

# Bachelorarbeit

Tobias Schreier

Auswertung von Pharmakovigilanzdaten auf Basis des  
FAERS Datensatzes

Tobias Schreier

# Auswertung von Pharmakovigilanzdaten auf Basis des FAERS Datensatzes

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung  
im Studiengang Bachelor of Science Wirtschaftsinformatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr.-Ing. Marina Tropmann-Frick  
Zweitgutachter: Prof. Dr. Ulrike Steffens

Eingereicht am: 22. November 2019

**Tobias Schreier**

**Thema der Arbeit**

Auswertung von Pharmakovigilanzdaten auf Basis des FAERS Datensatzes

**Stichworte**

Pharmakovigilanz, FAERS, Signalgenerierung

**Kurzzusammenfassung**

Diese Arbeit thematisiert die Auswertung von Pharmakovigilanzdaten aus dem Spontanmeldesystem der FDA mit Python. Es wurden Funktionalitäten zum Import und zur Aufbereitung der Daten, sowie zur Signalgenerierung und explorativen Datenanalyse entwickelt. Beispielhaft wurden für Medikamente und Nebenwirkungen aus einem Bericht des BfArM Signale generiert. Weiterhin wurde eine explorative Analyse der Daten durchgeführt. Die Implementierung ermöglicht weitere Auswertungen des Datensatzes und kann als Grundlage für ein neues Analysetool für Pharmakovigilanzdaten dienen.

**Tobias Schreier**

**Title of Thesis**

Analysis of pharmacovigilance data based on the FAERS dataset

**Keywords**

Pharmacovigilance, FAERS, Signal Generation

**Abstract**

This thesis addresses the analysis of pharmacovigilance data from the Adverse Event Reporting System of the FDA with Python. Features to import and process the data, as well as for signal generation and exploratory data analysis have been developed. Exemplary, signals have been generated for drugs and adverse drug reactions from a report of BfArM. The implementation allows for further evaluation of the dataset and may be the basis of a new analysis tool for pharmacovigilance data.

# Inhaltsverzeichnis

Abbildungsverzeichnis	vi
Tabellenverzeichnis	viii
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Ziel der Arbeit . . . . .	3
1.3 Aufbau der Arbeit . . . . .	4
<b>2 Datenlage</b>	<b>5</b>
2.1 Pharmakovigilanz . . . . .	5
2.2 Unerwünschte Ereignisse und Nebenwirkungen . . . . .	6
2.3 U.S. Food and Drug Administration . . . . .	7
2.4 FDA Adverse Event Reporting System . . . . .	7
2.5 Einschränkungen . . . . .	8
<b>3 Datenaufbereitung</b>	<b>10</b>
3.1 Parsen der XML-Dateien . . . . .	10
3.2 Aufbereitung der Daten . . . . .	11
3.3 Weitere Anpassungen . . . . .	17
<b>4 Signalgenerierung</b>	<b>21</b>
4.1 Grundlagen . . . . .	21
4.2 Beispiel . . . . .	28
<b>5 Explorative Datenanalyse</b>	<b>31</b>
5.1 Implementierung . . . . .	31
5.2 Berichte . . . . .	33
5.3 Patienten . . . . .	44
5.4 Unerwünschte Ereignisse . . . . .	52

5.5	Medikamente . . . . .	54
<b>6</b>	<b>Zusammenfassung</b>	<b>61</b>
6.1	Fazit . . . . .	62
6.2	Ausblick . . . . .	63
<b>A</b>	<b>Anhang</b>	<b>68</b>
A.1	Übersetzungen der Nebenwirkungen aus dem BfArM-Bericht in MedDRA PTs . . . . .	68
A.2	ME-, Chi-Quadrat- und PRR-Werte nach Evans et al. [18] für Medika- mente und Nebenwirkungen aus dem BfArM-Bericht . . . . .	69
	<b>Selbstständigkeitserklärung</b>	<b>70</b>

# Abbildungsverzeichnis

3.1	Veränderung der Verteilung der Altersgruppen . . . . .	16
4.1	Nach Evans et al. [18] generierte Signale bei Kombinationen von Medikamenten und Nebenwirkungen aus dem BfArM-Bericht . . . . .	30
5.1	Berichte der Gesamtpopulation pro Quartal . . . . .	34
5.2	Berichte der Teilpopulation pro Quartal . . . . .	34
5.3	Neue und alte Berichte der Gesamtpopulation . . . . .	35
5.4	Neue und alte Berichte der Teilpopulation . . . . .	36
5.5	Zehn häufigste Herkunftsländer der Berichte . . . . .	37
5.6	Herkunftsländer der Berichte der Gesamtpopulation . . . . .	37
5.7	Herkunftsländer der Berichte der Teilpopulation . . . . .	38
5.8	Berichtstypen . . . . .	39
5.9	Ernsthaftigkeit der Berichte . . . . .	39
5.10	Sterblichkeitsrate der Patienten pro Land . . . . .	41
5.11	Sterblichkeitsrate der Patienten pro Geschlecht . . . . .	41
5.12	Dringlichkeit der Berichte . . . . .	43
5.13	Qualifikation der Berichterstatter . . . . .	44
5.14	Geschlecht der Patienten . . . . .	45
5.15	Geschlecht der Patienten pro Land . . . . .	46
5.16	Durchschnittsalter der Patienten der Gesamtpopulation . . . . .	48
5.17	Durchschnittsalter der Patienten der Teilpopulation . . . . .	48
5.18	Durchschnittliches Todesalter der Patienten pro Land . . . . .	49
5.19	Altersgruppen der Patienten der Gesamtpopulation . . . . .	50
5.20	Altersgruppen der Patienten der Teilpopulation . . . . .	51
5.21	Gewicht der Patienten . . . . .	52
5.22	Ausgang der unerwünschten Ereignisse . . . . .	53
5.23	Rolle der Medikamente beim Auftreten von unerwünschten Ereignissen . . . . .	57
5.24	Anpassung der Dosen . . . . .	58

5.25	Rückkehr von unerwünschten Ereignissen nach Wiedereinnahme des Medikaments . . . . .	59
5.26	Nachlassen von unerwünschten Ereignissen nach Reduzierung der Dosis oder Absetzung des Medikaments . . . . .	60

# Tabellenverzeichnis

4.1	2 x 2 Kontingenztafel nach Böhm [12]	24
4.2	2 x 2 Kontingenztafel von Ciprofloxacin und Sehnenentzündung	24
5.1	Fünf am häufigsten gemeldete unerwünschte Ereignisse	53
5.2	Fünf am häufigsten gemeldete Medikamente	55
5.3	Fünf am häufigsten gemeldete Indikationen	55
5.4	Fünf am häufigsten gemeldete Indikationen von Ciprofloxacin bei Sehnenentzündung	56
A.1	ME-, Chi-Quadrat- und PRR-Werte nach Evans et al. [18] für Medikamente und Nebenwirkungen aus dem BfArM-Bericht	69

# 1 Einleitung

Medikamente werden aufgrund ihrer heilenden Wirkung zur Behandlung von Krankheiten eingesetzt. Neben diesen heilenden Wirkungen können jedoch auch unerwünschte Wirkungen auftreten, die die Gesundheit des Patienten beeinträchtigen können. Bei diesen unerwünschten Wirkungen handelt es sich um Nebenwirkungen. Die Behandlung mit Medikamenten stellt daher immer eine Abwägung zwischen Nutzen und möglichen Schäden durch Nebenwirkungen dar. Eine besondere Gefahr geht durch bisher unbekannte Nebenwirkungen von Medikamenten aus [33]. Die Identifizierung von zuvor nicht erkannten unerwünschten Arzneimittelwirkungen ist ein Ziel der Pharmakovigilanz [24].

Zur Identifizierung dieser Nebenwirkungen werden Spontanmeldesysteme eingesetzt. Praktisch überall auf der Welt ist die Meldung unerwünschter Arzneimittelwirkungen für Angehörige der Gesundheitsberufe obligatorisch, aber auch andere Personen können Ereignisse an die zuständigen Aufsichtsbehörden melden. Diese Einrichtungen verwenden regelmäßig Algorithmen des Data-Mining, um Daten periodisch zu verarbeiten und mögliche unbekannte Assoziationen zwischen Medikamenten und unerwünschten Ereignissen zu finden. Diese Algorithmen identifizieren Kombinationen von Medikamenten und unerwünschten Ereignissen, die mit einer signifikanten Disproportionalität verglichen mit allen anderen Paaren auftreten [24].

Klinische Pharmakologiekenntnisse sind dann erforderlich, um diese Signale auszuwerten und zu entscheiden, ob weitere Untersuchungen notwendig sind, oder ob bestimmte Verzerrungen die Gültigkeit eines Signals beeinflussen. Sobald ein Signal verifiziert wurde, kann eine Reihe von Regulierungsmaßnahmen in Betracht gezogen werden, die von der Prioritätsstufe abhängig sind, die dem Signal zugewiesen wurde. Dieser Priorisierungsprozess basiert hauptsächlich auf der Bewertung der Auswirkungen auf die öffentliche Gesundheit, der Schwere des unerwünschten Ereignisses und der Stärke der Disproportionalität. Die regulatorischen Maßnahmen können von einer genauen Überwachung des Signals über die Zeit bis hin zum Rückzug von Medikamenten aus dem Markt aufgrund eines negativen Nutzen-Risiko-Verhältnisses reichen [24].

## 1.1 Motivation

Leider bieten nur wenige Spontanmeldesysteme einen öffentlichen Zugang für externe Forscher und Verbraucher [24]. Eines davon ist das FDA Adverse Event Reporting System (FAERS) [28] der U.S. Food and Drug Administration (FDA) [25]. Doch auch die FDA stellt nur die empfangenen Berichte öffentlich zur Verfügung. Einblick in die Verarbeitung der Daten und die Generierung von Signalen wird nicht geboten. Um das FAERS transparenter zu gestalten, hat die FDA das *FDA Public Dashboard* [26] entwickelt. Dieses interaktive, webbasierte Tool erlaubt eine benutzerfreundliche Auswertung der FAERS-Daten. Das *FDA Public Dashboard* ermöglicht jedoch zur Zeit nur die Analyse einiger weniger Attribute und stellt keine Funktionen zur Signalgenerierung zur Verfügung. Da die FAERS-Daten öffentlich zugänglich sind, können jedoch eigene Auswertungen vorgenommen werden.

An der Christian-Albrechts-Universität zu Kiel wurde mit *OpenVigil* [13] ein Analysetool für Pharmakovigilanzdaten entwickelt. Dieses importiert die Daten der FDA, bereitet sie auf und bietet dem Benutzer verschiedene Funktionen zur Auswertung der Daten an. Das System ist öffentlich zugänglich und seine Funktionsweise dank umfangreicher Dokumentation gut nachvollziehbar. Zwar werden dem Benutzer die Ergebnisse der Signalgenerierung angezeigt und die Berichte können eingesehen werden, Statistiken über diese Berichte werden von *OpenVigil* jedoch nicht visualisiert. Die Charakteristiken auf Basis generierter Signale ausgewählter Teilpopulationen lassen sich somit nur schwer erfassen. Die Analyse von Teilpopulationen, bei denen bestimmte unerwünschte Ereignisse aufgetreten sind, stellt jedoch einen wichtigen Schritt nach der Generierung von Signalen dar, um diese zu verstärken. Die Identifizierung möglicherweise besonders häufig von bestimmten unerwünschten Ereignissen betroffener Teilpopulationen ist außerdem ein weiteres wichtiges Ziel der Pharmakovigilanz [24].

In dieser Arbeit soll daher eine Analyse des FAERS-Datensatzes mit Fokus auf der Visualisierung der Daten durchgeführt werden. Aufgrund des umfangreichen Angebots an Bibliotheken zur Visualisierung von Daten bietet sich dafür Python an. Um das Vorgehen zu dokumentieren, können *IPython Jupyter Notebooks* [22] verwendet werden, welche das Kommentieren von Python Quellcode mit *Markdown* ermöglichen. Bei der Analyse sollen auch Attribute visualisiert werden, die bisher nicht vom *FDA Public Dashboard* dargestellt werden. Zudem sollen die Visualisierungen angemessener für die jeweils dargestellten Daten gewählt werden und damit vielfältiger ausfallen.

## 1.2 Ziel der Arbeit

Ziel dieser Arbeit ist es, eine Analyse des FAERS-Datensatzes mit Python durchzuführen. Die dabei entwickelten Funktionalitäten und das gewählte Vorgehen sollen weitere Analysen ermöglichen und anregen. Die Implementierung soll außerdem als mögliche Basis für ein neues System dienen können. Dieses könnte, im Gegensatz zu *OpenVigil*, auch eine Visualisierung der Daten anbieten. Um die Daten analysieren zu können, müssen diese zunächst in Python importiert und aufbereitet werden. Wichtig ist dabei, dass der Import und die Aufbereitung der Daten gut nachvollziehbar und einfach anpassbar ist. Neue Daten sollen ohne Änderungen importiert und aufbereitet werden können, sofern sich diese in ihrer Struktur nicht von den bisherigen Daten unterscheiden. Auf Änderungen in der Struktur der Daten soll aufgrund der einfachen Anpassbarkeit des Imports und der Aufbereitung der Daten entsprechend reagiert werden können.

Eines der Hauptziele von Pharmakovigilanz ist die Signalerkennung [24]. Daher sollen verschiedene Methoden zur Generierung von Signalen implementiert werden. Diese sollen auf Häufigkeitsmethoden beschränkt werden. Bayes'sche Methoden sollen nicht integriert werden. Signale müssen außerdem nur für Assoziationen von einem Medikament und einem unerwünschten Ereignis generiert werden können. Die Funktionsweise der implementierten Häufigkeitsmethoden soll anhand eines Beispiels demonstriert werden. Dafür werden Medikamente und Nebenwirkungen aus einem Bericht des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM) vom 2. Mai 2019 ausgewählt. In diesem wird vor schweren und langanhaltenden Nebenwirkungen im Bereich Muskeln, Gelenke und Nervensystem bei der Einnahme von Fluorchinolonen gewarnt [10]. Die Verwendung eines realen Beispiels soll den praktischen Nutzen der implementierten Funktionalitäten aufzeigen. Auf Basis der Ergebnisse der Signalgenerierung soll eine untersuchenswerte Kombination von einem Medikament und einem unerwünschten Ereignis ausgewählt werden.

Weiterhin soll eine explorative Analyse des Datensatzes durchgeführt werden. Dabei sollen Statistiken über den Datensatz gesammelt und visualisiert werden. Die Teilpopulation, bei der das ausgewählte Medikament und das unerwünschte Ereignis gemeldet wurde, soll außerdem mit der Gesamtpopulation verglichen und die Unterschiede zwischen den beiden Populationen untersucht werden. Dabei soll das Potenzial des Vergleichs von Teilpopulationen gezeigt werden. Gleichzeitig soll geprüft werden, ob der durch die Generierung eines Signals suggerierte Zusammenhang zwischen dem Medikament und dem unerwünschten Ereignis durch die Statistiken bekräftigt wird.

## 1.3 Aufbau der Arbeit

Diese Arbeit ist in fünf Kapitel untergliedert. In Kapitel 2 werden die Hintergründe des FAERS-Datensatzes erläutert. Abschnitt 2.1 erklärt die Grundlagen der Pharmakovigilanz und in Abschnitt 2.2 wird der wichtige formale Unterschied zwischen unerwünschten Ereignissen und Nebenwirkungen geklärt. Abschnitt 2.3 stellt die Aufgaben der FDA bei der Überwachung von Medikamenten nach deren Markteinführung dar. In Abschnitt 2.4 werden das FAERS und die darin enthaltenen Daten beschrieben. Abschnitt 2.5 diskutiert bei der Analyse dieses und anderer Pharmakovigilanzdatensätze zu beachtende Einschränkungen.

Kapitel 3 behandelt den Import und die Aufbereitung der Daten in Python. In Abschnitt 3.1 wird der Import der Daten und dabei aufgetretene Probleme beschrieben und in Abschnitt 3.2 wird auf die Aufbereitung der Daten eingegangen. Dabei werden die durch die Aufbereitung der Daten entstandenen Unterschiede zu den Rohdaten aufgezeigt. Anschließend wird in Abschnitt 3.3 auf weitere nachträgliche Anpassungen nach der Datenaufbereitung und auf die Speicherung der Daten eingegangen.

In Kapitel 4 werden zunächst in Abschnitt 4.1 die Grundlagen der Signalgenerierung erklärt. Anschließend werden in Abschnitt 4.2 die Ergebnisse der Signalgenerierung für Medikamente und Nebenwirkungen aus einem BfArM-Bericht präsentiert. Auf Basis des Ergebnisses der Signalgenerierung wird eine Teilpopulation ausgewählt, bei der das unerwünschte Ereignis Sehnenentzündung nach der Einnahme des Medikaments Ciprofloxacin aufgetreten ist. Die Kombination von Ciprofloxacin und Sehnenentzündung hat ein untersuchenswertes Signal ausgelöst.

Kapitel 5 behandelt die explorative Datenanalyse der Gesamt- und Teilpopulation. Dabei wird zum einen ein Überblick über die Eigenschaften des Datensatzes gegeben und zum anderen Unterschiede zwischen der Gesamt- und Teilpopulation aufgezeigt. Dabei wird geklärt, ob die Teilpopulation aufgrund ihrer Charakteristiken möglicherweise besonders häufig von Sehnenentzündung als unerwünschtes Ereignis nach der Einnahme von Ciprofloxacin betroffen ist.

In Kapitel 6 werden die Ergebnisse der Signalgenerierung und der explorativen Datenanalyse zusammengefasst. In Abschnitt 6.1 wird ein Fazit gezogen, inwiefern die in Abschnitt 1.2 gesteckten Ziele dieser Arbeit erreicht wurden. Abschließend wird in Abschnitt 6.2 ein Ausblick darauf gegeben, wie die entwickelten Funktionalitäten verbessert und weiterentwickelt werden können.

## 2 Datenlage

In diesem Kapitel werden die Hintergründe des FAERS-Datensatzes erläutert. In Abschnitt 2.1 wird dazu zunächst Pharmakovigilanz und damit verbundene Tätigkeiten beschrieben. In 2.2 wird der wichtige formale Unterschied zwischen unerwünschten Ereignissen und Nebenwirkungen geklärt. Anschließend wird in Abschnitt 2.3 mit der FDA auf die hinter dem FAERS stehende Organisation eingegangen und der Hintergrund der öffentlichen Verfügbarkeit der FAERS-Daten erläutert. In Abschnitt 2.4 wird das FAERS und die Herkunft und Beschaffenheit der FAERS-Daten beschrieben. In Abschnitt 2.5 werden Einschränkungen diskutiert, die bei der Analyse der FAERS- und anderer Pharmakovigilanzdaten zu beachten sind.

### 2.1 Pharmakovigilanz

Laut Evans stammt der Begriff Pharmakovigilanz aus dem Französischen und wurde von Professor Bernard Bégaud als eine Disziplin beschrieben, die die Erkennung, Auswertung und Prävention von unerwünschten Wirkungen von Arzneimitteln beinhaltet [17]. Die WHO definiert Pharmakovigilanz als die Wissenschaft und Tätigkeit im Zusammenhang mit der Erkennung, Bewertung, dem Verständnis und der Prävention von unerwünschten Ereignissen oder anderen möglichen arzneimittelbedingten Problemen [32]. Ausgelöst wurde die systematische Überwachung von Medikamenten durch den Contergan-Skandal 1961. Zu dieser Zeit wurden viele tausend Säuglinge mit Deformationen geboren, weil sie in der Gebärmutter einem unsicheren Medikament ausgesetzt waren, das für die Anwendung bei schwangeren Müttern empfohlen wurde.

## 2.2 Unerwünschte Ereignisse und Nebenwirkungen

Zentraler Gegenstand der Pharmakovigilanz sind die Begriffe „Medikament“ und „unerwünschtes Ereignis“. Unerwünschte Ereignisse sind dabei nicht mit Nebenwirkungen gleichzusetzen. Nach §3 Absatz 1 Satz 6 der Guten Klinischen Praxis Verordnung (GCP-V) ist ein unerwünschtes Ereignis jedes nachteilige Vorkommnis, das einer betroffenen Person widerfährt, der ein Medikament verabreicht wurde, und das nicht notwendigerweise in ursächlichem Zusammenhang mit dieser Behandlung steht [11]. Es beschreibt also lediglich das Auftreten eines Ereignisses nach der Einnahme eines Medikaments. Das unerwünschte Ereignis muss dabei nicht zwangsläufig von dem Medikament verursacht worden sein [33].

Bei einer Nebenwirkung ist das Medikament der Auslöser des unerwünschten Ereignisses. Nach §3 Absatz 1 Satz 7 GCP-V ist eine Nebenwirkung jede nachteilige und unbeabsichtigte Reaktion auf ein Medikament, unabhängig von dessen Dosierung [11]. Ziel der Pharmakovigilanz ist es, unerwartete Nebenwirkungen zu finden. Eine unerwartete Nebenwirkung ist laut §3 Absatz 1 Satz 9 GCP-V eine Nebenwirkung, die nach Art oder Schweregrad nicht mit der vorliegenden Information über das Medikament übereinstimmt [11]. Es wird also nach Nebenwirkungen gesucht, die noch nicht als Nebenwirkung des Medikaments bekannt sind oder in ihrer Schwere von den bisher bekannten Nebenwirkungen abweichen [33].

In dieser Arbeit wird der Begriff „unerwünschtes Ereignis“ verwendet, um nach der Einnahme von Medikamenten aufgetretene nachteilige Vorkommnisse zu beschreiben. Die später vorgestellten statistischen Methoden und die Ergebnisse der explorativen Datenanalyse liefern nur Hinweise auf mögliche Zusammenhänge zwischen Medikamenten und unerwünschten Ereignissen. Um ein unerwünschtes Ereignis als Nebenwirkung eines Medikaments bezeichnen zu können, muss nachgewiesen werden, dass dieses unerwünschte Ereignis durch das Medikament verursacht wurde [33]. Dies ist nur mit klinischen Pharmakologiekenntnissen möglich und daher nicht Gegenstand dieser Arbeit.

## 2.3 U.S. Food and Drug Administration

Die U.S. Food and Drug Administration (FDA) ist unter anderem dafür verantwortlich, die öffentliche Gesundheit zu schützen, indem sie die Sicherheit und Wirksamkeit von Medikamenten garantiert [25]. Als eine Bundesbehörde unterliegt die FDA dem Freedom of Information Act (FOIA) von 1967. Dieses Gesetz verpflichtet amerikanische Bundesbehörden alle Informationen offenzulegen, die unter Berufung auf den FOIA angefragt werden. Ausnahmen bilden Anfragen, die die Privatsphäre, nationale Sicherheit oder Rechtsdurchsetzung beeinträchtigen [6]. Da die FAERS-Daten anonymisiert sind, treffen diese Einschränkungen nicht zu. Sie sind daher von der FDA zu veröffentlichen.

## 2.4 FDA Adverse Event Reporting System

Die FDA beschreibt das FDA Adverse Event Reporting System (FAERS) als eine Datenbank, die Berichte über unerwünschte Ereignisse, Medikationsfehler und Beschwerden über die Qualität von Produkten beinhaltet, die zu unerwünschten Ereignissen geführt haben, die an die FDA gesendet wurden. Die Berichte werden von Gesundheitsexperten, Konsumenten und Pharmakonzernen an die FDA übermittelt. Die FDA empfängt freiwillige Berichte direkt von Gesundheitsexperten wie Ärzten, Apothekern und Krankenpflegern und Konsumenten wie Patienten, Familienangehörigen und Anwälten. Gesundheitsexperten und Konsumenten können außerdem an die Pharmakonzerne berichten. Wenn ein Pharmakonzern einen Bericht von einem Gesundheitsexperten oder Konsumenten empfängt, ist dieser verpflichtet, den Bericht an die FDA weiterzuleiten [28].

Die Berichte werden quartalsweise aus dem FAERS exportiert und von der FDA hier [27] zum Download zur Verfügung gestellt. Die Berichte können als ASCII- oder XML-Dateien heruntergeladen werden. In dieser Arbeit wurde mit den XML-Dateien gearbeitet. Die Berichte sind nicht kumulativ. Ein Export beinhaltet nur die Berichte, die in dem entsprechenden Quartal an die FDA gesendet wurden. Ein Export kann allerdings durchaus Berichte enthalten, die zuvor bereits exportiert wurden. Dies passiert, wenn ein Bericht in einem neuen Export aktualisiert wird. Da bei der Analyse alle Exporte verwendet wurden, können mehrere Versionen desselben Berichts in den Daten vorhanden sein. Um die Statistiken nicht zu verfälschen, wurden die alten Versionen der Berichte entfernt. Dieser Schritt wird in Abschnitt 3.3.2 erklärt.

