengetragen, die für die Analyse von Spontanmeldedaten entwickelt worden sind. Die statistischen Methoden für Spontanmeldedaten wurden aus zweierlei Gründen intensiver untersucht:

- 1) Longitudinale Abrechnungsdaten der GKV können jederzeit in das Format von Spontanmeldedaten gebracht werden; eine Transformation in die andere Richtung ist jedoch nicht möglich. Dieses Vorgehen ist aktuell in der Pharmakovigilanzforschung auch üblich.
- 2) Die statistischen Methoden sollten gegen den Gamma-Poisson Shrinker verglichen werden, der für die Analyse von Spontanmeldedaten entwickelt wurde.

Es wurde eine umfassende Literaturrecherche zum Stichwort „Pharmakovigilanz“ durchgeführt und darauf aufbauend nach dem Prinzip „Citation pearl growing“ (punktuelle Suche) gesucht. Insgesamt wurden über 200 Artikel gelesen. Alle dort erwähnten Methoden bzw. Maße wurden weiter verfolgt. Basierend auf der Literaturrecherche war es jedoch unmöglich zu erkennen, welche Methoden für welche Szenarien zu bevorzugen sind und wie stark sich diese Methoden voneinander unterscheiden. Um die Signaldetektion zu optimieren, ist es jedoch wichtig zu verstehen, welche Methoden für die verschiedenen Situationen am besten geeignet sind. Zudem wurde in den bisher veröffentlichten Vergleichsstudien die eher unrealistische Annahme getroffen, dass weder Ereignisse noch Arzneimitteleinnahme untereinander korreliert sind. Es ist aber zu erwarten, dass diese Form von Confounding (in der Pharmakovigilanz auch bekannt als Innocent Bystander Effect) die Performanz der Methoden beeinflusst. Letztendlich wurden 27 statistische Methoden ausgewählt und hinsichtlich ihrer Fähigkeit miteinander verglichen, wahre Signale trotz Confounding in den Daten zu entdecken. Darunter war z.B. auch der Lasso (least absolute shrinkage and selection operator), der bisher noch nicht in der Pharmakovigilanz eingesetzt wird, aber in der genetischen Epidemiologie eine Standardmethode darstellt. Dies war das einzige Verfahren, das im Gegensatz zu den anderen Methoden, alle potenziellen UAWs multivariat analysieren konnte. Eine Auflistung der 27 Methoden ist in Dijkstra et al. (2020) zu finden (siehe Anlage 8).

Die Projektgruppe hat nach ausführlicher Beratung die alleinige Fokussierung auf falsch-positive Signale (FP) als nicht ausreichend empfunden. Für die Bewertung von Signalen wurde eine stärkere Objektivierung gefordert und daher ein Vorgehen festgelegt, das typischerweise im Bereich der genetischen Epidemiologie eingesetzt wird. Dort müssen in analoger Weise aus einer Vielzahl möglicher paarweiser Assoziationen (häufig mehr als 1

Millionen) die „wahren“ Assoziationen ausgewählt werden. In beiden Anwendungsbereichen besteht ein extremes Ungleichgewicht zwischen der Anzahl an tatsächlich nicht vorhandenen Assoziationen im Vergleich zu allen wahren Assoziationen (TA). Wir haben daher das Beurteilungskriterium – angelehnt an die Vorgehensweise in der genetischen Epidemiologie – angepasst und bewertet die Performanz der statistischen Methoden sowohl hinsichtlich ihrer Genauigkeit (Precision) als auch ihrer Trefferquote (Recall, Sensitivität). Diese Beurteilungskriterien bevorzugen die Methoden, die die meisten wahren Signale (TP) sowohl aus der Menge aller wahren Assoziationen (TA) als auch aus der Menge aller gefundenen Signale (TP + FP) entdecken. Das Klassifikationskriterium der Genauigkeit ist darüber hinaus das Gegenstück zur False Discovery Rate (FDR), d.h. Genauigkeit = 1 – FDR. Die FDR wird in der genetischen Epidemiologie i.d.R. angewendet, um für das vorliegende multiple Testproblem zu kontrollieren.

(2) Erkennung seltener Risiken mit maschinellen Lernverfahren

Das primäre Ziel dieser Task war es, seltene UAWs auf Basis von Routinedaten vorherzusagen. Dafür sollten maschinelle Lernverfahren eingesetzt werden, da diese durch ihre flexiblen und automatisierten Ansätze geeignet sind, Muster in Daten zu entdecken.

An der TK und am BIPS wurden aufgrund der unterschiedlichen Rechnerinfrastrukturen zwei unterschiedliche maschinelle Lernverfahren verfolgt, um nach seltenen UAWs zu suchen:

- 1) Deep Learning mit künstlichen neuronalen Netzen
- 2) Random Forests (siehe Anlage 2)

Im Projekt wurden einerseits Deep Learning Verfahren (Künstliches Neuronales Netz, KNN) verwendet, weil diese bei den meisten Benchmarkvergleichen zurzeit mit am besten abschneiden und bei guter Parameteroptimierung sehr gute Erfolge erzielen. Möglich wurden diese Erfolge zum einen durch hardwareseitige Fortschritte, zum anderen durch neue Wege, eine Überbestimmtheit (overfitting) beim Trainieren neuronaler Netze zu vermeiden. Gegenüber klassisch-statistischen Ansätzen wie Regressionsverfahren können KNN Korrelationen auch innerhalb der unabhängigen Variablen berücksichtigen. Medizinische Daten zeigen sehr häufig derartige Interdependenzen, nicht selten auch nicht-linearer Natur. Aufgrund der nicht-linearen Aktivierungsfunktion und der hierarchischen Repräsentation können alle diese Zusammenhänge hervorragend im KNN abgebildet werden. Random Forests sind in der genetischen Epidemiologie äußerst beliebt, da sie robust sind und aufgrund der intrinsischen Variablenselektion sehr gute Ergebnisse bei hochdimensionalen Daten liefern, selbst wenn die voreingestellten Parameterkonstellationen nicht optimal sind. Darüber hinaus sind Random Forests durch das Konzept der „Variable Importance“ i.d.R. interpretierbarer als andere maschinelle Lernverfahren. Andere maschinelle Lernverfahren wie z.B. Support Vector Machines wurden diskutiert, aber aus unterschiedlichen Gründen nicht weiter verfolgt (z.B. aufgrund ihrer schlechteren Performanz mit sehr komplexen Daten umzugehen oder der ungenügenden Interpretierbarkeit hinsichtlich welche Variablen für das Ergebnis besonders wichtig waren.)

Zu 1)

Aus Vorarbeiten war bekannt, dass die Güte des Verfahrens ganz maßgeblich von der Berücksichtigung des medizinischen Kontexts abhängt, insbesondere der Vielzahl an Begleiterkrankungen und Ko-Medikationen. Da gleichzeitig jedoch zahlreiche (auch nicht-

lineare) gegenseitige Abhängigkeiten (Korrelationen) zwischen den unabhängigen Variablen bestehen, haben wir uns für einen Deep Learning entschieden. Die besondere Herausforderung bestand darin, dass eine manuelle Optimierung der Parametrisierung des KNN angesichts der Größe der Matrix aus zu untersuchenden UAWs und Arzneimitteln (ca. 3.800 UAWs x 12 Antikoagulanzen = ca. 45.600 mögliche Zusammenhänge) nicht möglich war. Deswegen musste ein autarkes und damit extrem robustes KNN gewählt werden. Um dies zu gewährleisten, wurde mit Kreuzvalidierung und Ensemble-Strategien gearbeitet, die beide äußerst rechenintensiv sind. Dies wiederum erforderte die Entwicklung eines hochperformanten und stark parallelisierten Codes sowie eine rechenstarke Infrastruktur. Aus Eigenmitteln hat die TK dafür einen Server (Kosten: ca. 50.000 €) mit mehr als 10.000 Rechenkernen (GPGPUs) bereitgestellt.

Ausgehend von den Routinedaten der TK (10,2 Mio. Versicherte) wurden KNN darauf trainiert, potenziell unerwünschte Arzneimittelwirkungen vorherzusagen. Dies erfolgte mit dem Wissen um potenziell relevante Aspekte in der Vorgeschichte (v.a. ICD-, ATC-, OPS- und EBM-Codes) sowie der Information, ob eine Krankenhausaufnahme unter dem pharmakologischen Einfluss des untersuchten Arzneimittels stattfand. Schlussendlich war jedoch nicht die Klassifikationsgüte des so trainierten KNN von Interesse, sondern die Analyse der Stärke des Zusammenhangs zwischen untersuchtem Arzneimittel und untersuchter UAW. Dazu wurde im Rahmen einer sich anschließenden Sensitivitätsanalyse untersucht, inwieweit sich das Klassifikationsergebnis verschlechtert, wenn inputseitig auf die Information zum untersuchten Arzneimittel verzichtet wurde. Diese Information wurde in Form von Bedeutsamkeitswerten (Importance Values) ausgegeben. Dabei gilt, je höher der Importance Value ist, desto größer ist die Wahrscheinlichkeit für einen Zusammenhang zwischen untersuchtem Arzneimittel und untersuchter UAW.

Es wurde eine 6-schrittige Verarbeitungskette entwickelt, die den Datenabzug vom Data Warehouse der TK bis hin zur Generierung der Analyseergebnisse automatisiert (siehe Abbildung 2). Am Beispiel der DOAKs waren die Schritte wie folgt:

Schritt 1: Detektion möglicher Assoziationen von UAWs mit den untersuchten Arzneistoffen (hier: zwölf unterschiedliche orale Antikoagulanzen). Eingeschlossen wurden nur diejenigen UAWs, deren Prävalenz im Zusammenhang mit der Arzneimittelgabe gestiegen ist, die also durch die Arzneimittelgabe potenziell getriggert wurden.

Schritt 2: Datenbankabzug der UAWs plus Antikoagulanzen einschließlich vielfältiger Diagnose- und Behandlungsinformationen aus dem Vorjahr (Indexjahr) vor Eintritt der untersuchten potenziellen UAWs, so u.a. ICD-, OPS-, EBM- und ATC-Codes (potenzielle Confounder). Im Anschluss erfolgte ein Matching mit alters- und geschlechtsgleichen Versicherten ohne UAW sowie eine Alters- und Geschlechtsadjustierung an die bundesrepublikanische Bevölkerung des Jahres 2015.

Schritt 3: Ausschluss aller UAW-Arzneistoff Kombinationen, die seltener als 10-mal in den Daten enthalten waren.

Schritt 4: Variablenselektion mittels Random Forests. Neben UAWs und Arzneistoffen wurden nur die 100 Variablen mit den ranghöchsten Importance Values übernommen. Hierfür wurde die C++ Version von Ranger (<https://github.com/imbs-hl/ranger>) verwendet.

Schritt 5: Klassifikation mit Hilfe eines Deep Learning Algorithmus mit anschließender Sensitivitätsanalyse zur Signaldetektion, siehe Zwischenbericht 2017.

Schritt 6: Aufbereitung von mehreren Tausend Ergebnisdateien in Form einer Excel-Mappe.

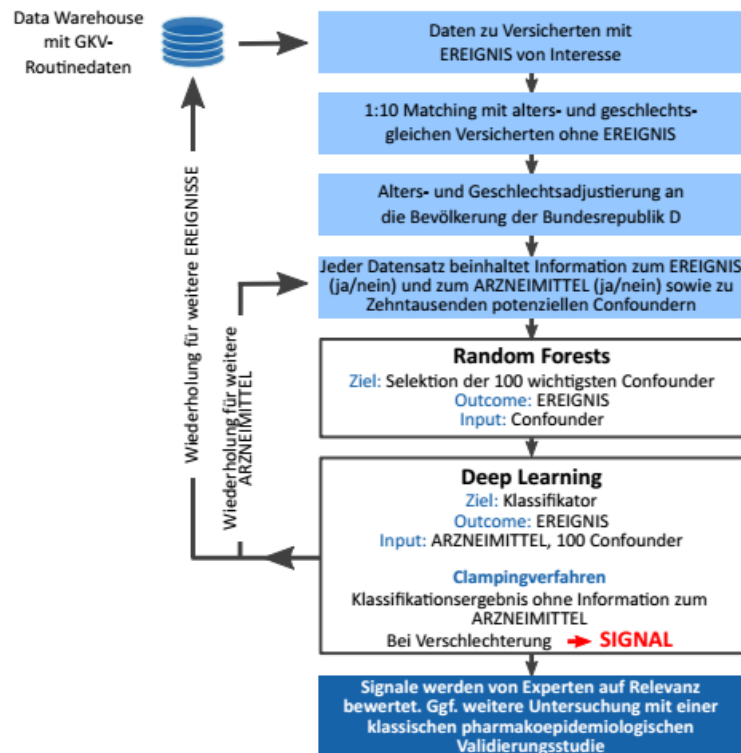


Abbildung 2. Schematische Darstellung des Deep Learning Ansatzes mit Verarbeitungskette.

(3) Ermittlung von Risikoprofilen von Patienten und Patientinnen

Ziel war es, mit Hilfe von Vorwissen Risikoprofile für das Auftreten bestimmter Ereignisse zu erstellen. Dafür wurden Arzneimittel und Krankheiten auf höheren Ebenen gruppiert. Als höhere Ebenen haben wir „Functional Targets“ (FT) gewählt. Ein FT ist ein Pfad von interagierenden Biomolekülen, wie z.B. metabolisierende Enzyme oder Rezeptoren, deren Interaktion durch einen Wirkstoff beeinflusst wird und sich dadurch auch auf die Entstehung von im selben FT befindlichen Krankheiten auswirkt. Agieren Arzneimittel in mehr als in einer FT, kann dies zu UAWs führen, die in diesen FTs verankert sind.

Nach einer ausführlichen Recherche hat sich die Projektgruppe für die Online-Datenbanken Kyoto Encyclopedia of Genes and Genomes (KEGG, Kanehisa et al. 2017), Search Tool for Interacting Chemicals (STITCH, Szklarczyk 2016) und die Therapeutic Targets Database (TTD, Yang 2016) entschieden. TTD ist unserer Ansicht nach aktuell die beste kostenlose Datenbank, die Arzneimittel und Diagnosen im ICD-Format in einem FT verbinden kann. KEGG ist hier selektiver, weil die Datenbank hauptsächlich Informationen über Krankheiten beinhaltet, die einen genetischen Einfluss haben. STITCH bietet einen Network-Korrelationsscore, den wir zusätzlich als einen topologischen Ensemble-Score für die Analyse verwenden wollten. Wir mussten aber feststellen, dass der Einbezug von Krankheiten hier nicht möglich war. Dies ist aber entscheidend, um Nebenwirkungsprofile zu erstellen, so dass STITCH nicht weiter in Betracht gezogen wurde. In der von uns entwickelten Software *rgp* wurde ein Schnittstelle zu TTD implementiert, mit der Arzneimittel und Diagnosen in FTs gruppiert und mit GKV-Daten verknüpft.

Die von uns durchgeführte, methodische Weiterentwicklung basiert auf dem „Adaptive Runc Truncated Product“ (ARTP) von Yu et al. (2009). Es ist eine der populärsten Ansätze in der genetischen Epidemiologie, wenn für eine Enrichment-Analyse keine kontinuierlich

gemessenen Genexpressionsdaten, sondern kategoriale SNP-Daten zur Verfügung stehen. Wir haben hier eine ähnliche Datengrundlage, da in den Versicherungsdaten i.d.R. nur binäre Variablen vorkommen. Um mit dem ARTP ein individuelles Risiko vorhersagen zu können, wurde das Verfahren zu einem Ensemble-Verfahren erweitert. Dieses Verfahren sollte mit statistischen Methoden verglichen werden, die ebenso in der Lage sind, eine übergeordnete Gruppenstruktur zu modellieren. Dabei zeigte sich zum einen, dass viele der publizierten Verfahren nicht als Software zur Verfügung standen und zum anderen konnten viele Verfahren keine hochdimensionalen Daten verarbeiten. Dies war aber notwendig, da circa 8,000 Kovariablen gleichzeitig in einem einzigen Modell analysiert werden mussten. Aus diesem Grund haben wir uns entschieden, hauptsächlich regularisierte Regressionsmethoden zu verwenden, die sich im Wesentlichen darin unterscheiden, wie Gruppen von Variablen die Parameterschätzung beeinflussen können. Darüber hinaus wurde das Verhalten eines neuen Verfahrens aus dem Bereich des maschinellen Lernens untersucht, der sogenannte Block Forest (Hornung, Wright 2019), der erstmalig eine vorher definierte Gruppenstruktur berücksichtigen kann.

Auch hier wurden zunächst die statistischen Methoden auf Basis von simulierten GKV-Daten miteinander verglichen. Da die simulierten Daten zusätzlich die komplexe FT-Struktur widerspiegeln mussten, bedurfte die Simulation noch zusätzlicher Parametervoreinstellungen, die aufgrund des fehlenden Vorwissens schwer festzulegen waren. Wir haben uns stattdessen dazu entschieden, die Methoden basierend auf einer Vorhersagestudie zu vergleichen.

Es wurde eine eingebettete Fall-Kontroll-Studie mit gastrointestinaler Blutung (GB) als Endpunkt aufgesetzt und untersucht, ob die Vorhersage einer GB auf Basis von FTs besser ist als auf Basis der einzelnen Variablen. Für diese Studie konnten Versicherte der TK vom 1.1.2015 bis zum 31.12.2016 eingeschlossen werden, sofern diese über 18 Jahre alt, kontinuierlich versichert waren, in Deutschland lebten und frühestens 90 Tage nach Studienbeginn eine GB hatten. Zudem wurden nur Versicherte berücksichtigt, die ein Jahr vor Studieneintritt keine GB hatten. Versicherte mit GB (Fälle) wurden am Tag des Ereignisses im Verhältnis 1:4 hinsichtlich Alter und Geschlecht mit Kontrollen gematcht. Alle Fälle mit geraden Versicherungsnummern und deren Kontrollen wurden für das Training der Modelle verwendet, alle anderen für die Vorhersage. Als statistische Methoden wurden das ARTP, der Block Forest, ein naiver Lasso, der standardmäßige Lasso sowie Random Forests hinsichtlich diverser Klassifikationsmaße miteinander verglichen. Die ersten vier Methoden haben die Fähigkeit, eine Gruppenstruktur zu berücksichtigen, während die letzten zwei dies nicht erlauben. Der naive Lasso ist hier ein gewöhnlicher Lasso, der als Kovariaten für jeden FT einen Summenscore aus all seinen zugehörigen Kovariaten bildet. Dies reduzierte die Dimension der Kovariaten von ca. 8.500 Kovariaten auf 260 FTs.

