universität wien

MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„Analyse der Verrechnungsdaten der österreichischen Krankenanstalten mit Bezug auf unerwünschte Arzneimittelereignisse"

verfasst von / submitted by

Paulina Ratajczak, Bakk. rer. soc. oec.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 951
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Magisterstudium Statistik
Betreut von / Supervisor:	ao. Univ.-Prof. Mag. Dr. Wilfried Grossmann

Abkürzungsverzeichnis

ATC ..... Anatomisch-Therapeutisch-Chemischer Code
MBDS ..... Minimal Basic Data Set
UAE ..... Unerwünschte Arzneimittelereignisse
ADE ..... Adverse Drug Events
COPD ..... Chronic obstructive pulmonary disease
DM ..... Diabetes Mellitus
ICD ..... International Statistical Classification of Diseases and Related Health Problems
MEL ..... medizinische Einzelleistungen

Inhalt

1 Einleitung

Die administrativen Daten in der österreichischen Gesundheitsversorgung bieten ein breites Spektrum für Analysen, die für die Zwecke der evidenzbasierten Medizin genutzt werden. Durch die Verpflichtung der Krankenhäuser genaue Dokumentationen zu führen, entstehen Datenbanken mit Informationen wie durchgeführten Leistungen und gestellten Diagnosen, bei jedem stationären oder teilstationären Krankenhausaufenthalt in Österreich. Solche Datenressourcen sind nützlich zur Beantwortung sowohl medizinischer als auch gesundheitspolitischer Fragestellungen. Die Auswertungen aus dem ersten Teil dieser Arbeit wurden in dem Projekt "Drug safety 3 - Analysis of Adverse Drug Events in Austrian hospital discharge diagnoses between 2001 and 2011" (Christoph Rinner, 2015) veröffentlicht. Im Rahmen dieser Kooperation, konnte der Zugang zu zwei wichtigen Datenbeständen genutzt werden, um weitere, ausführliche und interessante Untersuchungen durchführen zu können. Die erste Datenquelle ist das Minimal Basic Data Set (MBDS), das für die Jahre 2001-2011 vorhanden ist. Bei dem zweiten Datensatz, handelt es sich um die Verrechnungsdaten der österreichischen Krankenhausanstalten für Jahre 2006-2007, die im Rahmen des Projekts zur Verfügung gestellt worden sind. Die beiden Datenbestände unterscheiden sich in mehreren, bedeutenden Aspekten (siehe dazu den Abschnitt Datenquellen). Ziel dieser Arbeit ist es die Methoden, die zu Auswertung der tatsächlich vorhandenen Datenressourcen genützt werden, näher darzustellen und mit Hilfe verschiedener Datenquellen zu evaluieren. Der medizinische Aspekt in dieser Arbeit betrifft die unerwünschten Arzneimittelereignisse.

1.1 Unerwünschte Arzneimittelereignisse

Krankenhausaufenthalte die auf unerwünschte Arzneimittelereignisse (kurz UAE) zurück zu führen sind, werden als UAE-Aufenthalte bezeichnet. Bei jedem Spitalaufenthalt werden eine Hauptdiagnose und mehrere Nebendiagnosen (falls vorhanden) dokumentiert. Die Kodierung der Diagnosen erfolgt gemäß dem ICD-10 Klassifikation System (eng. „International Statistical Classification of Diseases and Related Health Problems")[1]. Bei jeder Diagnose die gestellt wird, ist es ebenfalls zu vermerken, ob eine UAE-Diagnose vorliegt. Bei unerwünschten Arzneimittelereignissen handelt es sich um

[1] Die offizielle Suchmaschine für ICD-10 Codes ist unter http://www.icd-code.de zugänglich.

Ereignisse, die unmittelbar nach einer medikamentösen Behandlung auftreten können. Oft führen solche Ereignisse zu einem Aufenthalt im Krankenhaus und können somit erkannt und mittels eines ICD-10-Codes vermerkt werden. Es wurden 505 ICD-10 Kodierungen identifiziert die auf unerwünschte Arzneimittelereignisse hinweisen (Stausberg Jürgen, 2010). Es handelt sich meistens um Vergiftungen oder allergische Reaktionen bis zu anaphylaktischen Schocks, die durch die Einnahme eines Medikamentes ausgelöst werden können. Die Kategorisierung der UAE-Diagnosen ist in Tabelle 1 ersichtlich. Zu beachten ist die hierarchische Einstufung der Diagnosen. Die Kategorien A1 und B1 sind die Hauptkategorien und werden hinsichtlich ihrer Definition als gleichwertig betrachtet, die Kategorien A2 und B2 sind jeweils die Abschwächungen davon. Danach folgen wie in der Tabelle dargestellt die C, D und E Diagnosen. Die ersten 5 Kategorien werden auch als K5 bezeichnet. Die Klassifizierung von Stausberg ist bereits in zahlreichen Publikationen im deutschsprachigen Raum zur Kategorisierung von Diagnosen verwendet worden. Auch auf der internationalen Ebene ist diese Klassifizierung anerkannt, was internationale Vergleiche, wie in der Publikation „International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA“ (Stausberg, 2014) oder übergreifende Übersichtsarbeiten (Hohl CM, 2013) ermöglicht.

Tabelle 1 Kategorisierung der UAE-Diagnosen (Christoph Rinner, 2015, S. 9)

UAE-Kategorie	Definition
A1	Induzierung durch Arzneimittel
A2	Induzierung durch Arzneimittel oder andere Ursachen
B1	Vergiftung durch Arzneimittel
B2	Vergiftung oder schädlicher Gebrauch durch Arzneimittel oder andere Ursachen
C	Unerwünschte Arzneimittelereignisse sehr wahrscheinlich
D	Unerwünschte Arzneimittelereignisse mäßig wahrscheinlich
E	Unerwünschte Arzneimittelereignisse weniger wahrscheinlich

2 Fragestellung

2.1 Zeitliche Entwicklung von Diagnosen

Der erste Teil dieser Arbeit ist die Analyse, die auf dem Mini Basic Data Set basiert. Hier soll die Gelegenheit genützt werden, dass die Daten für die Zeitspanne von 11 Jahren vorliegen. Damit sind die zeitliche Entwicklung, Häufigkeiten sowie der Einfluss demografischen Faktoren auf Diagnosen, die auf unerwünschte Arzneimittelereignisse hindeuten hinreichend erfasst. Bei einem Krankenhausaufenthalt werden nicht selten mehrere Diagnosen gestellt, dabei unterscheidet man zwischen der Haupt- und Nebendiagnose. Analysiert werden Aufenthalte, bei denen eine UAE-Diagnose entweder als Haupt- oder Nebendiagnose gestellt wurde. Unter anderem werden die häufigsten Diagnosen in jeder Stausberg Kategorie ermittelt. Nach Betrachtung dieser Kategorien konnte beobachtet werden, dass manche Nebenwirkungen eines Arzneimittels unmittelbar nach der Einnahme des Medikamentes auftreten, andere wiederum als Langzeitfolgen eingestuft werden können. Der zweite Teil der Auswertungen ist der Assoziation der Begleiterkrankungen mit den UAE-Diagnosen/Aufenthalten gewidmet. Präzise gesagt werden in diesem Abschnitt die häufigsten Kombinationen der UAE-Diagnosen mit anderen Diagnosen beschrieben. Somit konnten Erkrankungen ausfindig gemacht worden, bei denen ein unerwünschtes Arzneimittelereignis besonders häufig vorkommt.

2.2 Ereigniszeitanalyse

Die dritte Fragestellung wird mithilfe der Verrechnungsdaten der österreichischen Krankenanstalten für die Jahre 2006-2007 analysiert. Diese liegen zwar nur für den Zeitraum von 2 Jahren vor, enthalten jedoch wichtige Zusatzinformationen, die spezifischere Fragestellungen ermöglichen. Durch die Analyse der Kombinationen der UAE-Diagnosen mit weiteren Diagnosen mittels MBDS-Daten, sind über die gesamte Zeitspanne von 11 Jahren koronare Herzkrankheit und Herzinfarkt als Grunderkrankung sichtbar geworden. Weiter konnten unerwünschte Arzneimittelereignisse mit gewissen Begleiterkrankungen in Verbindung gebracht werden. Somit wird die weitere Auswertung auf einem explizit vorbestimmten Patientenkollektiv durchgeführt. Die Verrechnungsdaten ermöglichen eine Extraktion aller Patienten, die an der Grunderkrankung

leiden und enthalten weitere Informationen hinsichtlich der Komorbiditäten (Begleiterkrankungen) und gegebenenfalls auch bezüglich medikamentöser Behandlung. Somit kann die weitere Analyse der unerwünschten Arzneimittelereignisse im gewissen medizinischen Kontext stattfinden. Ergänzend wird eine Ereigniszeitanalyse durchgeführt, die das Risiko einer UAE-Diagnose differenziert und zu drei verschiedenen Zeitpunkten quantifiziert.

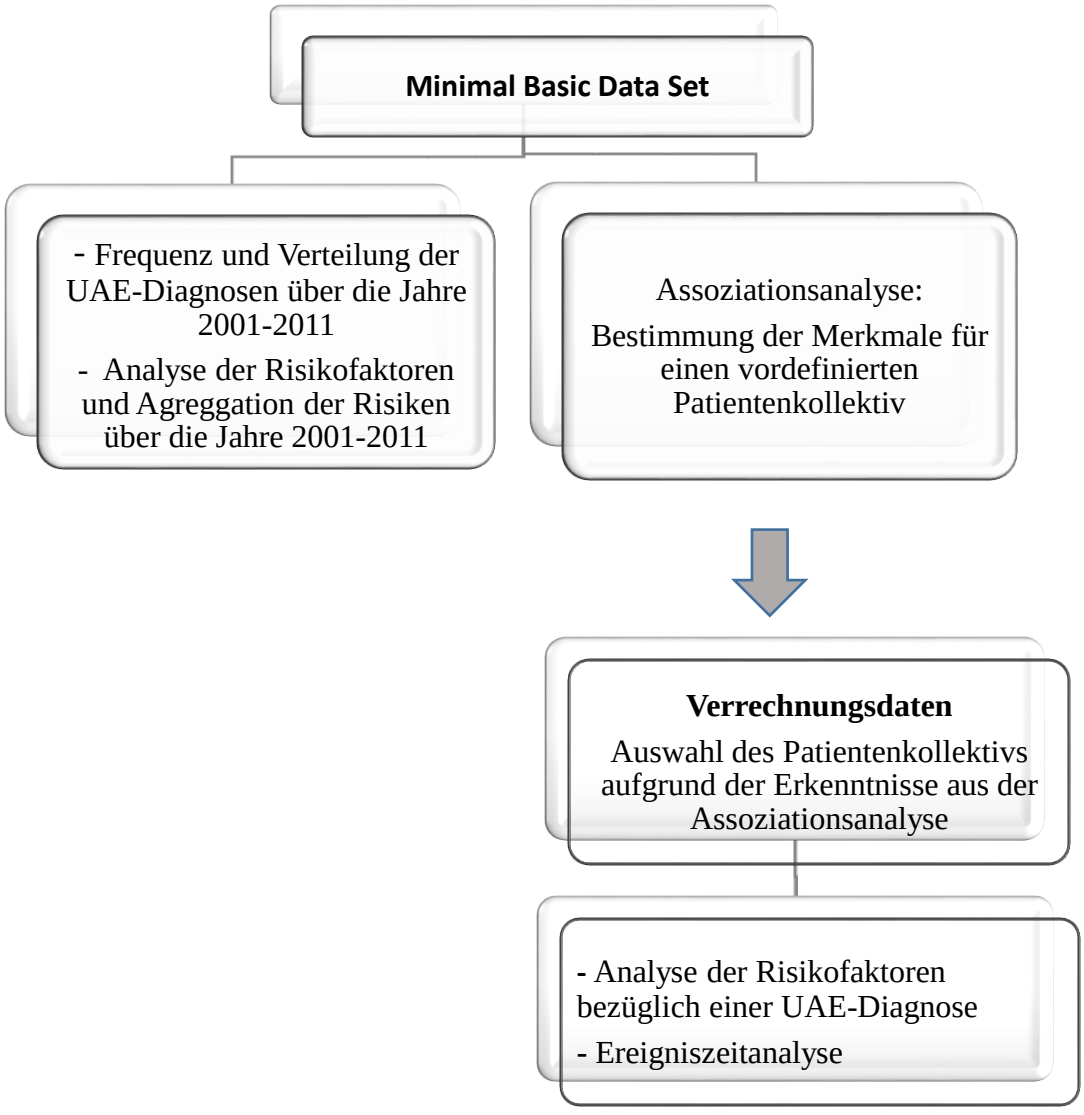

Abbildung 1 Verlauf der Analysen

3 Datenquellen

Für die statistische Analyse wurde mit zwei verschiedenen Datenquellen gearbeitet. Die Struktur der beiden Datensätze wird in folgenden Abschnitten näher beschrieben.