Ein Bericht enthält Informationen über den Bericht selbst, sowie über den Patienten aus diesem Bericht. Einem Bericht ist dabei immer genau ein Patient zugeordnet. Ob demselben Patienten mehrere Berichte zugeordnet sind, lässt sich auf Basis der Daten nicht feststellen. Es ist jedoch fest davon auszugehen, dass es Patienten gibt, denen mehrere Berichte zugeordnet sind. In einem Bericht können beliebig viele unerwünschte Ereignisse und Medikamente gelistet werden. Jeder Bericht muss jedoch mindestens ein unerwünschtes Ereignis und ein Medikament beinhalten. Zwischen den Medikamenten und unerwünschten Ereignissen besteht kein bekannter kausaler Zusammenhang. Bei der Analyse müssen daher alle Kombinationen von in einem Bericht gelisteten Medikamenten und unerwünschten Ereignissen beachtet werden.

## 2.5 Einschränkungen

Es gibt verschiedene Einschränkungen, die bei der Auswertung der FAERS- und anderen Pharmakovigilanzdaten zu beachten sind. Daher weist die FDA darauf hin, dass die Informationen aus den Berichten nicht dazu verwendet werden können, die Häufigkeiten der gemeldeten unerwünschten Ereignisse zu beurteilen [28].

### 2.5.1 Duplikate

Für Bate stellen Duplikate, also Berichte, die dieselben unerwünschten Ereignisse beschreiben, ein enormes Problem für alle großen Pharmakovigilanzdatensätze dar. Das Ausmaß der Duplikate ist nicht genau bekannt. Ihre Anwesenheit verringert jedoch die Signalerkennungsfähigkeit deutlich. Anonymisierte Berichte erschweren das Auffinden von Duplikaten. Diese können durch verschiedene Mechanismen auftreten. Derselbe Bericht kann sowohl von Gesundheitsexperten, als auch von Pharmakonzernen übermittelt werden. Es wird erwartet, dass das Aufkommen der Patientenberichterstattung dieses Problem verschlimmert. Nachfolgenden Berichten kann es außerdem an ausreichender Verknüpfung zu den Originalberichten mangeln [8].

Für die FDA führt die Tatsache, dass Pharmakonzerne gesetzlich dazu verpflichtet sind, empfangene Berichte an die FDA weiterzuleiten, dazu, dass Berichte doppelt im FAERS enthalten sein können, wenn Gesundheitsexperten oder Konsumenten Berichte sowohl direkt an die FDA, als auch an die Pharmakonzerne senden [28]. Da die Berichte vom

FAERS unterschiedliche Identifikationsnummern erhalten, können Duplikate nicht anhand dieser als solche erkannt werden. Im Rahmen dieser Arbeit konnten daher doppelte Berichte mit unterschiedlichen Identifikationsnummern nicht entfernt werden.

Eine Möglichkeit um diesem Problem in Zukunft zu begegnen, ist der Vergleich von zwei Berichten anhand ihrer Eigenschaften. Norén et al. haben eine Methode zur Identifikation von Duplikaten entwickelt, die auf einem „Hit-Miss-Modell“ basiert [21]. „Hits“ erhalten ein positives Gewicht, wenn z.B. das Medikament in beiden Berichten gelistet ist. „Misses“ erhalten ein negatives Gewicht, wenn z.B. dasselbe Medikament nicht in beiden Berichten gelistet ist. Wenn Informationen fehlen, ist das Gewicht null. Das Gewicht für jedes Feld wird berechnet und die Gesamtsummen addiert, um eine Gesamtbewertung für das Bericht-Paar zu erhalten. Die Punktzahl berücksichtigt sowohl den Übereinstimmungsgrad, als auch die Informationsmenge in den Berichten. Diese Methode könnte auch auf die Berichte aus dem FAERS angewendet werden und stellt eine Möglichkeit der Weiterentwicklung zur Erhöhung der Datenqualität dar.

### 2.5.2 Kausalität

Es gibt es keine Sicherheit, dass ein unerwünschtes Ereignis von einem Medikament verursacht wurde. Die FDA verlangt nicht, dass ein kausaler Zusammenhang zwischen einem Medikament und einem unerwünschten Ereignis besteht. Das unerwünschte Ereignis kann auch von der Krankheit ausgelöst worden sein, die behandelt wurde. Das unerwünschte Ereignis kann außerdem von einem anderen Medikament ausgelöst worden sein, dass zeitgleich eingenommen wurde. Oder das unerwünschte Ereignis ist aus anderen Gründen aufgetreten [28].

### 2.5.3 Underreporting

Die FDA empfängt nicht für jedes, mit einem Medikament aufgetretenes, unerwünschtes Ereignis, einen Bericht. Dieses Problem wird als Underreporting bezeichnet. Viele Faktoren können beeinflussen, ob ein unerwünschtes Ereignis gemeldet wird, wie der Zeitpunkt, zu dem ein Medikament in den Verkehr gebracht wurde, und die Publizität eines unerwünschten Ereignisses. Besonders bekannte Medikamente oder unerwünschte Ereignisse werden u.U. häufiger gemeldet als weniger bekannte [28].

## 3 Datenaufbereitung

Dieses Kapitel behandelt den Import und die Aufbereitung der Daten in Python. An dieser Stelle sei auf die für die Implementierung verwendete Hardware hingewiesen. Es wurde ein 15-Zoll MacBook Pro aus dem Jahr 2018 in der Basiskonfiguration mit 16 GB Hauptspeicher verwendet. Weitere technische Spezifikationen finden sich hier [7]. Sämtliche in dieser Arbeit angegebene Zeitmessungen wurden auf dieser Hardware durchgeführt. Die während des Imports und der Aufbereitung der Daten aufgetretenen Probleme sind auf die hohe Hauptspeicherauslastung zurückzuführen und wären auf einem System mit mehr Hauptspeicher möglicherweise nicht aufgetreten.

### 3.1 Parsen der XML-Dateien

Da der Export der Daten durch die FDA quartalsweise stattfindet, enthält eine Datei normalerweise die Daten aus einem Quartal. Seit dem zweiten Quartal 2018 werden die Daten aus einem Quartal allerdings auf drei Dateien aufgeteilt. Die Dateien wurden nach dem Schema `<Jahr><Quartal>[_<Dateinummer>]` umbenannt. ADR12Q4 wurde also z.B. in `2012Q4` und `1_ADR18Q2` in `2018Q2_1` umbenannt. Dadurch liegen die Dateien in chronologischer Reihenfolge vor und weisen eine einheitliche Benennung auf.

Die Dateien sind überwiegend im UTF-8-Format codiert. Eine Ausnahme bildet der erste Export, der die Daten aus dem vierten Quartal 2012 beinhaltet. Dieser ist im ISO-8859-1-Format codiert. Daher musste diese Datei beim Parsen gesondert behandelt werden. Um die XML-Dateien in Python zu parsen, wurde zunächst nur die Bibliothek *ElementTree* [2] verwendet. Allerdings konnte die Datei mit den Daten aus dem vierten Quartal 2012 nicht geparkt werden, da diese nicht im UTF-8-Format, sondern im ISO-8859-1-Format codiert ist.

Das liegt daran, dass *ElementTree* den Standard-XML-Parser verwendet, wenn zum Parsen kein dedizierter XML-Parser übergeben wird. Dieser ist mit UTF-8 als Format konfiguriert und kann damit die Datei mit den Daten aus dem vierten Quartal 2012 nicht parsen. Daher wurde der XML-Parser aus der Bibliothek *etree* [1] verwendet. Dieser wurde mit UTF-8 als Format konfiguriert. Für die Datei mit den Daten aus dem vierten Quartal 2012 wurde ein weiterer XML-Parser mit ISO-8859-1 als Format erstellt. Der entsprechende XML-Parser wurde *ElementTree* zum Parsen übergeben. Dadurch konnte auch die Datei mit den Daten aus dem vierten Quartal 2012 geparkt werden.

Da sich die Dateien nicht verändern, müssen diese nur einmal geparkt werden. In Python werden die Dateien durch *ElementTree*-Objekte repräsentiert. Diese sollten mit *pickle* gespeichert werden. *Pickle* ist eine Python-Bibliothek zur Serialisierung und Deserialisierung von Python-Datenstrukturen [5]. Die gespeicherten *ElementTree*-Objekte sollten zur Aufbereitung der Daten wieder geladen werden. Dadurch müssten die Dateien nicht erneut geparkt werden, wenn sich an der Aufbereitung der Daten etwas geändert hat. Allerdings wird das Speichern von *ElementTree*-Objekten durch *pickle* nicht unterstützt. Es wurde daher immer nur eine Datei zur Zeit geparkt und aufbereitet. Dadurch mussten die Dateien jedes mal neu geparkt werden, wenn sich an der Aufbereitung der Daten etwas geändert hatte.

## 3.2 Aufbereitung der Daten

Die einzelnen Attribute können anhand ihres XML-Tags mit *ElementTree* aus den XML-Dateien gelesen werden. *ElementTree* gibt den Inhalt eines XML-Tags als *String* zurück oder *Null*, wenn der XML-Tag nicht gefunden wurde. Wie in Abschnitt 2.4 beschrieben wurde, können die Berichte eindeutig anhand ihrer Identifikationsnummer erkannt werden. Problematisch ist, dass die Berichte beliebig viele unerwünschte Ereignisse und Medikamente beinhalten können. Die unerwünschten Ereignisse und Medikamente bestehen jeweils aus mehreren Attributen, welche diese charakterisieren. Um die Berichte nicht redundant zu speichern, hätten die unerwünschten Ereignisse und Medikamente aggregiert z.B. als *DataFrames* in den Berichten gespeichert werden müssen. Diese Art der Speicherung hätte jedoch die Analyse der Daten erschwert und zu schwer verständlichem Quellcode geführt. Es wurde sich daher für eine andere Vorgehensweise entschieden. Die Berichte wurden in acht verschiedenen *DataFrames* gespeichert, deren Aufbau im Folgenden näher erläutert wird. Die dabei angegebene Größe der *DataFrames* ist nur vorläufig

und wird durch das Entfernen der alten Versionen der Berichte in Abschnitt 3.3.2 noch reduziert.

Zur Analyse der Berichts- und Patientendaten wurden die Daten zu unerwünschten Ereignissen und Medikamenten nicht benötigt. Daher wurden die Berichts- und Patientendaten ohne die Daten zu unerwünschten Ereignissen und Medikamente im *DataFrame* „reports“ abgespeichert. Ein Eintrag entspricht dabei einem Bericht. Der „reports“-*DataFrame* umfasst 8.488.189 Berichte.

Zur Analyse der unerwünschten Ereignisse wurden die Berichts- und Patientendaten, sowie alle Daten über unerwünschte Ereignisse benötigt. Diese wurden im *DataFrame* „aes“ gespeichert. Ein Eintrag lässt sich als das Auftreten eines unerwünschten Ereignisses bei einem Patienten aus einem Bericht verstehen. Ein Bericht hat folglich so viele Einträge, wie er unerwünschte Ereignisse beinhaltet. Die Berichts- und Patientendaten wurden demnach redundant gespeichert. Der „aes“-*DataFrame* beinhaltet 25.683.689 unerwünschte Ereignisse.

Für die Analyse der Medikamente wurden ebenfalls die Berichts- und Patientendaten benötigt. Diese wurden mit den Medikamentendaten im *DataFrame* „drugs“ abgespeichert. Ein Eintrag ist dabei als die Einnahme eines Medikaments durch den Patienten in diesem Bericht zu verstehen. Ein Bericht hat dementsprechend so viele Einträge, wie Medikamente in ihm aufgeführt sind. Auch hier wurden die Berichts- und Patientendaten redundant gespeichert. Der „drugs“-*DataFrame* besteht aus 30.815.184 Medikamenteneinnahmen.

Die in Abschnitt 2.4 angesprochenen Kombinationen von unerwünschten Ereignissen und Medikamenten aus einem Bericht wurden im *DataFrame* „combinations“ gespeichert. Dieser enthält die Daten zu Berichten, Patienten, unerwünschten Ereignissen und Medikamenten und damit alle Daten. Ein Eintrag ist eine mögliche Kombination eines unerwünschten Ereignisses mit einem Medikament aus einem Bericht. Dabei steht nicht fest, dass das unerwünschte Ereignis in Zusammenhang mit diesem Medikament, oder einem anderen Medikament aus diesem Bericht aufgetreten ist. Ein Bericht hat so viele Einträge, wie es mögliche Kombinationen von unerwünschten Ereignissen und Medikamenten aus diesem Bericht gibt. Dadurch wurden die Berichts- und Patientendaten, sowie auch Daten zu unerwünschten Ereignissen und Medikamenten redundant gespeichert. Der „combinations“-*DataFrame* beinhaltet 135.489.143 Kombinationen.

Zu jedem dieser *DataFrames* existiert ein weiterer „Roh“-*DataFrame*, der die ursprünglichen Werte von Attributen enthält, die während der Aufbereitung der Daten verän-

dert wurden. Außerdem enthält er Attribute, die während der Aufbereitung der Daten benutzt, aber nicht in den *DataFrame* aufgenommen wurden. Unveränderte Attribute wurden nicht abgespeichert, da diese bereits in den *DataFrames* enthalten sind. Die „Roh“-*DataFrames* tragen den Namen des zugehörigen *DataFrames* mit dem Präfix „raw\_“. Der „Roh“-*DataFrame* zum *DataFrame* „reports“ heißt also z.B. „raw\_reports“. Die „Roh“-*DataFrames* haben die gleiche Größe wie die dazugehörigen *DataFrames*. Außerdem wurden auch bei diesen *DataFrames*, wie in Abschnitt 3.3.2 beschrieben wird, die alten Versionen der Berichte entfernt. Die „Roh“-*DataFrames* wurden dazu verwendet, die durch die Aufbereitung der Daten entstandenen Unterschiede zu den Rohdaten zu dokumentieren. Im Folgenden wird auf die Attribute eingegangen, die während der Aufbereitung der Daten verändert wurden.

#### 3.2.1 Herkunftsländer

Es gibt drei Attribute, die eine geografische Zuordnung der Berichte zulassen. Alle drei sind als ISO-3166 ALPHA-2 (i.F. ALPHA-2) Ländercodes codiert. Die verwendete ISO-3166-Tabelle beinhaltet 249 unterschiedliche Länder [3].

Das erste Attribut, welches eine geografische Zuordnung der Berichte ermöglicht, ist das Land des primären Berichterstatters (*primarysourcecountry* [29]). Dieses wurde in 98,58% der Berichte angegeben. Dabei wurden 232 unterschiedliche Länder aufgeführt. Ein weiteres Attribut, das eine geografische Zuordnung der Berichte zulässt, ist das Land, in dem die unerwünschten Ereignisse aufgetreten sind (*occurcountry* [29]). Es wurde in 95,76% der Berichte angegeben und weist 228 unterschiedliche Länder auf. Das dritte geografische Attribut ist das Land des Berichterstatters der letzten Version eines Berichts (*reporttercountry* [29]). Dieses wurde in 93,72% der Berichte angegeben. Dabei wurden 234 unterschiedliche Länder gemeldet.

Die ALPHA-2 Ländercodes bestehen aus zwei Zeichen, z.B. „US“. Die Ländercodes werden daher als *Strings* gespeichert. Der Speicherverbrauch liegt bei 51 Bytes pro Eintrag. 49 Bytes für einen leeren *String* und ein Byte je weiteres Zeichen. Aufgrund dieses hohen Speicherverbrauchs wurde sich dafür entschieden, die geografischen Attribute in einem Attribut „reportorigin“ zusammenzufassen.

Dazu wurde zunächst das „occurcountry“ als „reportorigin“ festgelegt. War kein „occurcountry“ angegeben, wurde das „primarysourcecountry“ als „reportorigin“ verwendet. Ist

dieses auch nicht angegeben worden, wurde das „reportercountry“ als „reportorigin“ genommen. Dieses kann nicht *Null* sein, hat aber die mögliche Ausprägung „COUNTRY NOT SPECIFIED“. Lag diese Ausprägung vor, wurde *Null* für den „reportorigin“ eingetragen. Durch dieses Vorgehen konnte eine geografische Zuordnung für 99,94% der Berichte vorgenommen werden.

Der ermittelte ALPHA-2 Ländercode wurde anschließend in einen ISO-3166 ALPHA-3 (i.F. ALPHA-3) Ländercode umgewandelt. „US“ z.B. wurde in „USA“ umgewandelt. Dies ist dadurch motiviert, dass die bei der Analyse verwendete Funktion zur Erstellung von Choroplethen-Karten keine ALPHA-2 aber dafür ALPHA-3 Ländercodes unterstützt. Um die spätere Umwandlung bei der Analyse zu vermeiden, wurde diese direkt beim Import der Daten vorgenommen. Ein kleiner Nachteil dieser Vorgehensweise ist der etwas höhere Speicherverbrauch. ALPHA-2 Ländercodes bestehen aus zwei Zeichen, ALPHA-3 aus drei. Der Speicherverbrauch steigt dadurch von 51 auf 52 Bytes pro Eintrag.

Um Speicherplatz zu sparen, könnten die ALPHA-2 Ländercodes in numerische ISO-3166 Ländercodes umgewandelt werden. Dies würde die Verwendung von *uint16* als Datentyp zulassen. Dieser hat einen Speicherverbrauch von zwei Bytes pro Eintrag. Die verwendete Funktion zur Erstellung von Choroplethen-Karten unterstützt allerdings keine numerischen ISO-3166 Ländercodes. Die numerischen ISO-3166 Ländercodes müssten also bei der Analyse in ALPHA-3 Ländercodes umgewandelt werden.

#### 3.2.2 Ernsthaftigkeit der Berichte

Die Ernsthaftigkeit der Berichte bezieht sich auf die Ernsthaftigkeit des medizinischen Zustands der Patienten. Der medizinische Zustand der Patienten kann ernsthaft oder nicht ernsthaft sein. Ist der medizinische Zustand der Patienten ernsthaft, wird zusätzlich der Grad der Ernsthaftigkeit angegeben. Es gibt sechs verschiedene Grade von Ernsthaftigkeit, von denen maximal einer pro Bericht angegeben werden kann [29]. Daher wurde sich dafür entschieden, die Ernsthaftigkeit des medizinischen Zustands der Patienten in einem Attribut zusammenzufassen. Für nicht ernsthafte Berichte wurde null eingetragen. Für ernsthafte Berichte wurde ein Wert zwischen eins und sechs für den Grad der Ernsthaftigkeit zugewiesen. Dadurch wurde die Anzahl notwendiger Spalten für die Ernsthaftigkeit des medizinischen Zustands der Patienten von sieben auf eine Spalte reduziert. Der Speicherverbrauch wurde dadurch von sieben auf ein Byte reduziert. Außerdem wurde die Anzahl möglicher *Nullwerte* pro Bericht von fünf bzw. sechs auf null reduziert.

#### 3.2.3 Datumsangaben

Die Berichte beinhalten jeweils zwei Datumsangaben. Der Tag, an dem die erste Version des Berichts empfangen wurde und der Tag, an dem die letzte Version des Berichts empfangen wurde. Die Daten wurden als *Strings* nach dem Schema  $\langle JJJJ \rangle [\langle MM \rangle] [\langle TT \rangle]$  gespeichert und mussten daher noch in Datumsobjekte umgewandelt werden. Zu den Datumsangaben wurden dazu die Zeiteinheiten angegeben, in denen die Daten gespeichert wurden. Mögliche Formate sind 102 für Jahr, Monat und Tag (JJJJMMTT), 610 für Jahr und Monat (JJJJMM) und 602 wenn nur das Jahr angegeben wurde (JJJJ) [29]. Da die Codes für die Zeiteinheiten nicht in den Wertebereich eines *uint8* passen, wurden stattdessen die Werte eins, zwei und drei zugewiesen. Dadurch wurde der Speicherverbrauch von zwei auf ein Byte pro Eintrag reduziert. Alle vier Attribute wurden in allen Berichten angegeben. Dadurch konnten alle Daten mit Hilfe des jeweiligen Formats in Datumsobjekte umgewandelt werden. Als Format wurde nur 102 (JJJJMMTT) angegeben.

#### 3.2.4 Alter der Patienten

Das Alter wurde bei 58,2% der Patienten angegeben. Bei 58,03% der Patienten wurde außerdem die Zeiteinheit aufgeführt, in der das Alter angegeben wurde. Das Alter wurde mit Hilfe der Zeiteinheit in Jahre umgewandelt. War keine Zeiteinheit angegeben, wurde *Null* für das Alter eingetragen. Ohne Zeiteinheit ist das Alter nutzlos. Für 58,03% der Patienten konnte so das Alter in Jahre umgewandelt werden. Nach der Umwandlung in Jahre wurde für Patienten jünger als null Jahre oder älter als 125 Jahre *Null* für das Alter eingetragen. Patienten jünger als null Jahre sind unmöglich und Patienten älter als 125 Jahre unrealistisch.

#### 3.2.5 Altersgruppen der Patienten

Die Altersgruppe wurde nur bei 16,4% der Patienten angegeben. Die Altersgruppe konnte jedoch aus dem Alter des Patienten berechnet werden, sofern dieses angegeben wurde. Da die FDA nicht offenlegt, welche Zeitintervalle sie für die Altersgruppen verwendet, wurden folgende Zeitintervalle für die Altersgruppen verwendet:

- Neugeborene: 0 bis 1 Monat
- Kleinkinder: > 1 Monat bis 2 Jahre
- Kinder: > 2 bis 12 Jahre
- Jugendliche: > 12 bis 18 Jahre
- Erwachsene: > 18 bis 65 Jahre
- Senioren: > 65 Jahre

Wenn die Altersgruppe und das Alter des Patienten nicht angegeben wurden, konnte die Altersgruppe des Patienten nicht bestimmt werden. War die Altersgruppe des Patienten aufgeführt, wurde diese trotzdem aus dem Alter des Patienten berechnet, wenn dieses angegeben wurde. Dadurch wurden die Zeitintervalle für die Altersgruppen eingehalten. Die angegebene Altersgruppe wurde nur verwendet, wenn das Alter des Patienten nicht aufgeführt wurde. Dadurch konnte für 62,07% der Patienten die Altersgruppe bestimmt werden. Wie in Abbildung 3.1 zu sehen ist, hat sich durch dieses Vorgehen die Verteilung der Altersgruppen leicht verändert.

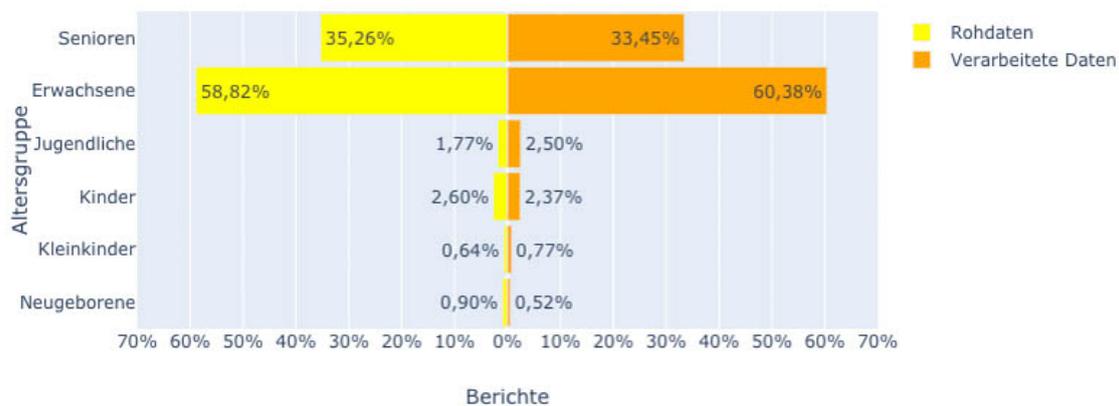


Abbildung 3.1: Veränderung der Verteilung der Altersgruppen

#### 3.2.6 Gewicht der Patienten

Das Gewicht der Patienten ist in Kilogramm mit einer Genauigkeit von zwei Nachkommastellen aufgeführt [29]. Es wurde bei 20,23% der Patienten angegeben. Bei Patienten mit einem Gewicht von unter null Kilogramm oder über 750 Kilogramm wurde *Null* für das Gewicht eingetragen. Patienten mit einem Gewicht von unter null Kilogramm sind

unmöglich und Patienten mit einem Gewicht von über 750 Kilogramm unrealistisch. Von dieser Einschränkung waren 47 Patienten betroffen.