B: Machbarkeitsstudie / Signaldetektion

Einer Pilotstudie auf Basis der TK-GKV-Daten sollte untersucht werden, inwieweit GKV-Daten zur Untersuchung von UAWs genutzt werden können (Klose et al. 2019). Um sich weiter der Signalgenerierung anzunähern, wurden DOAKs untersucht, deren Verordnungszahlen in den letzten Jahren stark angestiegen sind und die mittlerweile von einem relevanten Anteil der etwa 1,7 Millionen Patienten und Patientinnen mit Vorhofflimmern in Deutschland eingenommen werden. Neben der vergleichsweise hohen Prävalenz sowie unter Wirtschaftlichkeitsgesichtspunkten (etwa 15-fach höhere Behandlungskosten gegenüber der bewährten Therapie mit Vitamin K-Antagonisten) war für die Wahl des

Untersuchungsgegenstands ausschlaggebend, dass sowohl ein Vergleich mit Versicherten mit der untersuchten UAW aber ohne Einnahme von DOAK möglich war als auch mit Patienten und Patientinnen, die alternativ mit Vitamin K-Antagonisten therapiert wurden.

TK und BIPS haben parallel zwei Ansätze verfolgt, die sich trotz ungleicher Rechnerstruktur miteinander vergleichen ließen: Da GePaRD u.a. mit TK-Routinedaten gespeist wird, konnten die Algorithmen auf einer gemeinsamen Schnittmenge an Daten laufen, die beiden Einrichtungen zur Verfügung standen. Wie oben bereits beschrieben, waren dies aufgrund der unterschiedlichen technischen und gesetzlichen Rahmenbedingungen die Jahre 2015 und 2016 mit ca. 10,2 Mio. Versicherten.

Verarbeitungskette 1 (Deep Learning)

Auf Basis der TK-Routinedaten wurden Deep Learning Algorithmen trainiert, um mögliche UAWs vorherzusagen. Dies erfolgte mit dem Wissen um potenziell relevante Aspekte in der Vorgeschichte (v.a. ICD-, ATC-, OPS- und EBM-Codes) sowie der Information, ob eine Krankenhausaufnahme unter dem pharmakologischen Einfluss des untersuchten Arzneimittels stattfand. Das Vorgehen basiert auf den sechs Schritten, die auf Seite 11 vorgestellt wurden. Eine detaillierte Beschreibung des Vorgehens ist in Anlage 3 zu finden.

Verarbeitungskette 2 (Methoden-Mix)

Als Grundlage für die Signaldetektion wurde ein Self-Controlled Case-Series (SCCS)-Studiendesign verwendet (siehe Abbildung 3), um das Risiko für eine UAW nach einer DOAK-Dispensation zu bewerten. Hierfür wird die Inzidenz der UAW in "Risikoperioden", d.h. nach der Dispensation, mit der Inzidenz der UAW während Kontrollperioden, d.h. ohne jegliche DOAK-Exposition, innerhalb derselben Person miteinander verglichen (Whitaker et al. 2009). Durch die Schätzung relativer Inzidenzen wird es dann möglich, die Assoziation zwischen einer Exposition und einer UAW zu bewerten. Es wurde für die Signalgenerierung ein vereinfachtes SCCS-Design angewandt, das die Inzidenz innerhalb einer einzigen Risikoperiode (ab der ersten Dispensation bis zum Studienaustritt) mit der innerhalb einer einzigen Kontrollperiode (von Studienbeginn bis zur ersten DOAK-Dispensation) vergleicht.

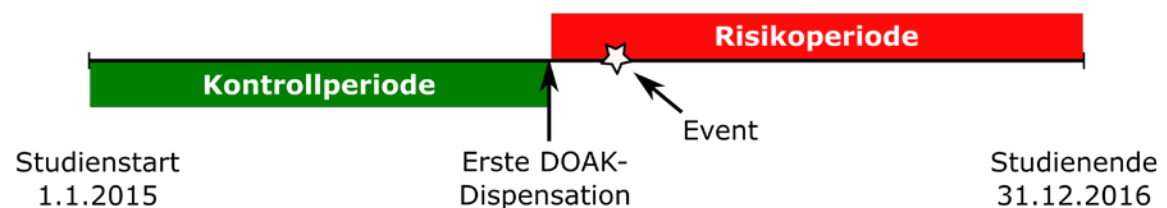


Abbildung 3: Vereinfachtes Self-Controlled Case-Series (SCCS) Studiendesign.
DOAK: direkte orale Antikoagulanzen

Als Grundlage dienen auch hier die TK-Daten der vollständigen Jahre 2015 und 2016. Einschlusskriterien für die Signaldetektion waren, dass der/die Versicherte kontinuierlich im Studienzeitraum versichert war, in Deutschland lebte und frühestens 90 Tage nach Studienbeginn und spätestens 90 Tage vor Studienende eine DOAK-Dispensation aufwies. Zudem wurden nur Versicherte berücksichtigt, die ein Jahr vor Studieneintritt keine DOAK-Dispensationen hatten. Für alle Nutzer und Nutzerinnen wurden zwei Datensätze erstellt, und zwar jeweils für die Zeit vor der Verschreibung und nach der Verschreibung.

Für diese Datensätze wurden alle Diagnosen (ICD-10) und Verschreibungen (ATC) sowie mögliche Störvariablen gespeichert. Die ICD- und ATC-Codes wurden jeweils als Indikatorvariablen gespeichert. Auf diesem Datensatz wurden ein Random Forest trainiert, die sogenannte „bias-corrected variable importance“ berechnet sowie ein p -Wert (Nullhypothese: Keine Assoziation zwischen ICD und Verschreibung) geschätzt. Alle signifikant assoziierten ICDs (5% Signifikanzniveau ohne Adjustierung für multiples Testen) wurden nach der Effektstärke sortiert. Da ein Random Forest keine Richtung der Effekte bestimmt, wurde diese aus Vierfeldertafeln geschätzt. Das Ergebnis dieser Analyse ist eine Rangliste detektierter Signale möglicher UAWs. Zudem wurde die Idee der Kollabierung der seltenen Varianten aus der genetischen Epidemiologie umgesetzt. Dort werden Regionen von Interesse definiert und die Informationen seltener Varianten in dieser Region zusammengefasst (engl.: collapse). Dies führt zu einer Erhöhung der statistischen Power. Bei unseren Analysen fassten wir Erkrankungen in funktionale Obereinheiten zusammen, um so das schwache Signal einer sehr seltenen Erkrankung zu verstärken. Bei der Random Forest Analyse wurde in den Simulationsstudien zudem untersucht, wie gut zusätzliche Gewichtungen an den Splitknoten den Gedanken der seltenen Variantenanalyse umsetzen können, um das Auffinden seltener Ereignisse zu ermöglichen. Zur Aufdeckung seltener UAWs durch Random Forests wurde als Beurteilungsmaß die sogenannte Impurity Importance verwendet. Dabei haben wir statt der Standardeinstellung die neu entwickelte und unverzerrte Impurity Importance von Nembrini et al. (2018) eingesetzt.

Es wurde ein Methoden-Mix aus den folgenden neun statistischen Verfahren für die Signalgenerierung angewandt, die mit zu den besten Methoden zur Vermeidung falsch-positiver Signale gehört haben (siehe Seite 19), und Random Forest als maschinelles Lernverfahren, das zusätzlich mögliche Confounder und Nichtlinearitäten berücksichtigen kann: Bayesian Confidence Propagation Neural Network, Information Component Alternative (BCPNN, ICA), Bayesian Confidence Propagation Neural Network, Information Component Original (BCPNN, ICO), Longitudinal Gamma-Poisson Shrinker (LGPS), Test of the Poisson Mean (POIS), Number of Reports (NREP), Proportional Relative Risk (PRR), Random Forest Impurity Importance (RF_IMPUI), Random Forest Corrected Impurity Importance (RF_CORR) und Least Absolute Shrinkage and Selection Operator (LASSO). Dabei sind BCPNN, LGPS und PRR in der Pharmakovigilanz gut etabliert und verwenden als Grundlage für jedes UAW eine Vierfeldertafel (siehe Tabelle 1). Ebenso basieren PRR, POIS und NREP auf Vierfeldertafeln. Für jede Kombination aus einem Arzneimittel und einem Ereignis wird basierend auf der entsprechenden Vierfeldertafel und der verwendeten Methode ein „Risikomaß“ berechnet, das als Indikator für die Stärke einer potenziellen UAW dient.

Für die Bewertung der Assoziation wurde bei den multivariaten Ansätzen (LASSO, RFs) anders verfahren. Beim LASSO wurde der niedrigste Maximal-Wert des Bestrafungsparameters λ verwendet, ab dem ein Ereignis das erste Mal in das Modell selektiert wurde. Für die RFs wurden die Variablenwichtigkeit sowie ein p -Wert (Nullhypothese: Keine Assoziation zwischen ICD und Verschreibung) ermittelt. Alle signifikant assoziierten ICDs (5% Signifikanzniveau ohne Adjustierung für multiples Testen) wurden nach der Effektstärke sortiert. Da ein Random Forest keine Richtung der Effekte bestimmt, wurde diese aus Vierfeldertafeln geschätzt. In den multivariaten Methoden wurde zudem für die Kovariaten Geschlecht und Gemeindekennziffer sowie für die zeitveränderlichen Kovariaten Alter und alle Diagnosen und Verschreibungen während der Kontroll- und Risikoperioden adjustiert. Siehe Anlage 4 für das detaillierte Studienprotokoll.

Table 1: Vierfeldertafel – Anzahlen der Versicherten mit einer bestimmten Kombination aus Ereignis (ja/nein) und Arzneimittel (ja/nein).

Kombination		Ereignis		Gesamt
		Ja	Nein	
DOAK Dispensation	Ja (Risikoperiode)	a	b	a + b
	Nein (Kontrollperiode)	c	d	c + d
Gesamt		a + c	b + d	a + b + c + d

DOAK: direkte orale Antikoagulanzen

C: Validierungsstudie

Die Signaldetektion hat die Signale akute Blutungen, akute Leberschädigungen, akutes Nierenversagen, Epilepsie, Glaukom, Sepsis und Zystitis identifiziert. Diese wurden in der Validierungsphase überprüft. Zusätzlich wurden die beiden bereits bekannten Nebenwirkungen der gastrointestinalen sowie intrakraniellen Blutung als Positivkontrolle analysiert. Das folgende Studiendesign wurde auf alle Endpunkte („Signale“) angewandt.

Für die Studie kamen alle Personen im GePaRD in Frage, die

- (i) zwischen dem 1. Januar 2011 und dem 31. Dezember 2017 mindestens 12 aufeinander folgende Monate versichert waren,
- (ii) mindestens eine Abgabe des Studienmedikaments (Phenprocoumon, Apixaban, Dabigatran, Edoxaban, Rivaroxaban) während des Studienzeitraums erhalten haben,
- (iii) mindestens eine ambulante oder stationäre Diagnose von Vorhofflimmern als Indikation innerhalb von 12 Monaten vor dem potenziellen Kohorteneintrittsdatum hatten (zur Reduktion des Confounding by Indication und der Indikationsabhängigkeit der Dosierung),
- (iv) keine Abgabe von mehr als einem Studienmedikament zum potenziellen Kohorteneintrittsdatum hatten,
- (v) keine Abgabe eines oralen Antikoagulans in dem Jahr vor dem potenziellen Kohorteneintritt hatten,
- (vi) keine Krebsdiagnose zu irgendeinem Zeitpunkt vor dem potenziellen Kohorteneintritt und keine Schwangerschaft zum Zeitpunkt des potenziellen Kohorteneintritts hatten.

Während die Signalgenerierung auf einzelnen ICD-Codes beruhte, wurde das Event in der Validierungsstudie aus mehreren ICD-Codes operationalisiert. Fälle von Sepsis wurden beispielsweise identifiziert, indem in den Hauptentlassungsdiagnosen der Krankenhauseinweisungen nach 24 unterschiedlichen ICD-10 Codes gesucht wurde.

Für jeden Endpunkt wurden potenzielle Confounder (Risikofaktoren, protektive Faktoren, Komorbidität, Komedikation, Lebensstil) basierend auf Diagnosen, Prozeduren und/oder Arzneimitteln in verschiedenen Zeitfenstern vor Kohorteneintritt erfasst. Zusätzlich wurden für alle Endpunkte die Risikoscores für Blutung (HAS-BLED) und Schlaganfall (CHA2DS2-VASC) berechnet.

Es wurde eine Kohorte von Neu-Nutzern oraler Antikoagulanzen aufgesetzt und eine eingebettete Fall-Kontroll-Analyse unter Verwendung eines Active-Comparator-Designs durchgeführt. Dabei wurden alle DOAKs gemeinsam als auch die einzelnen DOAKs mit dem Vitamin-K-Antagonisten Phenprocoumon verglichen. Die Fälle wurden definiert als Kohortenmitglieder mit einer inzidenten Diagnose, d.h. Fälle mit entsprechenden Diagnosen wurden vorher ausgeschlossen. Bei den Endpunkten Epilepsie und Glaukom wurden alle ausgeschlossen, die eine solche Diagnose bereits vor Kohorteneintritt aufwiesen, bei Zystitis 365 Tage vorher und bei allen anderen Endpunkten 180 Tage vorher. Für die Endpunkte Sepsis, akutes Nierenversagen und akutes Leberversagen wurden nur Krankenhausdiagnosen berücksichtigt (Indexdatum = Aufnahmetag), bei Epilepsie, Glaukom und Zystitis wurden sowohl Krankenhausdiagnosen als auch gesicherte ambulante Diagnosen (Indexdatum = Diagnosetag) berücksichtigt. Für jeden Fall wurden bis zu 10 zufällig ausgewählte Kontrollen nach Geschlecht, Alter zum Indexdatum (+/- 1 Jahr) und GKV unter Verwendung von Risk-Set-Sampling und unter Berücksichtigung der Zeit in der Kohorte ermittelt. Mögliche Kontrollen, die zum Indexdatum des Falles aus irgendeinem Grund ins Krankenhaus eingewiesen wurden, wurden aus der Gruppe der potenziellen Kontrollen ausgeschlossen. Fälle kamen für eine Auswahl als Kontrolle vor ihrem Indexdatum in Frage und Kontrollen konnten einmal für jeden Fall, aber insgesamt mehr als einmal ausgewählt werden.

Für alle Personen in den Fall-Kontroll-Sets wurde das zeitliche Verhältnis zwischen dem Vorratsende der letzten Dispensation und dem Indextag für alle interessierenden Studienmedikamente individuell nach den gültigen Verschreibungsrichtlinien ermittelt. Bei Phenprocoumon hängt die Dosierung vom Wirkgehalt der zuletzt ausgegebenen Packung geteilt durch die empfohlene Tagesdosis ab. Standardmäßig wird hier eine mittlere Tagesdosis von 3 mg angenommen. In Sensitivitätsanalysen wurden auch eine niedrige (1,5 mg) und höhere Dosis (4,5 mg) zugrunde gelegt. Bei den DOAKs hängt die Dosis insbesondere vom Alter und der Nierenfunktion ab (siehe Tabelle 2). Ein Kohortenmitglied wurde als aktuell exponiert betrachtet, wenn am Indextag die Studienmedikation seit der letzten Verordnung rechnerisch noch nicht aufgebraucht war (aktuelle Nutzung, engl. current use). Siehe Abbildung 4 für eine graphische Illustration.

Tabelle 2: Dosierung von DOAKs in mg.

	Empfohlene Dosis	Alter ≥ 80	Indikation Niereninsuffizienz	Alter ≥ 80 und Indikation Niereninsuffizienz
Apixaban	10	10	10	5
Dabigatran	300	220	220	220
Edoxaban	60	60	30	30
Rivaroxaban	20	20	15	15

Um die Häufigkeit des Auftretens des jeweiligen Signals zwischen Fällen und Kontrollen zu vergleichen, wurden neben deskriptiven Analysen auch konditionale logistische Regressionsanalysen unter Berücksichtigung potenzieller Confounder durchgeführt. Darüber hinaus wurden für einige Endpunkte (Epilepsie, Glaukom, Zystitis) Sensitivitätsanalysen ergänzt.

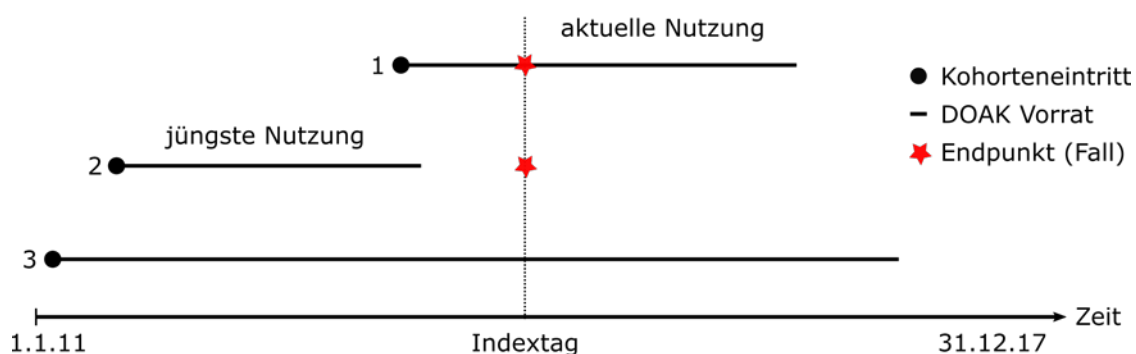


Abbildung 4: Expositions- und Falldefinition. DOAK Nutzer 1 nimmt das Medikament zeitgleich zur UAW und wird zum aktuell exponierten („current use“) Fall. DOAK Nutzer 2 wird zum Fall nach Absetzen des DOAKs („recent use“, wenn die Zeitspanne zwischen UAW und Vorratsende weniger als ein halbes Jahr beträgt, ansonsten bezeichnet als „past use“). DOAK Nutzer 3 erleidet die UAW nicht und dient als potenzielle Kontrolle am Indextag für DOAK Nutzer 1 und 2.