3.1 MBDS – Daten

Die Analyse der Diagnosen mit Bezug auf unerwünschte Arzneimittelereignisse basierte auf den Mini Basic Data Set (MBDS-Daten) für die Jahre 2001 - 2011. Die entsprechend lange Zeitspanne von 11 Jahren, bietet die Möglichkeit zeitliche Entwicklung der UAE-Aufenthalte zu analysieren. Das Mini Basic Data Set ist aus wichtigen administrativen Gründen entstanden. Es ist ein auf nationalen Ebene aufgebaute Datensatz, in dem jeder stationäre oder teilstationäre Aufenthalt in einer Krankenanstalt des jeweiligen Landes dokumentiert wird. Das Mini Basic Data Set bildet Grundlage für eine Vielzahl der Analysen in medizinischen Bereich (Trybou J, 2014) (Sarria-Santamera, 2013). Die wichtigste Charakteristik des Datensatzes ist, dass die Daten Aufenthalts bezogen sind. Der MBDS-Datensatz enthält unter anderem folgende Angaben:

- Administrative Informationen bezüglich eines Aufenthaltes wie: Krankenanstaltsnummer (dadurch wird die genaue Identifizierung der Krankenanstalt möglich), Aufenthalt ID, Aufnahme- und Entlassungsdatum, Aufnahme- sowie Entlassungsart.
- Informationen über den Patienten: Geschlecht, Geburtsdatum, Wohnadresse, Staatsbürgerschaft, Versicherungsträger.
- Medizinische Angaben: Hauptdiagnose, Nebendiagnosen (falls vorhanden) sowie durchgeführten Leistungen aus dem Katalog medizinischen Einzelleistungen, kurz MEL[2] genannt.
- Die Daten werden auf Aufenthaltsebene erhoben

Das Konzept der MBDS als Datenquelle und deren Kompatibilität auf europäischer Ebene sind im (Roger, 1981) beschrieben.

[2] Das MEL-Katalog umfasst 1.506 medizinische Leistungen und ist unter http://www.bmgf.gv.at/home/Gesundheit/Krankenanstalten/LKF_Modell_2015/Kataloge frei zugänglich

3.1.1 MBDS-Limitationen

1) Es wurden ausschließlich die Daten aus den Fonds-(öffentlichen) Krankenanstalten herangezogen.
2) Krankenhausaufenthalte mit einer Dauer von 0 - Tagen wurden ebenfalls in die Analyse eingeschlossen.
3) Zwischen der Art der Aufnahme und Art der Entlassung wird nicht differenziert, jeder Aufenthalt der ein Transfer ist wird als ein neuer Aufenthalt herausgelesen.
4) Es wurden nur Patienten mit der österreichischen Staatsbürgerschaft extrahiert.
   Punkt 1-4 vgl. (Christoph Rinner, 2015, S. 72).

3.2 Verrechnungsdaten

Verrechnungsdaten enthalten im Vergleich zur MBDS-Daten zusätzliche Informationen wie zum Beispiel medikamentöse Behandlung des Patienten oder die Informationen über den behandelnden Arzt. Der wesentliche Unterschied besteht jedoch darin, dass bei den Verrechnungsdaten durch die *Patienten ID,* dem Patienten mehrere Aufenthalte zugeordnet werden können, während bei den MBDS-Daten kein Patientenbezug möglich ist. Die Medikamentenverordnungen wurden mit dem ATC-Code kodiert. Für die Extraktion der Variablen und deren Kodierung wurden bestimmte Auswahlkriterien getroffen, diese sind im Abschnitt Limitationen beschrieben. Um die Daten analysieren zu können, mussten zusätzliche Variablen gebildet werden, dieser Vorgang ist in dem Abschnitt Vorbereitung und Bereinigung erläutert.

3.2.1 Verrechnungsdaten-Limitationen

Die genaue Extraktion fand aufgrund folgender Spezifikationen statt: Patientenkollektiv mit Diagnose Herzinfarkt oder koronare Herzkrankheit wurde mittels der ICD10-Codes I21, I22, I25 und deren Untercodes identifiziert. Die COPD Erkrankung wurde entweder mit dem ICD10-Code J44 und dessen Untercodes oder mit dem ATC-Code R03 identifiziert. Behandlung mit Betablocker ist mittels des ATC-Codes C07 bestimmt worden. Diabetes mellitus Erkrankung wurde aufgrund der ICD10-Codes: E10-E14 und ATC-Codes: A10AB, A10AC, A10AD, A10AE, A10BA, A10BB, A10BD, A10BG, A10BX bestimmt. (Christoph Rinner, 2015, S. 12-13)

3.3 Vorbereitung und Bereinigung

3.3.1 MBDS-Daten

Für das Jahr 2001 gab es für manche politischen Bezirke keine Bewohneranzahlen. Diese sind vor der Analyse, durch eine Schätzung ersetzt worden.

Für die Assoziationsanalyse mussten die Diagnosen des Patienten, die bei einem Aufenthalt gestellt worden sind, zusammengefasst werden. Somit ist eine Variable erzeugt worden, die die entsprechende Form einer Liste, mit Elementen die zusammen vorkommen (hier Diagnosen) besitzt.

3.3.2 Verrechnungsdaten

Für die Verrechnungsdaten mussten aufgrund des ausgewählten Patientenkollektivs adäquate Altersgruppen gebildet werden. Als Referenzgruppe diente hier die Altersgruppe 2 (46 - 60 Jahre).

Altersgruppe 1: 0 - 45 Jahre

Altersgruppe 2: 46 - 60 Jahre

Altersgruppe 3: 61 - 75 Jahre

Altersgruppe 4: 76 - 90 Jahre

Altersgruppe 5: 90 +

Alle Analysen in dieser Arbeit wurden auf Datensätzen durchgeführt, bei denen erhobene Merkmale vollständig vorhanden waren. Aufgrund der fehlenden Werte bei den demografischen Faktoren Alter und Geschlecht in den Verrechnungsdaten, wurden Patienten, bei denen diese Merkmale nicht vorhanden waren erstmal aus der Analyse ausgeschlossen. Vollständigkeitshalber wurden die fehlenden Ausprägungen im Nachhinein imputiert und die Analyse wurde wiederholt.

Die gesamten Auswertungen sind mit der Statistik-Software R (Version 3.3.1) (Team, 2016) durchgeführt worden. Das Signifikanzniveau α ist auf 0,05 festgesetzt worden.

Folgende Variablen wurden für die Analyse aus der Datenbank extrahiert:

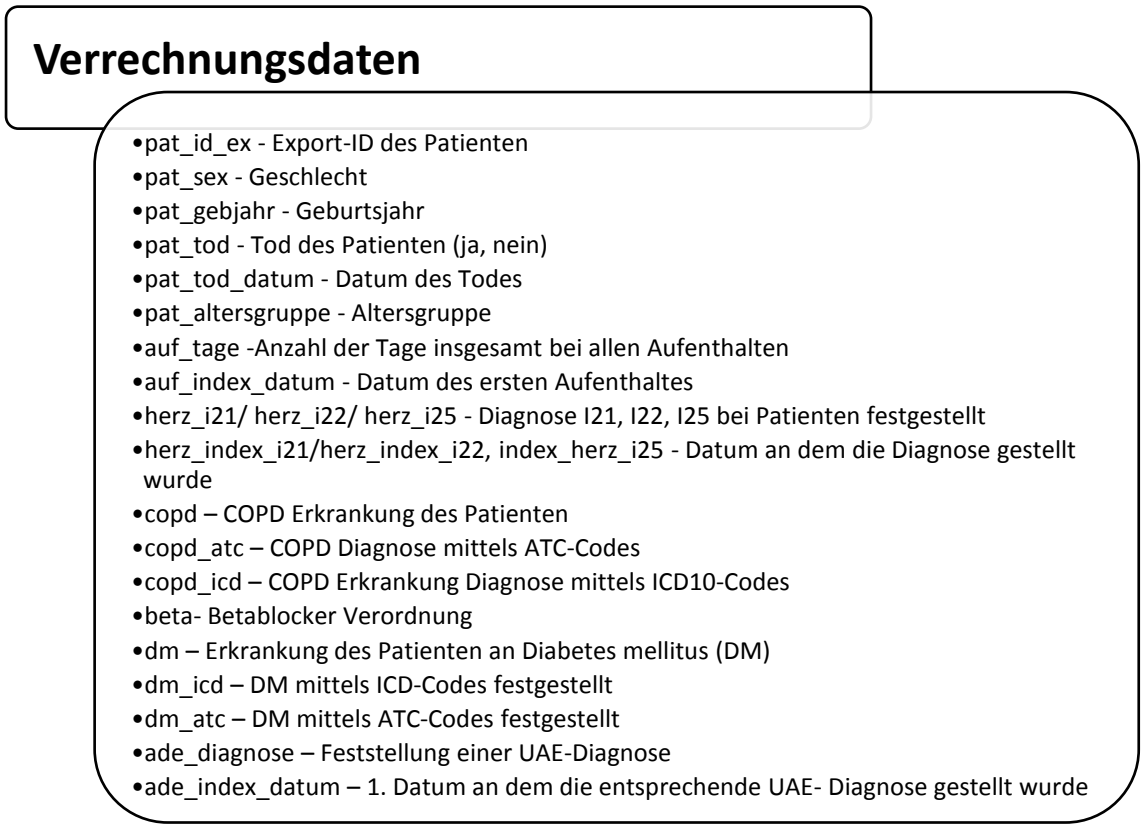

Abbildung 2 Variablen die aus Verrechnungsdaten extrahiert worden sind.

4 Methoden

4.1 Standardisierung

Das Ziel der Standardisierung ist die Vereinheitlichung der Kennzahlen einer Stichprobe, um sie vergleichbar mit Kennzahlen aus einer/mehreren anderen Stichproben zu machen. Dies ist notwendig, wenn die demografischen Merkmale in den zugrundeliegenden Grundgesamtheiten nicht identisch verteilt sind und somit der direkte Vergleich nicht korrekt wäre. Möchte man Auswertungen epidemiologischer Daten auf internationaler Ebene (Länder innerhalb der EU) miteinander vergleichen, so bietet die Standardisierung der statistischen Kennzahlen der jeweiligen Länder, ein geeignetes Verfahren dafür. Dies geschieht indem man bezüglich einer übergreifenden Referenzpopulation standardisiert. Bei der Analyse der österreichischen Daten in Hinblick auf Aufenthalte mit Diagnosen, die auf unerwünschte Arzneimittelereignisse deuten, wurde eine Alters- und Geschlechtsstandardisierung durchgeführt. Dadurch werden die Verteilung des Geschlechts sowie die Altersverteilung, die mittlerweile aufgrund der Alterung der Gesellschaft größere Auswirkungen haben kann, berücksichtigt. Neben der oben erwähnten Referenzpopulation, werden ebenfalls die Einwohnerzahlen in Österreich (unterteilt nach den Ausprägungen der Standardisierungsfaktoren) benötigt.