#### 3.2.7 Beschreibungen und Bezeichnungen

Die unerwünschten Ereignisse und die Indikationen der Medikamente sollen als MedDRA Preferred Terms (PTs) beschrieben werden [29]. Das MedDRA (Medical Dictionary for Regulatory Activities) ist eine umfassende und hochspezifische standardisierte medizinische Terminologie, um den internationalen Austausch von behördlichen Informationen für vom Menschen verwendete Medizinprodukte zu erleichtern [4].

Da es sich bei den Attributen für die unerwünschten Ereignisse und Indikationen jedoch um Freitextfelder handelt, kommen auch Beschreibungen außerhalb der MedDRA PTs vor. Die Beschreibungen wurden in Kleinschreibung umgewandelt, um Duplikate durch unterschiedliche Groß- und Kleinschreibung auszuschließen und späteren Fehlern bei der Analyse durch unterschiedliche Groß- und Kleinschreibung vorzubeugen. Die Anzahl unterschiedlicher unerwünschter Ereignisse wurde dadurch um 1.683 von 20.964 auf 19.281 reduziert. Die Anzahl unterschiedlicher Indikationen wurde um 5 von 15.318 auf 15.313 reduziert.

Die Medikamentenbezeichnungen sind keine MedDRA PTs, wurden jedoch aus denselben Gründen in Kleinschreibung umgewandelt. Am Ende einiger Medikamentenbezeichnungen taucht ein Punkt auf. Dieser wurde entfernt, ebenfalls um Duplikate auszuschließen. Dadurch wurde die Anzahl unterschiedlicher Medikamente um 2.100 von 401.525 auf 399.425 reduziert.

### 3.3 Weitere Anpassungen

Nachdem die *DataFrames* erstellt und befüllt worden sind, wurden noch weitere nachträgliche Anpassungen vorgenommen. Abschnitt 3.3.1 beschreibt die Anpassung der Datentypen zur Reduzierung des Speicherverbrauchs. In Abschnitt 3.3.2 wird auf das Entfernen der alten Versionen der Berichte eingegangen. Abschnitt 3.3.3 behandelt kurz die Speicherung der *DataFrames*.

#### 3.3.1 Anpassung der Datentypen

Sämtliche kategorische Attribute liegen in codierter Form vor. Beim Geschlecht des Patienten z.B. steht null für „unbekannt“, eins für „männlich“ und zwei für „weiblich“ [29]. Um das Interpretieren der Daten zu vereinfachen, wurden diese Attribute zunächst decodiert. Beim Geschlecht des Patienten z.B. wurde null in „unbekannt“, zwei in „männlich“ und drei in „weiblich“ umgewandelt. Dadurch wurden die Daten jedoch als *Strings* und nicht als numerische Datentypen abgespeichert.

Beim Erstellen der *DataFrames* trat daher das Problem auf, dass diese nicht in den Hauptspeicher passten. *Strings* haben mit mindestens 49 Bytes einen höheren Speicherbedarf als numerische Datentypen (maximal acht Bytes bei *float64*, *int64* und *uint64*). Daher wurde darauf verzichtet, kategorische Attribute zu decodieren und als *Strings* zu speichern.

Beim Erzeugen eines *DataFrames* verwendet *pandas* für eine Spalte jeweils die Variante eines Datentyps mit dem größten Wertebereich. Für ganzzahlige Werte wird z.B. *int64* verwendet. Die Verwendung von Datentypen mit so großen Wertebereichen ist allerdings für die meisten Spalten überflüssig.

Zunächst hat keine der Spalten negative Werte. Dadurch konnte bei numerischen Spalten mit ganzzahligen Werten die Version eines Datentyps verwendet werden, die einen positiven Wertebereich hat. Bei *int64* z.B. *uint64*. Beide Versionen eines Datentyps haben jedoch denselben Speicherbedarf. Bei *int64* und *uint64* z.B. haben beide Datentypen einen Speicherbedarf von acht Bytes. Dadurch alleine konnte der Speicherbedarf also noch nicht verringert werden. Die Anpassung wurde jedoch trotzdem vorgenommen, da es schlüssiger ist, einen Datentyp mit einem positiven Wertebereich für Attribute zu verwenden, die ausschließlich positive Werte annehmen können.

Die meisten Spalten nutzen den Wertebereich des jeweiligen Datentyps zudem nicht aus. Das Geschlecht des Patienten z.B. hat nur die Ausprägungen null, eins, zwei und *Null*. Daher wurde an Stelle des von *pandas* verwendeten Datentyps *int64* der Datentyp *uint8* zugewiesen. Dieser hat einen kleineren Wertebereich und verbraucht daher statt acht nur ein Byte Speicherplatz.

Das Vorkommen von *Nullwerte* in einer numerischen Spalte mit ganzzahligen Werten führt dazu, dass *pandas* eine Fließkommazahl als Datentyp für diese Spalte verwendet.

Das liegt daran, dass *NaN* die Repräsentation von *Nullwerte* bei numerischen Datentypen und nur bei Fließkommazahlen möglich ist. Bei Spalten, bei denen eigentlich der Datentyp *uint8* möglich wäre, bedeutet dies einen höheren Speicherverbrauch. Denn der kleinste Fließkommazahl Datentyp ist *float16*. Dieser verbraucht zwei Bytes gegenüber dem ein Byte Speicherverbrauch von *uint8*. Daher wurde für *Nullwerte* ein ganzzahliger numerischer Wert eingesetzt. Dies verhindert, dass *pandas* eine Fließkommazahl als Datentyp für Spalten verwendet, die nur ganzzahlige Werte beinhalten.

Diese Vorgehensweise hat die Konsequenz, dass *Nullwerte* nicht mehr als solche erkennbar sind. Daher wurde im Modul *analysis* die Funktion *null* implementiert, die die Anzahl der *Nullwerte* eines Attributs bestimmt. Wird dieser Funktion kein *Nullwert* übergeben, zählt diese die Anzahl der *None*- bzw. *NaN*-Werte. Wird der Funktion ein *Nullwert* übergeben, zählt diese stattdessen die Häufigkeit dieses Werts.

#### 3.3.2 Entfernen alter Versionen der Berichte

In einem Export ist jeder Bericht einzigartig und lässt sich eindeutig anhand seiner Identifikationsnummer erkennen. Ein Export beinhaltet außerdem immer nur die aktuellste Version eines Berichts. Die Version wird durch eine Versionsnummer bestimmt, welche bei eins startet und bei jeder Aktualisierung des Berichts um eins inkrementiert wird [29]. Da die *DataFrames* die Berichte aller Exporte beinhalten, sind auch die alten Versionen der Berichte in diesen enthalten. Um die Statistiken später nicht zu verfälschen, wurden die alten Versionen der Berichte entfernt.

Dazu wurden die Berichte nach Identifikationsnummer in aufsteigender und Versionsnummer in absteigender Reihenfolge sortiert. Dadurch stand jeweils die aktuellste Version der Berichte an oberster Stelle. Die Reihenfolge der Identifikationsnummern spielte dabei keine Rolle, diese hätten auch in absteigender Reihenfolge sortiert werden können. Anschließend wurden Duplikate entfernt. Dadurch blieben nur die aktuellsten Versionen der Berichte übrig. Die Kriterien für Duplikate unterscheiden sich zwischen den *DataFrames*, da diese, wie in Abschnitt 3.2 beschrieben, eine unterschiedliche Struktur aufweisen. Die Kriterien für Duplikate sind im Folgenden aufgeführt. Die Angaben gelten jeweils für einen *DataFrame* und den dazugehörigen „Roh“-*DataFrame*.

- reports: Identifikationsnummer
- aes: Identifikationsnummer und Beschreibung des unerwünschten Ereignisses
- drugs: Identifikationsnummer und Medikamentenbezeichnung
- combinations: Identifikationsnummer, Beschreibung des unerwünschten Ereignisses und Medikamentenbezeichnung

Die Anzahl der Einträge der *DataFrames* wurde dadurch wie folgt reduziert:

- reports: 8.488.189 auf 7.724.790
- aes: 25.683.689 auf 22.447.707
- drugs: 30.815.184 auf 23.298.217
- combinations: 135.489.143 auf 90.500.758

Wenn in einem Bericht ein unerwünschtes Ereignis und/oder ein Medikament mehrfach angegeben worden ist, wurden die entsprechenden Einträge in den *DataFrames* „aes“, „drugs“ und „combinations“ durch die Entfernung der alten Versionen der Berichte ebenfalls beseitigt. Dadurch wurde sichergestellt, dass ein Bericht bei der Signalgenerierung nicht doppelt gezählt wird. Allerdings könnten dadurch auch Informationen zu unerwünschten Ereignissen und/oder Medikamenten verloren gegangen sein.

#### 3.3.3 Speicherung der *DataFrames*

Nachdem die *DataFrames* erstellt und befüllt, die Datentypen der Spalten angepasst und die alten Versionen der Berichte entfernt worden sind, wurden die *DataFrames* mit *pickle* gespeichert. Die *DataFrames* konnten dann zur Analyse mit *pickle* wieder geladen werden. Dadurch musste der Import der Daten nicht erneut durchgeführt werden, wenn die *DataFrames* bei der Analyse verwendet wurden.

## 4 Signalgenerierung

Poluzzi et al. sehen die Signalerkennung als eines der Hauptziele der Pharmakovigilanz. D.h. die Identifikation potenzieller Medikament-Ereignis-Assoziationen, die aufgrund ihrer Art, Schwere und/oder Häufigkeit neu sein könnten [24]. Delamothe definiert ein Signal als die gemeldete Information über einen möglichen kausalen Zusammenhang zwischen einem unerwünschten Ereignis und einem Medikament, wobei der Zusammenhang zuvor unbekannt oder unvollständig dokumentiert war. Normalerweise ist mehr als ein einzelner Bericht erforderlich, um ein Signal zu generieren, abhängig von der Schwere des Ereignisses und der Qualität der Informationen [14]. Bate et al. betonen, dass ein Signal nicht richtig oder falsch ist. Es ist lediglich ein Hinweis auf ein mögliches Problem mit dem Ziel, potenzielle Probleme von Medikamenten hervorzuheben, die in klinischen Studien nicht entdeckt wurden [9].

### 4.1 Grundlagen

Die Grundlagen der Signalgenerierung werden anhand der Medikament-Ereignis-Assoziation Ciprofloxacin und Sehnenentzündung erklärt. Der Hintergrund der Auswahl dieser Kombination wird in Abschnitt 4.2 erläutert. Die Signalgenerierung wurde in dem Notebook *sg\_ciprofloxacin\_tendonitis* durchgeführt. Die dabei verwendeten Funktionen zur Signalgenerierung wurden im Modul *contingency\_table* implementiert. Um zu beurteilen, ob die Anzahl der gemeldeten Berichte die zu erwarteten Werte übersteigt, gibt es für Evans et al. zwei akzeptierte Ansätze [18]. Der erste Ansatz beinhaltet die Verwendung von Denominator-Daten bezogen auf die Verwendung des Medikaments und die darauf basierende Berechnung von Meldequoten. Dazu wird oft die Anzahl der Verschreibungen oder Verkäufe verwendet. Formel 4.1 zeigt exemplarisch die Berechnung einer Meldequote mit der Anzahl von Verschreibungen als Denominator. Die Anzahl der Berichte bezieht sich auf die Anzahl der Berichte, die das Medikament und das unerwünschte Ereignis beinhalten.

$$\text{Meldequote} = \text{Anzahl Berichte} / \text{Anzahl Verschreibungen} \quad (4.1)$$

Der Vergleich von Medikamenten auf Basis von Meldequoten kann laut Evans et al. durch verschiedene Faktoren beeinflusst werden, wie die erhöhte Meldung neuer Medikamente und die Effekte von Jahreszeit und der Publizität eines unerwünschten Ereignisses. Kleinen Unterschieden zwischen Medikamenten bei Meldequoten kann nur wenig Gewicht beigemessen werden, während große Unterschiede ein untersuchenswertes Signal darstellen können. Auf Basis des FAERS-Datensatzes kann nur das Verhältnis der Berichte berechnet werden, in denen das Medikament und das unerwünschte Ereignis gemeldet wurde, zu den Berichten, in denen das Medikament gemeldet wurde. Informationen über die Verschreibungs- oder Verkaufszahlen der Medikamente beinhaltet das FAERS nicht. Formel 4.2 zeigt die Berechnung dieser Meldequote. ME steht dabei für die Anzahl der Berichte mit dem unerwünschten Ereignis und dem Medikament und M für die Anzahl der Berichte mit dem Medikament.

$$\text{Meldequote} = \text{ME} / \text{M} \quad (4.2)$$

In Python kann die Meldequote mit der Funktion *rate* berechnet werden. Ciprofloxacin und Sehnenentzündung wurden in 621 Berichten und Ciprofloxacin in 25.829 Berichten gemeldet. Die Meldequote von Sehnenentzündung bei Ciprofloxacin beträgt somit 0,024. In 2,4% der Berichte, in denen Ciprofloxacin gemeldet worden ist, wurde also Sehnenentzündung als unerwünschtes Ereignis gemeldet.

Der zweite Ansatz besteht nach Evans et al. darin, die Gesamtzahl der Berichte für ein Medikament als Denominator zu verwenden und den Anteil der Berichte mit dem unerwünschten Ereignis zu berechnen. Dieser Ansatz hat einige Vorteile gegenüber Meldequoten. Erstens werden keine externen Daten benötigt und die Beschränkungen solcher Daten treffen nicht zu. Dies erlaubt die Berechnung auf Basis des FAERS-Datensatzes. Zweitens kann dieser Ansatz der Verzerrung von Meldequoten durch unterschiedliche Meldung entgegenwirken. Wenn z.B. die Gesamtzahl der Meldungen aufgrund der Verzerrung durch die Neuheit eines Medikaments hoch ist, muss dies nicht unbedingt den Anteil des unerwünschten Ereignisses bei diesem Medikament beeinflussen.

Wie Poluzzi et al. beschreiben, durchsuchen alle diese Methoden zur Signalgenerierung die Daten nach Disproportionalitäten und geben an, inwieweit das gemeldete unerwünschte

Ereignis mit dem verdächtigen Medikament in Verbindung steht, verglichen mit allen anderen Medikamenten (oder einer Untergruppe von Medikamenten) in den Daten [24]. Poluzzi et al. betonen dabei, dass das Vorhandensein eines statistisch signifikanten Ergebnisses nicht notwendigerweise einen tatsächlichen kausalen Zusammenhang zwischen dem unerwünschten Ereignis und dem Medikament impliziert, und das Fehlen eines statistisch signifikanten Ergebnisses nicht notwendigerweise den möglichen Zusammenhang widerlegt.

Poluzzi et al. unterteilen die Methoden in Häufigkeitsmethoden und Bayes'sche Methoden. Häufigkeitsmethoden sind besonders ansprechend und daher weit verbreitet, da sie relativ einfach zu verstehen, zu interpretieren und zu berechnen sind. Die Häufigkeitsmethoden basieren alle auf einer  $2 \times 2$  Kontingenztafel. Bayes'sche Methoden wie der Multi-item Gamma Poisson Shrinker (MGPS) von DuMouchel [15] und das Bayesian Confidence Propagation Neural Network (BCPNN) von Bate et al. [9] basieren, auf dem Bayes'schen Gesetz zur Abschätzung der Wahrscheinlichkeit (a-posteriori-Wahrscheinlichkeit), dass das verdächtige Ereignis bei der Anwendung des verdächtigen Medikaments auftritt. Der MGPS wird von der FDA im FAERS zur Signalgenerierung eingesetzt. Die WHO verwendet in VigiBase das BCPNN zur Generierung von Signalen. Die Bayes'schen Methoden wurden nicht implementiert. Auf deren Interpretation und Berechnung wird daher hier nicht genauer eingegangen.

### 4.1.1 Kontingenztafel

Jede Analyse einer Verbindung zwischen der Einnahme eines Medikaments und eines unerwünschten Ereignisses basiert auf einer  $2 \times 2$  Kontingenztafel. Anstatt sich nur auf den Zusammenhang zwischen einem Medikament und einem unerwünschten Ereignis zu konzentrieren, können auch Assoziationen mehrerer Elemente analysiert werden. Z.B. zwei Medikamente und ein unerwünschtes Ereignis oder ein Medikament und zwei unerwünschte Ereignisse. Dazu muss die Kontingenztafel erweitert werden. In dieser Arbeit wurden nur  $2 \times 2$  Kontingenztafeln implementiert und daher nur Assoziationen von einem Medikament und einem unerwünschten Ereignis analysiert. Tabelle 4.1 zeigt die  $2 \times 2$  Kontingenztafel nach Böhm [12], die dafür verwendet wurde. Großbuchstaben kennzeichnen die Einnahme des Medikaments (M) oder das Auftreten des unerwünschten Ereignisses (E). Kleinbuchstaben signalisieren, dass das Medikament nicht eingenommen wurde (m), oder das unerwünschte Ereignis nicht aufgetreten ist (e). N ist die Summe aller Berichte.

	Medikament eingenommen	Medikament nicht eingenommen	Summen
Unerwünschtes Ereignis aufgetreten	ME	mE	E
Unerwünschtes Ereignis nicht aufgetreten	Me	me	e
Summen	M	m	N

Tabelle 4.1: 2 x 2 Kontingenztabelle nach Böhm [12]

In Python kann eine 2 x 2 Kontingenztabelle mit der Klasse *ContingencyTable* aus dem Modul *contingency\_table* erstellt werden. Dazu muss das Medikament, das unerwünschte Ereignis und der *DataFrame* mit den Kombinationen übergeben werden. Unter Nutzung aller Spalten des Kombinations-*DataFrames* dauerte das Erstellen der Kontingenztabelle für Ciprofloxacin und Sehnenentzündung drei Minuten und zwölf Sekunden. Daher wurde der Kombinations-*DataFrame* auf die drei für das Erstellen der Kontingenztabelle notwendigen Spalten reduziert. Die Spalten mit den Identifikationsnummern der Berichte, den Medikamenten und den unerwünschten Ereignissen. Dadurch wurde die Rechenzeit auf eine Minute und 33 Sekunden verkürzt. Die Kontingenztabelle kann mit der Funktion *show* als *DataFrame* angezeigt werden lassen. Das Attribut *time* gibt die Zeit in Sekunden an, die das Erstellen der Kontingenztabelle gedauert hat. Tabelle 4.2 zeigt die Kontingenztabelle von Ciprofloxacin und Sehnenentzündung.

	Ciprofloxacin eingenommen	Ciprofloxacin nicht eingenommen	Summen
Sehnenentzündung aufgetreten	621	5.583	6.204
Sehnenentzündung nicht aufgetreten	25.208	7.693.378	7.718.586
Summen	25.829	7.698.961	7.724.790

Tabelle 4.2: 2 x 2 Kontingenztabelle von Ciprofloxacin und Sehnenentzündung

Die erste Wahl, um solch eine Kontingenztabelle zu analysieren, sind für Böhm die Häufigkeitsmethoden der Disproportionalitätsanalyse (DPA), welche alle auf der erwarteten und beobachteten Anzahl an Berichten basieren [12]. Auf Basis einer Häufigkeitsmethode

alleine wird normalerweise kein Signal generiert. Zusätzlich muss getestet werden, ob die Häufigkeitsmethode statistisch signifikant ist. Dazu kann der Chi-Quadrat-Test genutzt werden, der in Abschnitt 4.1.5 erläutert wird. Alternativ kann auch ein Konfidenzintervall um die Häufigkeitsmethode berechnet werden. Dessen Berechnung basiert auf der Standardabweichung der Häufigkeitsmethode. Die Berechnung der Standardabweichung unterscheidet sich zwischen den Häufigkeitsmethoden. Sie wird daher zusätzlich bei jeder Häufigkeitsmethode neben der Berechnung der Häufigkeitsmethode angegeben. Die Berechnung eines Konfidenzintervalls auf Basis der Standardabweichung unterscheidet sich hingegen nicht zwischen den Häufigkeitsmethoden. Dessen Berechnung wird daher in Abschnitt 4.1.6 erläutert, nachdem alle Häufigkeitsmethoden vorgestellt wurden.

### 4.1.2 Relative Reporting Ratio

Eine Häufigkeitsmethode zur Signalgenerierung ist die Relative Reporting Ratio (RRR). Böhm beschreibt sie als die Wahrscheinlichkeit des Auftretens eines unerwünschten Ereignisses in einer Teilpopulation im Verhältnis zur Wahrscheinlichkeit des Auftretens des unerwünschten Ereignisses in der Gesamtpopulation an [12]. Formel 4.3 zeigt die Berechnung der RRR und Formel 4.4 die Berechnung der Standardabweichung der RRR. In Python kann die RRR mit der Funktion `rrr` und die Standardabweichung der RRR mit der Funktion `rrr_sd` berechnet werden. Bei Ciprofloxacin und Sehnenentzündung beträgt die RRR 29,94.

$$RRR = ME * N / (M / E) \quad (4.3)$$

$$S(RRR) = \sqrt{Me / (ME * M) + e / (E * N)} \quad (4.4)$$

### 4.1.3 Proportional Reporting Ratio

Die Proportional Reporting Ratio (PRR) wurde von Evans et al. als eine statistische Methode zur Signalgenerierung entwickelt [18]. Sie beschreibt das Verhältnis des Auftretens eines unerwünschten Ereignisses in der Population, die das Medikament eingenommen hat, zur Population, die das Medikament nicht eingenommen hat. Die PRR wird z.B. in der UK Yellow Card Database zur Generierung von Signalen eingesetzt. Formel 4.5 zeigt

die Berechnung der PRR und Formel 4.6 die Berechnung der Standardabweichung der PRR. In Python kann mit der Funktion *pr*r die PRR und mit der Funktion *pr*r\_*sd* die Standardabweichung der PRR berechnet werden. Bei Ciprofloxacin und Sehnenentzündung beträgt die PRR 33,15. Sie ist also deutlich größer als die RRR (29,94).

$$PRR = (ME/M)/(mE/m) \quad (4.5)$$

$$S(PRR) = \sqrt{Me/(ME * M) + me/(mE * m)} \quad (4.6)$$

### 4.1.4 Reporting Odds Ratio

Eine weitere Häufigkeitsmethode zur Signalgenerierung ist die Reporting Odds Ratio (ROR). Nach Böhm gibt sie das Verhältnis der Chancen (Odds) an, dass ein unerwünschtes Ereignis bei der Population auftritt, die das Medikament eingenommen hat, zur Population, die das Medikament nicht eingenommen hat. Die Chance ist das Verhältnis der Wahrscheinlichkeit, dass ein unerwünschtes Ereignis auftritt, zur Wahrscheinlichkeit, dass das Ereignis nicht auftritt [12]. Laut Egberts et al. wird die ROR z.B. von der Netherlands Pharmacovigilance Foundation Lareb zur Generierung von Signalen eingesetzt [16]. Formel 4.7 zeigt die Berechnung der ROR und Formel 4.8 die Berechnung der Standardabweichung der ROR. In Python kann die ROR mit der Funktion *ror* und die Standardabweichung der ROR mit der Funktion *ror\_sd* berechnet werden. Bei Ciprofloxacin und Sehnenentzündung beträgt die ROR 33,95. Sie ist also ebenfalls deutlich größer als die RRR (29,94) und etwas größer als die PRR (33,15).

$$ROR = (ME/Me)/(mE/me) \quad (4.7)$$

$$S(ROR) = \sqrt{1/ME + 1/Me + 1/mE + 1/me} \quad (4.8)$$

### 4.1.5 Chi-Quadrat-Test mit Yates-Korrektur

Für Böhm dient der Chi-Quadrat-Test dazu, Unabhängigkeit in Kontingenztafeln zu testen. Er kann alternativ zur Berechnung eines Konfidenzintervalls zum Testen der Signifikanz der Häufigkeitsmethode verwendet werden. Je höher das Chi-Quadrat, desto mehr weicht die beobachtete Anzahl an Berichten von der erwarteten Anzahl an Berichten ab. Chi-Quadrat-Werte größer als 3,841 geben eine Signifikanz von  $p \leq 0,05$  an. D.h., es besteht eine Chance von mehr als 95%, dass die beobachtete Anzahl an Berichten tatsächlich von der erwarteten Anzahl an Berichten abweicht. In der Praxis wird ein Chi-Quadrat-Wert von vier ( $p = 0,045$ ) als Grenzwert verwendet. Die Yates-Korrektur ist notwendig, da 2 x 2 Kontingenztafeln nur einen Freiheitsgrad von eins haben [12]. Formel 4.9 zeigt die Berechnung von Chi-Quadrat mit Yates-Korrektur. In Python kann das Chi-Quadrat mit Yates-Korrektur mit der Funktion `χ2` berechnet werden. Bei Ciprofloxacin und Sehnenentzündung beträgt es 17.412,45. Es liegt damit deutlich über dem Grenzwert von vier. Die beobachtete Anzahl an Berichten, in der Sehnenentzündung zusammen mit Ciprofloxacin gemeldet wurde, weicht also stark von der erwarteten Anzahl an Berichten ab.

$$\chi^2\text{-Yates} = \frac{N * (|ME * me - mE * Me| - N/2)^2}{M * m * E * e} \quad (4.9)$$

### 4.1.6 Konfidenzintervall

Die Berechnung eines Konfidenzintervalls kann alternativ zum Chi-Quadrat-Test verwendet werden, um die statistische Signifikanz einer Häufigkeitsmethode zu testen. Wie Böhm beschreibt, ist die Stichprobenverteilung der Häufigkeitsmethoden positiv verzerrt aber nach einer z-Transformation annähernd normalverteilt. Die Konvertierung mit dem natürlichen Logarithmus, die Anwendung von 1,96 mal der Standardabweichung und die Rückkonvertierung auf die Ursprungsskalierung liefert das Konfidenzintervall. Dieses beinhaltet den wahren Wert mit  $p \leq 0,05$ . Für statistische Signifikanz darf das Konfidenzintervall nicht den Wert eins einschließen [12]. Formel 4.10 zeigt die Berechnung eines Konfidenzintervalls um eine Häufigkeitsmethode. H steht für eine der Häufigkeitsmethoden RRR, PRR oder ROR und S(H) für die Standardabweichung der Häufigkeitsmethode H. Um in Python ein Konfidenzintervall um eine Häufigkeitsmethode zu berechnen, muss

der Funktion  $ci$  die Häufigkeitsmethode und die Standardabweichung der Häufigkeitsmethode übergeben werden. Die Funktion berechnet dann die untere und obere Grenze des Konfidenzintervalls und gibt diese als Tupel zurück.

$$KI(H) = e^{\ln H \pm 1,96S(H)} \quad (4.10)$$

### 4.1.7 Signal

Laut Evans et al. basiert die Beurteilung, ob ein Signal vorliegt und seiner Stärke, auf drei Teilinformationen. Der PRR, Chi-Quadrat und der Gesamtzahl der Berichte mit dem Medikament und dem unerwünschten Ereignis. Damit ein Signal generiert wird, muss die PRR mindestens zwei, Chi-Quadrat mindestens vier und die Anzahl der Berichten mit dem Medikament und dem unerwünschten Ereignis mindestens drei betragen. Als eine äquivalente Alternative kann ein Konfidenzintervall um die PRR berechnet werden [18]. Diese Kriterien können auch auf die RRR und ROR angewendet werden. In dieser Arbeit wurde die PRR mit dem Chi-Quadrat-Test verwendet, um Signale zu generieren. Ciprofloxacin und Sehnenentzündung haben ein Signal ausgelöst, da die PRR mit 33,15 über dem Grenzwert von zwei, Chi-Quadrat mit 17.412,45 über dem Grenzwert von vier und die Anzahl der Berichte mit dem Medikament und dem unerwünschten Ereignis mit 621 über dem Grenzwert von drei liegt. Damit sind alle Kriterien von Evans et al. für ein Signal erfüllt. Die alternative Berechnung eines Konfidenzintervalls um die PRR hätte ebenfalls ein Signal generiert, da das Konfidenzintervall um die PRR von 30,54 bis 35,99 geht und damit nicht den Wert eins einschließt.