DOAK: direkte orale Antikoagulanzen

6. Projektergebnisse und Schlussfolgerungen

A: Statistische Methodenentwicklung

(1) Vermeidung falsch-positiver Signale für Arzneimittelrisiken

Das Projekt hat erfolgreich methodische Konzepte und statistische Methoden aus dem Gebiet der genetischen Epidemiologie auf die Pharmakovigilanz übertragen. So wurde gezeigt, dass mit Lasso die große Anzahl von Kovariaten, die in GKV-Daten zur Verfügung stehen, in die Signaldetektion integriert werden kann. Dies war auch die einzige Methode im Vergleich, die sich als robust gegen den Innocent Bystander Effect erwiesen hat (Dijkstra et al. 2020). Die einzelnen Methoden können interaktiv auf <https://srs.bips.eu> hinsichtlich ihrer Performanz miteinander verglichen werden. Hier lässt sich das Abschneiden der einzelnen Methoden im Vergleich zum Gamma-Poisson-Shrinker gut erfassen. Es zeigte sich, dass bei einer schwachen Assoziation zwischen Ereignis und

Arzneimittleinnahme ohne Innocent Bystander Effekt der Gamma-Poisson-Shrinker im oberen Mittelfeld lag, die Performanz im Vergleich zu anderen Methoden jedoch wesentlich besser wurde, wenn Innocent Bystander Effekte betrachtet wurden. Die Simulationsstudie zeigte auf, dass keine Methode in allen Szenarien am besten abschnitt, sondern die Performanz der Methoden von den Effektgrößen und der in den Daten vorhandenen Abhängigkeitsstruktur abhängt (siehe Dijkstra et al. in Anhang 8). Da dies i.d.R. unbekannt ist, wurde in Verarbeitungskette 2 ein Mix verschiedener Methoden verwendet.

(2) Erkennung seltener Risiken mit maschinellen Lernverfahren

Ferner wurden zwei verschiedene maschinelle Lernverfahren zur Identifikation seltener Signale untersucht. Dabei handelte es sich zum einen um ein Deep Learning Ansatz basierend auf künstlichen neuronalen Netzen und zum anderen um Random Forests und wiederkehrende neuronale Netze (RNN). Der Benchmarktest zeigte, dass die erzielten Klassifikationsergebnisse des Deep Learnings im Mittel um 23% besser waren als die Ergebnisse entsprechender logistischer Regressionen (siehe Anlage 1). Mittels einer Vorhersagestudie wurde zusätzlich der Nutzen von maschinellen Lernverfahren für die Prognose von UAWs untersucht. Dabei zeigte sich, dass weder Random Survival Forests noch RNNs geeignet sind, die Vorhersagekraft für seltene unerwünschte Ereignisse auf Basis von longitudinalen GKV-Daten zu verbessern (siehe Anlage 2).

(3) Ermittlung von Risikoprofilen von Patienten und Patientinnen

Für die Ermittlung von Risikoprofilen für Patienten und Patientinnen konnte hingegen gezeigt werden, dass die Vorhersage von UAWs basierend auf FTs vielversprechend ist. Es wurden fünf verschiedene statistische Verfahren miteinander verglichen. Mit intelligenten statistischen Methoden wie dem Bock Forest oder auch dem Lasso können alle Diagnosen und Verschreibungen für eine verbesserte Vorhersage verwendet werden. Dies würde erlauben, individuelle Risikovorhersagen basierend auf Krankenkassendaten zu erstellen und so vulnerable Gruppen von Patienten und Patientinnen besser schützen zu können (siehe Anlage 5).

B: Machbarkeitsstudie

Für eine Signalgenerierung auf Basis von GKV-Daten wurden zwei Verarbeitungsketten entwickelt und diese am Beispiel der DOAKs angewandt. Der verwendete Datensnapshot von GePaRD umfasste im Zeitraum 2015-2016 insgesamt 12.203.348 TK-Versicherte. Von diesen interessierten uns die Versicherten, die in diesem Zeitraum ein DOAK neu verschrieben bekamen und mindestens ein neues Ereignis hatten (siehe Tabelle 3). Die Anzahl der betrachteten ATC-Codes betrug $N_{ATC} = 1.767$.

Tabelle 3: DOAK-Neuverschreibungen (Apixaban, Dabigatran, Edoxaban, Rivaroxaban) zwischen 2015-2016 gruppiert nach ICD-10-GM Aggregationsniveau (Verarbeitungskette „Methoden-Mix“).

ICD-10 Aggregation	N _{ICD}	Apixaban	Dabigatran	Edoxaban	Rivaroxaban
3-Steller	1.142	19.468	3.659	2.558	32.787
4-Steller	3.795	19.162	3.610	2.522	32.194
5-Steller	4.815	19.181	3.611	2.526	32.217
HMG	189	15.969	3.014	1.997	25.436

HMG: Hauptmorbiditätsgruppen gemäß morbiditätsorientiertem Risikostrukturausgleich;
 ICD-10: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme

Die Ergebnisse wurden für jedes DOAK getrennt nach ICD-10-GM Aggregationsniveau und statistischer Methode ausgegeben. Um die Ergebnisse zwischen den Methoden vergleichbar zu machen, wurden die jeweiligen Schätzer auf den Wertebereich [-1, 1] normiert. Ein hoher positiver Wert spricht für ein häufiges Auftreten von einem Signal nach Verschreibung des DOAKs, ein Wert von null entspricht keiner Assoziation und negative Werte weisen darauf hin, dass das Ergebnis eine potenzielle Indikation und eher keine UAW darstellt. Zusätzlich wurden die Ereignisse in eine geordnete Rangfolge gebracht: je höher der Rang, desto stärker scheint die Assoziation zwischen dem Arzneimittel und dem Ereignis zu sein. Höhere Ränge sprechen demnach eher für ein potenzielles Sicherheitsrisiko.

Die Ergebnisse der verwendeten Verarbeitungskette Deep Learning Ansatzes stimmen im Wesentlichen mit den Signalen überein, die mit dem Methoden-Mix gefunden wurden. Die Ergebnisse der statistischen Methoden in der Verarbeitungskette „Methoden-Mix“ können in einzelnen Fällen allerdings auch sehr unterschiedlich sein. Ein Vergleich der Rangfolgen der statistischen Methoden zeigte beispielsweise, dass die ICA-Methode und das Auszählen der Reports (NREP) sehr ähnliche Signalarangfolgen erzeugten. Die Ergebnisse der multivariaten Methoden (Random Forest, LASSO) führten jedoch eher zu anderen Signalarangfolgen als die klassischen univariaten Verfahren (siehe Anlage 6). Um die Plausibilität der Ergebnisse zu untersuchen, wurden die erzeugten Signale mit bereits bekannten UAWs für das jeweilige DOAK verglichen. Die Liste wurde vom IPT aus der Fachinformation des jeweiligen Arzneimittels zusammengestellt. Tabelle 4 zeigt exemplarisch das Ergebnis für Rivaroxaban. Der von einer statistischen Methode gefundene stärkste positive Zusammenhang zwischen einer Diagnose und einem DOAK erhält den Rang 1 (grüne Zelle), währenddessen der schwächste positive Zusammenhang (dies kann auch ein negativer Zusammenhang sein) den Rang 1142 (rote Zelle) erhält. Obwohl die Ränge selber nicht direkt interpretiert werden können, ermöglichen sie den Vergleich zwischen den Methoden. Hier konnten alle statistischen Methoden die bekannten Blutungsnebenwirkungen und die Negativkontrolle (S71, Fraktur des Femurs) gut auffinden, während die Detektion der toxischen Leberkrankheit (K71) durch die meisten Methoden nicht gelang. Da die Fraktur des Femurs eine Indikation für Rivaroxaban ist, wurde ein hoher Rang erwartet.

Tabelle 4: Heatmap-Vergleich der gefundenen Signale mit bekannten UAWs für Rivaroxaban (Verarbeitungskette „Methoden-Mix“). Das stärkste positive Signal (Rang 1) könnte eine potenzielle UAW sein, hohe Ränge sprechen für eine mögliche Indikation (insgesamt 1.142 Ränge).

Methode	Rang					
	I60	I61	I62	N02	K71	S71
ICA	76	46	80	86	548	822
ICO	17	69	54	112	577	834
LGPS	109	39	93	84	545	800
POIS	22	76	51	127	458	999
NREP	112	42	94	84	423	782
PRR	8	150	76	131	410	932
RF_IMPU	95	29	79	103	358	815
RF_CORR	41	49	160	313	564	867
LASSO	84	183	128	216	372	440

Skala	1	228	457	685	914	1142
-------	---	-----	-----	-----	-----	------

Stärkster (Rang 1) bis schwächster (Rang 1142) positiver Zusammenhang

Je grüner die Zellfarbe, desto stärker ist die Übereinstimmung zwischen bekannten UAWs und dem Signal. I60 = Subarachnoidalblutung; I61 = Intrazerebrale Blutung, I62 = Sonstige nichttraumatische intrakranielle Blutung; N02 = Rezidivierende und persistierende Hämaturie, K71 = Toxische Leberkrankheit, S71 = Fraktur des Femurs (Kontrolle).

ICA: Bayesian Confidence Propagation Neural Network, Information Component Alternative; ICO: Bayesian Confidence Propagation Neural Network, Information Component Original; LASSO: Least absolute shrinkage and selection operator; LGPS: Longitudinal Gamma-Poisson Shrinker; NREP: Number of Reports; PRR: Proportional Relative Risk; POIS: Test of the Poisson Mean; RF_CORR: Random Forest Corrected Impurity Importance; RF_IMPU: Random Forest Impurity Importance

Die Ergebnisse der Signalgenerierung sind in Anlage 6 aufgeführt. Die erzeugten Listen mit Signalen (Verarbeitungskette „Deep Learning“) und Rangfolgen (Verarbeitungskette „Methoden-Mix“) wurden in Projekttreffen verglichen und mehrfach hinsichtlich ihrer biologisch-pharmakologischen Relevanz diskutiert. Die Ergebnisse der verwendeten Verarbeitungskette Deep Learning stimmten im Wesentlichen mit den Signalen überein, die mit dem „Methoden-Mix“ gefunden wurden, z.B. bei der Detektion von Sepsis (siehe unten). Die Ergebnisse der statistischen Methoden können in einzelnen Fällen allerdings auch sehr unterschiedlich sein (siehe Abbildung 6). Beide Verarbeitungsketten waren zudem in der Lage, bekannte UAWs zu identifizieren. Dieses Erkenntnis war wichtig für die interne Plausibilisierung unserer entwickelten Methoden, da diese nicht gegen einen Goldstandard abgeglichen werden konnten. Aufgrund der unterschiedlichen Ergebnisausgaben beider Verarbeitungsketten war leider auch ein direkter Vergleich nicht möglich. Dies erlaubte keine abschließende Beurteilung mit daraus resultierenden Empfehlungen für die Methodenwahl. Beide Verarbeitungsketten decken jedoch UAW-Arzneimittel Assoziationen auf und haben den Vorteil, dass sie problemlos auf andere Fragestellungen übertragen werden können. Jedoch sollten unabhängig von der Qualität des statistischen Verfahrens die gefundenen Signale stets von einem Gremium aus Experten und Expertinnen begutachtet und bewertet werden. Zwar liefern die in diesem Projekt verwendeten multivariaten statistischen Methoden, die bei der Signaldetektion auch Komedikationen und Komorbiditäten berücksichtigen können, eine eher bessere

Genauigkeit und Trefferquote als die bekannten statistischen Methoden zur Auswertung der Spontanmelderegister; ohne genaues Vorwissen über das Patientenkontext, Dosierungen oder mögliche Störvariablen bieten die Ergebnisse jedoch nur wertvolle Hinweise für spezifisch zu definierende Forschungshypothesen. Daher ist eine abschließende Überprüfung von potenziell interessanten Signalen mit Hilfe einer pharmakoepidemiologischen Studie in jedem Einzelfall zwingend notwendig.

Für die Bewertung wurden die erzeugten Listen mit Signalen und Rangfolgen potenzieller UAWs beider Verarbeitungsketten vom Projektteam gemeinsam verglichen und diskutiert. Als Grundlage dienten die 4-stelligen ICD-Codes. Kriterien für die Auswahl der Signale zur weiteren Untersuchung waren ihre biologisch-pharmakologische Relevanz als auch die Neuartigkeit des Signals. Als zusätzliche medizinische Betrachtung wurden die Ränge der ICD-Codes innerhalb derselben 3-stelligen Krankheitsklasse betrachtet (bspw. A41.- Sonstige Sepsis). Letztendlich wurden die Signale Sepsis, Glaukom, akute Blutungen, Zystitis, akute Leberschädigungen, Epilepsie und akutes Nierenversagen für die anschließende Validierung ausgewählt.

C: Validierungsstudie

Die Überprüfung der gefundenen Signale wurde mittels einer Active-Comparator Studie zu DOAKs durchgeführt. Neben den gefunden Signalen aus der Machbarkeitsstudie wurden zusätzlich gastrointestinale Blutungen und intrakranielle Blutungen als Positivkontrollen untersucht. Wie aus der Literatur zu erwarten war, zeigte eine DOAK-Therapie im Vergleich zu einer Therapie mit Vitamin-K-Antagonisten ein höheres Odds Ratio für gastrointestinale Blutungen und ein niedrigeres Odds Ratio für intrakranielle Blutungen. Von den sieben Signalen führte lediglich ein Signal zu einer potenziell interessanten Neuentdeckung (siehe Anlage 7). Bevor dieses Signal allerdings als solches publiziert werden kann, ist der Einbezug weiterer Experten oder Expertinnen notwendig. Die im Anlage 7 bereitgestellten Ergebnisse sind demnach vorläufig.

Zur Illustration wird im Folgenden das Ergebnis für den Endpunkt Sepsis aufgeführt. Tabelle 5 zeigt die absoluten und relativen Häufigkeiten für die Exposition durch ein DOAK oder durch Phenprocoumon am Indextag auf.

Tabelle 5: Häufigkeiten der Exposition am Indextag.

Sepsis		
Exposition am Indextag	Cases N = 3,504	Controls N = 35,022
<i>Nicht exponiert</i>	1.435 (41,0%)	13.189 (37,7%)
<i>Mit nur einer Studienmedikation exponiert</i>	2.036 (58,1%)	21.651 (61,8%)
<i>Phenprocoumon, moderate Dosierung</i>	806 (39,6%)	8.744 (40,4%)
<i>Dabigatran</i>	165 (8,1%)	1.818 (8,4%)
<i>Rivaroxaban</i>	670 (32,9%)	7.231 (33,4%)
<i>Apixaban</i>	358 (17,6%)	3.455 (16,0%)
<i>Edoxaban</i>	37 (1,8%)	403 (1,9%)

Mit mehreren Studienmedikationen exponiert 33 (0,9%) 182 (0,4%)

Tabelle 6 zeigt, dass Personen, die am Indextag ein DOAK oder Phenprocoumon eingenommen haben, ein vergleichbares Risiko haben, an Sepsis zu erkranken, wie Personen, die mindestens ein Jahr vorher das Studienmedikament abgesetzt haben (Referenzgruppe 1). Das Active-Comparator Design (Referenzgruppe 2) ergab ein ähnliches Bild. Es zeigte sich jedoch ein erhöhtes Odds Ratio für Versicherte, die am Indextag mehrere Studienmedikamente einnahmen, im Vergleich zu den Referenzgruppen 1 und 2.

Tabelle 6: Unadjustierte und adjustierte Odds Ratios mit 95%-Konfidenzintervall (CI).

Sepsis				
Aktuelle Einzelanwendung von	Unadjustiertes Odds Ratio¹ (95% CI)	Adjustiertes Odds Ratio¹ (95% CI)	Unadjustiertes Odds Ratio² (95% CI)	Adjustiertes Odds Ratio² (95% CI)
<i>Dabigatran</i>	0,94 (0,76; 1,16)	1,00 (0,80; 1,25)	0,99 (0,83; 1,18)	1,13 (0,94; 1,35)
<i>Rivaroxaban</i>	0,96 (0,81; 1,13)	0,98 (0,82; 1,16)	1,01 (0,91; 1,12)	1,10 (0,98; 1,23)
<i>Apixaban</i>	1,07 (0,89; 1,28)	0,98 (0,81; 1,19)	1,13 (0,99; 1,28)	1,11 (0,96; 1,27)
<i>Edoxaban</i>	0,94 (0,65; 1,37)	0,98 (0,67; 1,44)	0,99 (0,70; 1,40)	1,10 (0,77; 1,58)
<i>Phenprocoumon</i>	0,95 (0,80; 1,12)	0,88 (0,75; 1,05)	-	-
<i>Mehrere Studienmedikamente</i>	1,87 (1,25; 2,79)	1,46 (0,95; 2,23)	1,97 (1,35; 2,87)	1,64 (1,10; 2,45)

¹ Referenzgruppe: Versicherte/r war mind. 12 Monate vor dem Indextag weder mit einem DOAK noch mit Phenprocoumon exponiert.

² Referenzgruppe: Versicherte/r war am Indextag mit Phenprocoumon exponiert.

Zusätzlich zu den oben beschriebenen, im Fokus des Projekts liegenden Arbeiten wurden flankierend weitere Beiträge erarbeitet. (1) So haben wir zunächst eine Übersicht über unser Projekt und die sich daraus ergebenden methodischen Ansätze bereits sehr früh im Bundesgesundheitsblatt publiziert (Foraita et al. 2018), um in der Fachcommunity auf das Projekt aufmerksam zu machen. (2) Zudem haben Klose et al. (2020) den bekannten Zusammenhang zwischen einer Metamizol-Therapie und dem Auftreten einer Neutropenie oder Agranulozytose auf der Basis von GKV-Daten untersucht, um ein besseres Verständnis von methodischen Problemen in der Analyse und Interpretation von GKV-Daten zu erlangen. Der bekannte Zusammenhang konnte bestätigt werden. In der Studie wurden 630.285 Versicherte identifiziert, denen Metamizol verschrieben worden war. Die Metamizol-Gruppe wurde anhand eines Propensity Scores mit 390.830 Kontrollen gematcht, die kein Metamizol erhalten hatten. Die Studie zeigte ein beträchtliches Risiko von 1:1602 für die Entwicklung einer arzneimittelinduzierten Neutropenie oder Agranulozytose pro Metamizol-Verschreibung. Allerdings konnten durch das Matching auf Basis des Propensity Scores nicht alle konkurrierenden Ursachen für eine Neutropenie und Agranulozytose kontrolliert werden. Insbesondere fanden sich in der Metamizol-behandelten Gruppe auch mehr Patienten, die sich einer Tumorbehandlung mit

Chemotherapeutika unterziehen mussten. Da Chemotherapeutika eine Neutropenie oder Agranulozytose auslösen können, ist es vorstellbar, dass das Metamizol zugeschriebene Risiko teilweise durch die Begleitmedikation bedingt ist. Diese Punkte wurden in der Veröffentlichung diskutiert, um mögliche Störfaktoren bei der Analyse von Routinedaten zu illustrieren. (3) Schließlich wurden im Rahmen des Aufbaus einer Kooperation mit dem BfArM und zur Vorbereitung der Validierung von Risikosignalen anhand von Spontanmeldungen Spontanberichte von Nebenwirkungen unter Dimethylfumarat und Nikotinsäure verglichen (Dubrall et al. 2020). In der speziellen Untersuchung zeigte sich, dass Dimethylfumarat und Nikotinsäure teilweise ähnliche Nebenwirkungen verursachen. So wurden in Assoziation mit beiden Pharmaka gastrointestinale Nebenwirkungen und Angstsymptome berichtet. Andere Nebenwirkungen waren allerdings spezifisch für Dimethylfumarat oder Nikotinsäure. So wurde eine Lymphopenie häufiger unter Dimethylfumarat als unter Nikotinsäure berichtet.