4.1.1 Bevölkerungszahlen

Für die Bevölkerungszahlen aus Österreich wurden die Daten von der Statistik Austria verwendet. Die statistische Tabelle mit dem Namen „Bevölkerung zu Jahresbeginn ab 1982“[3], für die Jahre 2001-2011 war die genaue Quelle für die Bevölkerungsanzahlen. Die Auswahl dieser Population musste folgende Spezifikationen erfüllen: Kompatibilität der Altersgruppen mit den Gruppen in der Referenz-Standardbevölkerung, die Möglichkeit der genauen Aufsplittung auf die geografischen Versorgungszonen (also Bewohneranzahlen auf der Ebene politischen Bezirke) und gleichzeitig das Vorhandensein der Daten für den entsprechenden Zeitraum (die Jahre 2001-2011).

[3] Statistik Austria http://statcube.at/statcube/opendatabase?id=<database_id>

4.1.2 Referenzpopulation

Die Referenzpopulation kann sich auf geografisch oder politisch vordefinierte Bereiche beziehen sowie verschiedene Faktoren berücksichtigen. Sie gibt eine sogenannte Standardbevölkerung an, die für Korrektur möglicher Verteilungsunterschiede der Population innerhalb eines Landes, innerhalb Europas oder global auf der Weltebene verwendet werden kann. Sie wird von den internationalen Behörden veröffentlicht und dient als Vorgabe für Standardisierungen. Einige Beispiele für solche Standardpopulationen sind: die Europäische Standardbevölkerung der WHO (vom Jahr 1976), die Standardbevölkerung der OECD-Länder vom Jahr 1980 oder die Weltstandardbevölkerung (O.Jacke, 2013, S. 43). Für die Standardisierung in dieser Arbeit, wird die Neue Europa Standardbevölkerung (1990) verwendet (siehe Tabelle 2). Diese von der UNO vorgeschlagene Referenzpopulation, ersetzt die mittlerweile veraltete europäische Standardbevölkerung (Jahr 1966). Epidemiologische Kennzahlen und Häufigkeitsaufzählungen bei Studien auf gleichem Forschungsgebiet, die bezüglich der gleichen Referenzpopulation standardisiert worden sind, also in diesem Fall die Neue Europa-Standardbevölkerung (Jahr 1990), sind miteinander vergleichbar. Somit können Vergleiche auf europäischer Ebene durchgeführt werden.

4.1.3 Berechnung und Anwendung

Als erstes muss die Anpassung der Altersgruppen geschaffen werden und zwar: in den Daten, Bevölkerungszahlen und der Referenzpopulation. Dazu wurden folgende Altersgruppen gebildet:

- Altersgruppe 0: 0-14 Jahre
- Altersgruppe 1: 15-29 Jahre
- Altersgruppe 2: 30-44 Jahre
- Altersgruppe 3: 45-59 Jahre
- Altersgruppe 4: 60-74 Jahre
- Altersgruppe 5: 75 Jahre und älter (Christoph Rinner, 2015, S. 25)

Tabelle 2 Neue Europa-Standardbevölkerung (Statistisches-Bundesamt, 2006)

Alter von...bis... Jahren	"Neue" Europa-Standard-Bevölkerung		
	männlich	weiblich	insgesamt
0 - 1	1.345	1.218	1.305
1 - 4	5.303	4.800	5.021
5 - 9	6.800	6.160	6.472
10 - 14	7.108	6.452	6.772
15 - 19	7.570	6.863	7.208
20 - 24	8.163	7.438	7.292
25 - 29	8.209	7.552	7.871
30 - 34	7.811	7.258	7.528
35 - 39	7.448	6.986	7.212
40 - 44	7.068	6.661	6.860
45 - 49	5.997	5.739	5.865
50 - 54	5.937	5.817	5.876
55 - 59	5.521	5.585	5.553
60 - 64	5.015	5.463	5.245
65 - 69	4.139	5.196	4.680
70 - 74	2.449	3.392	2.932
75 - 79	2.228	3.536	2.897
80 - 84	1.094	2.076	1.606
85 u. mehr	798	1.808	1.305
Insgesamt	100.000	100.000	100.000

Bei den Bevölkerungsanzahlen für Österreich waren exakt die gleichen Altersgruppen vorgebeben. Die Referenzpopulation ist in feinere Abschnitte unterteilt aber durch die Zusammenfassung mehrere Stufen (abwechselnd graue und weise Markierung in der Tabelle 2), konnten ebenfalls Standardpopulationen für gleiche Altersgruppen bestimmt werden. Die Berechnung eines Standardisierungsfaktors für Frauen in der Altersgruppe x und für das Jahr 20xx:

$$\text{Faktor} = \frac{\text{Anteil der Frauen in der Altersgruppe x an allen Frauen in der Referenzpopulation}}{\text{Anteil der Frauen in der Altersgruppe x an allen Frauen in der österreichischen Population}}$$

Nach der obigen Formel können Standardisierungsfaktoren für jede Altersgruppe, jedes Jahr und ebenfalls für Männer analog berechnet werden. Die Bevölkerungsanzahlen Österreichs, sind für jedes Jahr (von 2001 bis 2011) in der Datenbank von Statistik Austria verfügbar. Die Referenzstandardbevölkerung ist dagegen einheitlich vorgegeben, also für alle Jahre werden gleiche Referenzanzahlen für die Standardisierung verwendet. Um die UAE-Aufenthalte quantifizieren zu können, wurden Anzahlen pro 100.000 Einwohner berechnet. Diese Vorgehensweise ist in der Epidemiologie weit verbreitet. Somit werden die Häufigkeiten der Aufenthalte bezüglich der Anzahlen der Bevölkerungsgruppen in den Patientenkollektiven (Männer, Frauen, Altersgruppen) entsprechend skaliert:

$$\text{Anzahl der UAE-Aufenthalte pro 100.000 Einwohner} = \frac{\text{Anzahl der UAE-Aufenthalte}}{\text{Anzahl der Einwohner}} * 100.000$$

(Christoph Rinner, 2015, S. 47)

Zuletzt werden die Anzahlen der UAE-Aufenthalte pro 100.000 Einwohner mit den entsprechenden Standardisierungsfaktoren für die jeweiligen Patientenkollektive multipliziert. Die so ausgerechneten Kennzahlen sind somit von der Verteilung innerhalb der Kollektive unabhängig aber auch von der Bevölkerungsverteilung bereinigt.

4.2 Logistische Regression

4.2.1 Einleitung

Seien $y_1, y_2, \dots, y_n$ Ergebnisse eines Experiments mit zwei möglichen Ausprägungen 0 und 1. Dabei nimmt die Variable Y den Wert 0 mit der Wahrscheinlichkeit p = P (Y = 1) und den Wert 1 mit Gegenwahrscheinlichkeit $(1 - p)$ = P (Y = 0). Möchte man solch eine Größe als abhängige Variable modellieren, wäre die Zielvariable Y, im Unterschied zur linearen Regression binär skaliert, was bedeutet, dass es sich um Schätzung bezüglich der Wahrscheinlichkeiten handelt. Die logistische Regression ermöglicht solche Schätzung und zwar mit der zusätzlichen Nutzung der Informationen von

erklärenden Variablen. Die unabhängigen Variablen können sowohl metrisch, ordinal oder nominal skaliert sein.

4.2.2 Logit-Model

Um den Einfluss der Kovariablen erfassen zu können, würde die Schätzung der Wahrscheinlichkeiten durch eine affine lineare Funktion

$$p_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

dazu führen, dass der geschätzte Wert nicht in dem Wertebereich zwischen 0 und 1 liegt. Um das Problem zu lösen benötigt es eine Transformation mithilfe einer Linkfunktion. Für die Darstellung dieser Linkfunktion betrachtet man zunächst die zentrale Größe bei der Analyse dichotome Merkmale, die Odds. Diese sind definiert als das Verhältnis der Wahrscheinlichkeit zur der Gegenwahrscheinlichkeit dafür, dass die abhängige Variable den Wert 1 annimmt.

$$\text{Odds} = \frac{\text{P(Y=1)}}{\text{P(Y=0)}}$$

Die Odds sind nach oben nicht begrenzt, aber von unten nähern sie sich approximativ dem Nullpunkt. Bildet man in weiterer Folge den natürlichen Logarithmus der Odds genannt *logit*, erhält man eine Größe, die Werte im unbeschränkten Wertebereich annimmt. Die logit-Linkfunktion ist folgendermaßen definiert (Liu, 2015, S. Chapter 3.1.1):

$$logit(\frac{p_i}{1-p_i}) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}$$

Wendet man die Exponentialfunktion auf beiden Seiten der obigen Gleichung an, erhält man den Ausdruck für die Odds:

$$\frac{p_i}{1-p_i} = e^{\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}$$

Eine weitere Rücktransformation erlaubt die Berechnung des exakten Ausdrucks der Wahrscheinlichkeit für das Eintreten eines Ereignisses (das logistische Modell) (Liu, 2015, S. Chapter 3.1.1):

$$p_i = \frac{e^{\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}}{(1 + e^{\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}})}$$

Die folgende Abbildung stellt das logistische Model für eine binär skalierte abhängige Variable dar.

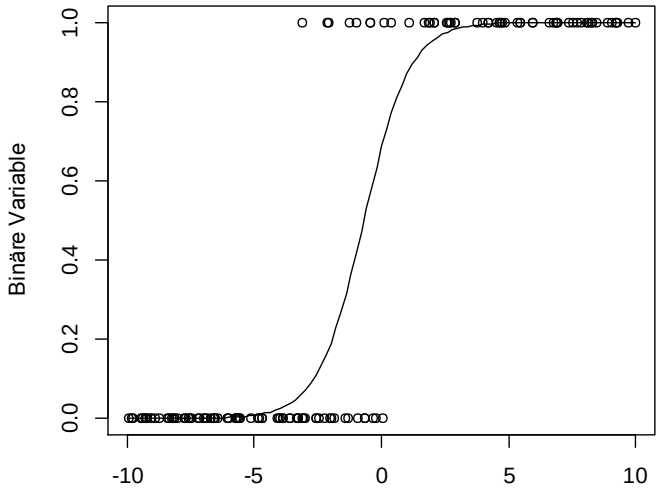

Abbildung 3 Schätzung im logistischen Model[4]

[4] Abbildung stellt das logistische Modell für simulierte Daten dar, erzeugt mit R Version 3.3.1

4.2.3 Maximum Likelihood - Schätzung

Die Schätzung der unbekannten Parameter im logistischen Modell erfolgt mittels Maximum Likelihood-Methode. Das Prinzip dieser Schätzung basiert auf der Maximierung der sogenannten Maximum Likelihood-Funktion.

Angenommen die Daten $y_1,\ y_2, \dots, y_n$ folgen einer bekannten Verteilung, mit einer Wahrscheinlichkeitsfunktion, beziehungsweise Dichte f (y_i; ϴ) wobei ϴ für den/die unbekannten Parameter steht. Die Likelihood-Funktion für n unabhängige, identisch verteilte Beobachtungen ist definiert als (Schlittgen, 2009, S. 97):

$$L(y_i; \theta) = \prod_{i=1}^{n} \mathrm{f}\,(y_i; \theta)$$

In weiterer Folge wird diese Funktion logarithmiert um die Maximierung zur erleichtern. Die log-Likelihood-Funktion wird gebildet (Schlittgen, 2009, S. 97) :

$$l(y_i; \theta) = lnL(y_i; \theta) = ln \prod_{i=1}^{n} \mathrm{f}\,(y_i; \theta) = \sum_{i=1}^{n} ln\ \mathrm{f}\,(y_i; \theta)$$

Für die Berechnung des Maximierers $\hat{\theta}$ = argmax $l(y_i; \theta)$ wird die erste Ableitung der log-Likelihood-Funktion nach ϴ, gleich Null gesetzt. Für einen eindimensionalen, unbekannten Parameter betrachte man daher den Ausdruck (Schlittgen, 2009, S. 97-98):

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta} ln\ \mathrm{f}(y_i; \theta) = \sum_{i=1}^{n} \frac{\partial \mathrm{f}\,(y_i; \theta) / \partial \theta}{\mathrm{f}(y_i; \theta)} = 0$$

Im Fall eines mehrdimensionalen Parameterraums entsteht folgendes Gleichungssystem:

$$\sum_{i=1}^{n} \frac{\partial \mathrm{f}\,(y_i; \Theta_{1,} \Theta_{2,} \dots, \Theta_k) / \partial \Theta_j}{\mathrm{f}(y_i; \Theta_{1,} \Theta_{2,} \dots, \Theta_k)} = 0$$

Wobei (j = 1, …, k) und das Symbol ∂ die Ableitung kennzeichnet (Schlittgen, 2009, S. 97-98).