## 4.2 Beispiel

Um die Verwendung der entwickelten Funktionalitäten zur Signalgenerierung zu demonstrieren, wurden Signale für einige Kombinationen von Medikamenten und Nebenwirkungen generiert. Auf Basis der generierten Signale wurde außerdem eine Nebenwirkung ausgewählt, die weiter analysiert werden sollte. Die Teilpopulation, bei der diese Nebenwirkung bei der Einnahme eines ausgewählten Medikaments aufgetreten ist, wurde in Kapitel 5 gemeinsam mit der Gesamtpopulation in einer explorativen Datenanalyse untersucht und mit dieser verglichen.

Die Medikamente und Nebenwirkungen stammen aus einem Bericht des BfArM vom 2. Mai 2019. In diesem wird vor schweren und langanhaltenden Nebenwirkungen im Bereich Muskeln, Gelenke und Nervensystem bei der Einnahme von Fluorchinolonen gewarnt [10]. Fluorchinolone sind eine Gruppe von Antibiotika, die zur Behandlung von bakteriellen Infektionen, vor allem Harnwegs- und Atemwegsinfektionen eingesetzt werden. Als betroffene Fluorchinolone wurden Ciprofloxacin, Levofloxacin, Lomefloxacin, Moxifloxacin, Norfloxacin, Ofloxacin, Pefloxacin, Prulifloxacin und Rufloxacin aufgeführt. Als Nebenwirkungen wurden Entzündungen oder Risse der Sehnen, Muskelschmerzen oder Muskelschwäche, Gelenkschmerzen oder Gelenkschwellungen, Schwierigkeiten beim Gehen, Gefühle von Nadelstichen oder Kribbeln, brennende Schmerzen, Müdigkeit, Depressionen, Gedächtnisstörungen, Schlafstörungen, Probleme beim Sehen oder Hören und veränderter Geschmacks- oder Geruchssinn genannt [10].

In diesem Kontext kann der Begriff „Nebenwirkung“ an Stelle von „unerwünschtes Ereignis“ verwendet werden, da ein kausaler Zusammenhang zwischen der Einnahme der aufgeführten Fluorchinolone und dem Auftreten der genannten Nebenwirkungen festgestellt wurde. Es wurde bewusst ein Beispiel gewählt, bei dem dieser Zusammenhang bereits belegt wurde, da zum Nachweis eines kausalen Zusammenhangs zwischen der Einnahme eines Medikaments und dem Auftreten einer Nebenwirkung, wie in Abschnitt 2.2 betont, klinische Pharmakologiekenntnisse notwendig sind.

Laut der Dokumentation der FDA werden die unerwünschten Ereignisse im FAERS mit englischen MedDRA PTs beschrieben [29]. Die auf Deutsch beschriebenen Nebenwirkungen aus dem Bericht des BfArM mussten daher in passende englische MedDRA PTs übersetzt werden. Für elf der 16 Nebenwirkungen wurde ein passender MedDRA PT gefunden. Für die anderen fünf Nebenwirkungen konnte hingegen kein passender MedDRA PT gefunden werden. Dementsprechend wurden elf Nebenwirkungen und neun Fluorchinolone analysiert. Die Nebenwirkungen und die entsprechenden Übersetzungen finden sich in Anhang A.1. Nach der Übersetzung wurde für alle 99 Kombinationen von Fluorchinolonen und Nebenwirkungen eine Kontingenztafel erstellt und geprüft, ob die Kombinationen die Kriterien von Evans et al. [18] für ein Signal erfüllen. Das Ergebnis ist in Abbildung 4.1 zu sehen. Tabelle A.1 in Anhang A.2 zeigt die ME-, Chi-Quadrat- und PRR-Werte nach Evans et al. der 24 Kombinationen, die ein Signal ausgelöst haben.

	ciprofloxacin	levofloxacin	lomefloxacin	moxifloxacin	norfloxacin	ofloxacin	pefloxacin	prulifloxacin	rufloxacin
tendonitis	True	True	False	True	True	True	False	False	False
tendon rupture	True	True	False	True	False	True	False	False	False
myalgia	True	True	False	False	True	False	False	False	False
muscle fatigue	True	True	False	False	False	False	False	False	False
arthralgia	True	True	False	False	True	False	False	False	False
joint swelling	True	False	False	False	False	False	False	False	False
burning sensation	True	False	False	False	False	True	False	False	False
fatigue	False	False	False	False	True	False	False	False	False
depression	True	False	False	False	False	True	False	False	False
amnesia	False	False	False	False	False	False	False	False	False
insomnia	False	False	False	False	True	False	False	False	False

Abbildung 4.1: Nach Evans et al. [18] generierte Signale bei Kombinationen von Medikamenten und Nebenwirkungen aus dem BfArM-Bericht

Wie in Abbildung 4.1 zu sehen ist, hat bei den Fluorchinolonen Ciprofloxacin mit acht Signalen die meisten Signale generiert. Levofloxacin und Norfloxacin haben mit jeweils fünf Signalen die zweitmeisten Signale ausgelöst. Ofloxacin hat vier und Moxifloxacin zwei Signale verursacht. Lomefloxacin, Pefloxacin, Prulifloxacin und Rufloxacin haben zu gar keinen Signalen geführt.

Bei den Nebenwirkungen hat mit fünf Signalen Sehnenentzündung (tendonitis) die meisten Signale generiert. Die zweitmeisten Signale hat Sehnenriss (tendon rupture) mit vier Signalen ausgelöst. Muskelschmerzen (myalgia) und Gelenkschmerzen (arthralgia) haben jeweils drei Signale verursacht. Zu jeweils zwei Signalen haben Muskelschwäche (muscle fatigue), brennende Schmerzen (burning sensation) und Depression (depression) geführt. Nur ein Signal haben jeweils Gelenkschwellung (joint swelling), Müdigkeit (fatigue) und Schlafstörung (insomnia) generiert. Gedächtnisstörung (amnesia) hat als einzige Nebenwirkung gar kein Signal ausgelöst.

Da Ciprofloxacin bei den Fluorchinolonen und Sehnenentzündung bei den Nebenwirkungen am meisten Signale generiert hat, wurden für die Teilpopulation Berichte ausgewählt, in denen diese zusammen gemeldet wurden. Diese Auswahl ist letztendlich arbiträr und es hätte auch jede andere Kombination von einem Fluorchinolon mit einer Nebenwirkung aus Abbildung 4.1 ausgewählt werden können, die ein untersuchenswertes Signal ausgelöst hat.

## 5 Explorative Datenanalyse

Bei der explorativen Datenanalyse (EDA) wurden Statistiken über den FAERS-Datensatz erstellt und visualisiert. Dabei sollten die Eigenschaften des Datensatzes veranschaulicht werden. Ausgewählte Methoden wurden auch auf die in Abschnitt 4.2 ausgewählte Teilpopulation angewendet, bei der das unerwünschte Ereignis Sehnenentzündung bei der Einnahme des Medikaments Ciprofloxacin aufgetreten ist. Ziel davon war es, zu prüfen, ob der durch das Signal suggerierte Zusammenhang zwischen Ciprofloxacin und Sehnenentzündung durch die Statistiken über einzelne Attribute bekräftigt wird. Dafür wurde die Teilpopulation auch mit der Gesamtpopulation verglichen und Unterschiede zwischen den beiden diskutiert. Weiterhin sollte der Vergleich zeigen, ob bestimmte Untergruppen von Patienten möglicherweise besonders häufig von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen sind. Am Beispiel des Geschlechts der Patienten wurde außerdem verdeutlicht, warum Teilpopulationen nicht isoliert betrachtet werden dürfen, um fundierte Aussagen über diese treffen zu können.

### 5.1 Implementierung

Die EDA wurde getrennt vom Import und der Aufbereitung der Daten und von der Signalgenerierung implementiert. Sie wurde in den Notebooks *reports*, *patients*, *aes* und *drugs* durchgeführt. Dabei wurde jeweils ein Notebook für die Gesamtpopulation und ein Notebook für die Teilpopulation erstellt. Die Notebooks nutzen für die EDA verschiedene externe Bibliotheken, darunter vor allem *pandas* [20]. Neben den externen Bibliotheken wurden auch Module verwendet, die im Zuge dieser Arbeit entwickelt wurden. Diese Module und weitere verwendete externe Bibliotheken werden im Folgenden beschrieben.

Das Modul *analysis* stellt verschiedene Funktionen zur Analyse der Daten zur Verfügung. Dazu gehört die bereits in Abschnitt 3.3.1 beschriebene Funktion *null* zur Analyse von *Nullwerten*. Es gibt außerdem eine Klasse *Timer*, die die Messung der Rechenzeit von Codeabschnitten ermöglicht. Diese wurde für sämtliche Zeitmessungen in diesem Projekt verwendet. Weitere Funktionen umfassen u.a. die Berechnung von Quartalen auf Basis von Datumsangaben (*quarter*, *only\_quarter* und *pretty\_quarter*).

Das Modul *charts* ermöglicht das Erstellen von Balkendiagrammen, die die Anzahl der Berichte pro Monat zeigen (*reports\_per\_quarter\_barplot*) und das Erstellen von generischen gestapelten Balkendiagrammen (*stacked\_barplot*). Weitere Diagramme wurden in den Notebooks erstellt und sind noch nicht generisch. Eine zukünftige Integration der Diagramme aus den Notebooks in das Modul *charts* ist aber geplant. Für die Erstellung der Diagramme wurden drei verschiedene Bibliotheken verwendet. Die Balken- und Liniendiagramme wurden mit *seaborn* [30] erstellt. Für die Kreis- und Donut-Diagramme wurde *matplotlib* [19] verwendet. Die divergierenden Balkendiagramme und die Choroplethenkarten wurden mit *plotly* [23] erzeugt.

Das Modul *charts* bietet außerdem die drei Funktionen *savefig*, *savemap* und *save\_diverging\_barplot* zum Speichern von Diagrammen an. Diese werden auch von den Notebooks zum Speichern von Diagrammen genutzt. Diese Funktionen stellen u.a. sicher, dass Diagramme gleicher Art mit dem gleichen Seitenverhältnis und der gleichen Auflösung gespeichert werden. Das Modul *formatting* bietet verschiedene Funktionen zum Formatieren einzelner Werte, sowie von Diagrammen und *DataFrames* an.

Wie in Abschnitt 3.3.1 beschrieben wurde, sind codierte Attribute beim Import aus Speicherplatzgründen nicht in ihre wörtliche Bedeutung übersetzt worden. Da die Bedeutung der codierten Attribute aber bei der EDA gebraucht wurde, bietet das Modul *decoding* Funktionen zum Decodieren codierter Attribute an. Bei den Bedeutungen der codierten Werte wurde sich größtenteils an das XML\_NTS-Dokument für Exporte aus dem FAERS der FDA gehalten [29]. Die Bedeutungen wurden teilweise vereinfacht und um Ausprägungen für *Nullwerte* ergänzt. Da die Implementierung auf Englisch durchgeführt, diese Arbeit jedoch auf Deutsch verfasst wurde, mussten die Bedeutungen für die Diagramme von Englisch auf Deutsch übersetzt werden. Dazu stehen Funktionen im Modul *labelmappings* zur Verfügung. Das Modul *colormappings* beinhaltet die Zuordnung von Farben zu Attributausprägungen. Diese werden zur Visualisierung der Attribute in den Diagrammen verwendet.

## 5.2 Berichte

Der analysierte Datensatz umfasst die Exporte vom vierten Quartal 2012 bis zum zweiten Quartal 2019 und beinhaltet 7.724.790 Berichte. Bei der Interpretation dieser Zahl sei auf die Entfernung der alten Versionen der Berichte in Abschnitt 3.3.2 und die damit einhergehenden Veränderungen des Datensatzes hingewiesen. Ein Bericht kann mehrere Versionen haben. Jeder Bericht beinhaltet daher zwei Datumsangaben, die eine zeitliche Einordnung des Berichts ermöglichen. Das erste Datum ist der Tag, an dem die erste Version des Berichts empfangen wurde [29]. Dieses Datum kann weit vor der Inbetriebnahme des FAERS liegen, wenn dieser Bericht nach Einführung des FAERS noch einmal aktualisiert wurde. Als zweites Datum wird daher der Tag angegeben, an dem der Bericht das letzte Mal aktualisiert wurde [29]. Bei den Analysen wurde ausschließlich das Datum verwendet, an dem die Berichte das letzte Mal aktualisiert wurden. Dadurch sollten eigentlich nur Berichte vorliegen, die nach Inbetriebnahme des FAERS das erste Mal empfangen oder aktualisiert wurden.

Die FDA hat im September 2012 vom Legacy Adverse Event Reporting System (LAERS) auf das neue FDA Adverse Reporting System (FAERS) gewechselt. Laut FDA beinhaltet der erste Export die Daten für September 2012 und das vierte Quartal 2012 im Zeitraum vom 28. August bis zum 31. Dezember 2012 [29]. Dies ist allerdings offenbar nicht ganz korrekt, da der erste Bericht bereits am 6. Januar 2012 empfangen wurde. Der letzte Bericht wurde am 30. Juni 2019 übermittelt.

Durchschnittlich wurden 965.599 Berichte pro Jahr und 257.493 Berichte pro Quartal empfangen. Im Durchschnitt die meisten Berichte wurden mit 267.698 in den ersten Quartalen übermittelt. In den zweiten und dritten Quartalen wurden mit durchschnittlich 252.070 und 251.448 ähnlich viele Berichte empfangen. In den vierten Quartalen wurden im Durchschnitt 258.073 Berichte übermittelt. In den Wintermonaten wurden also durchschnittlich mehr Berichte empfangen als in den Sommermonaten. Mögliche Erklärungen dafür sind, dass die Menschen in den Wintermonaten aufgrund der kälteren Temperaturen anfälliger für Krankheiten sind. Dementsprechend werden mehr Medikamente konsumiert, was zu mehr Berichten über unerwünschte Ereignisse führt. Eine weitere Erklärung ist, dass die Patienten in den Wintermonaten eher dazu geneigt sind, unerwünschte Ereignisse zu melden, als in den Sommermonaten. Dies könnte zu einer höheren Anzahl übermittelter Berichte führen, obwohl nicht zwangsläufig mehr Medikamente eingenommen wurden. Auch eine Möglichkeit ist, dass in der Wintermonaten mehr unerwünschte Ereignisse aufgetreten sind.

In Abbildung 5.1 sind die empfangenen Berichte pro Quartal dargestellt. Wie darin zu erkennen ist, wurden vor dem vierten Quartal 2012 nur sehr wenige Berichte empfangen. Erst mit Einführung des FAERS in vierten Quartal 2012 stieg die Anzahl übermittelter Berichte deutlich an. Der Höhepunkt wurde im zweiten Quartal 2019 mit 431.000 Berichten erreicht. Die wenigsten Berichte wurden mit unter 1.000 im ersten Quartal 2012 empfangen. Nach Einführung des FAERS wurden in zweiten Quartal 2013 mit 163.000 die wenigsten Berichte übermittelt. Insgesamt lässt sich ein leichter Aufwärtstrend in der Anzahl der empfangenen Berichte pro Quartal beobachten. Dieser wird vermutlich durch eine steigende Meldequote verursacht. Eher unwahrscheinlich ist das Auftreten von mehr unerwünschten Ereignissen.

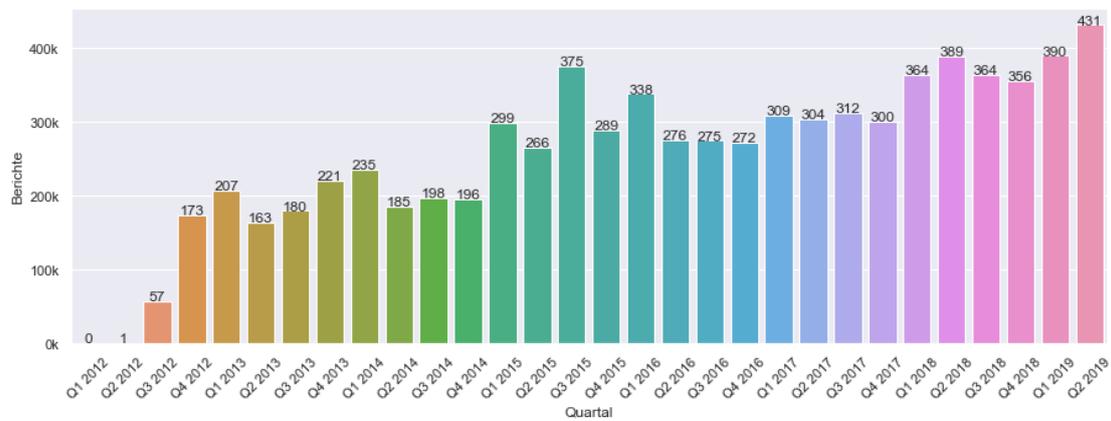


Abbildung 5.1: Berichte der Gesamtpopulation pro Quartal

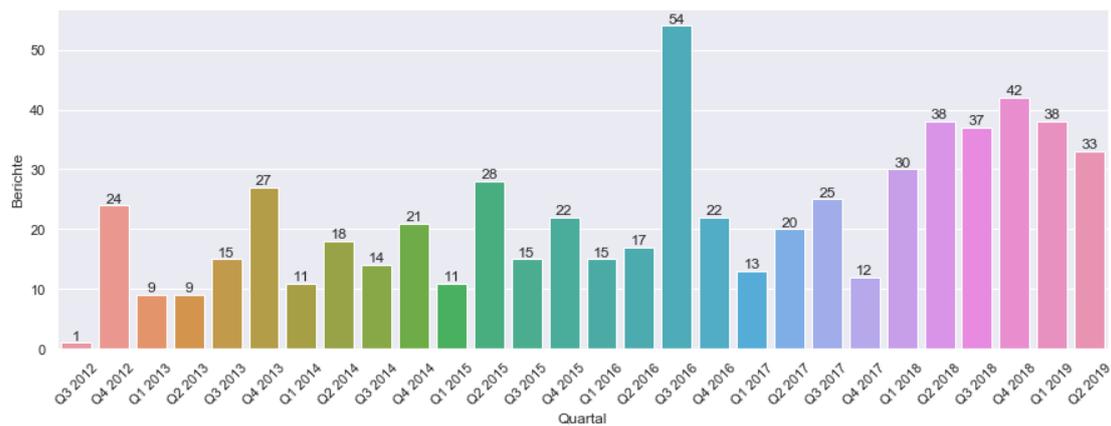


Abbildung 5.2: Berichte der Teilpopulation pro Quartal

Abbildung 5.2 zeigt die Anzahl empfangener Berichte der Teilpopulation pro Quartal. Auffällig viele Berichte wurden im dritten Quartal 2016 empfangen. Viele Berichte wurden außerdem im Zeitraum des ersten Quartals 2018 bis zum zweiten Quartal 2019 übermittelt. Auf den möglichen Einfluss des BfArM-Berichts auf die Anzahl der aus Deutschland übermittelten Berichte wird in Abschnitt 5.2.2 eingegangen.

### 5.2.1 Neue und alte Berichte

Wenn ein neuer Bericht eingeschickt wird, erhält er eine eindeutige Identifikationsnummer und eine Versionsnummer. Die Versionsnummer startet bei eins. Wenn ein vorhandener Bericht in einem neuen Export aktualisiert wird, wird die Versionsnummer um eins inkrementiert. Die Identifikationsnummer bleibt gleich. In einem Export ist immer nur die aktuellste Version eines Berichts enthalten. Ein neuer Bericht hat also die Versionsnummer eins und alte Berichte eine Versionsnummer größer eins [29]. Die meisten Versionen hat ein Bericht vom 23. August 2013. Dieser wurde zuletzt am 11. Juli 2017 zum 126. Mal aktualisiert. Durchschnittlich wird ein Bericht 1,46 mal aktualisiert. Die meisten Berichte haben nur eine Version.