Neben den zwei entwickelten Verarbeitungsketten zur Signaldetektion auf GKV-Daten war die Entwicklung von Software zentral in diesem Projekt. Die meiste Software kann auf <https://github.com/bips-hb> heruntergeladen werden und basiert auf einer GPL-3 Lizenz. Das R-Paket *pvm* bündelt beispielsweise die bis dato bekannten statistischen Methoden, die für die Signalgenerierung eingesetzt werden. Damit ist es möglich, einen Datensatz (Spontanmeldungen oder GKV-Daten) sehr leicht mit mehreren bekannten Methoden auszuwerten. Ein großer Erfolg ist die Entwicklung des Softwarepakets *srsim*, mit dem sehr große binäre Korrelationsmatrizen auf Basis einer vorgegebenen Abhängigkeitsstruktur simuliert werden können. Dies erscheint zunächst trivial; aber bis dato gab es keine Funktion in den bekannten Programmiersprachen C++, Python, R, die es ermöglichte, realistische Daten zur Untersuchung von Arzneimittelnebenwirkungen zu simulieren. Zudem erlaubt die Software durch die Vorgabe einer bestimmten Abhängigkeitsstruktur in den Daten auch explizit, den Innocent-Bystander Effekt zu modellieren. Für die Simulationsstudie wurde zudem das Softwarepaket *batchlasso* am BIPS programmiert, mit dem für Lasso eine Precision-Recall Area Under the Curve berechnet werden kann. Mit der Software *ARTPredict* kann eine Vorhersagewahrscheinlichkeit für eine UAW basierend auf FTs geschätzt werden. Da *ARTPredict* auf Permutationstests basiert, ist das Verfahren sehr zeitintensiv. Als Alternative bietet sich hier der Block Forest an, der nur ungefähr ein Fünftel der Zeit benötigt. Die Software *rgp* beinhaltet die notwendigen Funktionen, um FTs aus Onlinedatenbanken mit GKV-Daten zu verknüpfen. Der Datensimulator für GKV-Daten (*hccdsim*) sowie der Softwaredemonstrator mit dem Deep Learning Algorithmus werden auf Anfrage zur Verfügung gestellt.

Schlussfolgerung

Signaldetektion basierend auf Routinedaten bietet ein enormes Potenzial, um die Arzneytherapiesicherheit zu verbessern. Eine umfassende Nutzung dieser Informationsquelle ist im Interesse der Patienten und Patientinnen sowie des gesamten Gesundheitssystems dringend geboten. Dieses Projekt diente dazu, als Grundlage für den vermehrten Einsatz von GKV-Daten in der Pharmakovigilanz eine verbesserte Methodik zu entwickeln. Zur Illustration dieser methodischen Grundlagenforschung wurden DOAKs und die mit ihnen bekannter- und möglicherweise in Zusammenhang stehenden Arzneimittelnebenwirkungen untersucht. Im Ergebnis spiegeln die anhand der neuen Methoden generierten Signale die aus der Arzneimittelinformation bekannten Signale wider. Dies zeigt eindrücklich, dass der in diesem Projekt verfolgte Ansatz erfolgreich eingesetzt werden

kann. Darüber hinaus wurden weitere, bisher unbekannte Signale identifiziert, die allerdings bis auf ein Signal nicht in entsprechenden pharmakoepidemiologischen Validierungsstudien verifiziert werden konnten. Der endgültige Nachweis steht noch aus; hier erfolgen zurzeit die oben angesprochenen Expertenkonsultationen. Damit ermöglichen die in diesem Projekt entwickelten Methoden, weitere potenzielle Signale zu identifizieren, die natürlich in jedem Einzelfall im Detail abgeklärt werden müssen. Aber auch dann, wenn für DOAKs keine zuvor noch nicht beschriebenen Zusammenhänge zu UAWs gefunden werden sollten, sondern nur die bekannten, sprechen die von uns erzielten Ergebnisse für unser generelles Vorgehen.

Unsere Arbeit hat gezeigt, welche Fallstricke beim Design, der Durchführung und der statistischen Analyse solcher Studien in Deutschland bestehen. Wir konnten Lösungen erarbeiten und interessante Signale identifizieren.

Zitierte Literatur

- European Medicines Agency (2014). One-year report on human medicines pharmacovigilance tasks of the European Medicines Agency. EMA/171322/2014. URL: https://ec.europa.eu/health/files/pharmacovigilance/202014_ema_oneyear_pharmacov_en.pdf.
- Goldman S (1998). Limitations and strengths of spontaneous reports data. *Clinical Therapeutics* 20S: C40-C44.
- Hornung R, Wright MN (2019). Block forests: Random forests for blocks of clinical and omics co-variate data. *BMC Bioinformatics* 20: 358.
- Linder R (2006). Lernstrategien zur automatisierten Anwendung künstlicher neuronaler Netzwerke in der Medizin. Logos-Verlag.
- Nembrini S, König IR, Wright MN (2018). The revival of the Gini Importance? *Bioinformatics* 34:3711–3718
- Suling M, Pigeot I (2012). Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics* 4:607-640.
- Yu K, Li Q, Bergen AW, Pfeiffer RM, Rosenberg PS, Caporaso N, Kraft P, Chatterjee N (2009). Pathway analysis by adaptive combination of p-values. *Genetic Epidemiology*, 33(8):700-709.
- Whitaker HJ, Hocine MN, Farrington CP (2009). The methodology of self-controlled case series studies. *Stat Methods Med Res* 18:7-26.

7. Beitrag für die Weiterentwicklung der GKV-Versorgung und Fortführung nach Ende der Förderung

Das in Deutschland etablierte Spontanmeldesystem leistet einen wichtigen Beitrag zur Pharmakovigilanz, weist aber auch, wie oben erwähnt, massive Limitationen auf, wie beispielsweise lückenhafte Informationen, die den gesundheitlichen Gesamtkontext des Patienten nur unzureichend abbilden. Daher lag es nahe, auf die in den GKV-Routinedaten bereits vorhandene, nahezu umfassende Dokumentation der Diagnosen und Leistungsanspruchnahmen für die Pharmakovigilanz zurückzugreifen. In anderen Ländern wird diese Option bereits genutzt, beispielsweise in den USA von der FDA.

Unser Projekt hat gezeigt, dass Signaldetektion auf Basis von GKV-Routinedaten möglich ist und leicht angewandt werden kann. Die notwendige Software dafür ist entwickelt. Das BfArM, hier insbesondere Herr Prof. Bernard Sachs, hat Interesse an unseren Vorarbeiten gezeigt und seine Unterstützung bei der Validierung unserer Ergebnisse zugesagt. Nach diesem Proof of Concept ist es wichtig, das Potenzial für weitere Arzneimittelgruppen auszuloten. Die Methodik dazu steht bereit.

8. Erfolgte bzw. geplante Veröffentlichungen

Erfolge Veröffentlichungen (peer-review)

Foraita R, Dijkstra L, Falkenberg F, Garling M, Linder R, Pflock R, Rizkallah MR, Schwaninger M, Wright MW, Pigeot I (2018). Aufdeckung von Arzneimittelrisiken nach der Zulassung. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 61: 1075-1081, URL: <https://doi.org/10.1007/s00103-018-2786-z> (Green Access: <https://repository.publisso.de/resource/fr!%3A6421679>)

Dijkstra LJ, Garling M, Foraita R, Pigeot I (2020). Adverse event or innocent bystander? A systematic comparison of statistical discovery methods for spontaneous reporting systems. Pharmacoepidemiology and Drug Safety 29: 396-403, URL: <https://doi.org/10.1002/pds.4970> (Open Access)

Klose S, Pflock R, König IR, Linder R, Schwaninger M (2020). Metamizole and the risk of drug-induced agranulocytosis and neutropenia in statutory health insurance data. Naunyn-Schmiedeberg's Archives of Pharmacology 393: 681–690, URL: <https://doi.org/10.1007/s00210-019-01774-4>

Konferenzbeiträge

Dijkstra L, Garling M, Foraita R, Pigeot I. Adverse event discovery for spontaneous re-orting systems: A systematic comparison. 64. Biometrisches Kolloquium, 25.-28. März 2018, Frankfurt (Vortrag)

Foraita R, Dijkstra LJ, Linder R, Schwaninger M, Pigeot I. Nutzung von Routinedaten zur Aufdeckung von Arzneimittelrisiken in Deutschland: Methodenentwicklung. Kongress "Zwei Jahre Innovationsfonds – Impulsgeber für eine bessere Versorgung" des Gemeinsamen Bundesausschuss (G-BA)-Innovationsausschusses, 28. Mai 2018, Berlin (Poster)

Foraita R, Dijkstra LJ, Linder R, Schwaninger M, Pigeot I. Nutzung von Routinedaten zur Aufdeckung von Arzneimittelrisiken in Deutschland: Methodenentwicklung. 13. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi), 26.-28. September 2018, Bremen (Poster)

Foraita R, Dijkstra LJ, Linder R, Schwaninger M, Pigeot I. Nutzung von Routinedaten zur Aufdeckung von Arzneimittelrisiken in Deutschland: Methodenentwicklung. 5. Kongress für Patientensicherheit bei medikamentöser Therapie, 18.-19. Oktober 2018, Berlin (Poster)

Linder R, Falkenberg F, Pflock R, Schwaninger M, Thöne K, Foraita R. Innovationsfondsprojekt PV-Monitor: Nutzung von GKV-Routinedaten für eine bessere Pharmakovigilanz, 64. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS), 8.-11. September 2019, Dortmund (Vortrag)

Linder R, Pflock R, Schwaninger M, Thöne K, Foraita R, Pigeot I, Falkenberg F. Künstliche Intelligenz – Deep Learning mit GKV-Routinedaten für eine bessere Pharmakovigilanz. BMC Kongress 2020, 21.-22. Januar 2020, Berlin (Vortrag)

Linder R. Explainable AI. Arbeitsgemeinschaft für wirtschaftliche Verwaltung (AWV), Arbeitskreis „Organisatorische Aspekte des Einsatzes von Automatisierung und KI“, 6. Februar 2020, Berlin (Vorstellung von Möglichkeiten der XAI am Beispiel von PV-Monitor, Ziel der AWV ist Politikberatung)

Linder R, Falkenberg F, Pflock R, Schwaninger M, Thöne K, Foraita R. Innovationsfondsprojekt PV-Monitor: Nutzung von GKV-Routinedaten für eine bessere Pharmakovigilanz. Gesundheitsforen Leipzig (virtuelle Konferenz), 21. April 2020 (Vortrag, vorwiegend Vertreter der Analytikabteilungen der gesetzlichen Krankenkassen sowie Großunternehmen wie z.B. Siemens oder SAS)

Linder R, Falkenberg F, Pflock R, Schwaninger M, Thöne K, Foraita R. Innovationsfondsprojekt PV-Monitor: Nutzung von GKV-Routinedaten für eine bessere Pharmakovigilanz. I'M AI 2020, 21.-22. September 2020, Berlin (Vortrag)

Software (unter GPL-3.0 Lizenz veröffentlicht)

ARTPredict: Prediction for binary outcomes with the ARTP (<https://github.com/bips-hb/ARTPredict>; letztes Update: Dez 2020)

batchlasso: An R package for using the LASSO for ranking response-exposure pairs on the basis of the highest lambda for which they first appear in the active set (<https://github.com/bips-hb/batchlasso>; letztes Update: Dez 2020)

pvm: A collection of signal detection methods used in the field of pharmacovigilance (<https://github.com/bips-hb/pvm>; letztes Update: Sept 2019)

pvmcomparison: R package for comparing various methods used in the field of pharmacovigilance to detect associations between drugs and adverse events in spontaneous reporting data (<https://github.com/bips-hb/pvmcomparison>; letztes Update: Jan 2020)

rgp: Identification of risk groups in pharmacovigilance using penalized regression (RGP) (<https://github.com/bips-hb/rgp>; letztes Update: März 2019)

Akronym: PV-Monitor
Förderkennzeichen: 01VSF16020

srsim: R package for simulating spontaneous reports (<https://github.com/bipshb/srsim>; letztes Update: Feb 2019)

Software (auf Anfrage)

hccdsim: Simulator for health care claims data (letztes Update: Okt 2019; erhältlich auf Anfrage; Kontakt: dijkstra@leibniz-bips.de)

Deep Learning Signal Detector: Softwaredemonstrator der TK (letztes Update: März 2020, erhältlich auf Anfrage; Ansprechpartner: Prof. Dr. Roland Linder, prof.dr.roland.linder@tk.de)

Geplante Veröffentlichungen

Dijkstra LJ, Schink T, Linder R, Schwaninger M, Pigeot I, Foraita R*, Wright M* (2021). A structured approach to discovering and verifying adverse drug event in health care claims data. Manuskript in Vorbereitung.

Dubrall D*, Pflock R*, Kosinska J, Schmid M, Bleich M, Himmerkus N, Offermanns S, Schwaninger M*, Sachs B* (2020). Do dimethyl fumarate and nicotinic acid elicit common, potentially HCA2-mediated adverse reactions?, submitted to British Journal of Clinical Pharmacology.

Rizkallah MR, Dijkstra LD, Schwaninger M, Wilhelm AFX, Pigeot I, Foraita R (2021). Predicting patient risk for adverse drug events in health care claims data using functional targets. Manuskript in Vorbereitung.

PV-Monitor Group (2021). Do direct oral anticoagulants cause epilepsy? Manuskript in Vorbereitung.

* contributed equally to this work

Unveröffentlichte Forschungsmanuskripte

Dijkstra LJ, Foraita R, Pflock R, Linder R, Pigeot I, Wright MN (2019). Discovering rare side effects in health care claims data. Technischer Report. Erhältlich auf Anfrage.

Rizkallah MR, Dijkstra LJ, Pflock R, Wilhelm AFX, Pigeot I, Foraita R (2019). Identifying patient risk groups for adverse drug events in health care claims data using functional targets. Technischer Report. Erhältlich auf Anfrage.

9. Anlagen

- Anlage 1: Ergebnisse der Benchmarktests
- Anlage 2: Vorhersage seltener UAWs mit maschinellen Lernverfahren basierend auf GKV-Daten
- Anlage 3: Studienprotokoll zur Machbarkeitsstudie (Verarbeitungskette 1)
- Anlage 4: Studienprotokoll zu Verarbeitungskette 2
- Anlage 5: Ergebnisse der Vorhersagestudie zu Functional Targets
- Anlage 6: Ergebnisse der Signalgenerierung
- Anlage 7: Ergebnisse der Validierungsstudie
- Anlage 8: Veröffentlichte Publikationen

Anlage 1: Ergebnisse der Benchmarktests

Zur Bewertung und zur Verbesserung des Deep Learning Algorithmus wurden die bekannten Datensätze aus der Proben1-Benchmarksammlung von Lutz Prechelt (1994) verwendet. Darüber hinaus wurden weitere, unterschiedliche Benchmarkdatensätze aus dem Bereich der Life Sciences zusammengestellt, um die Klassifikationsleistung zu bewerten. Dazu gehören die folgende Datensätze des UCI Machine Learning Repository (Dua & Graff 2019): Adult, Arcene, Dorothea, Forest Type Mapping, PubChem Bioassay, Smartphone-Based Recognition of Human Activities and Postural Transitions, SPECT Heart, SPECTF Heart, Thyroid Disease, Wilt. Die genannten Benchmarks wurden in das für unseren Softwareprototyp erforderliche Datenformat überführt und untersucht. Die Performanz der entwickelten Software wurde u.a. an den medizinischen Zweiklassen-Klassifikationsproblemen Cancer, Diabetes und Heartc der Proben1-Benchmarksammlung untersucht. Tabelle 7 zeigt die Ergebnisse des Benchmarktests. Die erzielten Fehlklassifikationsraten vom Deep Learning Ansatz waren im Mittel um 9% besser als die von Prechelt erzielten Resultate sowie um 23% besser als die Ergebnisse entsprechender logistischer Regressionen.

Tabelle 7: Erzielte Fehlklassifikationsraten im Benchmarktest.

Benchmark-Datensatz	L. Prechelt, Fehlklassifikationen (%)	Deep Learning (%)	Logistische Regression (%)	Relative Verbesserung DL vs. LP (%)	Relative Verbesserung DL vs. LR (%)
<i>Cancer1</i>	1,38	1,72	2,30	-24,64	25,22
<i>Cancer2</i>	4,52	2,87	4,02	36,50	28,61
<i>Cancer3</i>	3,37	2,87	5,17	14,84	44,49
<i>Diabetes1</i>	24,10	21,88	23,96	9,21	8,68
<i>Diabetes2</i>	25,91	23,96	23,44	7,53	-2,22
<i>Diabetes3</i>	22,59	19,79	20,83	12,39	4,99
<i>Heartc1</i>	20,82	17,33	20,00	16,76	13,35
<i>Heartc2</i>	5,07	5,33	16,00	-5,13	66,69
<i>Heartc3</i>	15,40	13,33	16,00	13,44	16,69

DL: Deep Learning; LP: Lutz Prechelt; LR: Logistische Regression

Literatur

Dua D, Graff C (2019). UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine. URL: <http://archive.ics.uci.edu/ml> (Letzter Zugriff: 14.03.2019)

Anlage 2: Vorhersage seltener UAWs mit maschinellen Lernverfahren basierend auf GKV-Daten

Design der Vorhersagestudie

In die Studie wurden Versicherte der TK eingeschlossen, die in den Jahren 2014 bis 2016 eine erste DOAK-Verschreibung (Studieneintritt) erhalten haben, bis Studieneintritt eine durchgehende Versicherungszeit von mind. 10 Jahren aufweisen konnten sowie weiterhin durchgehend versichert waren bis zur UAW, zum Tod oder bis Studienende. Nach Studieneintritt wurden die Versicherten 365 Tage beobachtet. Alle nicht-verstorbenen Versicherten ohne UAW, die weniger als 365 Tagen Follow-up Zeit aufwiesen, wurden zensiert.