4.2.4 Schätzung der Koeffizienten

Seien $X_1, X_2, \dots, X_k$ die erklärenden Variablen, Y ~ B (n_i, p_i) eine binomialverteilte, abhängige Variable und $\beta = (\beta_{0,} \beta_1, \dots, \beta_k)$ die zu schätzenden Regressionskoeffizienten. Mit p_i bezeichne man die Erfolgswahrscheinlichkeit. Adaptiert man Wahrscheinlichkeitsfunktion der Binomialverteilung mithilfe obiger Bezeichnungen, nimmt die Likelihood-Funktion im logistischen Modell folgende Gestalt an:

$$L(\beta) = \prod_{i=1}^{n} \mathrm{p}_i{}^{\mathrm{y}_i}\ [1 - \mathrm{p}_i]^{1-\mathrm{y}_i}$$

Die log-Likelihood-Funktion erhält man durch das Bilden des Logarithmus (David W. Hosmer, 2004, S. 8-9):

$$l(\beta) = lnL(\beta) = \sum_{i=1}^{n} \mathrm{y}_i ln\, \mathrm{p}_i + (1 - \mathrm{y}_i) \ln(1 - \mathrm{p}_i)$$

Wobei die Erfolgswahrscheinlichkeit die vorhin im logistischen Modell definierte Form annimmt 4.2.2 (S.19).

Bei der Maximierung der log-Likelihood-Funktion müssen nichtlineare Gleichungssysteme gelöst werden, dies geschieht mithilfe des iterativen Newtonverfahrens (auf das Newtonverfahren wird hier nicht näher eingegangen). Durch das Maximieren der log-Likelihood-Funktion mit Hilfe der Daten, erhält man die unbekannten Regressionskoeffizienten.

In der folgenden Abbildung sind die Likelihood und log-Likelihood-Funktion für die Binomialverteilung dargestellt. Die Abbildung basiert auf einer Simulation von fiktiven Daten, die einer Binomialverteilung folgen, siehe auch (Andreas Behr, 2014, S. 163).

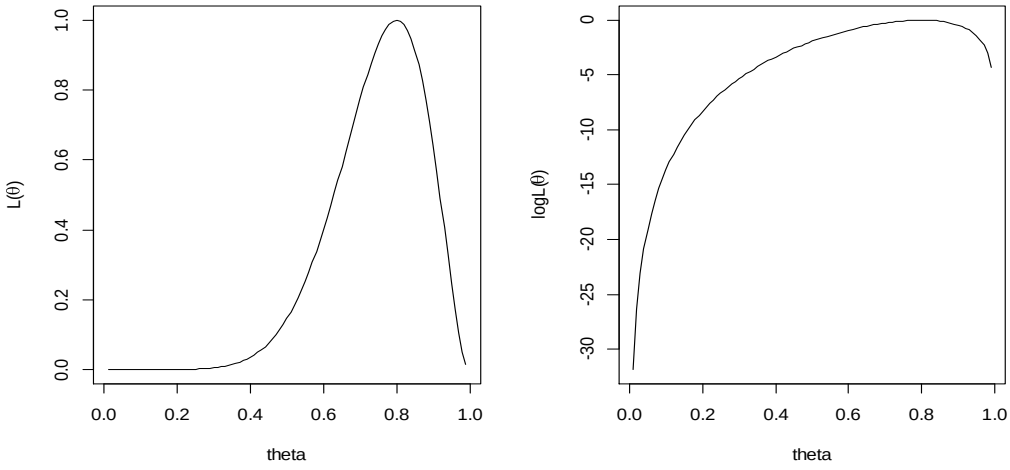

Abbildung 4 Likelihood und log-Likelihood-Funktion der Binomialverteilung

4.2.5 Testung und Interpretation der Koeffizienten

Nachdem die Regressionskoeffizienten und deren Standardfehler (SE für englisch: standard error) geschätzt worden sind, gilt das weitere Interesse der Frage, ob die entsprechenden erklärenden Variablen, eine signifikante Veränderung der abhängigen Zielgröße bewirken. Mithilfe des Wald-Tests kann die Signifikanz der einzelnen Koeffizienten $\beta_{0,}\, \beta_1 , \dots , \beta_k$ getestet werden.

Die Wald-Tests Statistik W ist definiert als die Differenz zwischen dem berechneten Koeffizienten und dessen theoretischem Wert, der unter der Nullhypothese angenom-

men wird (hier $\beta = 0$), dividiert durch den errechneten Standardfehler. Unter den Modellvoraussetzungen ist diese Statistik asymptotisch standardnormalverteilt. Die Wald-Test Statistik ist definiert als (David W. Hosmer, 2004, S. 16):

$$W = \frac{\widehat{\beta_\iota} - 0}{\widehat{SE}(\widehat{\beta_\iota})} \sim N(0,1)$$

Die H_0 Hypothese wird beibehalten, falls der Wert der Teststatistik folgende Ungleichung erfüllt: $z_{\alpha/2} < W \leq z_{1-\alpha/2}$ wobei $z_{\alpha/2}$ und $z_{1-\alpha/2}$ - Quantile der Standardnormalverteilung. (Michael Eid, 2014, S. 188)

Die asymptotischen Konfidenzintervalle (KI) können mit Hilfe des folgenden Ausdrucks geschätzt werden:

$$\text{KI} = (\widehat{\beta_\iota} \pm z_{1-\alpha/2}\, \widehat{SE}(\widehat{\beta_\iota}))$$ (Urban, 1993, S. 60)

Die Koeffizienten können direkt interpretiert werden. Eine Veränderung in X um eine Einheit bewirkt eine multiplikative Veränderung der Odds um dem Faktor e^{β_i}, wobei β_i der Koeffizient für die entsprechende erklärende Variable X_i ist (Klaus Backhaus, 2015, S. 311)

Die folgende Abbildung stellt bildlich dar, wie die Schätzung im logistischen Modell von dem Wert des Koeffizienten abhängt. Bei einer Veränderung der Konstanten erfolgt die Verschiebung der Schätzungskurve, bei der Veränderung eines Regressionskoeffizienten wird deren Steilheit beeinflusst (Klaus Backhaus, 2015, S. 308)

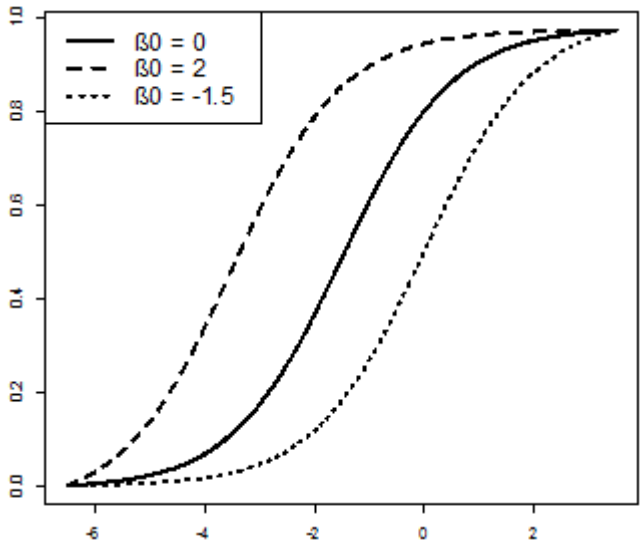

Abbildung 5 Schätzungskurve bei unterschiedlichen Konstanten β_0, nach (Marco Giesselmann, 2013, S. 133)

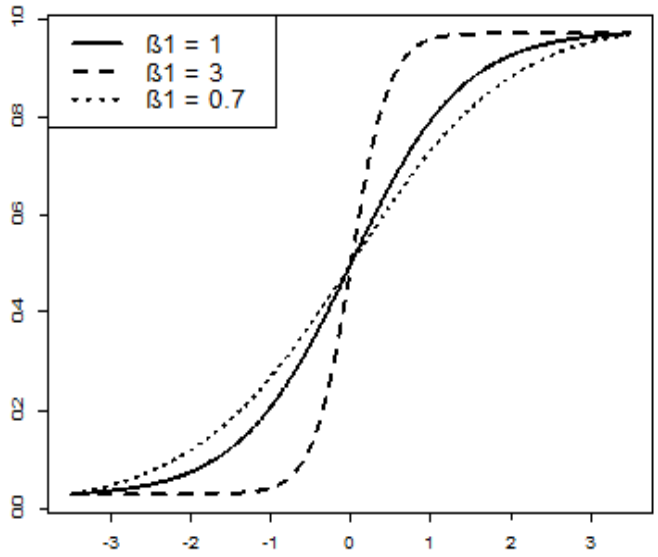

Abbildung 6 Schätzungskurve bei unterschiedlichen Koeffizienten β_i nach (Marco Giesselmann, 2013, S. 133)

4.2.6 Mantel-Haenszel – Methode

Um den Einfluss eines Risikofaktors auf die Wahrscheinlichkeit für das Eintreten eines Ereignisses zu berechnen, werden Odds-Ratios (kurz OR) also Odds-Quotienten der Gruppen (Gruppe 1: der Risikofaktor ist vorhanden, Gruppe 2: kein Risikofaktor) gebildet.

$$OR = Odds(Gruppe\ 1)/Odds\ (Gruppe\ 2)$$

Oft besteht jedoch die Notwendigkeit der Aggregation von Odds-Ratios über mehrere Schichten. Dazu wird der Mantel-Haenszel Schätzer verwendet.

Als Grundlage sei für jede Schicht i = 1, …, k, folgende Tabelle verfügbar, wobei a_i, b_i, c_i, d_i die Häufigkeiten mit entsprechenden Ausprägungskombinationen darstellen, dabei bezeichne man mit $n_i = a_i + b_i + c_i + d_i$ die Gesamtanzahl der Personen in der Schicht i.

Tabelle 3 Grundlage für die Berechnung des Mantel-Haenszel Schätzers (Woodwar, 1999, S. 168)

Risikofaktor	Ereignis eigetreten	Ereignis nicht eigetreten
vorhanden	a_i	b_i
nicht vorhanden	c_i	d_i

Nach (Woodwar, 1999, S. 168) ist der Mantel-Haenszel Schätzer definiert als:

$$OR_{MH} = \sum_{i=1}^{k} \frac{a_i d_i}{n_i} \Big/ \sum_{i=1}^{k} \frac{b_i c_i}{n_i}$$

wobei i = 1, … , k - Anzahl der Schichten.

Um die Konfidenzintervalle angeben zu können muss eine Schätzung des Standardfehlers vorgenommen werden. Die Methode für diese Schätzung wurde von James Robins, Norman Breslow and Sander Greenland [5]entwickelt.