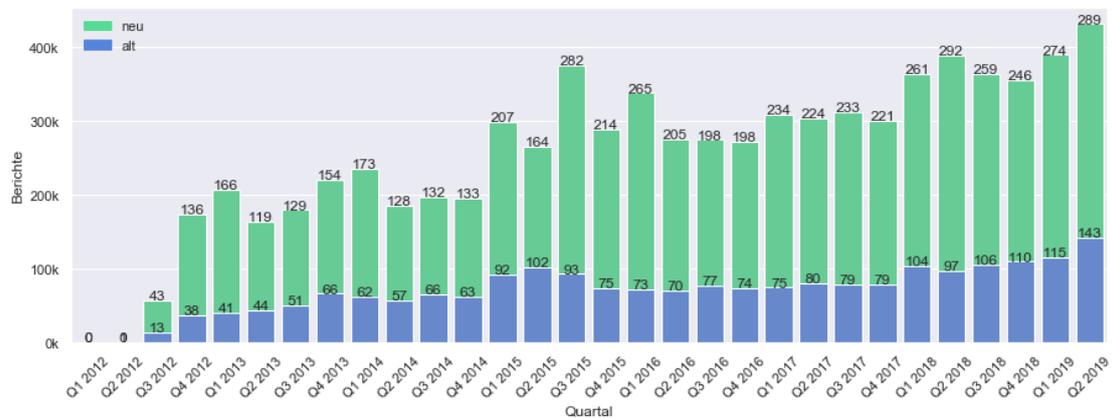


Abbildung 5.3: Neue und alte Berichte der Gesamtpopulation

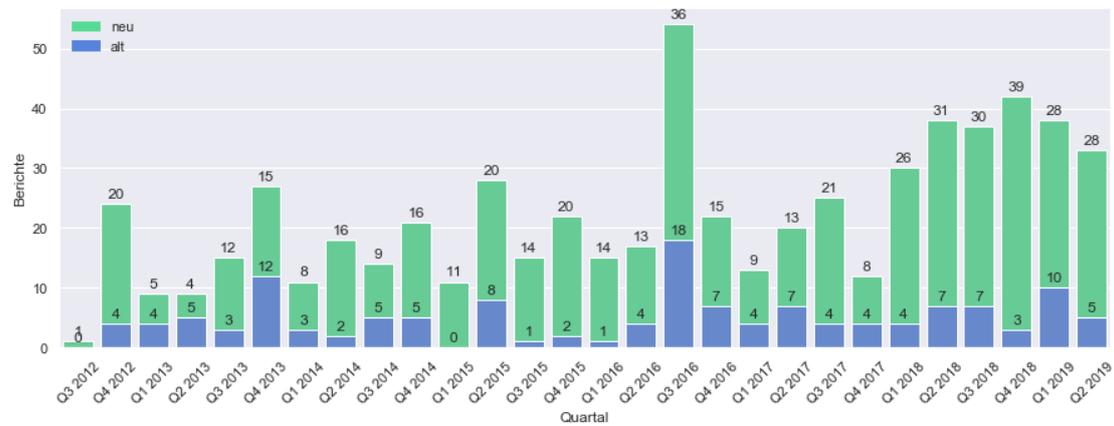


Abbildung 5.4: Neue und alte Berichte der Teilpopulation

In Abbildung 5.3 ist die Anzahl neuer und alter Berichte pro Quartal für die Gesamtpopulation dargestellt. Wie darin zu erkennen ist, überwiegt die Anzahl neuer Berichte in allen Quartalen deutlich. Bei der Teilpopulation zeigt sich ein ähnliches Bild, wie in Abbildung 5.4 zu erkennen ist. In einigen Quartalen überwiegt zwar der Anteil alter Berichte, dies kann allerdings der geringen Anzahl Berichte bei der Teilpopulation geschuldet sein. Auffällig ist, dass im dritten Quartal 2016 nicht nur sehr viele neue Berichte mit Ciprofloxacin und Sehnenentzündung empfangen wurden, sondern auch viele alte Berichte aktualisiert wurden.

### 5.2.2 Herkunftsländer der Berichte

Mit dem in Abschnitt 3.2.1 beschriebenen Verfahren konnte das Herkunftsland für 99,94% der Berichte bestimmt werden. Es wurden Berichte aus 227 unterschiedlichen Ländern empfangen. Insgesamt gibt es bei der verwendeten Codierung 249 mögliche Länder [3]. In 5.5a sind die zehn häufigsten Herkunftsländer der Berichte dargestellt. Wie darin zu erkennen ist, stammen mit 73,36% die meisten Berichte aus den Vereinigten Staaten. Dies war zu erwarten, da die FDA eine amerikanische Bundesbehörde ist. Unterschiede zwischen anderen Ländern sind Abbildung 5.5a kaum zu erkennen. Daher wurden in Abbildung 5.5b die zehn häufigsten Herkunftsländer ohne die Vereinigten Staaten dargestellt. Die Anteile beziehen sich darin nur auf Berichte aus Ländern außerhalb der Vereinigten Staaten.

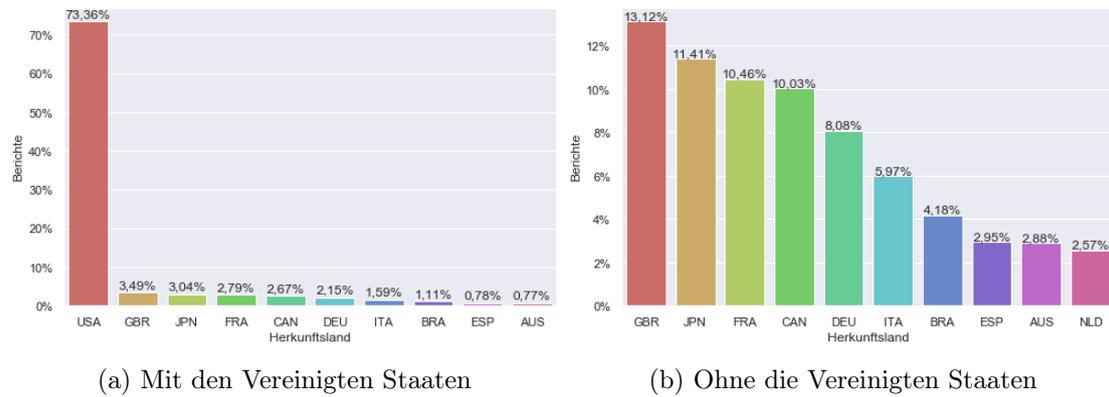


Abbildung 5.5: Zehn häufigste Herkunftsländer der Berichte

Wie in Abbildung 5.5b zu sehen ist, stammen viele ausländische Berichte aus westeuropäischen Ländern wie Großbritannien, Frankreich und Deutschland. Auch aus Japan, Kanada und Brasilien wurde ein erheblicher Teil der ausländischen Berichte übermittelt.

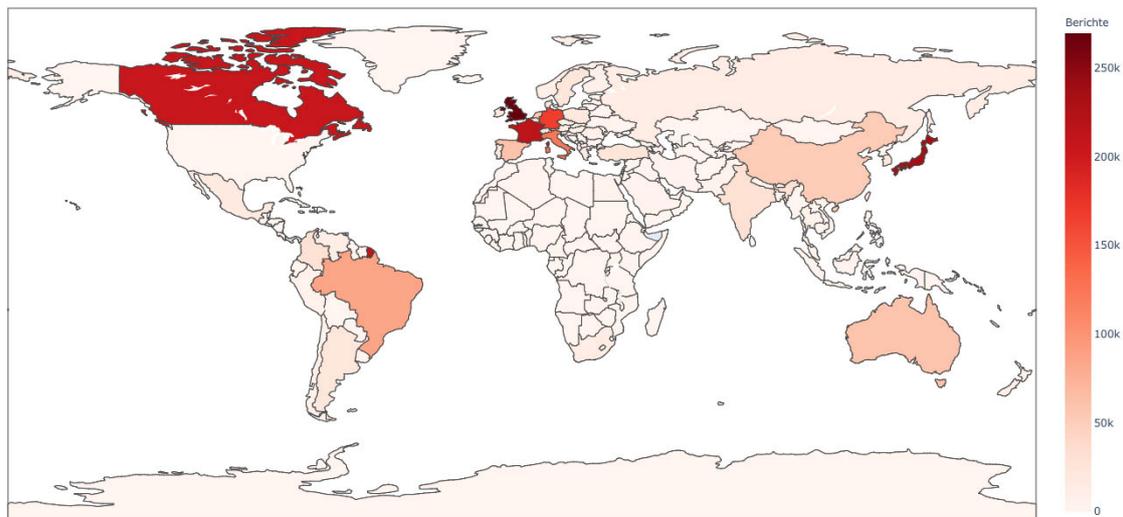


Abbildung 5.6: Herkunftsländer der Berichte der Gesamtpopulation

In Abbildung 5.6 ist die Anzahl übermittelter Berichten pro Land dargestellt. Die Berichte aus den Vereinigten Staaten wurden dabei nicht dargestellt, da aus diesen mit 73,36% der größte Anteil der Berichte stammt. Unterschiede zwischen anderen Ländern wären auf der Karte daher kaum ausmachbar gewesen. Neben den bereits genannten Ländern scheint auch eine große Anzahl Berichte aus Französisch-Guayana nördlich von Brasilien

zu stammen. Das liegt allerdings daran, dass Französisch-Guayana ein Überseedépartement von Frankreich ist. Französisch-Guayana erhält daher die gleiche Einfärbung wie Frankreich und aus dem stammt mit 10,46% ein großer Teil der ausländischen Berichte.



Abbildung 5.7: Herkunftsländer der Berichte der Teilpopulation

Abbildung 5.7 zeigt die Herkunftsländer der Berichte der Teilpopulation. Es wurden Berichte aus 19 unterschiedlichen Ländern empfangen. Die meisten Berichte der Teilpopulation stammen mit 42,56% aus Großbritannien und nicht aus den Vereinigten Staaten, wie aufgrund der Gesamtpopulation zu erwarten gewesen wäre. Aus den Vereinigten Staaten wurden mit 35,11% nur am zweitmeisten Berichte übermittelt. 11,49% der Berichte kommen aus Deutschland. Der BfArM-Bericht wurde am 2. Mai 2019, also im zweiten Quartal 2019 veröffentlicht. 45,45% der Berichte, die im zweiten Quartal 2019 empfangen wurden, stammen aus Deutschland. Im Quartal davor wurden nur 21,05% der Berichte aus Deutschland übermittelt. Die hohe Anzahl deutscher Berichte im zweiten Quartal 2019 wurde also wahrscheinlich durch den BfArM-Bericht initiiert.

### 5.2.3 Berichtstypen

Wie in Abbildung 5.8a und Abbildung 5.8b zu erkennen ist, sind sowohl bei der Gesamtpopulation, als auch bei der Teilpopulation die meisten Berichte spontan. Bei der Teilpopulation sind mit 91,95% fast alle Berichte spontan, bei der Gesamtpopulation liegt der Anteil der spontanen Berichte bei 75,18%. Während bei der Gesamtpopulation mit 20,36% etwa ein Fünftel der Berichte aus Studien stammt, kommen bei der Teilpopulation nur fünf Berichte aus Studien und machen somit nur einen Anteil von 0,81% aus.

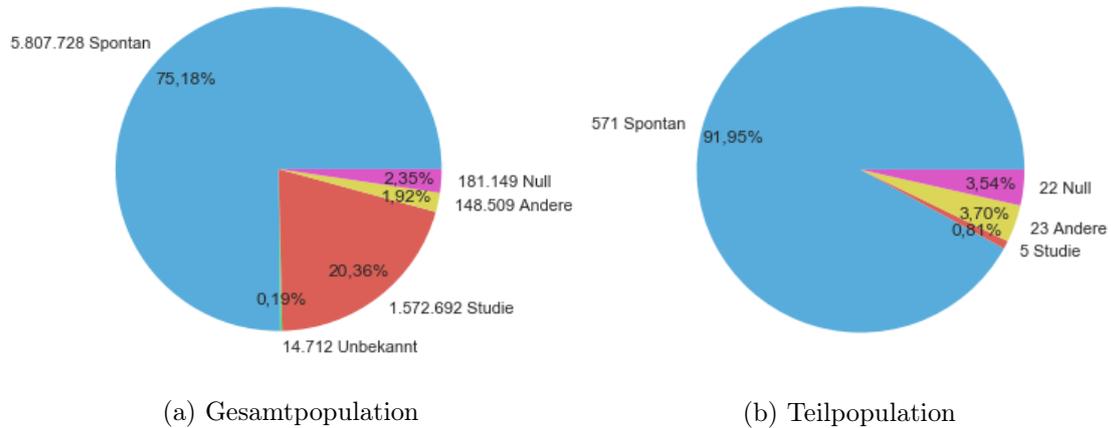


Abbildung 5.8: Berichtstypen

### 5.2.4 Ernsthaftigkeit der Berichte

Die Ernsthaftigkeit der Berichte bezieht sich auf die Ernsthaftigkeit des medizinischen Zustands der Patienten. Sie wurde in allen Berichten angegeben. Bei der Ernsthaftigkeit des medizinischen Zustands des Patienten wird zwischen ernsthaft und nicht ernsthaft unterschieden. Ernsthafte medizinische Zustände sind in „Tod“, „lebensbedrohlich“, „Krankenhausaufenthalt“, „Behinderung“, „Geburtsfehler“ und „anderer medizinisch wichtiger Zustand“ untergliedert [29]. Wie in Abbildung 5.9a zu sehen, ist der medizinische Zustand eines Großteils der Patienten nicht ernsthaft (45,59%). Selten sind Geburtsfehler (0,26%), Behinderungen (1,09%) und lebensbedrohliche Zustände (2,06%). Häufiger sind Krankenhausaufenthalte (18,45%) und andere medizinisch wichtige Zustände (23,18%). 9,38% der Patienten sind verstorben.

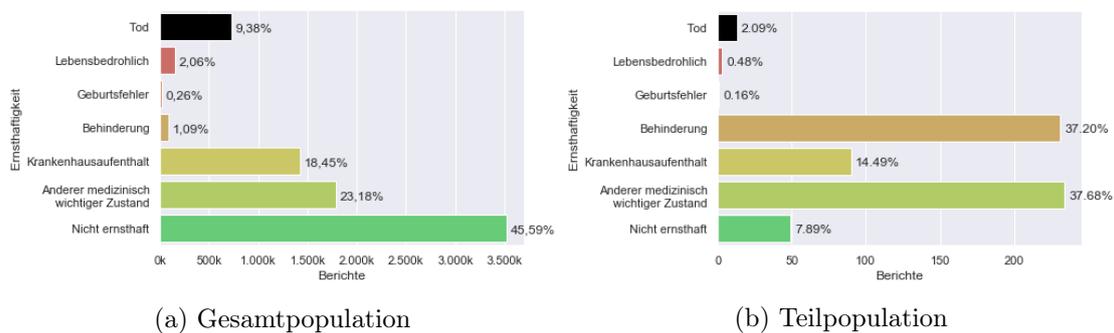


Abbildung 5.9: Ernsthaftigkeit der Berichte

Wie Abbildung 5.9b zeigt, sieht dies bei der Teilpopulation anders aus. Bei nur 7,89% der Patienten ist der medizinische Zustand nicht ernsthaft. 2,09% der Patienten sind verstorben und 0,48% der Patienten waren in einem lebensbedrohlichen Zustand. 0,18% der Patienten haben einen Geburtsfehler erlitten. Diese niedrigen Werte waren bei dem unerwünschten Ereignis Sehnenentzündung zu erwarten und sind wahrscheinlich auch nicht auf dieses zurückzuführen. Anders sieht dies bei Behinderungen aus. Diese machen mit 37,2% den zweitgrößten Teil der medizinischen Zustände bei der Teilpopulation aus und könnten mit Sehnenentzündung in Verbindung stehen. Der Anteil der Krankenhausaufenthalte ist mit 14,49% etwas geringer als bei der Gesamtpopulation. Krankenhausaufenthalte könnten ebenfalls auf Sehnenentzündung zurückzuführen sein. Der Anteil anderer medizinisch wichtiger Zustände ist mit 37,68% deutlich höher als bei der Gesamtpopulation. Da nicht klar ist, um was für medizinisch wichtige Zustände es sich handelt, kann keine Aussage darüber getroffen werden, ob diese mit Sehnenentzündung in Verbindung stehen könnten oder nicht.

### 5.2.5 Sterblichkeitsrate der Patienten

Wie in Abbildung 5.9a zu erkennen ist, liegt die Sterblichkeitsrate der Patienten bei 9,38%. Bei den männlichen Patienten liegt sie bei 12,37% und bei den weiblichen bei 6,77%. Da in der Teilpopulation nur 2,09% der Patienten verstorben sind, wurde die Sterblichkeitsrate der Patienten nur für die Gesamtpopulation berechnet. Abbildung 5.10 zeigt die Sterblichkeitsrate der Patienten pro Land. Für die hellblau gefärbten Länder wurde die Sterblichkeitsrate der Patienten nicht berechnet, da aus diesen Ländern weniger als 100 Berichte übermittelt wurden. Bei sehr wenigen Berichten besteht die Gefahr von Ausreißern, die den visuellen Eindruck verfälschen können.

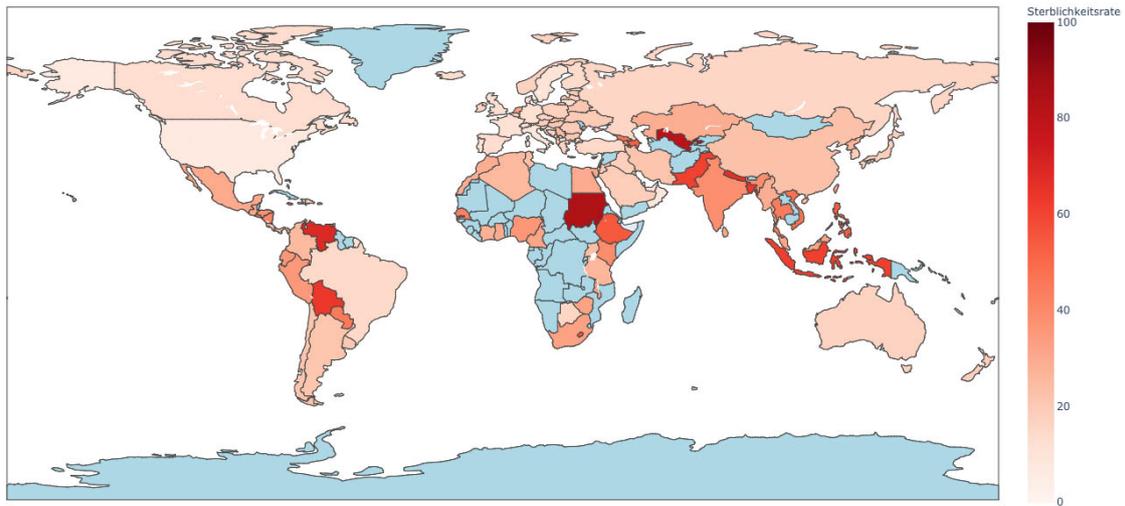


Abbildung 5.10: Sterblichkeitsrate der Patienten pro Land

Wie Abbildung 5.10 zeigt, ist die Sterblichkeitsrate der Patienten in einigen Süd- und mittelamerikanischen, sowie afrikanischen Ländern besonders hoch. Auch in einigen Mittel- und Südasiatischen Ländern, sowie Indonesien, ist sie auffällig hoch. Geringer ist die Sterblichkeitsrate der Patienten in nordamerikanischen und westeuropäischen Ländern. Eine mögliche Erklärung für diese geografischen Diskrepanzen ist, dass aus den Regionen, mit einer hohen Sterblichkeitsrate der Patienten, insgesamt nur wenige Berichte übermittelt wurden. Die Berichte, die übermittelt wurden, behandeln vermutlich Patienten, bei denen davon ausgegangen wurde, dass diese aufgrund eines unerwünschten Ereignisses verstorben sind.



Abbildung 5.11: Sterblichkeitsrate der Patienten pro Geschlecht

Abbildung 5.11 zeigt die Sterblichkeitsrate der Patienten pro Altersgruppe und Geschlecht. Patienten der jüngeren Altersgruppen Neugeborene und Kleinkinder, sowie der älteren Altersgruppen Erwachsene und Senioren, haben eine höhere Sterblichkeitsrate als Patienten der Altersgruppen Kinder und Jugendliche. Zwischen den Geschlechtern gibt es bei den Patienten der jüngeren Altersgruppen Neugeborene bis Jugendliche kaum Unterschiede. Bei Patienten der Altersgruppen Erwachsene und Senioren haben die männlichen Patienten mit 11,17% bei den Erwachsenen und 17,85% bei den Senioren eine deutlich höhere Sterblichkeitsrate als die weiblichen Patienten. Diese haben nur eine Sterblichkeitsrate von 5,62% bei den Erwachsenen und 11,41% bei den Senioren.

### 5.2.6 Dringlichkeit der Berichte

Die Dringlichkeit wurde in allen Berichten angegeben. Ein Bericht gilt als dringend, wenn der medizinische Zustand des Patienten ernsthaft ist und er ein unerwartetes unerwünschtes Ereignis enthält. Ein unerwünschtes Ereignis ist unerwartet, wenn es nicht als mögliche Nebenwirkung in den Produkthinweisen des Medikaments aufgeführt ist, mit dem es aufgetreten ist. Ein Bericht ist nicht dringend, wenn er eine der folgenden Bedingungen erfüllt. Der medizinische Zustand des Patienten ist ernsthaft, der Bericht enthält aber kein unerwartetes unerwünschtes Ereignis. Der Bericht enthält ein unerwartetes unerwünschtes Ereignis, aber der medizinische Zustand des Patienten ist nicht ernsthaft. Oder der medizinische Zustand des Patienten ist nicht ernsthaft und der Bericht enthält kein unerwartetes unerwünschtes Ereignis.

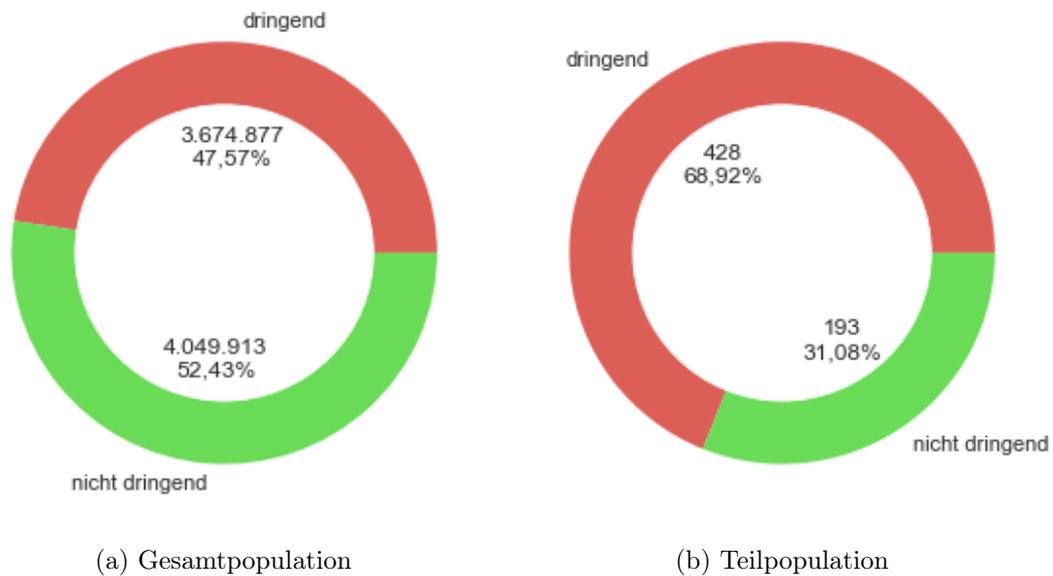


Abbildung 5.12: Dringlichkeit der Berichte

Wie Abbildung 5.12a zeigt, überwiegt der Anteil nicht dringender Berichte bei der Gesamtpopulation leicht. 47,57% der Berichte sind dringend, 52,43% nicht. Bei der Teilpopulation hingegen überwiegt der Anteil dringender Berichte mit 68,92% deutlich. Wie in Abbildung 5.12b zu erkennen ist, sind nur 31,08% der Berichte nicht dringend. Dies spricht dafür, dass Sehnenentzündung von Ciprofloxacin ausgelöst werden kann, da dieses unerwünschte Ereignis noch nicht als mögliche Nebenwirkung in den Produkthinweisen von Ciprofloxacin aufgeführt ist. Berichte, die dieses unerwünschte Ereignis in Kombination mit Ciprofloxacin enthalten, sollten daher von der FDA als dringend markiert werden.

### 5.2.7 Qualifikation der Berichterstatter

Ein Bericht hat einen Berichterstatter. Der Berichterstatter ist die Person, die den Bericht an die FDA gesendet hat. Der Patient kann selbst der Berichterstatter sein, muss er aber nicht. Bei der Gesamtpopulation wurde die Qualifikation des Berichterstatters in 97,24% der Berichte angegeben, bei der Teilpopulation nur bei 85,76%.