Als UAW wurde gastrointestinale Blutungen gewählt, wobei Tod als konkurrierendes Risiko betrachtet wurde. Es wurden fünf Modelle untersucht:

- *Nullmodell* (ohne Kovariaten): gefittet mit dem Aalen-Johanson Schätzer (Standard-schätzer für die kumulative Inzidenzfunktion bei konkurrierenden Risiken)
- *Modell 1* (mit Alter, Geschlecht): Ursachenspezifisches Cox-Modell (cause-specific Cox model, CSC)
- *Modell 2* (mit Alter, Geschlecht, DOAK): Ursachenspezifisches Cox-Modell
- *Modell 3* (mit Alter, Geschlecht, DOAK, alle ICD/ATC aus dem letzten Jahr vor Studieneintritt): Lasso
- *Modell 4*: (mit Alter, Geschlecht, DOAK, alle ICD/ATC aus dem letzten Jahr vor Studieneintritt): Random Survival Forests (RSF)
- *Modell 5* (mit Alter, Geschlecht, DOAK, vollständige Zeitreihen aller ICD/ATCs aus den 10 Jahren vor Studieneintritt): Recurrent Neural Network (RNN)

Die RNNs wurden implementiert, wie in Wright et al. (2021) beschrieben. Alle Versicherten, deren Studieneintritt im Jahr 2014 war, wurden für das Training der Modelle verwendet. Dafür wurde eine 10-fache Kreuzvalidierung verwendet. Versicherte mit Studieneintritt in 2015 oder 2016 wurden für die temporale Validierung der Modelle benutzt.

Es wurde untersucht, ob die Modelle 1 bis 5 eine bessere Vorhersage für die UAW liefern als das Nullmodell.

Ergebnisse der Vorhersagestudie

In der Vorhersagestudie wurde die Vorhersagekraft bekannter Nebenwirkungen durch u.a. Random Forest sowie RNNs basierend auf longitudinalen GKV-Daten evaluiert.

Insgesamt wurden 98.335 Versicherte in die Studie eingeschlossen. Von diesen haben 805 die UAW erfahren, 3.426 Versicherte sind während des Follow-ups verstorben und 94.104 wurden zensiert. Zum Trainieren der Modelle konnten 26.798 Versicherte (UAW: 257, Tod: 1.155) genutzt werden und zur temporalen Validierung 33.001 Versicherte (UAW: 320, Tod: 1.398). Mit diesen Fallzahlen konnte eine maximale kumulative Inzidenz von 0,01 für die UAW und 0,04 für Tod ermittelt werden. Dies zeigt deutlich, dass eine Vorhersage der UAW aufgrund ihrer Seltenheit extrem schwierig ist.

Tabelle 8 zeigt die Vorhersagekraft der einzelnen Modelle im Vergleich zum Nullmodell ausgedrückt als Index of Prediction Accuracy (Kattan & Gerds 2108) (höhere Werte deuten auf eine bessere Vorhersagekraft hin).

Tabelle 8: Vorhersagewahrscheinlichkeit für die UAW nach 12 Monaten nach Erstverschreibung.

Modell	Kovariaten	Index Prediction Accuracy nach 12 Monaten	
		UAW	Tod
Modell 1: CSC	Alter, Geschlecht	0,34 %	3,40 %
Modell 2: CSC	wie Modell 1 + DOAK	0,34 %	3,44 %
Modell 3: LASSO	wie Modell 2 + alle ICD/ATC aus letztem Jahr vor Studieneintritt	-0,52 %	7,65 %
Modell 4: RSF	wie Modell 2 + alle ICD/ATC aus letztem Jahr vor Studieneintritt	0,47%	8,89 %
Modell 5: RNN	wie Modell 2 + vollständige Zeitreihen aller ICD/ATCs aus den 10 Jahren vor Studieneintritt)	Fit misslungen	Fit misslungen

CSC: Cause-specific Cox model; LASSO: Least absolute shrinkage and selection operator; RSF: Random Survival Forest; RNN: Recurrent Neural Network.

Es ist deutlich zu erkennen, dass Modell 3 die UAW sogar schlechter als das Nullmodell vorhersagte. Leider hat das RNN nicht zu einer sinnvollen Modellanpassung geführt. Das mag daran liegen, dass für die Verarbeitung der Zeitreihe von über 10 Jahren das RNN ca. 1.3 Millionen Parameter anpassen musste, was für die angewandten statistischen Methoden ein äußerst hochdimensionales Problem darstellte. Zudem war der Anteil an UAW so gering, dass ein extrem unbalanciertes Design vorlag, das der Methode zusätzliche Schwierigkeiten bereitete. Abbildung 5 zeigt für Modell 1 bis 4 die monatsgenauen Vorhersagewerte.

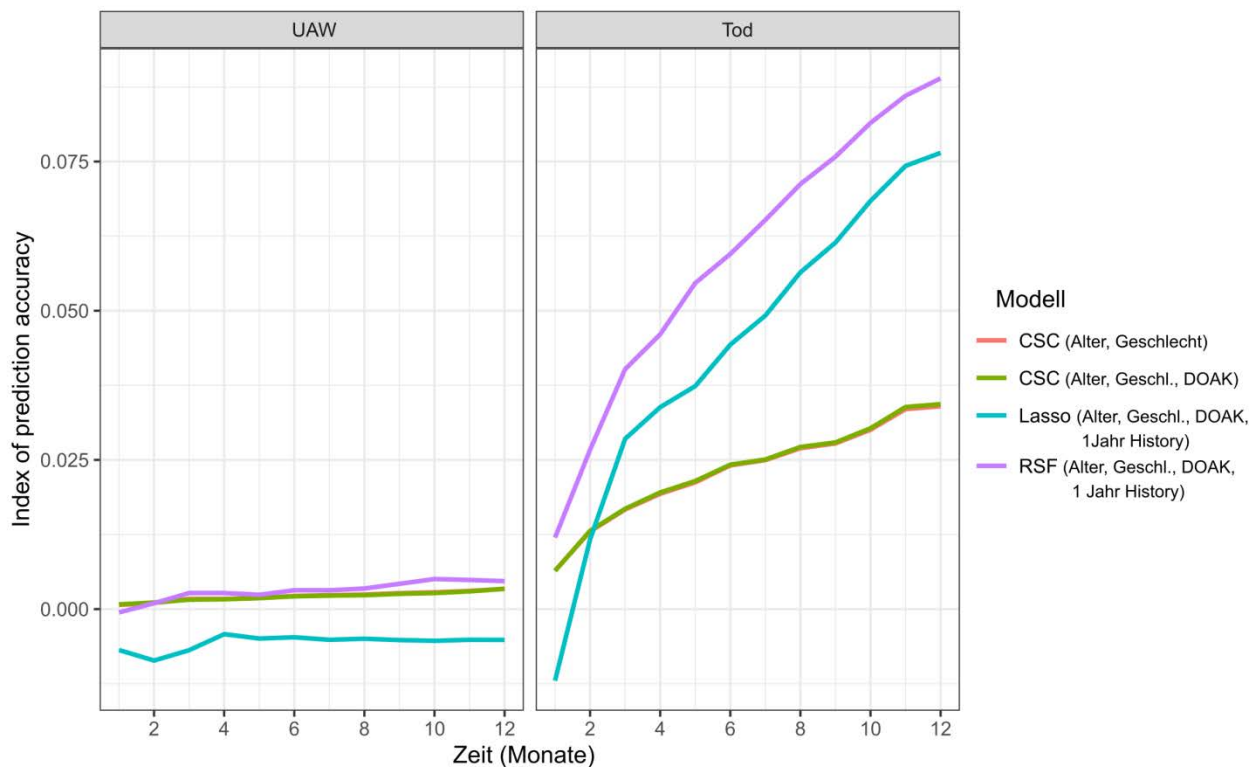


Abbildung 5: Index of Prediction Accuracy als Ergebnis der temporalen Validierung.

CSC: Cause-specific Cox model; LASSO: Least absolute shrinkage and selection operator; RSF: Random Survival Forest.

Zusammenfassend zeigen die Ergebnisse, dass die Vorhersage einer bekannten und eher häufigen Nebenwirkung wie die der gastrointestinalen Magenblutung nicht verbessert werden kann, indem vorhandenes Wissen über Diagnosen und Verschreibungen aus den letzten 10 Jahren berücksichtigt wird. Die Vorhersage des konkurrierenden Risikos Tod konnte mit den zusätzlichen Informationen ein wenig verbessert werden. Für seltene Nebenwirkungen erwarten wir ein noch unbalancierteres Verhältnis und damit keine Verbesserungen der Vorhersage.

Letztendlich kamen wir zu dem Schluss, dass die hier verwendeten maschinellen Lernverfahren nicht in der Lage waren, im Rahmen einer Signaldetektion seltene UAWs vorherzusagen. Die Gründe hierfür können vielfältig sein. Bei Verfahren zur Signaldetektion ist es wichtig, dass das Studiendesign schnell auf andere Ereignisse übertragbar ist, kein Vorwissen über das Ereignis oder das Arzneimittel notwendig ist und dass der Analysedatensatz zwar qualitätskontrolliert ist, aber darüber hinaus keine weitere, meist zeitintensive, Datenbearbeitung benötigt. Wir hatten die Hoffnung, dass die maschinellen Lernverfahren auf Basis des individuellen Versichertenprofils bedeutsame Muster entdecken werden. Es scheint jedoch so, dass die Informationen aus Abrechnungsdaten für die Vorhersage einer gastrointestinalen Blutung im Zusammenhang mit DOAK-Verschreibungen nicht ausreichten. Eine solche Vorhersage könnte bei anderen Ereignis-Arzneimittel-Kombinationen besser sein. Bei unserer Untersuchung war die Anzahl der Ereignisse recht gering. Auf den gesamten Zeitraum gesehen wurde weniger als eine von 100 DOAK-Nutzern und –Nutzerinnen aufgrund einer GB hospitalisiert.

Dadurch waren die Gruppengrößen der Erkrankten und Nicht-Erkrankten sehr unterschiedlich. Bei einer wirklich seltenen UAW wird dieses Ungleichgewicht jedoch noch viel ausgeprägter sein. Zwar gibt es statistische Methoden, um grundsätzlich mit unbalancierten Daten umgehen zu können. Jedoch stellen diese für den Einsatz bei RSF und RNN noch eine große methodische Herausforderung dar. Unserer Einschätzung nach benötigen RSF noch eine methodische Verbesserung, damit diese zufriedenstellend auf GKV-Daten zur Signaldetektion eingesetzt werden können. RNNs haben sich hingegen als ungeeignet für GKV-Daten erwiesen: Zur Modellierung der zeitlichen Versichertenverläufe müssen extrem viele Parameter geschätzt werden. Dies hat zur Folge, dass dafür sehr viele Daten notwendig sind (insbesondere auch in der kleineren Gruppe der Erkrankten) und die Berechnungen sehr lange dauern. Dies macht den Einsatz von RNNs für die Signaldetektion seltener Ereignisse unbrauchbar. Eventuell könnten bei den einzelnen maschinellen Lernverfahren die Parameter so getunt werden, dass damit einzelne, bereits bekannte UAWs besser vorhergesagt werden können. Aber für unbekannte UAWs, so wie es in der Signaldetektion der Fall wäre, sind wir aufgrund unserer Forschungsergebnisse skeptisch, dass unbekannte seltene Ereignisse auf diese Art aufgedeckt werden können.

Literatur

- Kattan MW, Gerds TA (2018). The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagnostic and Prognostic Research* 2: 7. URL: <https://doi.org/10.1186/s41512-018-0029-2>
- Wright MN, Kusumastuti S, Mortensen LH, Westendorp R, Gerds T (2021). Personalised need of care in an ageing society: The making of a prediction tool based on register data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. URL: <https://doi.org/10.1111/rssa.12644>

Anlage 3: Studienprotokoll zur Machbarkeitsstudie (Verarbeitungskette 1)

PV-Monitor: Methodisches Vorgehen im Rahmen der Machbarkeitsstudie

R. Linder, F. Falkenberg

Stand 12.07.2019

1. Zielsetzung

Die Machbarkeitsstudie hat zum Ziel, ausgehend von nahezu demselben Datenkörper in zwei separaten Arbeitsgruppen (BIPS und TK) mit unterschiedlichen Methoden und Herangehensweisen Signale zu generieren, zu vergleichen und schlussendlich diejenigen Signale zu identifizieren, für die eine weitergehende Abklärung in pharmakoepidemiologischen Validierungsstudien lohnenswert erscheint.

2. Prinzipielles Vorgehen

Ausgehend von den Routinedaten der Techniker Krankenkasse (10,2 Mio. Versicherte) werden Deep Learning Algorithmen darauf trainiert, unerwünschte Arzneimittelwirkungen (UAW) vorherzusagen. Dies erfolgt mit dem Wissen um potenziell relevante Aspekte in der Vorgeschichte (v.a. ICD-, ATC-, OPS- und EBM-Codes) sowie der Information, ob eine Krankenhausaufnahme unter dem pharmakologischen Einfluss des untersuchten Arzneimittels erfolgte. In einer sich anschließenden Sensitivitätsanalyse wird die Information, ob sich die stationäre Aufnahme unter dem Einfluss des DOAK ereignete, dahingehend verfremdet, dass die Merkmalsausprägungen 1 (Einfluss ja) und -1 (Einfluss nein) für die Variable " DOAK abgegeben j/n" vertauscht werden (Switched Sensitivity Analysis, SSA, vgl. Neural Net Clamping Technique von Wang et al. (1998)). Je stärker der Switch zu einer Verschlechterung der Klassifikationsleistung führt, desto größer ist der daraus berechnete Wert für die SSA-Importance, d.h. die Relevanz des Zusammenhangs zwischen DOAK und UAW. Auch wenn eine große SSA-Importance keine Kausalitätsaussage ermöglicht, generiert sie jedoch Arbeitshypothesen für weitergehende Validierungen/Plausibilisierungen, die sowohl durch das BIPS auf Basis der pharmakoepidemiologischen Datenbank GePaRD (25 Mio. Versicherte) als auch durch medizinische und pharmakologische Experten und Expertinnen am Institut für Experimentelle und Klinische Pharmakologie und Toxikologie der Universität zu Lübeck erfolgen können.

3. Daten und Methode

Im Folgenden soll auf einzelne Aspekte der Machbarkeitsstudie näher eingegangen werden.

Datenkörper

Das geplante Vorgehen dieser Machbarkeitsstudie, die jeweils von der TK und vom BIPS entwickelten Methoden auszutauschen und von der jeweils anderen Gruppe auf deren entsprechenden Datensatz anzuwenden, konnte so nicht umgesetzt werden. Die im BIPS in der Programmiersprache R geschriebenen Routinen können de facto nicht oder nur unter großen Schwierigkeiten in der Software-Umgebung der TK genutzt werden. Umgekehrt ist im BIPS keine Hardware verfügbar, die es erlaubt, die rechenintensiven, in C#.NET programmierten Deep Learning Algorithmen der TK auf einer Windows-Plattform auszuführen. Da zudem ein Austausch von Sozialdaten ebenfalls nicht ohne weiteres möglich ist, wird ein reduzierter, für beide Gruppen verfügbarer Datenkörper für die

Machbarkeitsstudie genutzt. Dieser überlappende Datenkörper entsteht dadurch, dass die TK jährlich einen breiten Datenkranz an Routinedaten zur Nutzung in GePaRD an deren Vertrauensstelle übermittelt, so dass Daten derselben Jahrgänge für die separaten Analysen zur Verfügung stehen.

Aufgrund der gesetzlich vorgegebenen Löschfristen nach §304 SGB V, denen die TK unterliegt, und des zeitlichen Verzugs durch die Datenaufbereitung in der Vertrauensstelle von GePaRD stehen de facto die Jahre 2015 und 2016 zur Verfügung. Da jährlich nicht alle verfügbaren Sozialdaten an das BIPS übermittelt werden, werden seitens der TK nur die tatsächlich übermittelten Daten verwendet.

Geringfügige Abweichungen zwischen den Datenbeständen ergeben sich dadurch, dass der in GePaRD verfügbare Datenbestand eingefroren ist, der Datenbestand der TK jedoch nachträglichen Korrekturen unterliegt.

Untersuchungsgegenstand

Die Machbarkeitsstudie beschränkt sich auf die direkten oralen Antikoagulanzen (DOAK), namentlich Apixaban (ATC B01AF02), Edoxaban (B01AF03), Dabigatran (B01AE07) und Rivaroxaban (ATC B01AF01). Darüber hinaus werden von der TK acht weitere Antikoagulanzen untersucht, darunter insbesondere Vitamin-K-Antagonisten.

Definition der UAW

Die Selektion der Teilnehmergruppe erfolgt über eine stationäre Aufnahme wegen bestimmter potenzieller UAW. Als solche gelten alle Hauptentlassungsdiagnosen (HDIA), da hier unterstellt werden kann, dass gemäß Deutscher Kodierrichtlinie die HDIA die Indikation für die stationäre Behandlung abbildet. Den Autoren ist dabei durchaus bewusst, dass auch manche Nebendiagnosen den Aufnahmegrund abbilden können, sei es im Fall von mehrfachen Indikationen oder bedingt durch DRG-Gaming. Letzteres meint eine unzu-lässige Erlösoptimierung im Regelwerk der diagnosebezogenen Diag-nosis Related Groups (DRG), dem momentan gültigen Fallpauschalsystem. In der Güterabwägung erscheint es jedoch gerechtfertigt, die Nebendiagnosen auszuschließen, um nicht vorbestehende Probleme als neu aufgetretene Ereignisse fehl zu deuten. Ambulante Diagnosen, Arbeitsunfähigkeitsdiagnosen oder Diagnosen im Zusammenhang mit Heil- und Hilfsmitteln werden nicht genutzt, da diese auch bei gegebenem Diagnosesicherheitskennzeichen als eher unsicher gelten. Erschwerend kommt im Fall der vertragsärztlichen Diagnosen hinzu, dass diese nur quartalsgenau, nicht aber tagesgenau dokumentiert sind (vgl. Abschnitt „Defined Daily Dose“). Auch die Behandlungsdiagnosen aus den Polikliniken werden nicht für das Aufgreifen potenzieller UAW genutzt.

Im weiteren Verlauf werden die ICDs auf unterschiedlichen Aggregationsniveaus verwendet: 5-Steller, 4-Steller, 3-Steller und HMG (Hauptmorbiditätsgruppen aus der Berechnung des morbiditätsorientierten Risikostrukturausgleichs).