(Woodwar, 1999, S. 169-170) beschreibt diese Methode im Detail. Demnach wird der Mantel-Haenszel Schätzer zunächst logarithmiert und dessen Standardfehler ist definiert als:

$$\widehat{se}(\ln(OR_{MH})) = \sqrt{\frac{\sum P_i R_i}{2(\sum R_i)^2} + \frac{\sum P_i S_i + \sum Q_i R_i}{2\sum R_i \sum S_i} + \frac{\sum Q_i S_i}{2(\sum S_i)^2}}$$

Wobei für jede Schicht i gilt:

$$P_i = \frac{(a_i + d_i)}{n_i}, R_i = \frac{a_i * d_i}{n_i}, Q_i = \frac{(b_i + c_i)}{n_i}, S_i = \frac{b_i * c_i}{n_i}$$

Mit der Berücksichtigung der Rücktransformation durch Exponentialfunktion, können die Konfidenzintervalle mit Hilfe des folgenden Ausdrucks bestimmt werden:

$$\exp(\ln(OR_{MH}) \pm 1{,}96 * \widehat{se}(\ln(OR_{MH})))$$

[5] James Robins, Norman Breslow and Sander Greenland, *Biometrics*,Vol. 42, No. 2 (Jun., 1986), pp. 311-323, Published by: International Biometric Society, DOI: 10.2307/2531052, Stable URL: http://www.jstor.org/stable/2531052

4.3 Assoziationsanalyse

Die Assoziationsanalyse wird oft in der Marketingbranche, hauptsächlich im Bereich des Kaufkraftsektors angewendet. Dennoch bei einer richtig konzipierten Fragestellung, kann sie in weit entfernten Bereichen wie zum Beispiel in der Medizin eingesetzt werden. Das Ziel der Assoziationsanalyse ist das Auffinden von Assoziationsregeln, also Regeln die auf ein bestimmtes Assoziationsmuster in den Daten hindeuten. Für das Auffinden von Assoziationsregeln werden sogenannte Transaktionen betrachtet, jeder Transaktion enthält verschiedene Elemente - *Items*. Zuerst werden unter allen Transaktionen, die meist frequentierte *Items* gesucht, dabei kann ein bestimmter Schwellenwert vorgegeben werden. Dann startet der A-Priori-Algorithmus mit der Überprüfung der Erfüllung dieser Bedingung für das erste *Item*. *Items* die den Schwellenwert überschreiten, sind sogenannte Haupt-*Items* und bilden die Grundlage für die Assoziationsregel. Nachdem die Haupt-*Items* gefunden sind, werden die Kombinationen dieser mit weiteren Items betrachtet, dabei wird der *Support* und *Confidence*[6] der Regel berechnet.

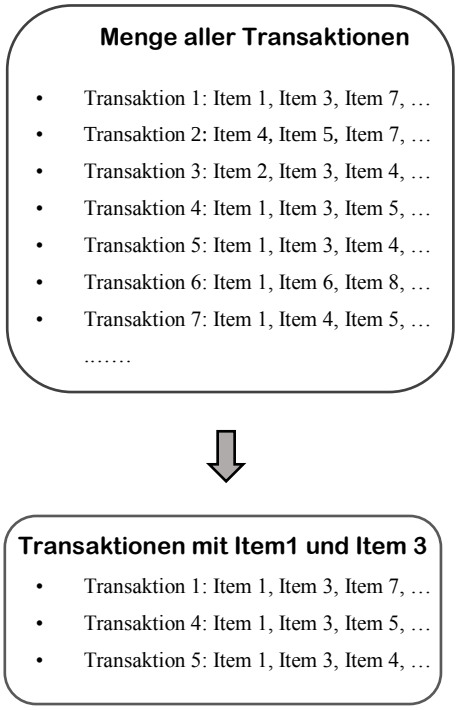

Abbildung 7 Schema für die Berechnung des Supports

[6] die Fachbegriffe Item, Support und Confidence werden nicht übersetzt, die Begriffe werden auch in der deutschen Fachliteratur im Original angeführt.

Der *Support* ist der prozentuale Anteil der Transaktionen, in denen zum Beispiel zwei bestimmte Items zusammen vorkommen, an der Anzahl aller untersuchten Transaktionen. (# bedeutet Anzahl)

$$Support\ \{Item1, Item3\} = \frac{\#\ Transaktionen\ mit\ \{Item1\ \&\ Item3\}}{\#\ aller\ Transaktionen}$$

Confidence ist eine Maßzahl welche die Assoziationsregel bewertet. Betrachtet wird die Situation mit zwei *Items*, die *Confidence* gibt den prozentualen Anteil der Transaktionen an, wo beide *Items* gleichzeitig vorkommen, an der Anzahl aller Transaktionen wo nur das erste *Item* vorkommt. Formell kann die *Confidence* aufgeschrieben werden als (Jürgen Cleve, 2016, S. 231-232):

$$Confidence\ \{Item1, Item3\} = \frac{\#\ Transaktion\ mit\ \{Item1\ \&\ Item3\}}{\#\ Transaktionen\ mit\ \{Item1\}}$$

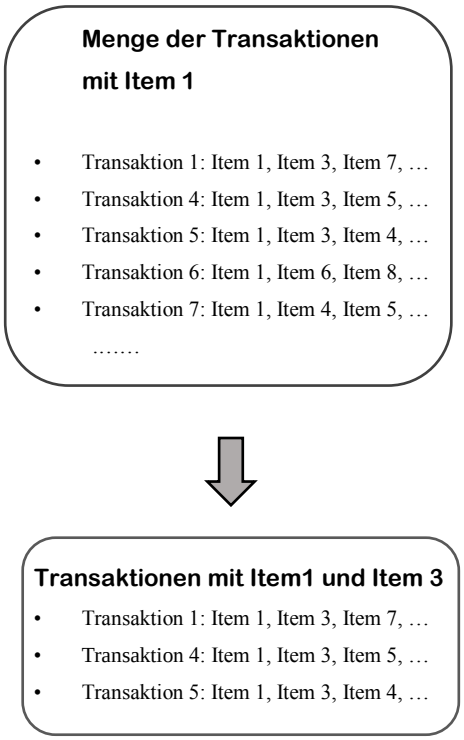

Abbildung 8 Schema für die Berechnung der *Confidence*

Confidence kann vorab vordefiniert werden und die ganze Prozedur kann ebenfalls iterativ für mehr als nur zwei *Items* durchgeführt werden. Eine Regel kann aus mehreren *Items* bestehen und wird dann verschachtelt interpretiert. Obwohl die *Confidence* eine Maßzahl ist, die die Regel bewertet, muss der Aspekt beachtet werden, dass zum Beispiel zwei *Items* die in der Regel enthalten sind und beide einen hohen Support aufweisen, zwangläufig eine hohe *Confidence* für die Regel erzeugen. Somit wird ein weiteres Maß benötigt, das die Bedeutung der gefundenen Assoziationen besser quantifiziert.

Solch ein Maß ist das *Lift* und ist definiert als:

$$Lift\{Item1 \rightarrow Item3\} = \frac{Confidence\ \{Item1 \rightarrow Item3\}}{Support\ \{Item3\}}$$

Lift beschreibt das Verhältnis zweier Häufigkeiten von *Item3*: wie oft kommt *Item3* in den Transaktionen wo bereits der *Item1* enthalten ist, im Vergleich wie oft kommt er vor in allen Transaktionen überhaupt (*Support*). Ein Lift größer 1 deutet auf einen positiven Einfluss des *Items1* auf die Häufigkeit des Auftretens von *Item3.* (Karsten Lübke, 2014, S. 161) Die verschiedenen Parameter wie minimaler *Support,* die *Confidence* sowie Anzahl der *Items* für Bildung einer Regel, können in Bezug auf die Forschungsfrage vorbestimmt werden, um aussagekräftigere Regeln zu bilden.

4.4 Chi^2 – Test

Die Testtheorie in der Statistik beinhaltet eine Reihe von Verfahren zur direkten Testung eines signifikanten Zusammenhangs zwischen zwei interessierenden Merkmalen. Abhängig von der Skalierung der erhobenen Daten, sowie Voraussetzungen die erfüllt sein müssen, kommen entsprechende Testverfahren in Frage. Weisen die erhobenen Merkmale ein kategoriales Datenniveau auf, wird die Darstellung der Häufigkeitsaufzählungen mittels Kontingenztafeln (eine 2 x 2 Kontingenztafel wird auch als Vierfeldertafel bezeichnet) dargestellt. Für die Analyse von Häufigkeitsaufzählungen wird der Chi^2 (χ^2) – Unabhängigkeitstest angewendet. Mit Hilfe des Tests, kann die Hypothese getestet werden, ob zwei Merkmale unabhängig voneinander sind. Besteht ein Zusam-

menhang zwischen den beiden Variablen, so wird das in der auffälligen Häufigkeitsverteilung sichtbar. Die folgende Tabelle zeigt eine typische Vierfeldertafel, die eine Berechnungsgrundlage für die Teststatistik (χ^2) bildet.

Tabelle 4 Vierfeldertafel

	Variable B		
Variable A	Ausprägung 1	Ausprägung 2	Zeilensummen
Ausprägung 1	n_{11}	n_{12}	z_1
Ausprägung 2	n_{21}	n_{22}	z_2
Spaltensummen	s_1	s_2	N= $n_{11} + n_{12} + n_{21} + n_{22}$

Seien $n_{11}, n_{12}, n_{21}, n_{22}$ die Häufigkeiten für die entsprechenden Ausprägungskombinationen der beiden interessierenden Merkmale, s_1 und s_2 die Summen für die erste und zweite Spalte, sowie z_1 und z_2 die Zeilensummen.

Man definiere zunächst die Begriffe wir beobachtete und erwartete Häufigkeiten. Die beobachteten Häufigkeiten sind die tatsächlich in den Daten vorhandene Anzahl der Personen/Objekte mit entsprechenden Merkmalkombinationen $n_{11}, n_{12}, n_{21}, n_{22}$. Die erwarteten Häufigkeiten sind definiert als:

$$e_{11} = \frac{s_1 * z_1}{N}\,,\; e_{12} = \frac{s_2 * z_1}{N},\; e_{21} = \frac{s_1 * z_2}{N}\,,\; e_{22} = \frac{s_2 * z_2}{N}.$$

Es sind Häufigkeiten die man erwarten würde, wenn die zu untersuchenden Merkmale unabhängig voneinander wären. Die Teststatistik bei Chi^2-Test ist definiert als (Devore, 2015, S. 621):

$$\chi^2 = \sum_{j=1}^{K} \frac{(n_j - e_j)^2}{e_j}$$

Wobei K – Anzahl der Zellen in der Kontingenztafel.

Es werden die Abweichungen der erwarteten von den beobachteten Häufigkeiten quadriert und in Relation zu den erwarteten Häufigkeiten gesetzt. Unter der Nullhypothese ist die Teststatistik, Chi^2 verteilt mit (K-1) Freiheitsgraden. Die folgende Abbildung zeigt die Dichte einer Chi^2 -Verteilung mit 3 Freiheitgraden (basierend auf einer Simulation).

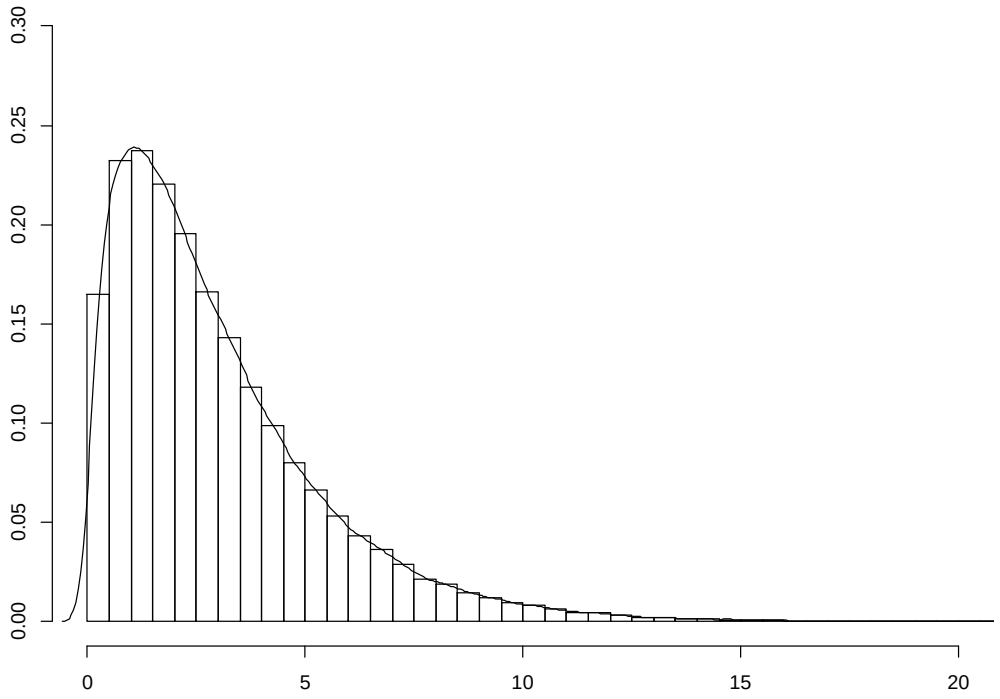

Abbildung 9 Dichte der Chi^2 Verteilung.