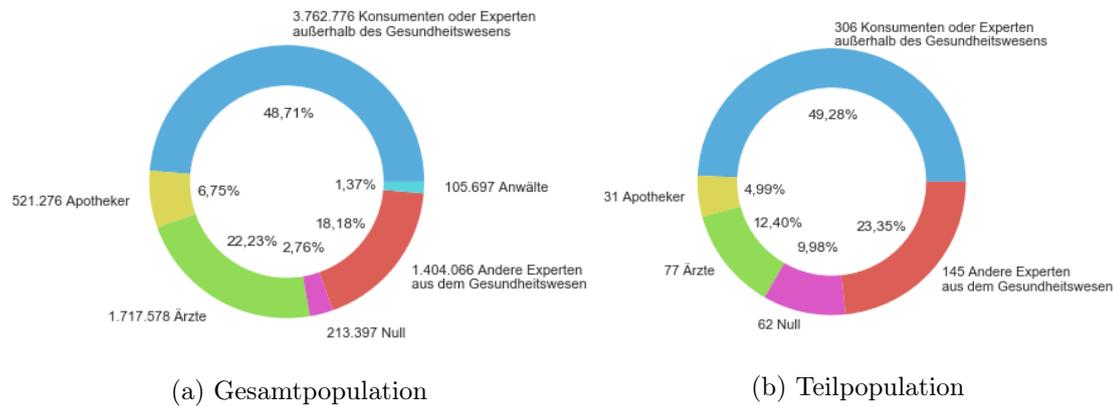


Abbildung 5.13: Qualifikation der Berichterstatter

Wie die Abbildungen 5.13a und 5.13b zeigen, ist bei beiden Populationen etwa die Hälfte der Berichterstatter ein Konsument oder Experte außerhalb des Gesundheitswesens. Bei der Teilpopulation ist der Anteil mit 49,28% etwas höher als bei der Gesamtpopulation (48,71%). Während bei der Gesamtpopulation mit 22,23% über ein Fünftel der Berichterstatter Ärzte sind, liegt deren Anteil bei der Teilpopulation nur bei 12,4%. Auch der Anteil der Apotheker ist bei der Gesamtpopulation mit 6,75% höher als bei der Teilpopulation (4,99%). Mit 23,35% sind bei der Teilpopulation mehr Berichterstatter andere Experten aus dem Gesundheitswesen, auf die keine der vorangegangenen Kategorien zutrifft. Bei der Gesamtpopulation sind es 18,18%. Bei der Gesamtpopulation sind 1,37% der Berichterstatter Anwälte, während bei der Teilpopulation keiner der Berichterstatter Anwalt ist.

### 5.3 Patienten

Bei der Analyse der Patienten besteht das in Abschnitt 2.5 beschriebene Problem, dass zwar mit jedem Bericht genau ein Patient assoziiert wird, einem Patienten aber beliebig viele Berichte zugeordnet sein können. Auf Basis des FAERS-Datensatzes lässt sich die tatsächliche Anzahl der Patienten aber nicht bestimmen. Es sei daher darauf hingewiesen, dass die tatsächliche Anzahl der Patienten von der bei den Statistiken und in den Abbildungen angegeben Anzahl an Patienten abweichen kann und dies höchstwahrscheinlich auch tut.

### 5.3.1 Geschlecht der Patienten

Wie Abbildung 5.14a zeigt, überwiegt der Anteil weiblicher Patienten mit 54,93% deutlich. Die wahrscheinlichste Erklärung dafür ist, dass weibliche Patienten häufiger unerwünschte Ereignisse melden, als männliche Patienten. Unwahrscheinlicher ist hingegen, dass unerwünschte Ereignisse bei weiblichen Patienten häufiger auftreten, als bei männlichen Patienten. Ebenfalls eine eher unwahrscheinliche Erklärung ist, dass weibliche Patienten mehr Medikamente konsumiert haben, als männliche Patienten. Dadurch wären bei weiblichen Patienten möglicherweise mehr unerwünschte Ereignisse aufgetreten, als bei männlichen Patienten.

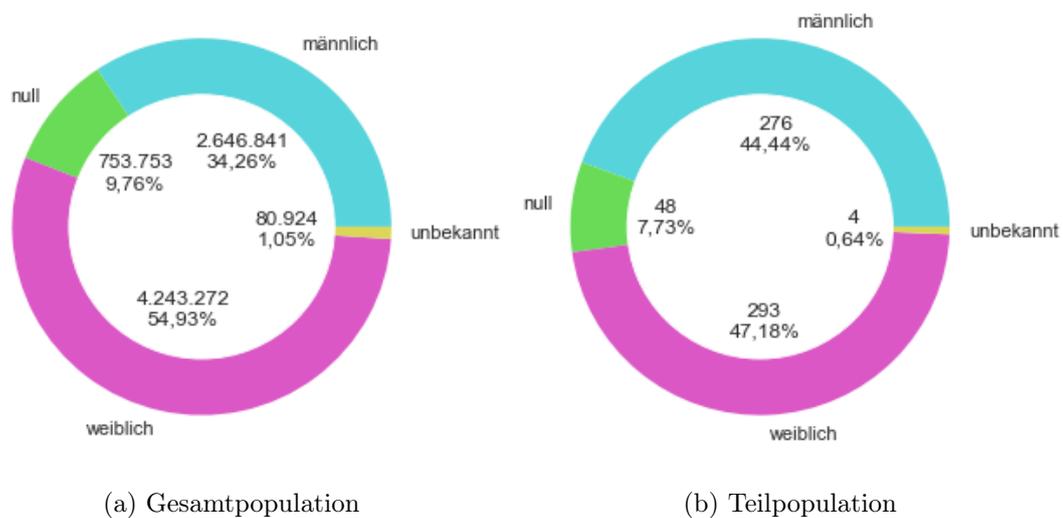


Abbildung 5.14: Geschlecht der Patienten

Wie Abbildung 5.14b zeigt, überwiegt der Anteil weiblicher Patienten bei der Teilpopulation mit 47,18% nicht ganz so stark. Eine mögliche Erklärung dafür ist, dass überdurchschnittlich viele männliche Patienten Ciprofloxacin konsumiert haben. Dadurch sind Sehnenentzündungen nach der Einnahme von Ciprofloxacin bei männlichen Patienten häufiger aufgetreten. Eine weitere Möglichkeit ist, dass häufiger männliche Patienten von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen sind. Auch eine Möglichkeit ist, dass Sehnenentzündungen von männlichen Patienten überdurchschnittlich häufig gemeldet wurden. Diese Annahme wird durch die Schwere von Sehnenentzündung gestützt. Ein Patient sollte bei einer Sehnenentzündung eher dazu geneigt sein,

diese zu melden, als bei einem harmloseren unerwünschten Ereignis wie Ermüdung oder Übelkeit.

An dieser Stelle sei einmalig darauf hingewiesen, weshalb Teilpopulationen nicht isoliert betrachtet werden dürfen, um fundierte Aussagen über diese treffen zu können. Würde z.B. das Geschlecht der Patienten der Teilpopulation isoliert von der Gesamtpopulation betrachtet werden, würde ein höherer Anteil weiblicher Patienten vermutlich damit assoziiert werden, dass möglicherweise mehr weibliche Patienten von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen sind, als männliche Patienten. Wird die Teilpopulation jedoch vor dem Hintergrund der Gesamtpopulation betrachtet, wird deutlich, dass der Anteil männlicher Patienten in der Teilpopulation in Wirklichkeit auffällig hoch ist, verglichen mit dem Anteil männlicher Patienten in der Gesamtpopulation. Die Beobachtung kann daher zu dem gegenteiligen Schluss führen, dass möglicherweise mehr männliche Patienten von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen sind, als weibliche Patienten. Das gleiche Prinzip trifft auch auf alle anderen Attribute zu, die bei der EDA analysiert wurden, beim Geschlecht der Patienten war dieses Problem jedoch besonders auffällig.

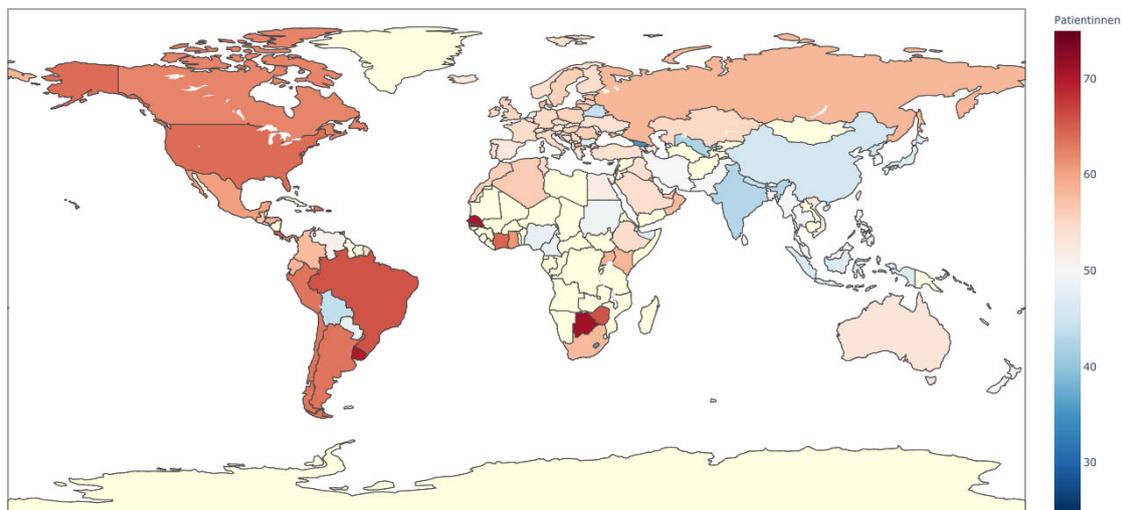


Abbildung 5.15: Geschlecht der Patienten pro Land

In Abbildung 5.15 ist das Geschlecht der Patienten pro Land für die Gesamtpopulation dargestellt. Für die Teilpopulation wurde diese Abbildung aufgrund von zu wenigen Berichten nicht erstellt. Ein rötlicher Farbton signalisiert einen höheren Anteil weiblicher und ein bläulicher Farbton einen höheren Anteil männlicher Patienten. Länder, für

die der Anteil weiblicher Patienten aufgrund von zu wenigen Berichten nicht berechnet worden ist, wurden gelb gefärbt. Der Anteil weiblicher Patienten überwiegt in den meisten Ländern. Besonders deutlich überwiegt er in Nord- und Südamerika, sowie in einigen afrikanischen Ländern. In Westeuropa und Australien ist der Anteil männlicher und weiblicher Patienten relativ ausgeglichen. In Indien, China, Südostasien und Indonesien überwiegt der Anteil männlicher Patienten. Auch in einigen wenigen afrikanischen und Südamerikanischen Ländern ist der Anteil männlicher Patienten größer als der Anteil weiblicher Patienten. Diese geografischen Unterschiede können verschiedene Ursachen haben. Denkbar wäre beispielsweise, dass weibliche Patienten in den Ländern mit einem höheren Anteil weiblicher Patienten eher dazu geneigt sind, unerwünschte Ereignisse zu melden, als männliche Patienten. In den Ländern mit einem höheren Anteil männlicher Patienten könnte das Gegenteil der Fall sein. Die Unterschiede beim Geschlecht der Patienten wären dann nicht unbedingt auf ein unterschiedliches Konsumverhalten bezogen auf Medikamente zurückzuführen. Dies stellt jedoch natürlich auch eine Möglichkeit dar.

### 5.3.2 Alter der Patienten

Das angegebene Alter ist das Alter des Patienten zum Zeitpunkt des Einsetzens des unerwünschten Ereignisses oder der unerwünschten Ereignisse [29]. Da in einem Bericht ein oder mehrere unerwünschte Ereignisse aufgeführt sein können, kann nicht bestimmt werden, auf das Alter des Patienten zum Zeitpunkt des Auftretens welches unerwünschten Ereignisses sich das Alter des Patienten bezieht. Das Alter wurde bei 58,03% der Patienten der Gesamtpopulation und 80,68% der Patienten der Teilpopulation angegeben. Das Durchschnittsalter der Gesamtpopulation beträgt 55,27 Jahre und der Teilpopulation 50,41 Jahre. Von Sehnenentzündungen nach der Einnahme von Ciprofloxacin sind also durchschnittlich etwas jüngere Patienten betroffen. In der Gesamtpopulation sind männliche Patienten durchschnittlich 56,17 Jahre alt und in der Teilpopulation 48,15 Jahre. Weibliche Patienten sind in der Gesamtpopulation im Durchschnitt 54,84 Jahre alt und in der Teilpopulation 52,3 Jahre. Bei beiden Geschlechtern wurden Sehnenentzündungen nach der Einnahme von Ciprofloxacin also durchschnittlich bei jüngeren Patienten gemeldet. Bei den männlichen Patienten fällt der Unterschied dabei stärker aus als bei den weiblichen. Während bei der Gesamtpopulation außerdem männliche Patienten durchschnittlich älter sind als weibliche, ist dies bei der Teilpopulation umgekehrt.

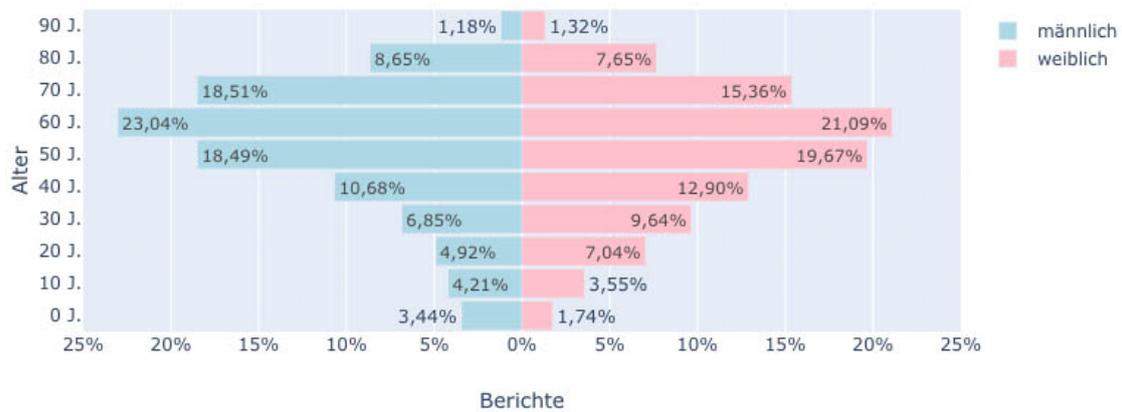


Abbildung 5.16: Durchschnittsalter der Patienten der Gesamtpopulation

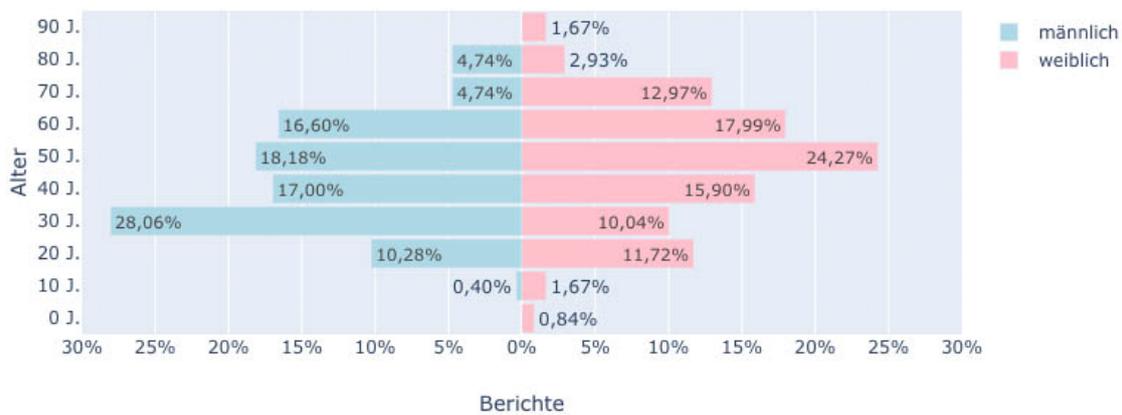


Abbildung 5.17: Durchschnittsalter der Patienten der Teilpopulation

Die Abbildungen 5.16 und 5.17 zeigen das Durchschnittsalter männlicher und weiblicher Patienten der Gesamt- und Teilpopulation in einer Alterspyramide. Patienten über 100 Jahre wurden nicht dargestellt, da diese nur einen Anteil von 0,02% bei den männlichen und 0,03% bei den weiblichen Patienten der Gesamtpopulation ausmachen. In der Teilpopulation sind gar keine Patienten über 100 Jahre alt. Bei der Gesamtpopulation gibt es zwischen den Geschlechtern bei der Altersverteilung nur geringe Unterschiede. Bei der Teilpopulation sind hingegen deutliche Unterschiede zu erkennen. Sowohl insgesamt, als auch zwischen den Geschlechtern. Bei den männlichen Patienten sind vor allem Erwachsene zwischen 30 und < 70 Jahren von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen. Bei den weiblichen Patienten eher ältere Patienten zwischen 40 und < 80 Jahren. Ab 80 Jahren sind bei beiden Geschlechtern nur wenige Patienten

von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen, selbiges gilt für Patienten unter 20 Jahren. Dies kann darauf zurückzuführen sein, dass Ciprofloxacin häufiger von Patienten in dem angegebenen Alter eingenommen wird. Daher werden Sehnenentzündungen bei Patienten in diesem Alter womöglich häufiger gemeldet. Eine weitere Erklärung ist, dass Patienten diesen Alters anfälliger für Sehnenentzündungen unabhängig von der Einnahme von Ciprofloxacin sind. Dies würde möglicherweise zu einer höheren Anzahl Meldungen für Patienten diesen Alters führen.

### 5.3.3 Todesalter der Patienten

Wie in Abschnitt 5.3.2 erklärt, ist das angegebene Alter des Patienten das Alter des Patienten zum Zeitpunkt des Auftretens des unerwünschten Ereignisses oder der unerwünschten Ereignisse [29]. Daher stimmt dieses bei verstorbenen Patienten nicht zwangsläufig mit dem Alter der Patienten zum Zeitpunkt des Todes überein. Es gibt keine Möglichkeit, auf Basis der vorliegenden Daten das tatsächliche Alter der Patienten zum Zeitpunkt des Todes zu bestimmen. Aufgrund dessen bezieht sich der Begriff „Todesalter“ in diesem Zusammenhang auf das Alter der Patienten aus Berichten, deren Ernsthaftigkeit „Tod“ ist und nicht auf das tatsächliche Alter der Patienten zum Zeitpunkt des Todes. Das durchschnittliche Todesalter der Patienten beträgt 61,67 Jahre. Männliche Patienten sind mit durchschnittlich 61,66 Jahren und weibliche Patienten mit 61,81 Jahren verstorben.

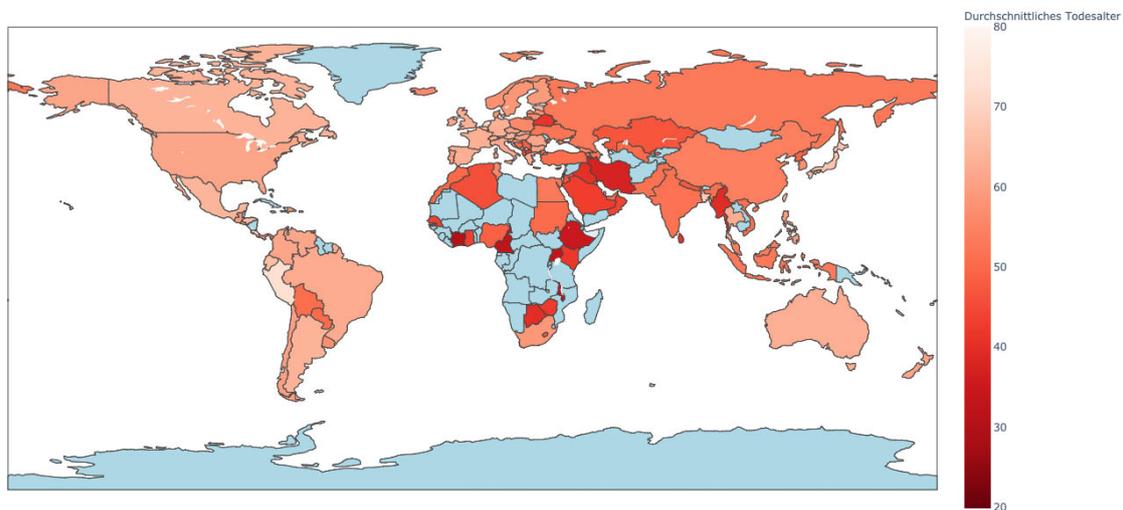


Abbildung 5.18: Durchschnittliches Todesalter der Patienten pro Land

In Abbildung 5.18 ist das durchschnittliche Todesalter der Patienten pro Land dargestellt. Die Abbildung zeigt dabei die Patienten der Gesamtpopulation. Für die Teilpopulation wurde diese Statistik nicht erstellt, da nur 2,09% der Patienten der Teilpopulation verstorben sind. Wie in Abbildung 5.18 zu erkennen, ist das durchschnittliche Todesalter der Patienten in einigen afrikanischen und arabischen Ländern besonders niedrig. Auch in Russland und Mittel- und Südasiatischen Ländern ist es eher gering. Höher ist das durchschnittliche Todesalter der Patienten in Nord- und Südamerika, sowie in Westeuropa und Australien. Für die blau gefärbten Länder wurde aufgrund von zu wenigen Berichten das durchschnittliche Todesalter der Patienten nicht berechnet. Eine mögliche Erklärung für diese geografischen Unterschiede ist, dass aus den Ländern mit einem niedrigen durchschnittlichen Todesalter der Patienten insgesamt weniger Berichte übermittelt wurden. Bei den Patienten, für die Berichte übermittelt worden sind, wurde vermutlich davon ausgegangen, dass ein durch die Einnahme eines Medikaments ausgelöstes unerwünschtes Ereignis zum Tod geführt hat.

### 5.3.4 Altersgruppen der Patienten

Neben dem Alter kann auch die Altersgruppe des Patienten angegeben werden. Bei der Gesamtpopulation wurde sie bei 62,07% und bei der Teilpopulation bei 81,32% der Patienten aufgeführt. Bei der Visualisierung in den Abbildungen 5.19 und 5.20 wurden nur die Anzahl der Patienten dargestellt, bei denen die Altersgruppe angegeben wurde. Die Anteile beziehen sich also auf den Anteil an den 62,07% der Patienten der Gesamtpopulation und 81,32% der Patienten der Teilpopulation, für die die Altersgruppe angegeben wurde und nicht auf den Anteil an allen Patienten.

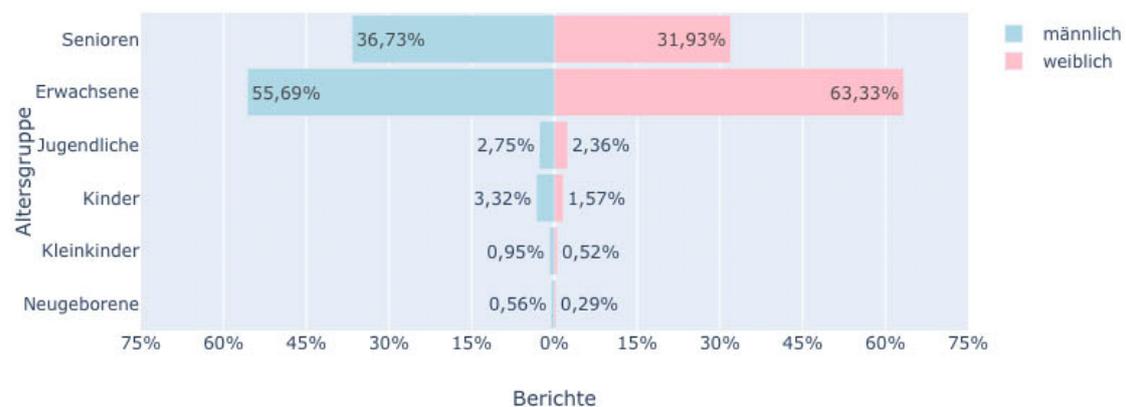


Abbildung 5.19: Altersgruppen der Patienten der Gesamtpopulation

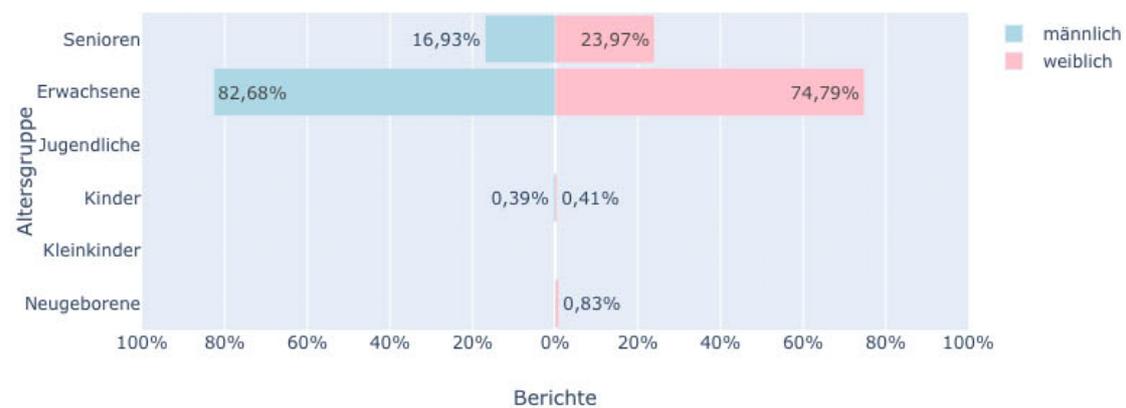


Abbildung 5.20: Altersgruppen der Patienten der Teilpopulation

Wie Abbildung 5.19 zeigt, sind bei der Gesamtpopulation die meisten Patienten Teil der Altersgruppen Erwachsene und Senioren. Nur ein geringer Teil der Patienten gehört den jüngeren Altersgruppen Neugeborene, Kleinkinder, Kinder und Jugendliche an. Bei diesen überwiegt außerdem der Anteil männlicher Patienten, während bei den Erwachsenen der Anteil weiblicher Patienten überwiegt. Bei den Senioren überwiegt wiederum der Anteil männlicher Patienten. Die Unterschiede zwischen den Altersgruppen sind vor allem auf die unterschiedlichen Zeitspannen der Altersgruppen zurückzuführen. Die Zeitspannen der Altersgruppen wurden in Abschnitt 3.2.5 angegeben.