Indexjahr

Zur Vorhersage einer unter DOAK-Einfluss aufgetretenen UAW werden Informationen aus den 365 Tagen vor der Krankenhauseinweisung verwendet, wir sprechen von einem versichertenspezifischen Indexjahr. Um dieses Indexjahr für alle Versicherten verfügbar zu haben, werden nur Versicherte mit einer stationären Aufnahme vom 1.1.2015 bis 31.12.2016 in die Teilnehmergruppe eingeschlossen.

Durchgängigkeit der Versicherung

Eingeschlossen werden ausschließlich Versicherte mit durchgängiger TK-Versicherung im versichertenspezifischen Indexjahr.

Defined Daily Dose (DDD)

Um zu ermitteln, ob Versicherte zum Zeitpunkt der stationären Aufnahme unter dem pharmakologischen Einfluss eines DOAK standen, wird vereinfachend mit der Defined Daily Dose (DDD) in Kombination mit den aus der Packungsgröße abgeleiteten Tagesdosen gearbeitet: Liegt der Tag der Krankenhausaufnahme innerhalb des Intervalls aus dem Tag der letzten Abgabe des DOAK und diesem Tag plus der Zahl der abgegebenen Tagesdosen, dann wird davon ausgegangen, dass die potenzielle UAW unter dem Einfluss des untersuchten DOAK erfolgt ist.

Teilnehmer - und Kontrollgruppen

Ausgehend von den gefundenen UAW ergibt sich für jede Paarung aus UAW und DOAK eine Teilnehmergruppe (innerhalb derer die potenzielle UAW teils unter dem Einfluss des DOAK erfolgt, teilweise aber auch nicht). Teilnehmergruppen mit weniger als 10 Fällen einer unter DOAK-Einfluss erfolgten potenziellen UAW werden von der weiteren Analyse ausgeschlossen.

Zu jeder der verbliebenen Teilnehmergruppen wird eine alters- und geschlechtsgleiche Kontrollgruppe gebildet, wobei für die Kontrollen eine fiktive tagesgleiche Krankenhausaufnahme unterstellt wird, um Verzerrungen hinsichtlich saisonaler Effekte auszuschließen. Dieses 1:1 Matching gilt für alle Teilnehmergruppen mit mehr als 10.000 Fällen. Bei weniger Fällen wird ein 1:10 Matching gewählt.

Unabhängige Variablen (Input)

Als unabhängige Variablen werden im versichertenspezifischen Indexjahr taggenau dokumentierte ICDs aus Krankenhäusern (Haupt- und Nebendiagnosen) sowie den Polikliniken (Behandlungsdiagnosen) verwendet, im Fall von Doppelklassifikationen neben dem Primär- auch der Sekundär-Code. Die von Vertragsärzten dokumentierten ambulanten ICDs werden auch genutzt, jedoch nur gesicherte Diagnosen (Diagnosesicherheitskennzeichen = "G") und Diagnosen, die vor dem Quartal der stationären Aufnahme dokumentiert sind.

Darüber hinaus werden die taggenau dokumentierten stationären und poliklinischen OPS- und EBM-Codes verwendet sowie die ATC-Codes der im Indexjahr abgegebenen Arzneimittel. Hier wird unterschieden zwischen Arzneimitteln, die im Indexjahr abgegeben wurden, und solchen, unter deren Einfluss der Versicherte am Tag der Krankenhausaufnahme stand.

Hinzu kommt die unabdingbare Information, ob der Versicherte zum Zeitpunkt der stationären Aufnahme unter dem Einfluss des untersuchten DOAK stand.

Alle zuvor genannten Variablen werden als Dummy-Variablen kodiert (1=trifft zu; 0=trifft nicht zu). Ergänzt werden die Dummy-Variablen durch das Geburtsjahr und die Geschlechtsangabe, wiederum eine dichotome Variable.

4. Literatur

Wang W, Jones P, Partridge D (1998) Ranking pattern recognition features for neural networks. Proc. of the 1st Int. Conf. on Advances in Pattern Recognition (ICAPR). Plymouth, England. pp. 232-241.



Leibniz Institute for Prevention Research and Epidemiology – BIPS GmbH

Study Design Feasibility Study (version 1.0)

Signal Detection Study for Directly Acting Oral Anticoagulants (DOACs)

(PV-Monitor: Rivaroxaban PASS)

Funded by GB-A

Principal Investigator: Prof. Dr. Iris Pigeot
pigeot@leibniz-bips.de, +49-421-218-56942
Author: Dr. Ronja Foraita
foraita@leibniz-bips.de, +49-421-218-56954

Address:
Leibniz Institute for Prevention Research and Epidemiology – BIPS GmbH
Achterstraße 30, 28359 Bremen, Germany
<https://www.bips-institut.de>

Document Information

Version	Author	Reason for Modification	Date
0.1	R. Foraita	First draft	20.02.2019
0.2	M. Wright	Update data flow	28.02.2019
0.3	L. Dijkstra	Add statistical methods	05.03.2019
0.4	M. Wright	After discussion with ClinEpi	17.04.2019
1.0	R. Foraita	Final version	18.04.2019

Responsibilities

Study Team	
Prof. Dr. rer. nat. Iris Pigeot	Principal investigator
Dr. rer. nat. Ronja Foraita	Project manager, senior statistician
Dr. rer. hum. biol. Marvin N. Wright	Senior statistician
Louis Dijkstra	Statistician
Mariam R. Rizkallah	Data engineer
Dr. rer. medic Tania Schink	Senior epidemiologist
Kathrin Krüger	Epidemiologist

Table of Contents

Study Design Feasibility Study (version 1.0).....	9
1 Introduction.....	14
2 Data Management	14
2.1 Data Flow	14
2.2 General Derivation Rules.....	15
2.2.1 Continuous Insurance Periods	15
2.2.2 Identification of In- and Outpatient Diagnoses	15
2.2.3 Assessment of Dispensation Data.....	15
2.2.4 Identification of OPS Codes	16
2.2.5 Identification of EBM Codes	16
2.2.6 Missing Data	16
2.3 Analysis Dataset.....	16
2.3.1 Cohort Data.....	16
2.3.2 Covariate Data	17
2.3.3 Dispensation Data.....	17
2.3.4 Diagnosis Data.....	17
3 Statistical Analyses	18
3.1 Main Analyses	18
3.2 Comparison of Methods.....	19
4 References	19

ADE	Adverse Drug Event
ADEI	Adverse Drug Event of Interest
ATC	Anatomical Therapeutic Chemical Classification System
BIPS	Leibniz Institute for Prevention Research and Epidemiology
CPR	Central Pharmaceutical Reference Database
DOAC	Directly Acting Oral Anticoagulants
DOACI	Directly Acting Oral Anticoagulants of Interest
GePaRD	German Pharmacoepidemiological Research Database
ICD-10-GM	International Classification of Diseases 10 th revision - German Modification
OPS	Operationen- und Prozedurenschlüssel (Operation- and Procedure Code)
PZN	Pharmazentralnummer (Central Pharmaceutical Number)
SCCS	Self-Controlled Case Series
SHI	Statutory Health Insurance
TK	Techniker Krankenkasse

1 Introduction

Study objective: The objective of this feasibility study is to detect adverse drug events associated with directly acting oral anticoagulants (DOAC). For each DOAC of interest (DOACI), a separate signal detection study is conducted. The following DOACI are considered:

- Apixaban ATC_WIDO = B01AF02
- Edoxaban ATC_WIDO = B01AF03
- Dabigatran ATC_WIDO = B01AE07
- Rivaroxaban ATC_WIDO = B01AF01

For this purpose a self-controlled case series method (SCCS) will be used.

Purpose of this document: This document describes the statistical analyses and data management activities of the study.

2 Data Management

2.1 Data Flow

The study will be based on administrative data from the Techniker Krankenkasse (TK), included in the German Pharmacoepidemiological Research Database (GePaRD). The raw data as delivered by the insurances to the third party trust center is pseudonymized and transferred to oracle tables which are stored at an oracle database server at the Leibniz Institute for Prevention Research and Epidemiology - BIPS GmbH (BIPS). Details on database description and transfer rules are documented in standard operating procedures (see G:\PHEpi_DB\20_Manuale_und_Dokumente\Arbeitsanweisungen). As the database is updated regularly, database snapshots are stored at the Oracle database server (path: PHDB2B) to fix the current status of the database. The analyses will be based on the database snapshot SNAP1801. The respective data used are displayed in **Table 1**.

Table 1: Snapshots and used data years

SNAP	AOK	hkk	DAK	TK
1801	-	-	-	2015-2016

Analysis datasets will be stored at the Oracle database server in the project scheme AntikoagPASS or on the linux cluster in /home/Gepard/pv_monitor/.

The Central Pharmaceutical Reference Database (CPR), which has been established by the BIPS, will be used for the identification of relevant drugs. Information from the CPR can be linked to the prescription data by reference IDs (REF_ID) which is based on the central pharmaceutical number (Pharmazentralnummer - PZN). Details on the CPR are described in the documentation available at G:\PHEpi_DB\35_ArzneimittelDB.

REF_IDs are required for the identification of study drugs and concurrent medication. REF_IDs are available from the reference data set, including a mapping to ATC (G:\

G:\PHEpi_DB\20_Manuale_und_Dokumente\HowToDoPE\Vorlagen\ATC_zugelassene_Arzn
eimittel\Ref_ID_ATC_180718.sas7bdat).

2.2 General Derivation Rules

2.2.1 Continuous Insurance Periods

Insurance periods will be selected by using the variables PERIOD_BEG and PERIOD_END from T_IPERIODS_COMB_TK. Adjacent periods will be combined to a single period of maximal length using the SAS-macro %combine_periods (path: G:\PHEpi_DB\25_SASTOOLS\macros) with a grace period of 30 days in-between single periods.

Information on death will only be considered from PERIOD_BEG onwards and will be selected from hospital as well as from core data, using tables T_INPATIENT (DIS_CAUSE) and T_IPERIODS (OFF_CAUSE). Hospital deaths will be selected from T_INPATIENT as sub-str(DIS_CAUSE,1,2)='07'. For deceased persons identified by means of hospital diagnoses the end of insurance is defined as mdy(HOSP_ENDM/D/Y) of the last coded death. Outpatient deaths will be selected from T_IPERIODS as OFF_CAUSE = 1 and mdy(INS_ENDM/D/Y). Out- and inpatient death is only valid if the end of insurance or date of discharge from the hospital is equal to the "calculated end" mdy(B_ENDM/D/Y) ± 14 days in table T_INSURANTS_TK. If INS_ENDM/D=99 then INS_END=31.12.min(INS_END, CLAIM_ENDY) and if B_ENDM/D=99, then B_END=31.12.min(B_END, CLAIM_ENDY). For the information on claims, the table T_CLAIM with max(CLAIM_ENDY) per IDNUM will be used. If different dates are found for a death in T_INPATIENT and T_IPERIODS, the date from T_INPATIENT is used. End of insurance period will be set to the minimum of date of end of insurance (PERIOD_END) and date of death.

2.2.2 Identification of In- and Outpatient Diagnoses

Only inpatient diagnoses are taken into account.

- Only codes valid on the respective day (as indicated by the columns "VAL_BEG/END_HOSP" and "VAL_BEG/END_AMB")
- Inpatient diagnoses
 - from T_DIAG_TK.DIAG_CLEAN, with T_DIAG_TK.DIAG_TYPE in ('HH','NH')
 - the date of diagnosis will be extracted from the HOSP_ID as follows
mdy(substr(HOSP_ID,11,2),substr(HOSP_ID,13,2),substr(HOSP_ID,7,4)).

2.2.3 Assessment of Dispensation Data

The date of the dispensation is set based on the variables DEL_DAT in T_PRESCRIPTION_TK. If one of these variables is missing, the prescription date (ISSUE_DAT) will be used instead.

2.2.4 Identification of OPS Codes

No OPS codes are used.

2.2.5 Identification of EBM Codes

No EBM codes are used

2.2.6 Missing Data

Missing data will only occur for core demographic variables (e.g. sex or age). Please note that missing values in these variables are very rare. Subjects with missing data will be deleted from the analysis if the statistical method includes demographic variables as covariates (RF, LASSO, see Section 3.1).

2.3 Analysis Dataset

A retrospective cohort study will be performed to investigate the risk ADEI associated with DOACI using a SCCS design. The structure of the analysis dataset is described in detail in the following.

2.3.1 Cohort Data

The cohort data contains one record per patient. Cohort members will be defined according to the following criteria.

Inclusion criteria

All insured persons enrolled in GePaRD during the study period between January 01, 2015 and December 31, 2016 are eligible for the cohort.

Cohort members have to fulfill all of the following inclusion criteria:

- At least one dispensation of DOACI during the study period
- At least 90 days before first dispensation of DOACI
- At least 90 days after first dispensation of DOACI
- No DOAC use in the baseline period
- No missing values regarding sex, year of birth and GKZ5
- Living in Germany according to GKZ5

Cohort Entry

All patients with a continuous insurance period of more than one year within the study period between January 1, 2015 and December 31, 2016 will be selected.

Cohort entry is defined as the last of the following dates

- begin of study period (January 1, 2015)
- begin of continuous insurance period
- first DOACI dispensation

Cohort Exit

Cohort exit is defined as the first of the following dates:

- end of study period (December 31, 2016)
- end of continuous insurance period
- death of any cause as defined in **Fehler! Verweisquelle konnte nicht gefunden werden.**
- switch to another study drug
- cohort entry + 12 months

Cohort re-entry is not possible.

Baseline Period

The start date of the baseline period is defined as the cohort entry – 12 months.

2.3.2 Covariate Data

The following variables are considered as covariates:

- Age (continuous)
- Sex (nominal)
- GKZ5 (nominal)

2.3.3 Dispensation Data

The dispensation data contains all prescriptions in the study period according to the criteria in 2.2.3.

In order to conduct an SCCS analysis, the observation time of each individual is subdivided into a control and a risk period. The control period is defined as the period from the beginning of the study period until cohort entry. The risk period is defined as the period from the cohort entry until the cohort exit.

2.3.4 Diagnosis Data

The diagnosis data contains all diagnoses (ICD-10-GM) in the study period according to the criteria in 2.2.2. The ICD-10-GM codes are aggregated according to

- three digits
- four digits
- five digits
- hierarchical morbidity groups (G:\pv-monitor\Ablage\40_WorkingPackages\Task2\Task21\ICD_Zuordnung_und_Gueltigkeitskriterien_AJ2018_FESTLEGUNG.xlsx)

3 Statistical Analyses

3.1 Main Analyses

An SCCS analysis will be used to estimate relative incidences for an ADE in DOACI dispensation periods vs. control periods. The SCCS method allows measuring the association between an exposure and an event by estimating relative incidences which compare the rate of events occurring in risk periods with the rate in control periods (Whitaker et al. 2009). Only cases, i.e., patients with DOACI dispensation are considered in the analysis. The method is based on within-person comparisons and as such a self-controlled approach, hence time-invariant confounders are controlled for (Whitaker et al. 2009). The time-varying confounder age will be taken into account in the random forest and the LASSO. The SCCS method is described in more detail in Whitaker et al. (2006) and Whitaker et al. (2009).

Statistical Methods

The following statistical methods will be applied:

- Bayesian Confidence Propagation Neural Network (BCPNN). More specifically, we will compute the lower limit of the 90% credible interval of the information component as described in the paper by Norén et al. (2006).
- Longitudinal Gamma-Poisson shrinker (LGPS). We use, specifically, the lower limit of the 95% credible interval (Schuemie 2011).
- Test of the Poisson mean (DuMouchel 1999).
- Proportional relative risk (Evans et al. 2001).
- Random forests with 500 trees and the impurity and corrected impurity importance measures (Wright et al. 2017; Nembrini et al. 2018).
- LASSO (Caster et al. 2008).

These statistical methods were implemented in R and are available through the packages `pvm`, `gepardsim` and `ranger`. They can be found at <https://github.com/bips-hb/pvm>, <https://github.com/bips-hb/gepardsim> and <https://cran.r-project.org/package=ranger>. We will use these packages as they were available on Github at the 1st of March, 2019 or the latest CRAN release. For each method we use the default parameter settings as given in the R functions in these packages, unless stated otherwise.

The following covariates are included in the random forest and the LASSO

- Age (continuous)
- Sex (nominal)
- GKZ5 (nominal)
- Diagnoses (dichotomous variable for every ICD in the study period, aggregated as described in 2.3.5)
- Dispensations (dichotomous variable for every ATC in the study period)

Signal Detection Cycles

The analysis will be conducted for each of the four DOACs and for each group of ICD-10-GM codes as described in 2.3.

Measuring of Run Time

The run time for each statistical method and for each ICD-10-GM code grouping is measured by using the R package *microbenchmark* (<https://cran.r-project.org/package=microbenchmark>, current version 1.4-6).

Presentation of Results

For each DOACI, a ranking of ADEs is presented stratified by ICD-10-GM code grouping (3-/ 4-/ 5-digit grouping) and statistical method.

3.2 Comparison of Methods

The performance of the statistical methods is assessed by comparing the detected signals for each DOAC with the known ADEs as listed in the “Fachinformation” (G:\pv-monitor\Ablage\40_WorkingPackages\Task2\Task21\ Liste_ICDs_Fachinfo.xlsx).

The usability of the statistical methods is assessed by comparing their individual run times.

4 References

- Caster O, Madigan D, Norén GN, Bate A. Large-scale regression-based pattern discovery in international adverse drug reaction surveillance. *Proc KDD Work Min Med Data KDD Cup 2008*;24-27
- DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999;53:177-190
- Evans S, Waller P, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001;10:483-486
- Nembrini S, König IR, Wright MN. The revival of the Gini Importance? *Bioinformatics* 2018;34:3711–3718
- Norén G, Bate A, Orre R, Edwards I. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med* 2006;25:3740-3757
- Schuemie M. Methods for drug safety signal detection in longitudinal observational database: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf* 2011;20:292-299
- Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768-1797
- Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res* 2009;18:7-26
- Wright MN, Ziegler A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 2017;77:1-17

Anlage 5: Ergebnisse der Vorhersagestudie zu Functional Targets¹

Es hatten in den zwei Studienjahren insgesamt 64.720 Versicherte eine GB. Fälle hatten im Durchschnitt 9,3 (+/- SD: 7,0) unterschiedlich verschriebene Arzneimittel und 39,0 (+/- 24,1) unterschiedliche Diagnosen. Bei den Kontrollen war dies mit 5,5 (+/- 5,2) Arzneimitteln und 21,6 (+/- 18,3) Diagnosen wesentlich seltener der Fall. Insgesamt wurden 8.577 Kovariaten betrachtet, die sich in 260 FTs gruppieren ließen. Die durchschnittliche Anzahl von Kovariaten in einer FT war 1.649 (+/- 1.243) mit einer Spannweite von einer Kovariate bis 5.165 Kovariaten pro FT. Insgesamt konnten 18% der Kovariaten keinem FT zugeordnet werden. Der Prozentsatz der Fälle, die im Studienzeitraum ein DOAK verschrieben bekamen, betrug 4,5% und bei den Kontrollen 12,3%. Die Performanz der statistischen Methoden bei der Vorhersage einer GB ist in Tabelle 9 dargestellt.