Anschließend wird der Wert der Teststatistik mit dem Quantil der Chi^2 -Verteilung für das entsprechende Signifikanzniveau verglichen. Ist der Wert der Teststatistik größer als das Quantil, ist das Testergebnis signifikant.

4.5 Mann-Whitney-U-Test

Bei Anwendung statistischer Tests müssen mehrere Aspekte betrachtet werden. Erstens, ist es zu beachten, dass die primäre Wahl eines Tests abhängig von der Skalierung der Daten erfolgt. Nachdem das entsprechende Verfahren gewählt ist, müssen wiederum die

Daten, die Voraussetzungen für das Verfahren erfüllen. Um ein metrisch skaliertes Merkmal zwischen zwei Gruppen auf signifikante Unterschiede untersuchen zu können, wurde hier der Mann-Whitney-U-Test angewandt. Aufgrund der Tatsache, dass die Daten keiner Normalverteilung folgen, ist für die Analyse dieses nichtparametrische Verfahren gewählt worden. Beim diesen Test handelt es sich ein parameterfreies Verfahren, die Beobachtungswerte werden mit Hilfe der Rangsummen quantifiziert.

Die Teststatistik des Mann-Whitney-U-Tests ist definiert, als das Minimum der beiden Statistiken aus den Vergleichsgruppen, dabei seien n_1, n_2 Anzahl der Beobachtungen in der jeweiligen Gruppe, und R_1 seien die Ränge, die den geordneten Beobachtungen in der entsprechenden Gruppe zugewiesen wurden.

$$U = \min\{U_{1,} U_{1,}\}$$

Wobei $U_1 = n_1 * n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1$ die Teststatistik für Gruppe 1.

Analog ist die Teststatistik für Gruppe 2 definiert: $U_2 = n_1 * n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$
(Gregory W. Corder, 2011, S. 17)

Für die Beurteilung der Signifikanz wird die Teststatik mit dem kritischen Wert für das vorgegebene Signifikanzniveau verglichen. Die H_0wird abgelehnt, wenn der Wert der Teststatistik den kritischen Wert nicht überschreitet.

4.6 Hotdeck - Imputation

Um Analysen auch für vollständige Daten durchführen zu können, wurden die fehlenden Werte mittels der sogenannten Hotdeck-Imputation vervollständigt. Aufgrund der Annahme, die aus Betrachtung der fehlenden Werte resultiert, ist dieses Verfahren für die Imputation gewählt worden, da das Fehlen der Daten anscheinend einem MCAR-Mechanismus folgt.

MCAR (eng. Missing completely at random) bedeutet, dass die Daten in der Variable X rein zufällig fehlend sind. Genau gesagt ist die Wahrscheinlichkeit für das Fehlen der Werte in der Variable X unabhängig von der Vollständigkeit der Daten in anderen Variablen und in der Variable X selbst ist. Es kann also kein kausaler Zusammenhang erkannt werden, der das Fehlen der Daten erklären würde. (Marwala, 2009, S. 5)

Die Hotdeck – Imputation ist ein Verfahren, dass für die Gewinnung der fehlenden Werte die restliche Stichprobe verwendet. Diese Methode kann für kontinuierliche als auch für kategoriale Variablen verwendet werden. Um den fehlenden Wert bei einem Merkmal für ein Individuum zu ersetzen, werden in der vorhandenen, restlichen Stichprobe, Individuen gesucht, die der Beobachtungseinheit mit dem fehlenden Wert in allen anderen vorhandenen Merkmalen am ähnlichsten sind und der fehlende Wert wird ausgefüllt. Es gibt eine Reihe an Optionen, die je nach Bedarf variabel eingestellt werden können, um die Suche je nach Wunsch zu optimieren. Die fehlenden Werte können ebenfalls aus Werten innerhalb von Teilmengen der Daten (sogenannten *subsets*) imputiert werden, die Unterteilung kann Aufgrund der Struktur der Daten eine sinnvolle Maßnahme sein, um die Varianz der Schätzung gering zu halten (Longford, 2006, S. 43-44).

4.7 Ereigniszeitanalyse

Im Folgenden soll die Überlebensanalyse näher beschrieben werden, wobei auf die Hazards Annahme im Hinblick auf die Cox-Regression und auf den Kaplan-Meier-Schätzer eingegangen wird.

Generell beinhaltet die Überlebensanalyse die Modellierung von Zeit-zu-Ereignis-Daten; in diesem Zusammenhang wird Tod oder Scheitern in der Literatur der Überlebensanalyse als Ereignis bezeichnet - üblicherweise tritt für jedes Subjekt nur ein einziges Ereignis auf. Das eintretende Event kann sein: Tod, Auftreten von Krankheiten, Wiederauftreten von Krankheiten, Genesung oder andere interessante Ereignis. Der Zeitparameter misst die Zeit vom Beginn des Beobachtungszeitraums zum Beispiel: Beginn der Behandlung oder Beginn der Studie bis: zum Ereignis / zum Ende der Studie / oder bis zur Zensierung der Beobachtung.

4.7.1 Konzept

Die Ereignisanalyse beschäftigt sich mit der Berechnung des Risikos, für das Auftreten eines interessierenden Ereignisses zu einem bestimmten Zeitpunkt. Das Ereignis, das von Interesse ist, muss nicht zwangsläufig der Tod des Patienten sein. Zunächst müssen die Grundlegenden Begriffe erläutert werden.

Daten die aus offiziellen Registern stammen und für eine definierte Zeitspanne vorhanden sind, eignen sich ideal für Ereignis- bzw. Überlebenszeitanalyse. Zunächst wird der Start- und Endpunkt der Beobachtungsphase festgelegt. Innerhalb dieses Zeitintervalls wird das Auftreten des interessierenden Ereignisses, das nominal oder kategorial skaliert ist, analysiert. Dies geschieht in den man eine Zeitvariable einführt, die die Zeit zwischen dem Start der Beobachtungsphase bis zum Eintreten des Ereignisses quantifiziert. Ist das Ereignis, innerhalb der vordefinierten Zeitspanne nicht beobachtet worden, oder ist die Beobachtungseinheit vor dem Ende der Beobachtungszeit rausgefallen, wird diese Beobachtung zensiert. Zensierungsmechanismus kann auf zwei Arten erfolgen:

Sei z^* eine latente Zufallsvariable und z ist beobachtet. Dann ist z bei T linkszensiert, wenn:

$$\mathrm{z} = \begin{cases} z^* & z^* > T \\ T & z^* \leq T \end{cases}$$

und analog erfolgt die Rechtszensierung:

$$\mathrm{z} = \begin{cases} z^* & z^* < T \\ T & z^* \geq T \end{cases}$$

Durch die Zensierung kann die unvollständige Information solch einer Beobachtungseinheit trotzdem in die Analyse eingeschlossen werden. Mithilfe der Cox-Regression kann die Analyse bezüglich eines Ereignisses nicht nur im Kontext der Zeit stattfinden, es kann ebenfalls der Einfluss der möglichen Faktoren berücksichtigt werden.

4.7.2 Begriffserklärung

Zunächst müssen ein paar Grundbegriffe erläutert werden. Sei T eine stetige, nicht negative Zufallsvariable und F die dazugehörige Verteilungsfunktion die folgendermaßen definiert ist: F(t) = P(T≤ t), sie beschreibt die Wahrscheinlichkeit, dass T den Wert kleiner/kleiner gleich t annimmt. Weiter sei die Funktion S(t), die Überlebensfunktion, die die Wahrscheinlichkeit beschreibt, dass T > t ist, S(t) = P(T>t).

Dann definiert man die sogenannte Hazard-Funktion als (Mills, 2011, S. 9):

$$\lambda(t) := \lim_{h \to 0} \frac{\Pr(t \leq T < t + h | T \geq t)}{h}$$

Beschrieben wird also die Todesrate zum Zeitpunkt t, bedingt darauf bis dorthin überlebt zu haben. Anders ausgedrückt kann man die Hazard-Funktion folgendermaßen anschreiben, wobei f(t) die Ableitung der Verteilungsfunktion ist, also die Dichte (Mills, 2011, S. 115):

$$\lambda(t) = \frac{f(t)}{S(t)}$$

4.7.3 Cox Regression mit Proportional Hazard Annahme

In diesem Kapitel wird das Cox-Regressionsmodell zur gleichzeitigen Bewertung der Beziehung zwischen mehreren Risikofaktoren und der Überlebenszeit des Patienten bzw. der Patientin beschrieben. Das Cox-Proportional-Hazardmodell ist eine der wichtigsten Methoden zur Modellierung von Überlebensdaten. Das Modell ist im Wesentlichen ein in der medizinischen Forschung gebräuchliches Regressionsmodell, zur Untersuchung des Zusammenhangs zwischen der Überlebenszeit von Patienten sowie Patientinnen und einer oder mehreren Prädiktorvariablen. Ziel des Modells ist es, gleichzeitig den Einfluss mehrerer Faktoren auf das Überleben zu bewerten. Mit anderen Worten, es erlaubt zu untersuchen, inwiefern die Einflussfaktoren das Risiko für das Eintreten eines bestimmten Ereignisses (z.B. Infektion, Tod) zu einem bestimmten Zeitpunkt beeinflussen.

Das Modell in der Cox-Regression für das i – te Individuum zum Zeitpunkt t lautet, gemäß der Definition (Mills, 2011, S. 87):

$$\lambda_i(t) := \lambda_0(\mathrm{t}) \exp\left(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik}\right)$$

Wobei die $\beta = (\beta_1, \ldots, \beta_k)$ der unbekannte Parametervektor ist, die X_i die erklärenden Variablen darstellen und λ_0 die Baselinehazard Funktion, also der Wert der Hazardfunktion wenn die erklärenden Variablen den Wert 0 annehmen. Durch Umformung und logarithmieren der Gleichung, erhalt man den Ausdruck für das Hazardratio (Mills, 2011, S. 87):

$$log\ \frac{\lambda_i(t)}{\lambda_0(t)} = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik}$$

Zusätzlich wird angenommen, dass der Einfluss der erklärenden Variablen auf die Hazardfunktion konstant und somit unabhängig von dem Zeitpunkt t ist. Auswirkungen dieser Annahme werden sichtbar, indem man die Hazardfunktionen von zwei Individuen i und j, vergleicht: (Guo, 2010, S. 75)

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\cancel{\lambda_0(t)} \exp\left(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik}\right)}{\cancel{\lambda_0(t)} \exp\left(\beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3} + \cdots + \beta_k x_{jk}\right)}$$

4.7.3.1 Schätzung der Koeffizienten

In dem Cox-Regressionsmodel mit Proportional Hazard Annahme werden die Koeffizienten für die unbekannten Parameter, durch das Maximieren der partiellen Likelihood-funktion geschätzt.