Wie in Abbildung 5.20 zu erkennen ist, sind bei der Teilpopulation bei beiden Geschlechtern vor allem Erwachsene von Sehnenentzündungen nach der Einnahme von Ciprofloxacin betroffen. Auch Senioren sind betroffen, allerdings nicht so viele, wie aufgrund der Gesamtpopulation zu erwarten gewesen wäre. Von den jüngeren Altersgruppen ist bei den männlichen Patienten nur ein kleiner Teil der Kinder betroffen, bei den weiblichen Patienten ein geringer Teil der Neugeborenen und Kinder. Kleinkinder und Jugendliche sind auch bei den weiblichen Patienten gar nicht betroffen. Der Grund dafür ist vermutlich, dass Ciprofloxacin nur von Patienten der Altersgruppen Erwachsene und Senioren konsumiert wird. Daher existieren auch nur Berichte über Sehnenentzündungen bei diesen Altersgruppen. Unwahrscheinlich ist, dass auch Patienten der jüngeren Altersgruppen Ciprofloxacin eingenommen haben, allerdings nicht so stark von Sehnenentzündungen betroffen sind, wie Patienten der Altersgruppen Erwachsene und Senioren.

### 5.3.5 Gewicht der Patienten

Das Gewicht der Patienten wurde in Kilogramm als Fließkommazahl mit einer Genauigkeit von zwei Dezimalstellen angegeben [29]. Es wurde nur bei 20,23% der Patienten aufgeführt. Das Durchschnittsgewicht beträgt 75,05 Kilogramm. Männliche Patienten wiegen durchschnittlich 81,31 Kilogramm und weibliche Patienten 71,15 Kilogramm. Abbildung 5.21 zeigt das Durchschnittsgewicht männlicher und weiblicher Patienten pro Altersgruppe. Wie darin zu erkennen ist, sind weibliche Patienten bis zum Kleinkindalter durchschnittlich schwerer als männliche Patienten. Ab dem Kindesalter sind männliche Patienten hingegen durchschnittlich schwerer als weibliche Patienten.

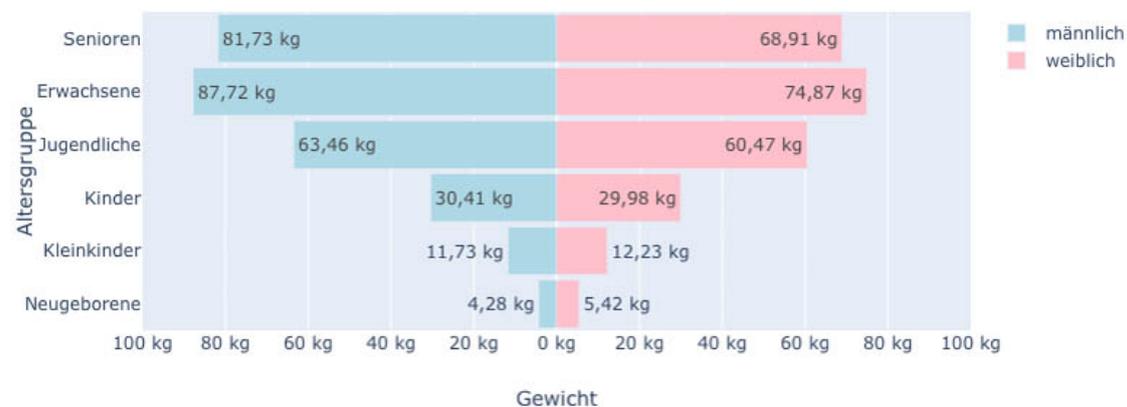


Abbildung 5.21: Gewicht der Patienten

Für die Teilpopulation wurde diese Statistik nicht erstellt, da das Gewicht der Patienten mit den anderen Informationen aus den Berichten nicht ausreicht, um Aussagen über die Gesundheit der Patienten auf Basis des Gewichts treffen zu können.

## 5.4 Unerwünschte Ereignisse

Es wurden 22.447.707 unerwünschte Ereignisse gemeldet. Das entspricht durchschnittlich 2,91 unerwünschten Ereignissen pro Bericht. In den meisten Berichten wurden zwei unerwünschte Ereignisse gemeldet. In dem Bericht mit den meisten unerwünschten Ereignissen wurden 367 unerwünschte Ereignisse aufgeführt. Pro Jahr wurden durchschnittlich 2.805.963 unerwünschte Ereignisse gemeldet.

### 5.4.1 Beschreibung der unerwünschten Ereignisse

Die unerwünschten Ereignisse sollen eigentlich mit MedDRA PTs beschrieben werden [29]. Bei dem entsprechenden Attribut handelt es sich jedoch um ein Freitextfeld. Es muss also davon ausgegangen werden, dass die unerwünschten Ereignisse nicht immer mit MedDRA PTs beschrieben wurden. 30 unerwünschte Ereignisse wurden gar nicht beschrieben. Insgesamt wurden 19.280 unterschiedliche unerwünschte Ereignisse gemeldet. Von diesen wurden 17.408 mehr als einmal beschrieben. Wenn ein unerwünschtes Ereignis mehr als einmal genannt wurde, ist davon auszugehen, dass es sich um einen MedDRA PT handelt. In Tabelle 5.1 sind die fünf am häufigsten gemeldeten unerwünschten Ereignisse dargestellt.

	Unerwünschtes Ereignis	Meldungen	Anteil
1.	Medikament unwirksam	508.060	2,26%
2.	Tod	338.264	1,51%
3.	Ermüdung	303.820	1,35%
4.	Übelkeit	301.812	1,34%
5.	Zulassungsüberschreitende Anwendung	264.710	1,20%

Tabelle 5.1: Fünf am häufigsten gemeldete unerwünschte Ereignisse

### 5.4.2 Ausgang der unerwünschten Ereignisse

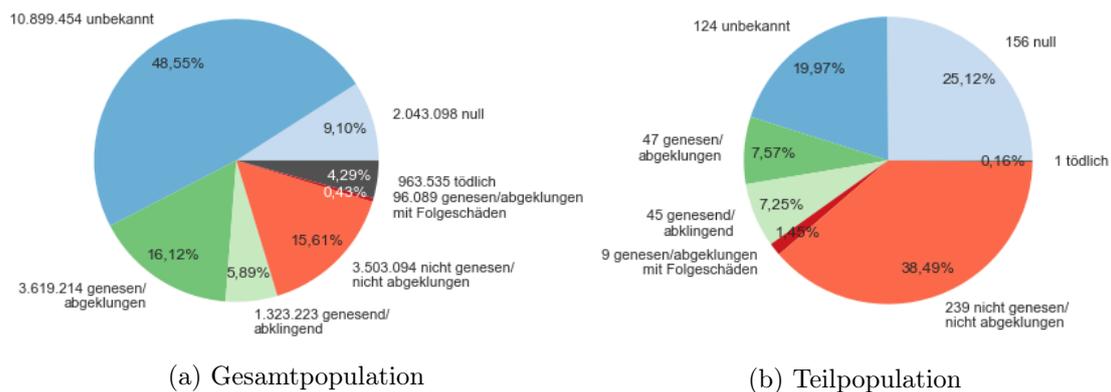


Abbildung 5.22: Ausgang der unerwünschten Ereignisse

Wie in Abbildung 5.22b zu sehen ist, sind mit 38,49% bei der Teilpopulation bei einem großen Teil der Patienten die Sehnenentzündungen noch nicht abgeklungen. Bei der Gesamtpopulation beträgt der Anteil der unerwünschten Ereignisse, die noch nicht abgeklungen sind, nur 15,61%, wie Abbildung 5.22a zeigt. Auch der Anteil der Patienten, bei denen die Sehnenentzündungen zwar abgeklungen sind, aber Folgeschäden hinterlassen haben, ist mit 1,45% deutlich höher als der Anteil der unerwünschten Ereignisse bei der Gesamtpopulation, die Folgeschäden hinterlassen haben. Dieser liegt bei gerade einmal 0,43%. Dies ist ein Hinweis darauf, dass durch Ciprofloxacin ausgelöste Sehnenentzündungen lang anhaltende oder sogar dauerhafte Schäden verursachen können. Dafür spricht auch der geringe Anteil der Patienten der Teilpopulation, bei denen die Sehnenentzündungen dabei sind abzuklingen oder bereits abgeklungen sind. Dieser liegt bei 7,25% und 7,57%. Bei der Gesamtpopulation liegt der Anteil der unerwünschten Ereignisse, die dabei sind abzuklingen bei 5,89% und der Anteil der unerwünschten Ereignisse, die bereits abgeklungen sind bei 16,12%. Bei der Gesamtpopulation sind 4,29% der unerwünschten Ereignisse tödlich verlaufen. Bei der Teilpopulation hingegen haben Sehnenentzündungen nur bei einem Patienten zum Tod geführt (0,16%).

## 5.5 Medikamente

Insgesamt wurden 23.298.217 Medikamenteneinnahmen an die FDA gemeldet. Das entspricht durchschnittlich 3,02 Medikamenteneinnahmen pro Bericht. In den meisten Berichten wurde jedoch nur eine Medikamenteneinnahme angegeben. In dem Bericht mit den meisten Medikamenteneinnahmen wurden 168 Medikamenteneinnahmen aufgeführt. Pro Jahr wurden durchschnittlich 2.912.277 Medikamenteneinnahmen gemeldet.

### 5.5.1 Medikamentenbezeichnungen

Die Medikamentenbezeichnungen sind die gültigen Handelsbezeichnungen der Medikamente, sofern diese in der Datenbank der FDA gespeichert sind. Ansonsten sind die Medikamentenbezeichnungen eine wörtliche Beschreibung des Medikaments durch den Berichterstatter [29]. Das eingenommene Medikament wurde nur bei vier Medikamenteneinnahmen nicht angegeben. Insgesamt wurden 399.424 unterschiedliche Medikamente gemeldet. 149.088 von diesen Medikamenten wurden dabei mehr als einmal genannt. Bei

diesen Medikamentenbezeichnungen ist davon auszugehen, dass es sich um standardisierte Bezeichnungen für die Medikamente handelt. In Tabelle 5.2 sind die fünf am häufigsten gemeldeten Medikamente aufgelistet.

	Medikament	Meldungen	Anteil
1.	Enbrel	352.142	1,51%
2.	Humira	316.268	1,36%
3.	Aspirin	174.744	0,75%
4.	Prednison	165.964	0,71%
5.	Methotrexat	152.498	0,65%

Tabelle 5.2: Fünf am häufigsten gemeldete Medikamente

### 5.5.2 Indikationen

Die Indikationen der Medikamente sollen eigentlich mit MedDRA PTs beschrieben werden [29]. Bei dem entsprechenden Attribut handelt es sich jedoch um ein Freitextfeld. Die Indikationen der Medikamente wurden daher nicht zwangsläufig mit MedDRA PTs beschrieben. Die Indikation wurde bei 69,21% der Medikamenteneinnahmen angegeben. Insgesamt wurden 15.312 unterschiedliche Indikationen beschrieben. 12.031 von diesen Indikationen wurden dabei mehr als einmal genannt. Bei diesen Indikationen handelt es sich vermutlich um MedDRA PTs oder andere allgemein anerkannte Beschreibungen dieser Indikationen. Tabelle 5.3 zeigt die fünf am häufigsten gemeldeten Indikationen.

	Indikation	Meldungen	Anteil
1.	Unbekannt	5.948.945	36,89%
2.	Schmerzen	643.288	3,99%
3.	Rheumatoide Arthritis	565.972	3,51%
4.	Multiple Sklerose	344.729	2,14%
5.	Bluthochdruck	332.739	2,06%

Tabelle 5.3: Fünf am häufigsten gemeldete Indikationen

In Tabelle 5.4 sind die fünf am häufigsten angegebene Indikationen aufgelistet, für die Ciprofloxacin eingenommen wurden, als eine Sehnenentzündung aufgetreten ist.

	Indikation	Meldungen	Anteil
1.	Harnwegsinfektion	126	22,78%
2.	Unbekannt	72	13,02%
3.	Prostataentzündung	55	9,95%
4.	Blasenentzündung	19	3,44%
5.	Nasennebenhöhlenentzündung	19	3,44%

Tabelle 5.4: Fünf am häufigsten gemeldete Indikationen von Ciprofloxacin bei Sehnenentzündung

### 5.5.3 Rolle der Medikamente beim Auftreten von unerwünschten Ereignissen

Ein Medikament kann verdächtig sein, eine oder mehrere unerwünschte Ereignisse auszulösen. Es kann außerdem vermutet werden, dass das Medikament mit einem oder mehreren anderen Medikamenten wechselwirkt und dadurch das unerwünschte Ereignis oder die unerwünschten Ereignisse auslöst. Außerdem kann angenommen werden, dass ein Medikament eine begleitende Rolle beim Auftreten eines unerwünschten Ereignisses oder mehrerer unerwünschter Ereignisse hat. D.h., dass das Medikament begleitend zu einem anderen Medikament eingenommen wurde, welches das unerwünschte Ereignis oder die unerwünschten Ereignisse auslöst. Die vermutete Rolle des Medikaments beim Auftreten des unerwünschten Ereignisses oder der unerwünschten Ereignisse wurde bei der Gesamtpopulation nur bei vier Medikamenteneinnahmen nicht angegeben. Bei der Teilpopulation wurde sie bei allen Medikamenteneinnahmen angegeben.

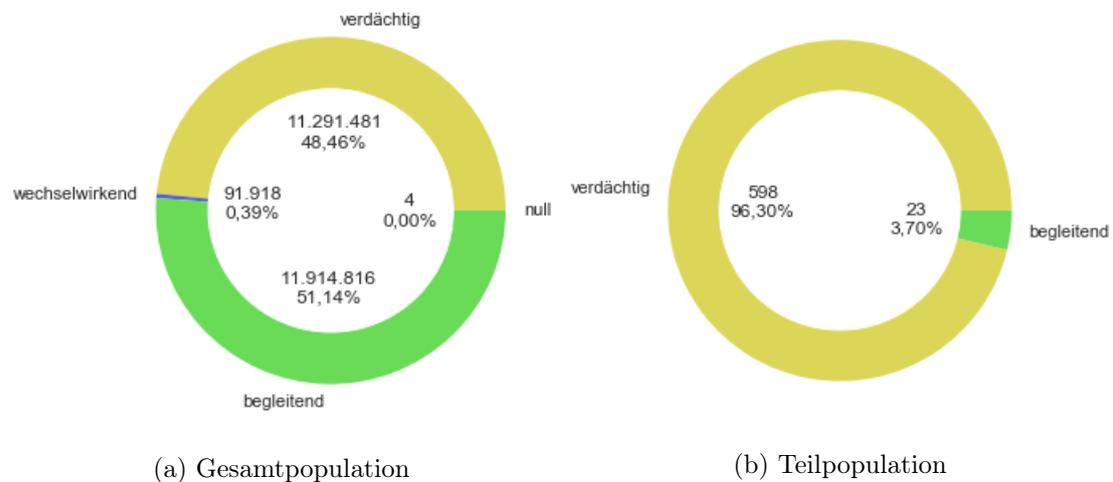


Abbildung 5.23: Rolle der Medikamente beim Auftreten von unerwünschten Ereignissen

Wie in Abbildung 5.23a zu erkennen ist, wird bei der Gesamtpopulation bei 48,46% der Medikamenteneinnahmen das Medikament verdächtig, das unerwünschte Ereignis oder die unerwünschten Ereignisse auszulösen. Abbildung 5.23b zeigt, dass bei der Teilpopulation bei fast allen Medikamenteneinnahmen (96,3%) Ciprofloxacin verdächtig wird, die Sehnenentzündung zu verursachen. Dies ist ein starkes Indiz dafür, dass Ciprofloxacin mit Sehnenentzündung in Zusammenhang steht. Bei der Gesamtpopulation wird bei 51,14% der Medikamenteneinnahmen dem Medikament eine begleitende Rolle beim Auftreten von unerwünschten Ereignissen zugeschrieben. Bei der Teilpopulation wird nur bei 3,7% der Medikamenteneinnahmen eine begleitende Rolle von Ciprofloxacin beim Auftreten von Sehnenentzündung vermutet. Bei der Gesamtpopulation wird bei nur 0,39% der Medikamenteneinnahmen das Medikament verdächtig, mit einem anderen Medikament zu wechselwirken und dadurch das unerwünschte Ereignis oder die unerwünschten Ereignisse auszulösen. Bei der Teilpopulation ist dies bei keiner Einnahme von Ciprofloxacin der Fall. Das ist ein Hinweis darauf, dass Sehnenentzündung von Ciprofloxacin alleine und nicht durch das Wechselwirken von Ciprofloxacin mit anderen Medikamenten ausgelöst wird.

### 5.5.4 Anpassung der Dosen

Abbildung 5.24a zeigt die Anpassung der Dosen bei der Gesamtpopulation und Abbildung 5.24b bei der Teilpopulation. Auffällig ist, dass bei der Teilpopulation der Anteil der Medikamenteneinnahmen, bei denen das Medikament abgesetzt wurde, mit 37,68% deutlich größer ist, als bei der Gesamtpopulation (10,64%). Das könnte daran liegen, dass bei diesen Medikamenteneinnahmen erkannt wurde, dass die Sehnenentzündung in Zusammenhang mit der Einnahme von Ciprofloxacin stehen könnte. Daher wurde Ciprofloxacin vorsichtshalber abgesetzt. Ein weiterer Grund könnte auch die Schwere des unerwünschten Ereignisses sein. Sehnenentzündung ist ein ernsthafteres unerwünschtes Ereignis als beispielsweise Ermüdung oder Übelkeit. Der Patient könnte bei einer Sehnenentzündung daher eher dazu geneigt sein, das Medikament abzusetzen, in der Hoffnung, dass diese nachlässt, als bei harmloseren unerwünschten Ereignissen wie Ermüdung oder Übelkeit.

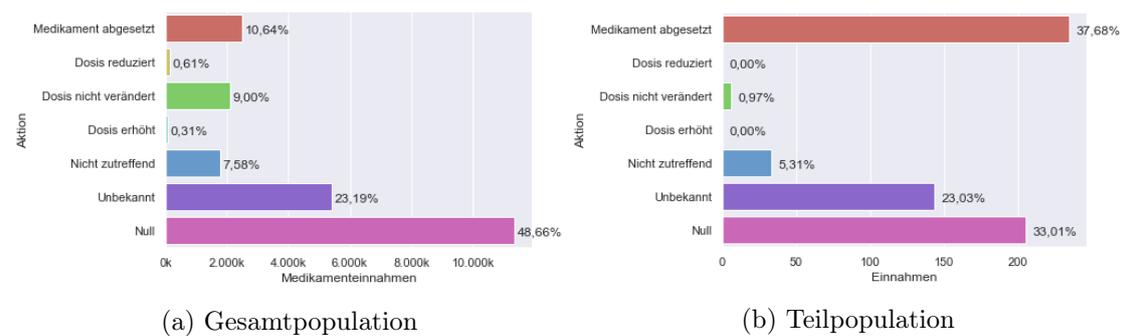


Abbildung 5.24: Anpassung der Dosen

### 5.5.5 Rückkehr von unerwünschten Ereignissen nach Wiedereinnahme des Medikaments

Ob die unerwünschten Ereignisse nach der Wiedereinnahme des Medikaments zurückgekehrt sind, wurde bei der Gesamtpopulation nur bei 1,49% der Medikamenteneinnahmen angegeben. Bei der Teilpopulation immerhin bei 3,06%. Da nur das Verhältnis von „ja“ zu „nein“ relevant ist, wurden in Abbildung 5.25 nur die Medikamenteneinnahmen dargestellt, bei denen angegeben wurde, ob die unerwünschten Ereignisse nach Wiedereinnahme des Medikaments zurückgekehrt sind oder nicht. Abbildung 5.25a zeigt dabei das Verhältnis von „ja“ zu „nein“ bei der Gesamtpopulation und Abbildung 5.25b bei der Teilpopulation.

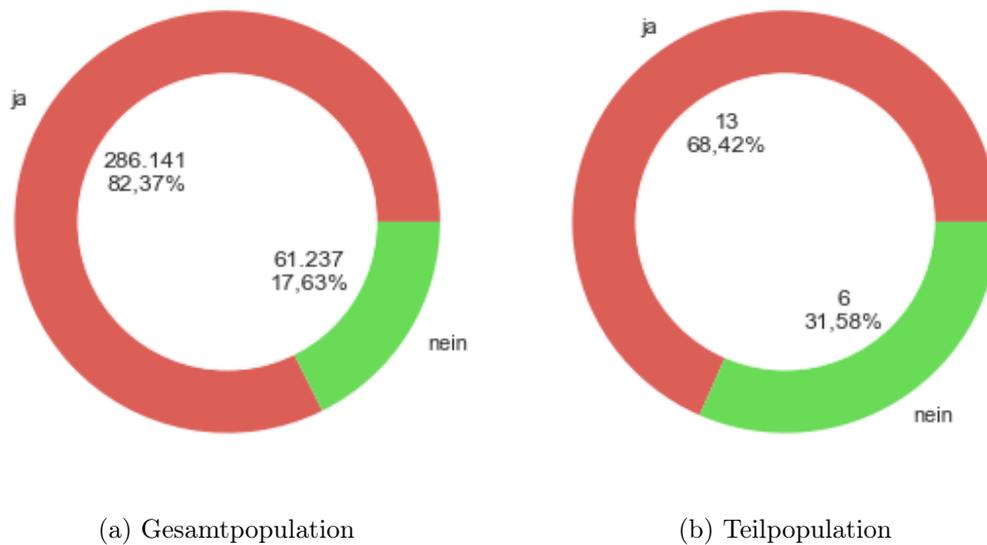


Abbildung 5.25: Rückkehr von unerwünschten Ereignissen nach Wiedereinnahme des Medikaments

Bei beiden Populationen überwiegt der Anteil der Medikamenteneinnahmen, bei denen die unerwünschten Ereignisse nach Wiedereinnahme des Medikaments zurückgekehrt sind, deutlich. Bei der Teilpopulation überwiegt er mit 68,42% allerdings nicht ganz so stark, wie bei der Gesamtpopulation mit 82,37%. Das Ergebnis ist aber dennoch ein Indiz dafür, dass Ciprofloxacin mit Sehnenentzündung in Zusammenhang steht.

### 5.5.6 Nachlassen von unerwünschten Ereignissen nach Reduzierung der Dosis oder Absetzung des Medikaments

Ob nach Reduzierung der Dosis oder Absetzung des Medikaments die unerwünschten Ereignisse nachgelassen haben, wurde bei der Gesamtpopulation nur bei 4,85% der Medikamenteneinnahmen angegeben. Bei der Teilpopulation bei 49,83%. Um das Verhältnis von „ja“ zu „nein“ bei beiden Populationen vergleichen zu können, wurden in Abbildung 5.26 nur die Medikamenteneinnahmen dargestellt, bei denen angegeben wurde, ob die unerwünschten Ereignisse nach Reduzierung der Dosis oder Absetzung des Medikaments verschwunden sind. Abbildung 5.26a zeigt dabei das Ergebnis für die Gesamtpopulation und Abbildung 5.26b für die Teilpopulation.