Tabelle 9: Sensitivität, Spezifität und False Discovery Rate (FDR) der Vorhersage.

	Sensitivität	Spezifität	FDR
<i>LASSO</i>	0,688	0,973	0,133
<i>Naiver LASSO</i>	0,531	0,959	0,235
<i>Random Forest</i>	0,723	0,965	0,162
<i>Block Forest</i>	0,677	0,972	0,142
<i>ARTP</i>	0,891	0,794	0,478

Sensitivität = TP / P, Spezifität = TN / N, FDR = FP / (FP + TP), P = Anzahl richtiger Vorhersagen, N = Anzahl falscher Vorhersagen, TP = Anzahl richtiger Vorhersagen einer GB, TN = Anzahl richtiger Vorhersagen keiner GB, FP = falsche Vorhersagen einer GB; ARTP: Adaptive Runc Truncated Product; LASSO: Least absolute shrinkage and selection operator.

Die Studie konnte zeigen, dass die Vorhersage von UAWs basierend auf FTs möglich ist. Der erwartete Powergewinn konnte jedoch nicht bestätigt werden. Die Studie hat aber gezeigt, dass eine naive Gruppierung (Summenscore für jedes FT gebildet aus all seinen zugehörigen Kovariaten) die Vorhersage verschlechtert, während eine intelligente Methode wie der Block Forest hinsichtlich Sensitivität, Spezifität und FDR so gut ist wie Lasso und Random Forest. Das ARTP zeigte von allen Methoden die höchste Sensitivität: Es hat bei 89% der Patienten, die das Ereignis GB bekommen haben, basierend auf ihren persönlichen GKV-Daten dieses Ereignis korrekt vorhergesagt. Allerdings ist die Positivrate der Methode so hoch, dass von allen vorhergesagten Positivfällen in etwa 48% falsch waren (FDR).

Darüber hinaus bieten Block Forest und das ARTP bei der Interpretation der Ergebnisse den Vorteil beurteilen zu können, welcher FT bei der Vorhersage besonders wichtig war. Beim Block Forest kann dies über die Block Importance erfolgen, die eine Rangordnung der Wichtigkeit einzelner FTs ausgibt. Beim ARTP erhält jedes FT einen p-Wert basierend auf einem Permutationstest, der auf Assoziation zwischen FT und Ereignis testet. Individuelle Risikoprofile können mit Hilfe des erstellten Vorhersagemodells erzeugt werden. Mit diesen können, basierend auf spezifischen GKV-Leistungsdaten, Vorhersagewahrscheinlichkeiten für das spezielle Ereignis geschätzt werden. Verschreibungen und Diagnosen, die sich in den als wichtig eingestuften FTs befinden, erhöhen das Risiko für das Ereignis.

¹ Ergebnisse werden zur Veröffentlichung bei Drug Safety eingereicht

Eine bekannte Limitation von GKV-Daten ist, dass sie wenig Informationen über mögliche Störvariablen oder auch andere relevante Prädiktoren beinhalten. Beispielsweise wären für unsere Analyse hier die Ernährung, Rauchen, psychischer Stress, undiagnostizierte Infektionen mit *Helicobacter pylori* und genetische Faktoren zu nennen.

Insgesamt konnte erfolgreich gezeigt werden, dass GKV-Daten genutzt werden können, um Risikoprofile zu erstellen und so eventuell vulnerable Gruppen von Patienten und Patientinnen identifizieren und besser schützen zu können.

Anlage 6: Ergebnisse der Signalgenerierung

Die Signalgenerierung erfolgte für jedes DOAK getrennt und wurde nach ICD-10-GM Aggregationsniveau und statistischer Methode ausgegeben. Zusätzlich wurden die Verfahren auch für die Hauptmorbiditätsgruppen (HMG) gemäß morbiditätsorientiertem Risikostrukturausgleich untersucht. Pro Aggregationsniveau hat die Verarbeitungskette „Deep Learning“ unterschiedlich viele Signale gefunden, wie in Tabelle 10 zu sehen ist.

Tabelle 10: Untersuchte Ereignisse und Signale entsprechend dem Aggregationsniveau der ICDs (Verarbeitungskette „Deep Learning“).

ICD-10 Aggregation	Untersuchte Ereignisse	davon Signale
	N _{ICD}	N (%)
3-Steller	563	60 (11%)
4-Steller	1.345	68 (5%)
5-Steller	448	23 (5%)
HMG	157	34 (22%)

HMG: Hauptmorbiditätsgruppen gemäß morbiditätsorientiertem Risikostrukturausgleich; ICD: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme.

In der Verarbeitungskette „Methoden-Mix“ wendeten wir die folgenden neun statistischen Methoden für die Signalgenerierung an:

- Bayesian Confidence Propagation Neural Network, Information Component Alternative (ICA)
- Bayesian Confidence Propagation Neural Network, Information Component Original (ICO)
- Longitudinal Gamma-Poisson Shrinker (LGPS)
- Test of the Poisson Mean (POIS)
- Number of Reports (NREP)
- Proportional Relative Risk (PRR)
- Random Forest Impurity Importance (RF_IMPU)
- Random Forest Corrected Impurity Importance (RF_CORR)
- Least Absolute Shrinkage and Selection Operator (LASSO)

Als Ergebnis werden Listen mit Signalen und Rangfolgen erzeugt. Die erzeugten Signalarangfolgen der statistischen Methoden können sich sehr stark unterscheiden. Ein Vergleich der Rangfolgen zeigte beispielsweise, dass die ICA-Methode und das Auszählen der Reports (NREP) sehr ähnliche Signalarangfolgen erzeugten; die Ergebnisse der multivariaten Methoden (Random Forest, LASSO) jedoch eher zu anderen Signalarangfolgen führten als die klassischen univariaten Verfahren (siehe beispielhaft Abbildung 6 für die Signallisten von Dabigatran). Eine Abstraktion der Ränge ist für die Beurteilung der biologisch-pharmakologischen Relevanz daher vorteilhaft. Dies ist in

Tabelle 11 exemplarisch für die Ergebnisse des Signals Sepsis dargestellt, die zu einer weiteren Überprüfung in der Validierungsstudie geführt haben.

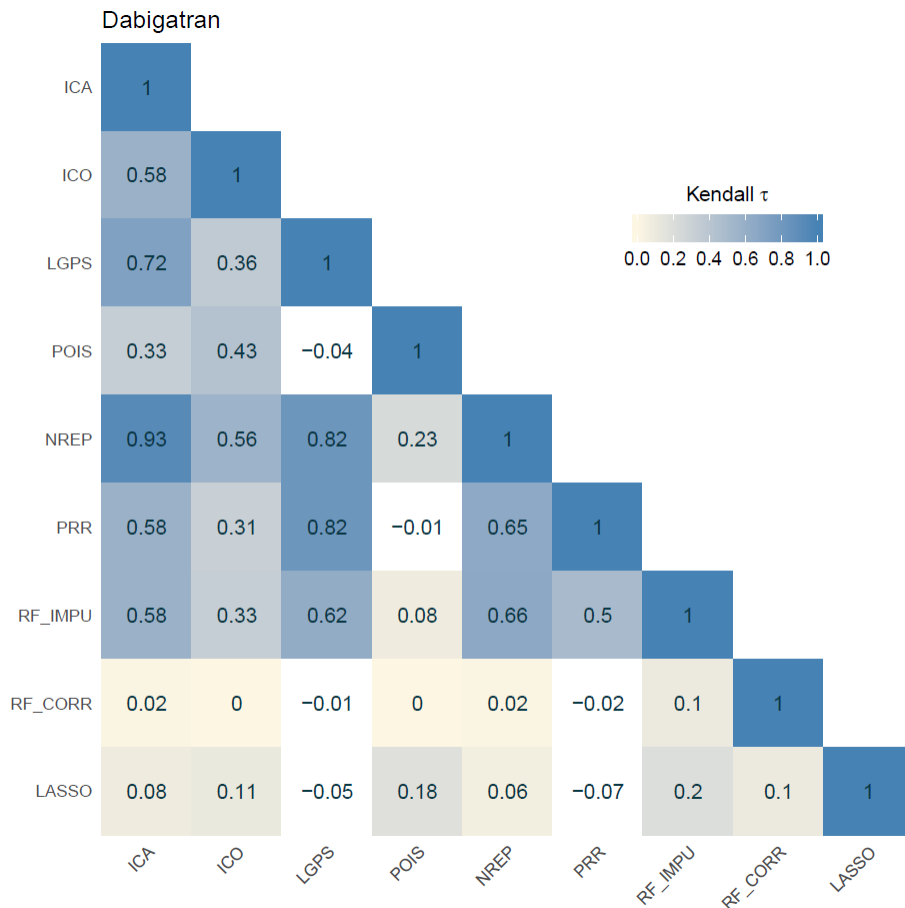


Abbildung 6: Heatmap-Analyse der Signalstärken nach statistischer Methode basierend auf Kendall'schen Rangkorrelationskoeffizienten.

ICA: Bayesian Confidence Propagation Neural Network, Information Component Alternative;
 ICO: Bayesian Confidence Propagation Neural Network, Information Component Original;
 LASSO: Least absolute shrinkage and selection operator; LGPS: Longitudinal Gamma-Poisson
 Shrinker; NREP: Number of Reports; PRR: Proportional Relative Risk; POIS: Test of the Poisson
 Mean; RF_CORR: Random Forest Corrected Impurity Importance; RF_IMPU: Random Forest
 Impurity Importance

Die Ergebnisse aus beiden Verarbeitungsketten sind große CSV-Dateien, die hier nicht dargestellt werden können. Sie können bei den Autoren angefragt werden.

Tabelle 11: Ergebnisse der Signaldetektion, die auf Sepsis als mögliche UAW hinweisen (Verarbeitungskette „Methoden-Mix“). Im Studienzeitraum sind zu sieben Sepsis Codes (A207, A227, A267, A392, A393, A394, A427) keine Einträge vorhanden.

ICD-10 GM	Sepsis	ICA	LASSO	LGPS	NREP	PIOS	PRR	RF
A021	Salmonellensepsis	+	+	-	-	+	-	+
A327	Listeriensepsis	-	-	-	-	-	-	-
A400	Sepsis durch Streptokokken, Gruppe A	+	+	+	+	+	+	+
A401	Sepsis durch Streptokokken, Gruppe B	+	+	-	-	+	-	+
A402	Sepsis durch Streptokokken, Gruppe D	+	+	+	+	+	+	+
A403	Sepsis durch Streptococcus pneumoniae	-	-	-	-	-	-	-
A408	Sonstige Sepsis durch Streptokokken	-	-	-	-	-	-	-
A409	Sepsis durch Streptokokken, nicht näher bezeichnet	+	+	+	+	+	+	+
A410	Sepsis durch Staphylococcus aureus	+	-	-	-	+	-	+
A411	Sepsis durch sonstige näher bezeichnete Staphylokokken	+	+	+	+	+	+	+
A412	Sepsis durch nicht näher bezeichnete Staphylokokken	+	-	+	+	+	+	+
A413	Sepsis durch Haemophilus influenzae	-	-	-	-	-	-	-
A414	Sepsis durch Anaerobier	+	+	-	-	+	-	+
A415	Sepsis durch sonstige gramnegative Erreger	+	+	+	+	+	+	+
A418	Sonstige näher bezeichnete Sepsis	+	+	+	+	+	+	+
A419	Sepsis, nicht näher bezeichnet	-	-	-	-	-	-	-
B377	Candida-Sepsis	-	-	-	-	-	-	-

+ positive Assoziation zwischen der Diagnose und einem DOAK

- keine oder negative Assoziation zwischen der Diagnose und einem DOAK

DOAK: direkte orale Antikoagulanzen; ICA: Bayesian Confidence Propagation Neural Network, Information Component Alternative; LASSO: Least absolute shrinkage and selection operator; LGPS: Longitudinal Gamma-Poisson Shrinker; NREP: Number of Reports; PRR: Proportional Relative Risk; PIOS: Test of the Poisson Mean; RF: Random Forest

Anlage 7: Ergebnisse der Validierungsstudie²

Tabelle 12 zeigt die Ergebnisse der Validierungsstudie für die Signale, die in der Machbarkeitsstudie ausgewählt wurden, sowie die Ergebnisse der bekannten Nebenwirkungen gastrointestinale und intrakranielle Blutungen als Positivkontrollen.

² Es ist beabsichtigt, die Ergebnisse bei Pharmacoepidemiology and Drug Safety einzureichen.

Tabelle 12: Unadjustierte und adjustierte Odds Ratios mit 95%-Konfidenzintervall (CI).

Endpunkt	Fälle	Kontrollen	Exposition am Indextag	in Fällen N (%)	in Kontrollen N (%)	Unadjustiertes Odds Ratio (95% CI)	Adjustiertes Odds Ratio (95% CI)	p-Wert
Gastrointestinale Blutung (Positivkontrolle)	6.705	67.019	Einzelnutzung eines DOAKs ¹	2,551 (38.0%)	24,930 (37.2%)	1.06 (0.99; 1.13)	1.17 (1.10; 1.26)	<.001
			Einzelnutzung von Apixaban ¹	467 (7.0%)	6,834 (10.2%)	0.71 (0.63; 0.79)	0.77 (0.69; 0.85)	<.001
			Einzelnutzung von Edoxaban ¹	85 (1.3%)	802 (1.2%)	1.10 (0.87; 1.38)	1.24 (0.98; 1.57)	0.072
			Einzelnutzung von Dabigatran ¹	364 (5.4%)	3,477 (5.2%)	1.08 (0.96; 1.22)	1.21 (1.07; 1.36)	0.003
			Einzelnutzung von Rivaroxaban ¹	1,635 (24.4%)	13,817 (20.6%)	1.22 (1.14; 1.31)	1.37 (1.27; 1.47)	<.001
			Gebrauch mehrerer SMs ²	113 (1.7%)	379 (0.6%)	3.12 (2.52; 3.88)	2.72 (2.17; 3.41)	<.001
			Jüngste Nutzung einer SM ³	1,837 (27.4%)	17,433 (26.0%)	1.08 (1.01; 1.17)	1.09 (1.01; 1.17)	0.026
			Bisherige Nutzung einer SM ⁴	513 (7.7%)	6,778 (10.1%)	0.76 (0.68; 0.85)	0.68 (0.61; 0.76)	<.001
Intrakranielle Blutung (Positivkontrolle)	2.974	29.720	Einzelnutzung eines DOAKs ¹	866 (29.1%)	10,757 (36.2%)	0.67 (0.60; 0.74)	0.68 (0.61; 0.75)	<.001
			Einzelnutzung von Apixaban ¹	232 (7.8%)	2,920 (9.8%)	0.66 (0.56; 0.77)	0.65 (0.56; 0.76)	<.001
			Einzelnutzung von Edoxaban ¹	29 (1.0%)	345 (1.2%)	0.70 (0.48; 1.03)	0.71 (0.48; 1.05)	0.084
			Einzelnutzung von Dabigatran ¹	90 (3.0%)	1,533 (5.2%)	0.48 (0.39; 0.61)	0.50 (0.40; 0.62)	<.001
			Einzelnutzung von Rivaroxaban ¹	515 (17.3%)	5,959 (20.1%)	0.72 (0.64; 0.80)	0.74 (0.66; 0.83)	<.001
			Gebrauch mehrerer SMs ²	22 (0.7%)	165 (0.6%)	1.11 (0.70; 1.74)	1.04 (0.66; 1.64)	0.864
			Jüngste Nutzung einer SM ³	921 (31.0%)	8,191 (27.6%)	0.93 (0.84; 1.03)	0.92 (0.83; 1.02)	0.112
			Bisherige Nutzung einer SM ⁴	277 (9.3%)	3,219 (10.8%)	0.70 (0.61; 0.82)	0.66 (0.57; 0.76)	<.001
Akute Leberschädigung	322	3.220	Einzelnutzung eines DOAKs ¹	86 (26.7%)	1,268 (39.4%)	0.45 (0.34; 0.61)	0.45 (0.33; 0.61)	<.001
			Einzelnutzung von Apixaban ¹	20 (6.2%)	351 (10.9%)	0.38 (0.23; 0.63)	0.37 (0.22; 0.61)	<.001
			Einzelnutzung von Edoxaban ¹	-	-	-	-	
			Einzelnutzung von Dabigatran ¹	12 (3.7%)	179 (5.6%)	0.45 (0.24; 0.83)	0.44 (0.23; 0.82)	0.010
			Einzelnutzung von Rivaroxaban ¹	52 (16.1%)	688 (21.4%)	0.50 (0.35; 0.71)	0.50 (0.35; 0.72)	<.001
			Gebrauch mehrerer SMs ²	9 (2.8%)	19 (0.6%)	3.22 (1.41; 7.39)	2.88 (1.20; 6.92)	0.018
			Jüngste Nutzung einer SM ³	81 (25.2%)	840 (26.1%)	0.63 (0.46; 0.87)	0.67 (0.48; 0.92)	0.014
			Bisherige Nutzung einer SM ⁴	33 (10.2%)	322 (10.0%)	0.67 (0.43; 1.04)	0.67 (0.43; 1.06)	0.084