Dazu betrachten wir zunächst das Risikoset also die Summe der Hazardfunktionen für die Individuen 1 bis n: $\lambda_1(t) + \lambda_2(t) + \lambda_3(t) + \cdots + \lambda_n(t)$. Die Wahrscheinlichkeit, dass das untersuchte Ereignis bei dem ersten Individuum, zum Zeitpunkt t eintreten

wird, ist der Quotient der Hazardfunktion über das Risikoset, also über die Summe der Hazardfunktionen aller Individuen, die zum Zeitpunkt t dem Risiko eines Ereignisses ausgesetzt sind.

$$L_1 = \frac{\lambda_1(t)}{\lambda_1(t)+\lambda_2(t)+\lambda_3(t)+\cdots+\lambda_n(t)} =$$

$$\frac{\cancel{\lambda_0(t)}\exp(\beta x_1)}{\cancel{\lambda_0(t)}\exp(\beta x_1)+\cancel{\lambda_0(t)}\exp(\beta x_2)+\cdots+\cancel{\lambda_0(t)}\exp(\beta x_n)} =$$

$$\frac{\exp(\beta x_1)}{\exp(\beta x_1)+\exp(\beta x_2)+\cdots+\exp(\beta x_n)} \quad \text{(Guo, 2010, S. 76)}$$

Man beachte auch hier kürzt sich die Baselinehazardfunktion weg, dank der Annahme über die Proportionalität der Hazardfunktionen. Zweiter, wichtiger Aspekt ist die Tatsache, dass die zensierten Beobachtungen in die Berechnungen der Likelihood-funktion ebenfalls einfließen und dadurch der Informationsgehalt nicht gemindert wird. Zusammenfassend bildet die Multiplikation der einzelnen Likelihood-funktionen für alle Beobachtungen, die partielle Stichprobenfunktion (Guo, 2010, S. 77):

$$L_{partial} = \prod_{i=1}^{n} L_i = L_1 * L_2 * L_3 * \ldots * L_n$$

4.7.3.2 Interpretation der Koeffizienten

Nachdem die β Koeffizienten durch das maximieren der partiellen Likelihood-Funktionen geschätzt worden sind, müssen sie richtig interpretiert werden.

Zunächst um die Interpretation zu erleichtern wird das sogenannte Hazardratio (HR) für jeden Prädiktor bildet. Dies geschieht in dem man die Exponentialfunktion bezüglich der entsprechenden Koeffizienten anwendet. $HR = e^{\beta_i}$ für die erklärende Variable X_i. Das HR gibt den multiplikativen Faktor an, mit dem sich das relative Risiko (für das Eintreten des Ereignisses) verändert und zwar bei einer Veränderung des Werts X_i um

eine Einheit, wobei die Werte aller anderen Prädiktoren festgehalten werden. Ist das Hazardratio für X_i größer 1, so steigt das relative Risiko, wenn der Wert von X_i ansteigt. Ist das Hazardratio für X_i kleiner 1 so wird das relative Risiko gemindert mit jeder Einheit, um die der Wert von X_i steigt. (Sullivan, 2017, S. 269)

In dem Cox-Proportional-Hazardmodell ist also der Effekt einer Einheitserhöhung des Werts der Kovariable multiplikativ in Bezug auf das relative Risiko, andere Arten von Überlebensmodellen wie z.B. beschleunigte Ausfallzeitmodelle weisen keine proportionalen Hazards auf, sondern beschreiben eine Situation wo die Hazardrate beschleunigt wird.

4.7.4 Kaplan-Meier Schätzer

Für die Schätzung der Überlebensfunktion in der Überlebensanalyse spielt der Kaplan-Meier Schätzer auch Produkt-Limit-Schätzer genannt, eine große Rolle. Es ist ein nicht-parametrischer Schätzer, dessen großer Vorteil ist es, dass auch zensierte Beobachtungen bei der Schätzung berücksichtigt werden. Bezeichne man mit $t_1, t_2, \dots, t_m$ Zeitpunkte, zu deren das interessierende Ereignis eingetreten ist, diese teilen die Zeitachse in Abschnitte der Art: $[0, t_1), [t_1, t_2), \dots, [t_m, \infty)$.

Bezeichne man mit n_i der Anzahl der Personen unter Risiko zum Zeitpunkt t_i – diese Personen haben den vorhergehenden Zeitpunkt t_{i-1} überlebt und keine Zensierung hatten, weiter bezeichne man mit d_i die Personen die zum Zeitpunkt t_i verstorben sind. Definiere man in weiteren Schritt:

$\widehat{q_i} = \frac{d_i}{n_i}$ also den Anteil der Verstorbenen an den unter Risiko Personen und

$\widehat{p_i} = 1 - \widehat{q_i} = 1 - \frac{d_i}{n_i} = \left(\frac{n_i - d_i}{n_i}\right)$

Der Kaplan-Meier Schätzer für die Überlebensfunktion lautet:

$$\widehat{S(t)} = \prod_{t_i \le t} \widehat{p_i} = \prod_{t_i \le t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i=1}^{k} \left(\frac{n_i - d_i}{n_i} \right)$$ (Mara Tableman, 2003, S. 28)

Die Varianz für den Kaplan-Meier Schätzer kann mit Hilfe der Greenwood Formel (1926) errechnet werden, und ist definiert als (Mara Tableman, 2003, S. 32):

$$\widehat{Var(\widehat{S(t)})} = \widehat{S^2}(t) = \sum_{t_i < t} \frac{d_i}{n_i(n_i - d_i)} = \widehat{S^2}(t) \sum_{i=1}^{k} \frac{d_i}{n_i(n_i - d_i)}$$

4.7.4.1 Darstellung des Kaplan-Meier Schätzers

Ein Plot des Kaplan-Meier-Schätzers ist eine Reihe von abnehmenden horizontalen Schritten, die sich mit einer ausreichend großen Stichprobe der wahren Überlebensfunktion für diese Population nähern. Der Wert der Überlebensfunktion zwischen aufeinanderfolgenden unterschiedlichen Beobachtungen wird als konstant angenommen. Ein wichtiger Vorteil der Kaplan-Meier-Kurve ist, dass die Methode einige Arten von zensierten Daten berücksichtigen kann, insbesondere die Rechtszensur, die auftritt, wenn ein Patient sich aus einer Studie zurückzieht, in der Nachsorge verloren geht oder am Leben ist, ohne dass ein Ereignis auftritt (Ton J. Cleophas, 2006, S. 94).

Die folgende Grafik zeigt ein Beispiel für den Kaplan-Meier Schätzer für die Überlebenswahrscheinlichkeit, basierend auf den R Datensatz *colon* (Überlebenswahrscheinlichkeit getrennt nach Geschlecht). Die vertikalen Striche zeigen die Zensierungen der Beobachtungen und die Überlebenswahrscheinlichkeit sinkt über die Zeit kontinuierlich für beide Geschlechter.

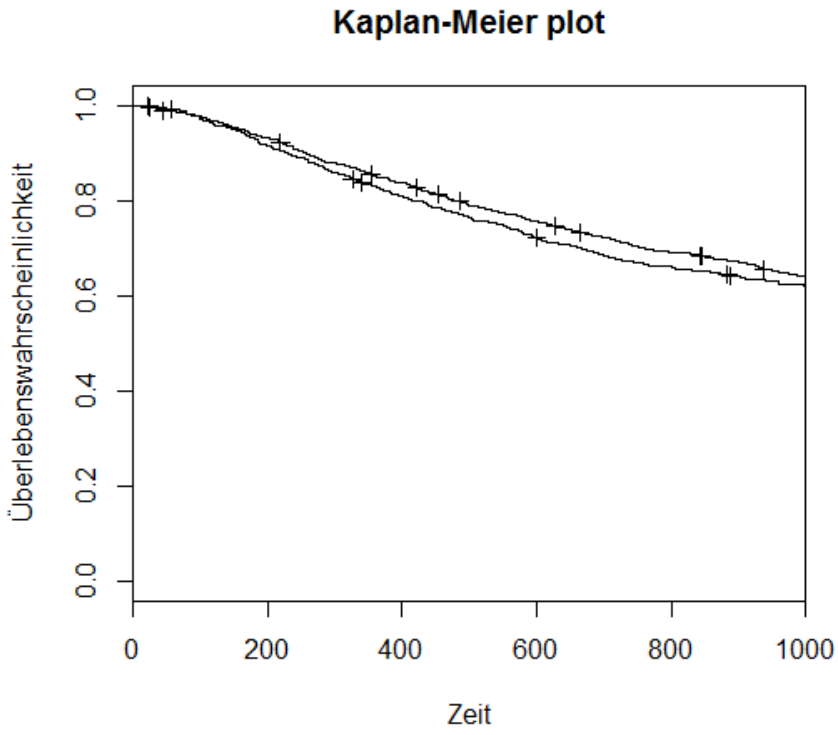

Abbildung 10 Kaplan - Meier Plot (*colon* Datensatz aus R)

5 Ergebnisse

In diesem Teil der Arbeit werden unter anderen die häufigsten Diagnosen von jeder Kategorie aufgelistet. Auch die prozentuellen Anteile, sowie die Entwicklung über die Zeit sind von Interesse. Die Unterteilung zwischen Aufenthalten und Diagnosen resultiert daraus, dass bei einem Aufenthalt mehrere UAE-Diagnosen festgestellt werden können (Haupt- und Nebendiagnose zum Beispiel).

5.1 Diagnosen und Krankenhausaufenthalte mit Bezug auf unerwünschte Arzneimittelereignisse.

Im Jahr 2001 wurden insgesamt 5.916.703 Diagnosen gestellt. Davon waren 267.689 UAE-Diagnosen. Das macht einen Prozentanteil von 4,5% aus. Bis zum Jahr 2009 war die Tendenz steigend sowohl bei allen Diagnosen als auch bei den UAE-Diagnosen, wobei die Anzahl der UAE-Diagnosen stärker anstieg. In dem Jahr 2010 sinkt die Anzahl aller gestellten Diagnosen von 6.978.794 (2009) auf 6.953.363, 2011 sinkt sie auf 6.900.644. Bei den Diagnosen, die auf unerwünschte Arzneimittelereignisse zurück zu führen sind, ist die Tendenz nach wie vor steigend. (Christoph Rinner, 2015, S. 43)

Diagnosen der Kategorie A1 machten einen durchschnittlichen prozentualen Anteil von 1,54% an allen UAE-Diagnosen aus (insgesamt 7 Kategorien). Der Anteil beträgt 2001, 1,8 % (4.777 Diagnosen) und sinkt bis zum Jahr 2011 auf 1,4% (5.093 Diagnosen) (Christoph Rinner, 2015, S. 44).

Die häufigsten ICD-10 Codes in dieser Kategorie im Jahr 2001 sind:

Tabelle 5 Die häufigsten A1-Diagnosen

M81.4 – Arzneimittelinduzierte Osteoporose: Mehrere Lokalisationen
G62.0 – Arzneimittelinduzierte Polyneuropathien
M80.4 – Arzneimittelinduzierte Osteoporose mit pathologischer Fraktur: mehrere Lokalisationen
G21.1 – Sonstiges arzneimittelinduziertes Parkinson-Syndrom
E16.0 – Arzneimittelinduzierte Hypoglykämie ohne Koma

Bei dem ICD-10 Code M80.4 wurde jedoch eine beachtliche Abnahme von 85% (von 501 - Jahr 2001 auf 72 Diagnosen im Jahr 2011) festgestellt. Ein Anstieg der A1-Diagnosen wurde unter anderen bei D61.1-Arzneimittelinduzierte aplastische Anämie, M10.2-Idiopathische Gicht: Oberarm [Humerus, Ellenbogengelenk] und I95.2-Hypotonie durch Arzneimittel beobachtet.