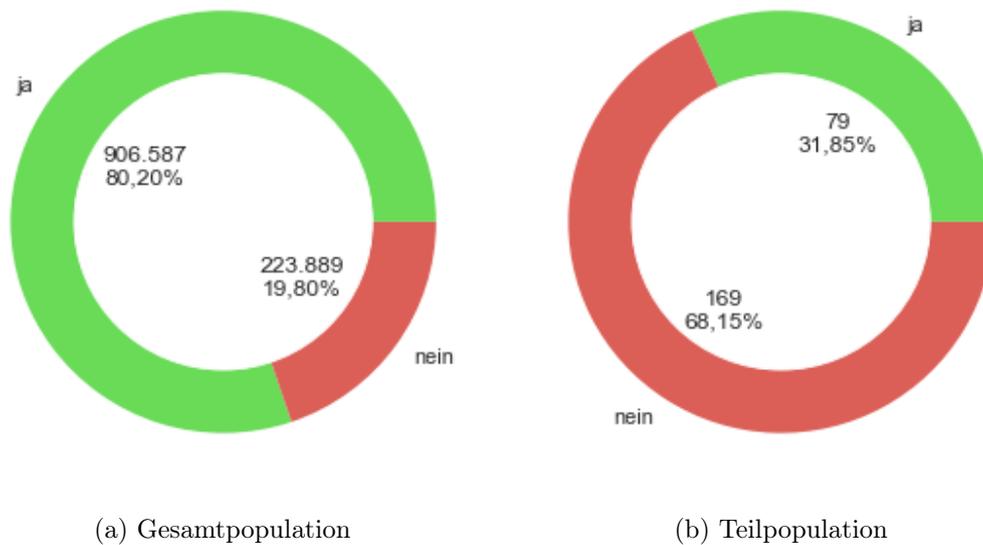


Abbildung 5.26: Nachlassen von unerwünschten Ereignissen nach Reduzierung der Dosis oder Absetzung des Medikaments

Wie in Abbildung 5.26a zu erkennen ist, sind die unerwünschten Ereignisse bei der Gesamtpopulation nach Reduzierung oder Absetzung des Medikaments mit 80,2% bei den meisten Medikamenteneinnahmen verschwunden. Wie Abbildung 5.26b zeigt, sieht dies bei der Teilpopulation anders aus. Bei der Teilpopulation ist die Sehnenentzündung mit 68,15% bei den meisten Patienten nicht verschwunden, nachdem die Dosis reduziert, oder Ciprofloxacin ganz abgesetzt wurde. Dies kann mehrere mögliche Erklärungen haben. Eine Möglichkeit ist, dass Ciprofloxacin nicht der Auslöser für die Sehnenentzündungen war. Daher würde eine Reduzierung der Dosis oder die Absetzung von Ciprofloxacin nicht dazu führen, dass die Sehnenentzündungen verschwinden. Aufgrund der Ergebnisse aus den vorherigen Analysen ist dies jedoch unwahrscheinlich. Eine wahrscheinlichere Erklärung ist, dass die Einnahme von Ciprofloxacin Folgeschäden verursacht hat, sodass die Sehnenentzündungen nicht mit Reduzierung der Dosis oder der Absetzung von Ciprofloxacin verschwunden sind. Diese Annahme wird gestützt durch die Ergebnisse aus Abschnitt 5.4.2.

## 6 Zusammenfassung

In Abschnitt 2.1 wurden zunächst die Grundlagen der Pharmakovigilanz erklärt. Anschließend wurde in Abschnitt 2.2 der wichtige formale Unterschied zwischen unerwünschten Ereignissen und Nebenwirkungen erläutert. Danach wurde in Abschnitt 2.3 die FDA und ihre Aufgaben bei der Überwachung von Medikamenten nach deren Markteinführung als die für das in Abschnitt 2.4 beschriebene FAERS verantwortliche Organisation vorgestellt. In Abschnitt 2.5 wurde auf Einschränkungen eingegangen, die bei der Analyse der FAERS- und anderer Pharmakovigilanzdaten zu beachten sind.

In Kapitel 3 wurde der Import und die Aufbereitung der Daten beschrieben. Dabei wurde in Abschnitt 3.1 auf Probleme durch eine zu hohe Hauptspeicherauslastung eingegangen und in Abschnitt 3.2 Unterschiede der aufbereiteten Daten zu den Rohdaten aufgezeigt. Die Grundlagen der Signalgenerierung wurden in Abschnitt 4.1 erklärt und verschiedene Häufigkeitsmethoden zur Generierung von Signalen implementiert. Um deren Funktionsweise zu demonstrieren, wurden diese in Abschnitt 4.2 auf Medikamente und Nebenwirkungen aus einem Bericht des BfArM angewendet. Die in dem Bericht beschriebenen Zusammenhänge haben sich teilweise durch die Generierung von Signalen in den FAERS-Daten widerspiegelt. Auf Basis der generierten Signale wurde eine Teilpopulation ausgewählt, bei der Sehnenentzündungen nach der Einnahme von Ciprofloxacin aufgetreten sind. Die Kombination von Ciprofloxacin und Sehnenentzündung hat ein untersuchenswertes Signal ausgelöst.

Die Gesamt- und Teilpopulation wurden in Kapitel 5 in einer explorativen Datenanalyse untersucht und Unterschiede zwischen den beiden Populationen aufgezeigt. Dabei wurden Statistiken über die Populationen gesammelt und visualisiert. Zwischen den Populationen wurden bei einigen Attributen deutliche Unterschiede festgestellt, die den durch das Signal suggerierten Zusammenhang zwischen Ciprofloxacin und Sehnenentzündung bekräftigen. Wie in dieser Arbeit jedoch wiederholt betont wurde, kann aufgrund der Natur von Pharmakovigilanzdaten keine Aussage darüber getroffen werden, ob zwischen Ciprofloxacin und Sehnenentzündung auch ein kausaler Zusammenhang besteht.

## 6.1 Fazit

Im Fazit wird geklärt, inwiefern die in Abschnitt 1.2 gesteckten Ziele dieser Arbeit erreicht wurden. Die Daten wurden erfolgreich in Python importiert und aufbereitet. Beim Import der Daten sind Probleme aufgetreten, die hauptsächlich durch den hohen Hauptspeicherverbrauch von *Strings* verursacht wurden. Um diesem Problem zu begegnen, wurde der Import speicherplatzsparender gestaltet. Mit den entwickelten Funktionen zum Import und zur Aufbereitung der Daten können zukünftige Exporte der FAERS-Daten importiert und aufbereitet werden, vorausgesetzt diese weisen die gleiche Struktur wie die bisherigen Daten auf. Auf Änderungen in der Struktur der Daten kann dank der einfachen Anpassbarkeit des Imports und der Aufbereitung der Daten entsprechend reagiert werden.

Auch für die Signalgenerierung wurden verschiedene Funktionen implementiert. Es können Kontingenztafeln für Kombinationen von einem Medikament mit einem unerwünschten Ereignis erstellt werden. Auf Basis der Kontingenztafeln lassen sich dann verschiedene Häufigkeitsmethoden berechnen. Anhand der Häufigkeitsmethoden können Signale generiert werden, wenn ein statistisch signifikanter Zusammenhang zwischen einem Medikament und einem unerwünschten Ereignis festgestellt wurde. Weiterhin wurden für die explorative Analyse der Daten zahlreiche Funktionen entwickelt. Mit Hilfe dieser Funktionalitäten wurden verschiedene Statistiken über den FAERS-Datensatz gesammelt und visualisiert. Dadurch wurde ein umfassender Überblick über die Eigenschaften des FAERS-Datensatzes gegeben. Weiterhin wurden Statistiken über die Teilpopulation erstellt und anhand von zahlreichen Abbildungen die Unterschiede zur Gesamtpopulation aufgezeigt.

Die entwickelten Funktionalitäten können als Grundlage für ein neues Analysetool für Pharmakovigilanzdaten dienen. Dieses könnte, im Gegensatz zu *OpenVigil*, auch eine Visualisierung der Daten anbieten. Um auf Basis der Implementierung ein neues System zu entwickeln, sind allerdings noch einige Verbesserungen und Weiterentwicklungen notwendig. Die Visualisierung der Daten für unterschiedliche Teilpopulationen ist größtenteils noch nicht generisch und müsste für ein solches System angepasst werden. Es sei jedoch darauf hingewiesen, dass eine generische Implementierung im Hinblick auf die Entwicklung eines neuen Systems nicht Ziel dieser Arbeit war. Es lässt sich daher festhalten, dass die in Abschnitt 1.2 gesteckten Ziele dieser Arbeit erreicht wurden. Die Probleme beim Import der Daten haben zwar zu einem höheren Zeitaufwand geführt, die Zielerreichung jedoch nicht beeinträchtigt.

## 6.2 Ausblick

Im Ausblick werden mögliche Verbesserungen und Weiterentwicklungen der Implementierung vorgeschlagen. Ein Problem des FAERS-Datensatzes sind Duplikate, die durch Berichte entstehen, die sowohl an die FDA direkt, als auch an die für die Herstellung der Medikamente verantwortlichen Pharmakonzerne gesendet werden. Diese sind gesetzlich dazu verpflichtet, empfangene Berichte über unerwünschte Ereignisse an die FDA weiterzuleiten. Vom FAERS erhalten diese Berichte unterschiedliche Identifikationsnummern und sind daher nicht mehr anhand dieser als Duplikate zu erkennen. Um doppelte Berichte zu identifizieren, könnte die von Norén et al. vorgeschlagene „Hit-Miss“-Methode angewendet werden [21].

Die Beschreibungen der unerwünschten Ereignisse könnten mit dem MedDRA [4] bereinigt werden. Auf Basis des MedDRA könnten die unerwünschten Ereignisse außerdem klassifiziert werden. Dies würde eine Analyse von Untergruppen von unerwünschten Ereignissen ermöglichen. Die Bezeichnungen der Medikamente könnten mit der *DrugBank* [31] bereinigt werden. Zu den Medikamenten könnten anhand der *DrugBank* außerdem die Wirkstoffe der Medikamente ergänzt werden. Zu beachten ist dabei, dass Medikamente nicht nur einen sondern mehrere Wirkstoffe beinhalten können. Die Wirkstoffe müssten daher entweder als Aggregate in den Berichten, oder der Bericht redundant mit jeder Kombination eines Wirkstoffs mit einem unerwünschten Ereignis gespeichert werden. Eine weitere Information, die mit der *DrugBank* ergänzt werden könnte, ist die Anatomisch-Therapeutisch-Chemische-Klassifikation (ATC-Klassifikation) der Medikamente und/oder Wirkstoffe. Diese klassifiziert Medikamente und Wirkstoffe nach dem Organ auf das sie einwirken, und nach ihren chemischen, pharmakologischen und therapeutischen Eigenschaften. Anhand der ATC-Klassifikation könnten Untergruppen von Medikamenten analysiert werden.

Die Methoden zur Signalgenerierung könnten um die Bayes'schen Methoden Multi-item Gamma Poisson Shrinker (MGPS) von DuMouchel [15] und das Bayesian Confidence Propagation Neural Network (BCPNN) von Bate et al. [9] erweitert werden. Außerdem könnte die Möglichkeit der Analyse von Assoziationen zwischen mehreren Medikamenten und einem unerwünschten Ereignis oder einem Medikament und mehreren unerwünschten Ereignissen ergänzt werden. Auf Basis der entwickelten Funktionalitäten könnte ein neues Analysetool für Pharmakovigilanzdaten aufgebaut werden. Dieses könnte, im Gegensatz zu *Open Vigil*, neben der Berechnung verschiedener Häufigkeitsmethoden, auch eine umfangreiche Visualisierung der Daten anbieten.

# Literaturverzeichnis

- [1] *Class XMLParser*. – URL <https://lxml.de/api/lxml.etree.XMLParser-class.html>. – Abgerufen am 7. September 2019.
- [2] *The ElementTree XML API*. – URL <https://docs.python.org/2/library/xml.etree.elementtree.html>. – Abgerufen am 7. September 2019.
- [3] *ISO-3166-Countries-with-Regional-Codes*. – URL <https://github.com/luke/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>. – Abgerufen am 7. September 2019.
- [4] *Medical Dictionary for Regulatory Activities (MedDRA)*. – URL <https://www.meddra.org/>. – Abgerufen am 25. Oktober 2019.
- [5] *pickle - Python object serialization*. – URL <https://docs.python.org/3/library/pickle.html>. – Abgerufen am 7. September 2019.
- [6] *What is FOIA?*. – URL <https://www.foia.gov/about.html>. – Abgerufen am 4. Oktober 2019.
- [7] APPLE INC.: *MacBook Pro (15-inch, 2018) - Technical Specifications*. 2019. – URL [https://support.apple.com/kb/SP776?locale=en\\_US](https://support.apple.com/kb/SP776?locale=en_US). – Abgerufen am 29. Oktober 2019.
- [8] BATE, Andrew: Bayesian Confidence Propagation Neural Network. In: *Drug Safety* 30 (2007), Nr. 7, S. 623–625. – URL <https://doi.org/10.2165/00002018-200730070-00011>. – Abgerufen am 31. Oktober 2019. – ISSN 1179-1942
- [9] BATE, Andrew ; LINDQUIST, Marie ; EDWARDS, Ivor R. u. a.: A Bayesian neural network method for adverse drug reaction signal generation. In: *European Journal of Clinical Pharmacology* 53 (1998), S. 315–321. – URL <https://link.springer.com/article/10.1007%2Fs002280050466>. – Abgerufen am 18. Oktober 2019.

- [10] BUNDESINSTITUT FÜR ARZNEIMITTEL UND MEDIZINPRODUKTE: *Fluorchinolone: Schwere und langanhaltende Nebenwirkungen im Bereich Muskeln, Gelenke und Nervensystem.* – URL [https://www.bfarm.de/SharedDocs/Risikoinformationen/Pharmakovigilanz/DE/RV\\_STP/a-f/fluorchinolone-bewegungsapparat.html](https://www.bfarm.de/SharedDocs/Risikoinformationen/Pharmakovigilanz/DE/RV_STP/a-f/fluorchinolone-bewegungsapparat.html). – Abgerufen am 20. Oktober 2019.
- [11] BUNDESMINISTERIUM DER JUSTIZ UND FÜR VERBRAUCHERSCHUTZ, BUNDESAMT FÜR JUSTIZ: *Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen (GCP-Verordnung - GCP-V).* – URL <https://www.gesetze-im-internet.de/gcp-v/GCP-V.pdf>. – Abgerufen am 11. Oktober 2019.
- [12] BÖHM, Ruwen: *Primer on Disproportionality Analysis.* 2018. – URL <http://openvigil.sourceforge.net/doc/DPA.pdf>. – Abgerufen am 2. Oktober 2019.
- [13] BÖHM, Ruwen ; HÖCKER, Jan ; CASCORBI, Ingolf u. a.: OpenVigil—free eyeballs on AERS pharmacovigilance data. In: *Nature biotechnology* 30 (2012), S. 137–138
- [14] DELAMOTHE, Tony: *Reporting adverse drug reactions.* 1992. – URL <https://www.who-umc.org/research-scientific-development/signal-detection/what-is-a-signal/>. – Abgerufen am 25. Oktober 2019. Definition vorgestellt beim Treffen der am WHO Programm für die internationale Überwachung von Medikamenten teilnehmenden nationalen Zentren 1991.
- [15] DUMOUCHEL, William: Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. In: *American Statistician* 53 (1999), S. 177–190
- [16] EGBERTS, Antoine C. ; MEYBOOM, Ronald H. ; PUIJENBROEK, Eugène P. van: Use of Measures of Disproportionality in Pharmacovigilance. In: *Drug Safety* 25 (2002), Nr. 6, S. 453–458. – URL <https://doi.org/10.2165/00002018-200225060-00010>. – Abgerufen am 12. November 2019. – ISSN 1179-1942
- [17] EVANS, Stephen J. W.: Pharmacovigilance: a science or fielding emergencies? In: *Statistics in Medicine* 19 (2000), Nr. 23, S. 3199–3209. – URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0258%2820001215%2919%3A23%3C3199%3A%3AAID-SIM621%3E3.0.CO%3B2-Q>. – Abgerufen am 11. November 2019.

- [18] EVANS, Stephen J. W. ; WALLER, Patrick C. ; DAVIS, S.: Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. In: *Pharmacoepidemiology and Drug Safety* 10 (2001), Nr. 6, S. 483–486. – URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/pds.677>. – Abgerufen am 25. Oktober 2019.
- [19] HUNTER, John D.: Matplotlib: A 2D graphics environment. In: *Computing in Science & Engineering* 9 (2007), Nr. 3, S. 90–95
- [20] MCKINNEY, Wes: Data Structures for Statistical Computing in Python. In: WALT, Stéfan van der (Hrsg.) ; MILLMAN, Jarrod (Hrsg.): *Proceedings of the 9th Python in Science Conference*, 2010, S. 51–56
- [21] NORÉN, Niklas ; ORRE, Roland ; BATE, Andrew: A hit-miss model for duplicate detection in the WHO drug safety database, URL [https://www.researchgate.net/publication/221653737\\_A\\_hit-miss\\_model\\_for\\_duplicate\\_detection\\_in\\_the\\_WHO\\_drug\\_safety\\_database](https://www.researchgate.net/publication/221653737_A_hit-miss_model_for_duplicate_detection_in_the_WHO_drug_safety_database), 2005, S. 459–468. – Abgerufen am 11. November 2019.
- [22] PÉREZ, Fernando ; GRANGER, Brian E.: IPython: a System for Interactive Scientific Computing. In: *Computing in Science and Engineering* 9 (2007), Nr. 3, S. 21–29. – URL <https://ipython.org>. – Abgerufen am 25. Oktober 2019. – ISSN 1521-9615
- [23] PLOTLY TECHNOLOGIES INC.: *Collaborative data science*. 2015. – URL <https://plot.ly>. – Abgerufen am 11. November 2019.
- [24] POLUZZI, Elisabetta ; RASCHI, Emanuel ; PICCINNI, Carlo u. a.: *Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS)*. In: *Data Mining Applications in Engineering and Medicine*, URL <https://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/data-mining-techniques-in-pharmacovigilance-analysis-of-the-publicly-accessible-fda-adverse-event-re>, 2012. – Abgerufen am 25. Oktober 2019.
- [25] U.S. FOOD AND DRUG ADMINISTRATION: *About FDA*. – URL <https://www.fda.gov/about-fda/what-we-do>. – Abgerufen am 29. September 2019.

- [26] U.S. FOOD AND DRUG ADMINISTRATION: *FDA Adverse Event Reporting System (FAERS) Public Dashboard*. – URL <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard>. – Abgerufen am 11. Oktober 2019.
- [27] U.S. FOOD AND DRUG ADMINISTRATION: *FDA Adverse Event Reporting System (FAERS) Quarterly Data Extract Files*. – URL <https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html>. – Abgerufen am 4. Oktober 2019.
- [28] U.S. FOOD AND DRUG ADMINISTRATION: *Questions and Answers on FDA's Adverse Event Reporting System (FAERS)*. – URL <https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers>. – Abgerufen am 28. September 2019.
- [29] U.S. FOOD AND DRUG ADMINISTRATION (FDA) CENTER FOR DRUG EVALUATION AND RESEARCH (CDER) OFFICE OF SURVEILLANCE AND EPIDEMIOLOGY (OSE): „XML\_NTS.doc“ file for the quarterly data extract (QDE) from the FDA Adverse Event Reporting System (FAERS). – URL <https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html>. – Abgerufen am 7. September 2019.
- [30] WASKOM, Michael ; BOTVINNIK, Olga ; O'KANE, Drew u. a.: *mwaskom/seaborn:v0.9.0*. 2018. – URL <https://doi.org/10.5281/zenodo.1313201>. – Abgerufen am 11. November 2019.
- [31] WISHART, David ; DJOUMBOU, Yannick ; GUO, An C. u. a.: DrugBank 5.0: A major update to the DrugBank database for 2018. In: *Nucleic acids research* 46 (2017)
- [32] WORLD HEALTH ORGANIZATION: *The Importance of Pharmacovigilance*. 2002. – URL [https://www.who.int/medicines/areas/quality\\_safety/safety\\_efficacy/pharmvigi/en/](https://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/). – Abgerufen am 11. Oktober 2019.
- [33] ZIEGER, Sören: *Masterarbeit - Statistische Methoden des Data Mining in der Pharmakovigilanz*. 2013

# A Anhang

## A.1 Übersetzungen der Nebenwirkungen aus dem BfArM-Bericht in MedDRA PTs

- Sehnenentzündung: tendonitis
- Sehnenriss: tendon rupture
- Muskelschmerzen: myalgia
- Muskelschwäche: muscle fatigue
- Gelenkschmerzen: arthralgia
- Gelenkschwellungen: joint swelling
- Schwierigkeiten beim Gehen: kein passender MedDRA PT gefunden
- Gefühle von Nadelstichen oder Kribbeln: kein passender MedDRA PT gefunden
- Brennende Schmerzen: burning sensation
- Müdigkeit: fatigue
- Depressionen: depression
- Gedächtnisstörungen: amnesia
- Schlafstörungen: insomnia
- Probleme beim Sehen: kein passender MedDRA PT gefunden
- Probleme beim Hören: kein passender MedDRA PT gefunden
- Veränderter Geschmacks- oder Geruchssinn: kein passender MedDRA PT gefunden

## A.2 ME-, Chi-Quadrat- und PRR-Werte nach Evans et al. [18] für Medikamente und Nebenwirkungen aus dem BfArM-Bericht

Medikament	Nebenwirkung	ME	Chi-Quadrat	PRR
Ciprofloxacin	Sehnenentzündung	621	17.412,45	33,15
Levofloxacin	Sehnenentzündung	461	11.729,72	29,49
Moxifloxacin	Sehnenentzündung	37	401,19	13,15
Norfloxacin	Sehnenentzündung	5	47,93	13,97
Ofloxacin	Sehnenentzündung	51	1.577,85	33,78
Ciprofloxacin	Sehnenriss	351	6.843,88	23
Levofloxacin	Sehnenriss	329	7.498,13	26,46
Moxifloxacin	Sehnenriss	30	333,18	13,5
Ofloxacin	Sehnenriss	13	106,34	10,85
Ciprofloxacin	Muskelschmerzen	825	1.631,16	3,72
Levofloxacin	Muskelschmerzen	471	464,14	2,6
Norfloxacin	Muskelschmerzen	11	11,49	2,85
Ciprofloxacin	Muskelschwäche	26	74,3	4,84
Levofloxacin	Muskelschwäche	19	45,09	4,34
Ciprofloxacin	Gelenkschmerzen	1.405	1.662,44	2,82
Levofloxacin	Gelenkschmerzen	904	618,3	2,23
Norfloxacin	Gelenkschmerzen	20	13,83	2,3
Ciprofloxacin	Gelenkschwellung	287	156,8	2,07
Ciprofloxacin	Brennende Schmerzen	348	649,11	3,61
Ofloxacin	Brennende Schmerzen	21	25,16	2,95
Norfloxacin	Müdigkeit	36	19,14	2,05
Ciprofloxacin	Depression	691	515,16	2,31
Ofloxacin	Depression	48	29,82	2,18
Norfloxacin	Schlafstörung	13	6,22	2,07

Tabelle A.1: ME-, Chi-Quadrat- und PRR-Werte nach Evans et al. [18] für Medikamente und Nebenwirkungen aus dem BfArM-Bericht

## Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Gemäß der Allgemeinen Prüfungs- und Studienordnung ist zusammen mit der Abschlussarbeit eine schriftliche Erklärung abzugeben, in der der Studierende bestätigt, dass die Abschlussarbeit „– bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit [(§ 18 Abs. 1 APSO-TI-BM bzw. § 21 Abs. 1 APSO-INGI)] – ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt wurden. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich zu machen.“

*Quelle: § 16 Abs. 5 APSO-TI-BM bzw. § 15 Abs. 6 APSO-INGI*

## Erklärung zur selbstständigen Bearbeitung der Arbeit

Hiermit versichere ich,

Name: \_\_\_\_\_

Vorname: \_\_\_\_\_

dass ich die vorliegende Bachelorarbeit – bzw. bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit – mit dem Thema:

### **Auswertung von Pharmakovigilanzdaten auf Basis des FAERS Datensatzes**

ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

\_\_\_\_\_  
Ort                      Datum                      Unterschrift im Original