Endpunkt	Fälle	Kontrollen	Exposition am Indextag	in Fällen N (%)	in Kontrollen N (%)	Unadjustiertes Odds Ratio (95% CI)	Adjustiertes Odds Ratio (95% CI)	p-Wert
Akute Nierenschädigung	3.221	32.195	Einzelnutzung eines DOAKs ¹	1,000 (31.0%)	11,835 (36.8%)	0.89 (0.80; 0.98)	0.97 (0.88; 1.08)	0.629
			Einzelnutzung von Apixaban ¹	351 (10.9%)	3,213 (10.0%)	1.14 (1.00; 1.31)	1.14 (0.98; 1.31)	0.083
			Einzelnutzung von Edoxaban ¹	30 (0.9%)	373 (1.2%)	0.83 (0.57; 1.22)	0.93 (0.62; 1.39)	0.724
			Einzelnutzung von Dabigatran ¹	102 (3.2%)	1,644 (5.1%)	0.65 (0.53; 0.81)	0.84 (0.67; 1.05)	0.117
			Einzelnutzung von Rivaroxaban ¹	517 (16.1%)	6,605 (20.5%)	0.82 (0.73; 0.92)	0.92 (0.81; 1.04)	0.196
			Gebrauch mehrerer SMs ²	41 (1.3%)	170 (0.5%)	2.49 (1.76; 3.53)	2.33 (1.58; 3.43)	<.001
			Jüngste Nutzung einer SM ³	990 (30.7%)	8,344 (25.9%)	1.27 (1.14; 1.40)	1.39 (1.25; 1.56)	<.001
			Bisherige Nutzung einer SM ⁴	406 (12.6%)	3,672 (11.4%)	1.18 (1.04; 1.35)	1.28 (1.11; 1.48)	<.001
Zysitis	29.698	296.532	Einzelnutzung eines DOAKs ¹	11,356 (38.2%)	112,306 (37.9%)	1.05 (1.02; 1.09)	1.05 (1.02; 1.09)	0.002
			Einzelnutzung von Apixaban ¹	3,083 (10.4%)	30,041 (10.1%)	1.08 (1.03; 1.13)	1.05 (1.00; 1.10)	0.037
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	1,481 (5.0%)	15,659 (5.3%)	0.99 (0.94; 1.05)	0.99 (0.93; 1.04)	0.616
			Einzelnutzung von Rivaroxaban ¹	6,476 (21.8%)	62,715 (21.1%)	1.09 (1.05; 1.12)	1.10 (1.06; 1.14)	<.001
			Gebrauch mehrerer SMs ²	261 (0.9%)	1,683 (0.6%)	1.62 (1.42; 1.85)	1.57 (1.37; 1.79)	<.001
			Jüngste Nutzung einer SM ³	7,795 (26.2%)	76,128 (25.7%)	1.07 (1.03; 1.11)	1.08 (1.04; 1.12)	<.001
			Bisherige Nutzung einer SM ⁴	3,114 (10.5%)	31,714 (10.7%)	1.02 (0.97; 1.07)	1.04 (1.00; 1.09)	0.069
Akute Zysitis	5.053	50.424	Einzelnutzung eines DOAKs ¹	1,976 (39.1%)	18,870 (37.4%)	1.08 (1.00; 1.16)	1.09 (1.00; 1.17)	0.038
			Einzelnutzung von Apixaban ¹	510 (10.1%)	5,043 (10.0%)	1.04 (0.93; 1.16)	1.03 (0.92; 1.15)	0.646
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	284 (5.6%)	2,641 (5.2%)	1.10 (0.96; 1.26)	1.10 (0.96; 1.27)	0.154
			Einzelnutzung von Rivaroxaban ¹	1,123 (22.2%)	10,590 (21.0%)	1.09 (1.00; 1.19)	1.11 (1.02; 1.21)	0.021
			Gebrauch mehrerer SMs ²	37 (0.7%)	253 (0.5%)	1.52 (1.07; 2.15)	1.50 (1.05; 2.13)	0.026
			Jüngste Nutzung einer SM ³	1,321 (26.1%)	13,438 (26.6%)	1.01 (0.93; 1.10)	1.02 (0.93; 1.11)	0.688
			Bisherige Nutzung einer SM ⁴	533 (10.5%)	5,643 (11.2%)	0.96 (0.86; 1.07)	0.99 (0.89; 1.11)	0.882

Endpunkt	Fälle	Kontrollen	Exposition am Indextag	in Fällen N (%)	in Kontrollen N (%)	Unadjustiertes Odds Ratio (95% CI)	Adjustiertes Odds Ratio (95% CI)	p-Wert
Akute und sonstige Zystitis	10.551	105.315	Einzelnutzung eines DOAKs ¹	4,035 (38.2%)	39,461 (37.5%)	1.04 (0.99; 1.09)	1.04 (0.98; 1.09)	0.202
			Einzelnutzung von Apixaban ¹	1,044 (9.9%)	10,595 (10.1%)	1.00 (0.93; 1.08)	0.97 (0.90; 1.05)	0.464
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	542 (5.1%)	5,482 (5.2%)	1.00 (0.91; 1.10)	0.99 (0.89; 1.09)	0.780
			Einzelnutzung von Rivaroxaban ¹	2,318 (22.0%)	22,137 (21.0%)	1.06 (1.00; 1.13)	1.07 (1.01; 1.13)	0.029
			Gebrauch mehrerer SMs ²	86 (0.8%)	609 (0.6%)	1.44 (1.15; 1.81)	1.37 (1.09; 1.73)	0.007
			Jüngste Nutzung einer SM ³	2,783 (26.4%)	27,438 (26.1%)	1.03 (0.97; 1.09)	1.03 (0.98; 1.10)	0.261
			Bisherige Nutzung einer SM ⁴	1,108 (10.5%)	12,010 (11.4%)	0.92 (0.86; 1.00)	0.95 (0.88; 1.02)	0.169
Epilepsie	3.070	30.676	Einzelnutzung eines DOAKs ¹	1,265 (41.2%)	11,380 (37.1%)	1.54 (1.38; 1.70)	1.39 (1.24; 1.55)	<.001
			Einzelnutzung von Apixaban ¹	433 (14.1%)	3,034 (9.9%)	1.97 (1.72; 2.25)	1.68 (1.46; 1.93)	<.001
			Einzelnutzung von Edoxaban ¹	27 (0.9%)	381 (1.2%)	0.98 (0.66; 1.46)	1.00 (0.66; 1.51)	0.991
			Einzelnutzung von Dabigatran ¹	188 (6.1%)	1,603 (5.2%)	1.62 (1.36; 1.92)	1.38 (1.15; 1.66)	<.001
			Einzelnutzung von Rivaroxaban ¹	617 (20.1%)	6,362 (20.7%)	1.34 (1.19; 1.51)	1.26 (1.11; 1.43)	<.001
			Gebrauch mehrerer SMs ²	30 (1.0%)	201 (0.7%)	2.04 (1.38; 3.03)	1.76 (1.16; 2.68)	0.008
			Jüngste Nutzung einer SM ³	856 (27.9%)	7,980 (26.0%)	1.50 (1.34; 1.68)	1.47 (1.30; 1.66)	<.001
			Bisherige Nutzung einer SM ⁴	356 (11.6%)	3,411 (11.1%)	1.46 (1.26; 1.69)	1.47 (1.26; 1.70)	<.001
Symptomatische Fokale Epilepsie	1.411	14.098	Einzelnutzung eines DOAKs ¹	645 (45.7%)	5,133 (36.4%)	1.97 (1.68; 2.30)	1.74 (1.47; 2.05)	<.001
			Einzelnutzung von Apixaban ¹	228 (16.2%)	1,348 (9.6%)	2.65 (2.18; 3.23)	2.09 (1.69; 2.58)	<.001
			Einzelnutzung von Edoxaban ¹	15 (1.1%)	166 (1.2%)	1.42 (0.82; 2.46)	1.57 (0.89; 2.78)	0.119
			Einzelnutzung von Dabigatran ¹	104 (7.4%)	731 (5.2%)	2.20 (1.72; 2.81)	1.84 (1.42; 2.40)	<.001
			Einzelnutzung von Rivaroxaban ¹	298 (21.1%)	2,888 (20.5%)	1.61 (1.35; 1.93)	1.52 (1.26; 1.85)	<.001
			Gebrauch mehrerer SMs ²	15 (1.1%)	77 (0.5%)	3.06 (1.72; 5.44)	2.59 (1.37; 4.89)	0.003
			Jüngste Nutzung einer SM ³	364 (25.8%)	3,758 (26.7%)	1.52 (1.27; 1.81)	1.56 (1.29; 1.88)	<.001
			Bisherige Nutzung einer SM ⁴	158 (11.2%)	1,585 (11.2%)	1.56 (1.25; 1.94)	1.60 (1.27; 2.02)	<.001

Endpunkt	Fälle	Kontrollen	Exposition am Indextag	in Fällen N (%)	in Kontrollen N (%)	Unadjustiertes Odds Ratio (95% CI)	Adjustiertes Odds Ratio (95% CI)	p-Wert
Idopathische Fokale Epilepsie	2.663	26.606	Einzelnutzung eines DOAKs ¹	1,069 (40.1%)	9,736 (36.6%)	1.52 (1.35; 1.70)	1.38 (1.22; 1.55)	<.001
			Einzelnutzung von Apixaban ¹	361 (13.6%)	2,537 (9.5%)	1.97 (1.70; 2.27)	1.67 (1.43; 1.94)	<.001
			Einzelnutzung von Edoxaban ¹	23 (0.9%)	302 (1.1%)	1.06 (0.68; 1.63)	1.11 (0.71; 1.73)	0.656
			Einzelnutzung von Dabigatran ¹	155 (5.8%)	1,344 (5.1%)	1.59 (1.31; 1.93)	1.38 (1.13; 1.68)	0.002
			Einzelnutzung von Rivaroxaban ¹	530 (19.9%)	5,553 (20.9%)	1.32 (1.16; 1.50)	1.25 (1.09; 1.43)	0.001
			Gebrauch mehrerer SMs ²	23 (0.9%)	150 (0.6%)	2.10 (1.34; 3.29)	2.00 (1.24; 3.22)	0.005
			Jüngste Nutzung einer SM ³	773 (29.0%)	6,955 (26.1%)	1.56 (1.38; 1.76)	1.50 (1.32; 1.70)	<.001
			Bisherige Nutzung einer SM ⁴	315 (11.8%)	3,153 (11.9%)	1.39 (1.19; 1.62)	1.38 (1.18; 1.62)	<.001
Glaukom	5.700	56.967	Einzelnutzung eines DOAKs ¹	2,127 (37.3%)	21,263 (37.3%)	0.95 (0.88 - 1.02)	0.95 (0.88 - 1.02)	0.143
			Einzelnutzung von Apixaban ¹	557 (9.8%)	5,497 (9.6%)	0.96 (0.87 - 1.06)	0.95 (0.85 - 1.05)	0.328
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	304 (5.3%)	3,042 (5.3%)	0.94 (0.83 - 1.07)	0.95 (0.83 - 1.09)	0.471
			Einzelnutzung von Rivaroxaban ¹	1,192 (20.9%)	12,016 (21.1%)	0.94 (0.87 - 1.02)	0.95 (0.88 - 1.03)	0.221
			Gebrauch mehrerer SMs ²	30 (0.5%)	314 (0.6%)	0.91 (0.62 - 1.33)	0.89 (0.60 - 1.32)	0.561
			Jüngste Nutzung einer SM ³	1,462 (25.6%)	14,751 (25.9%)	0.93 (0.86 - 1.01)	0.93 (0.86 - 1.01)	0.072
			Bisherige Nutzung einer SM ⁴	631 (11.1%)	6,909 (12.1%)	0.85 (0.77 - 0.94)	0.85 (0.77 - 0.95)	0.003
Primäres Weitwinkelglaukom	1.624	16.232	Einzelnutzung eines DOAKs ¹	602 (37.1%)	5,962 (36.7%)	0.96 (0.84 - 1.10)	0.94 (0.81 - 1.08)	0.355
			Einzelnutzung von Apixaban ¹	145 (8.9%)	1,521 (9.4%)	0.89 (0.73 - 1.09)	0.84 (0.68 - 1.04)	0.109
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	82 (5.0%)	847 (5.2%)	0.90 (0.70 - 1.15)	0.89 (0.68 - 1.16)	0.381
			Einzelnutzung von Rivaroxaban ¹	346 (21.3%)	3,395 (20.9%)	0.95 (0.82 - 1.10)	0.95 (0.81 - 1.11)	0.490
			Gebrauch mehrerer SMs ²	8 (0.5%)	74 (0.5%)	1.04 (0.50 - 2.17)	1.02 (0.47 - 2.24)	0.955
			Jüngste Nutzung einer SM ³	427 (26.3%)	4,238 (26.1%)	0.96 (0.83 - 1.11)	0.93 (0.80 - 1.08)	0.346
			Bisherige Nutzung einer SM ⁴	176 (10.8%)	2,049 (12.6%)	0.81 (0.67 - 0.97)	0.78 (0.64 - 0.95)	0.015
Primäres	239	2.390	Einzelnutzung eines DOAKs ¹	92 (38.5%)	865 (36.2%)	1.03 (0.73 - 1.45)	1.08 (0.76 - 1.55)	0.661

Endpoint	Fälle	Kontrollen	Exposition am Indextag	in Fällen N (%)	in Kontrollen N (%)	Unadjustiertes Odds Ratio (95% CI)	Adjustiertes Odds Ratio (95% CI)	p-Wert
Engwinkelglaukom			Einzelnutzung von Apixaban ¹	26 (10.9%)	214 (9.0%)	1.16 (0.71 - 1.89)	1.23 (0.74 - 2.05)	0.430
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	12 (5.0%)	141 (5.9%)	0.81 (0.42 - 1.54)	0.94 (0.49 - 1.82)	0.853
			Einzelnutzung von Rivaroxaban ¹	51 (21.3%)	490 (20.5%)	0.99 (0.67 - 1.46)	1.00 (0.66 - 1.49)	0.981
			Gebrauch mehrerer SMs ²	1 (0.4%)	15 (0.6%)	0.64 (0.08 - 4.97)	0.57 (0.07 - 4.60)	0.599
			Jüngste Nutzung einer SM ³	57 (23.8%)	640 (26.8%)	0.85 (0.58 - 1.26)	0.89 (0.60 - 1.34)	0.581
			Bisherige Nutzung einer SM ⁴	29 (12.1%)	291 (12.2%)	0.96 (0.59 - 1.55)	0.95 (0.57 - 1.58)	0.840
Sepsis	3.504	35.022	Einzelnutzung eines DOAKs ¹	1,230 (35.1%)	12,907 (36.9%)	1.04 (0.94 - 1.14)	1.11 (1.01 - 1.22)	0.039
			Einzelnutzung von Apixaban ¹	358 (10.2%)	3,455 (9.9%)	1.13 (0.99 - 1.28)	1.11 (0.96 - 1.27)	0.149
			Einzelnutzung von Edoxaban ¹	-	-	-	-	-
			Einzelnutzung von Dabigatran ¹	165 (4.7%)	1,818 (5.2%)	0.99 (0.83 - 1.18)	1.13 (0.94 - 1.35)	0.197
			Einzelnutzung von Rivaroxaban ¹	670 (19.1%)	7,231 (20.6%)	1.01 (0.91 - 1.12)	1.10 (0.98 - 1.23)	0.100
			Gebrauch mehrerer SMs ²	33 (0.9%)	182 (0.5%)	1.97 (1.35 - 2.87)	1.65 (1.10 - 2.46)	0.015
			Jüngste Nutzung einer SM ³	1,016 (29.0%)	9,258 (26.4%)	1.21 (1.09 - 1.33)	1.25 (1.12 - 1.38)	<.001
Bisherige Nutzung einer SM ⁴	419 (12.0%)	3,931 (11.2%)	1.17 (1.03 - 1.34)	1.21 (1.06 - 1.38)	0.006			

SM: Studienmedikation (darunter fallen alle DOAKs sowie Phenprocoumon)

Referenzgruppe: Versicherte/r war am Indextag ausschließlich mit Phenprocoumon exponiert.

¹ Versicherte/r war am Indextag mit irgendeinem DOAK exponiert.

² Versicherte/r war am Indextag mit mehr als einer SM exponiert.

³ Versicherte/r war im Zeitraum von 180 Tagen vor dem Indextag mit mehr als einer SM exponiert, aber nicht am Indextag.

⁴ Versicherte/r war bei Kohorteneintritt mit einem DOAK exponiert, aber war am Indextag weder aktuelle/r noch jüngste/r Nutzer/in

Anlage 8: Veröffentlichte Publikationen

Auf den nachfolgenden Seiten sind diese Publikationen zu finden:

- 1) Foraita R, Dijkstra L, Falkenberg F, Garling M, Linder R, Pflock R, Rizkallah MR, Schwaninger M, Wright MW, Pigeot I (2018). Aufdeckung von Arzneimittelrisiken nach der Zulassung. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 61: 1075-1081, URL: <https://doi.org/10.1007/s00103-018-2786-z>

Angehängt ist die *Green Access Version* der Publikation. Es ist die endgültig akzeptierte Version des Autors. Es kann jedoch Unterschiede zwischen dieser Version und der veröffentlichten Version geben. Es wird empfohlen, die Version des Herausgebers zu konsultieren, wenn Sie daraus zitieren möchten.

- 2) Dijkstra LJ, Garling M, Foraita R, Pigeot I (2020). Adverse event or innocent bystander? A systematic comparison of statistical discovery methods for spontaneous reporting systems. *Pharmacoepidemiology and Drug Safety* 29: 396-403, URL: <https://doi.org/10.1002/pds.4970> (Open Access)
- 3) Klose S, Pflock R, König IR, Linder R, Schwaninger M (2020). Metamizole and the risk of drug-induced agranulocytosis and neutropenia in statutory health insurance data. *Naunyn-Schmiedeberg's Archives of Pharmacology* 393: 681–690, URL: <https://doi.org/10.1007/s00210-019-01774-4>

Abstract:

The non-opioid analgesic metamizole (dipyrone) is used for the treatment of acute and chronic pain and fever. Agranulocytosis is known as a serious adverse drug reaction of metamizole with potentially fatal outcome. However, its frequency is controversially discussed. The aim of our study was to determine the risk of metamizole-associated agranulocytosis and neutropenia using statutory health insurance data. We analyzed data from a large German health insurance fund in the period from 2010 to 2013. Metamizole-exposed subjects were identified and compared to a propensity score-matched control cohort. A total of 630,285 metamizole-treated subjects and 390,830 matched control subjects were included. In the metamizole cohort, ICD codes for agranulocytosis and neutropenia appeared more often than in non-users. The relative risk for drug-induced agranulocytosis and neutropenia (D70.1) was 3.03 (95% confidence interval, 2.49 to 3.69). The risk for developing drug-induced agranulocytosis and neutropenia after metamizole prescription was 1: 1602 (CI 95%, 1:1926 to 1:1371). Our results confirm the risk estimation of previous studies. However, the outcome of our study may be confounded by an association of metamizole treatment and chemotherapy. Therefore, consequences for treatment have to be drawn with care.