Die Abschwächung der Kategorie A1 ist der Definition nach, die Kategorie A2. Anteil dieser Kategorien betrug durchschnittlich 11,7% an allen UAE-Diagnosen. Die häufigsten A2-Diagnosen im Jahr 2001 sind:

Tabelle 6 Die häufigsten A2-Diagnosen

T88.7 – Nicht näher bezeichnete unerwünschte Nebenwirkung eines Arzneimittels oder einer Droge
T78.4 – Allergie, nicht näher bezeichnet
F19.2 – Psychische und Verhaltensstörungen durch multiplen Substanzgebrauch und Konsum anderer psychotroper Substanzen: Abhängigkeitssyndrom
L27.0 – Generalisierte Hauteruption durch Drogen oder Arzneimittel
F13.2 – Psychische und Verhaltensstörungen durch Sedativa oder Hypnotika: Abhängigkeitssyndrom

Die Kategorie B1 (Definition: Vergiftung durch Arzneimittel: siehe Tabelle 1 Kategorisierung der UAE-Diagnosen), ist hierarchisch gleichwertig mit der Kategorie A1. Im Jahr 2001 betrug der prozentuelle Anteil der B1-Diagnosen an allen UAE-Diagnosen 1,8%, im Jahr 2011, waren es 0,8%. Der durchschnittliche prozentuale Anteil zwischen den Jahren 2001-2011 liegt bei 1,2%. B1-Diagnosen die im Jahr 2001 am häufigsten vorkommen sind in der folgenden Tabelle dargestellt.

Tabelle 7 Die häufigsten B1-Diagnosen

T46.0 – Vergiftung durch Herzglykoside und Arzneimittel mit ähnlicher Wirkung
T42.4 – Vergiftung durch Benzodiazepine
T42.7 – Vergiftung durch Antiepileptika, Sedativa und Hypnotika, nicht näher bezeichnet
T43.0 – Vergiftung durch Tri- und tetrazyklische Antidepressiva
T43.2 – Vergiftung durch Sonstige und nicht näher bezeichnete Antidepressiva

Bei den ICD-10 Codes T46.0, T42.4, T42.7 und T43.0 konnte aber eine deutliche Abnahme bis zum Jahr 2011 festgestellt werden. Analog, ist die Kategorie B2 die Abschwächung der Kategorie B1. Diese Kategorie macht den kleinsten Anteil aller UAE-Diagnosen aus. Der durchschnittliche prozentuale Anteil über die Jahre 2001-2011 betrug 0,88%. Die häufigsten ICD-10 Codes dieser Kategorie, die man im Jahr 2001 diagnostizierte, sind in der folgenden Tabelle ersichtlich.

Tabelle 8 Die häufigsten B2-Diagnosen

T50.9 − Vergiftung durch Diuretika und sonstige und nicht näher bezeichnete Arzneimittel, Drogen und biologisch aktive Substanzen
F55 - Schädlicher Gebrauch von nichtabhängigkeitserzeugenden Substanzen
T96 - Folgen einer Vergiftung durch Arzneimittel, Drogen und biologisch aktive Substanzen
T50.8 - Vergiftung durch Diagnostika
T50.3 - Vergiftung durch auf den Elektrolyt-, Kalorien- und Wasserhaushalt wirkende Mittel

Der durchschnittliche Anteil der Kategorie C betrug über die Jahre 2001-2011, 2,3%.

Tabelle 9 Die häufigsten C-Diagnosen

K71.9 - Toxische Leberkrankheit, nicht näher bezeichnet
D69.5 - Sekundäre Thrombozytopenie
A04.7 - Infektiöse Darmkrankheiten; Enterokolitis durch Clostridium difficile
K71.0 - Toxische Leberkrankheit mit Cholestase
D69.0 - Purpura anaphylactoides

Weitere zwei Kategorien D und E werden aus den weiteren Auswertungen ausgeschlossen. Die durchschnittlichen prozentualen Anteile dieser Kategorien an allen UAE-Diagnosen machten zwischen den Jahren 2001-2011, 31,4% beziehungsweise 51% aus. Alle sieben Kategorien werden auch kurz als K7 zusammengefasst.

5.2 UAE-Aufenthalte

Im Jahr 2001 wurden 2.170.460 Krankenhausaufenthalte dokumentiert. Die Krankenhausaufenthalte mit einer Hauptdiagnose, die auf unerwünschte Arzneimittelereignisse hinweist, machten im Jahr 2001 rund 10,6% (229.935 Aufenthalte) aller Krankenhausaufenthalte aus. Bis zum Jahr 2011 stieg dieser Anteil auf 12,4% (311.920 von 2.524.209). (Christoph Rinner, 2015, S. 39).

Bei der Analyse bezüglich des Geschlechts konnte beobachtet werden, dass Frauen viel öfter wegen unerwünschten Arzneimittelereignissen hospitalisiert wurden. In der Kategorie A1 steigt die Anzahl der Aufenthalte bei Männern über die Jahre leicht (von 41 auf 49 Aufenthalte pro 100.000 Einwohner), bei Frauen sinkt sie nur leicht (von 77 im Jahr 2001 auf 71 im Jahr 2011). Bei der Kategorie B1 wurde zwischen den Jahren 2001-2011 ein deutlicher Rückgang bei beiden Geschlechtern festgestellt, und zwar um 50 % bei den Frauen und um 37% bei den Männern. Bei den Kategorien B1 und B2 wurden bei Frauen, kontinuierlich über die 11 Jahre fast doppelt so viele UAE-Aufenthalte als

bei den Männern dokumentiert. Die standardisierte Anzahl der UAE-Aufenthalte pro 100.000 Einwohner insgesamt (über alle Kategorien) stieg zwischen den Jahren 2001-2011 aber bei beiden Geschlechtern an.

In Hinsicht auf die Altersgruppen, die meisten UAE-Aufenthalte wurden bei Patienten der Altersgruppe 5 (über 75 Jahre alt), die wenigsten in der Gruppe der (30-44) - Jährigen verzeichnet. Für die Altersgruppen 0, 1, 2, 3 und 4 ändert sich die standardisierte Anzahl der Aufenthalte, im Vergleich zu der steigenden Tendenz bei Geschlechtern nur gering. Bei der Gruppe 5 steigt sie bis zum Jahr 2009, in den folgenden 2 Jahren klingt sie dann leicht ab. Bei den Kategorien A1 und A2 wurden den Patienten der Altersgruppen 4 und 5 die meisten Aufenthalte zugeordnet, wobei bei den A1-Aufenthalten in beiden Gruppen ein Rückgang über die Jahre beobachtet wurde:

2001: Gruppe 4: 133, 2011: Gruppe 4: 121 Aufenthalte

2001: Gruppe 5: 193, 2011: Gruppe 5: 130 Aufenthalte

Bezüglich der B1-Kategorie verzeichnet die Altersgruppe 5 die meisten Aufenthalte, wobei eine starke Abnahme über die Jahre beobachtet worden ist (von 2001: 229 auf 2011: 84 Aufenthalte pro 100.000 Einwohner). Bei Patienten zwischen 15 und 29 Jahren wurden die zweithäufigsten B1-Aufentahalte dokumentiert. Die Analyse der B2-Kategorie bekräftigt das Ergebnis deutlich. Unzweifelhaft wurden die meisten Aufenthalte in dieser Kategorie den Patienten der Altersgruppe 1 (15-29 Jahre alt) zugeordnet. 2001 waren es 57, im Jahr 2011: 60 Aufenthalte pro 100.000 Einwohner. Bei der Kategorie C sind bei den Patienten, die über 74 Jahre alt sind, ab dem Jahr 2006 mit Abstand die meisten Krankenhausaufenthalte festgestellt worden.

5.3 Einfluss der demografischen Faktoren

In der Analyse bezüglich Risikofaktoren für Aufenthalte die auf unerwünschte Arzneimittelereignisse hindeuten, wurden Patientenkollektive gebildet und gegenübergestellt. Das Ziel war es, die binären Ereignisse wie zum Beispiel: Dokumentation eines UAE-Aufenthaltes (kodiert als: 0-nein und 1-ja), innerhalb eines Kollektivs relativiert zur Referenzgruppe zu quantifizieren. Mittels univariaten logistischen Regressionen

wurde der Einfluss demografischer Faktoren untersucht. Diese Faktoren waren das Geschlecht und die Altersgruppe. Für die Aggregation der Odds-Ratios und die Berechnung der dazugehörigen Konfidenzintervalle wurde der Mantel-Haenszel Schätzer als eine Methode der Metaanalyse angewendet. Nachdem die Odds-Ratios für die Risikofaktoren für die Jahre 2001-2011 einzeln berechnet worden sind, wurden diese, für eine allumfassende Quantifizierung des Einflusses zusammengefasst. Dabei ist die zeitliche Veränderung berücksichtigt worden. Bei der Analyse bezüglich des Geschlechts sind männliche Patienten als Referenzgruppe spezifiziert worden, beim Alter war es die Altersgruppe 5 (75+).

Bei der Kategorie A1 hatten Frauen (OR=1,24; KI (1,19-1,3)) und die Patienten aus der Altersgruppe 4 (OR=1,18, KI (1,12-1,24)) erhöhtes Risiko einer UAE-Diagnose. Ein signifikant kleineres Risiko hatten die Patienten aus der Altersgruppe 0 (OR=0,31; KI (0,27-0,35), Altersgruppe 1 (OR=0,33; KI (0,3-0,37) und aus der Gruppe 2 (OR=0,51; KI (0,47-0,55)). (Christoph Rinner, 2015, S. 58,60)

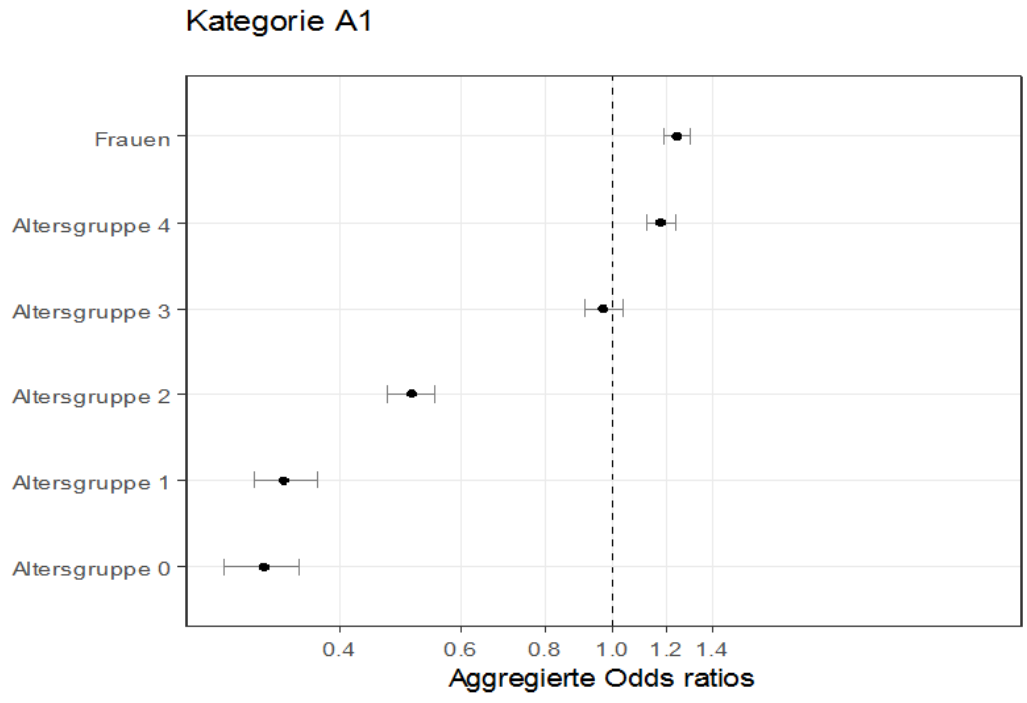

Abbildung 11 Einfluss der demografischen Faktoren bei der Kategorie A1

Die Analyse in der Kategorie A2 ergab, dass nur Personen aus der Altersgruppe 0 ein niedrigeres Risiko hatten. Für die Altersgruppen 1, 2, 3 und 4 war das Risiko einer UAE-Diagnose erhöht. Für die Patienten zwischen 15-29 Jahren (Gruppe 1) war das Risiko um den Faktor 2,1 und für die Patienten zwischen 30-44 Jahren um den Faktor 1,8 höher. (Christoph Rinner, 2015, S. 58,60)

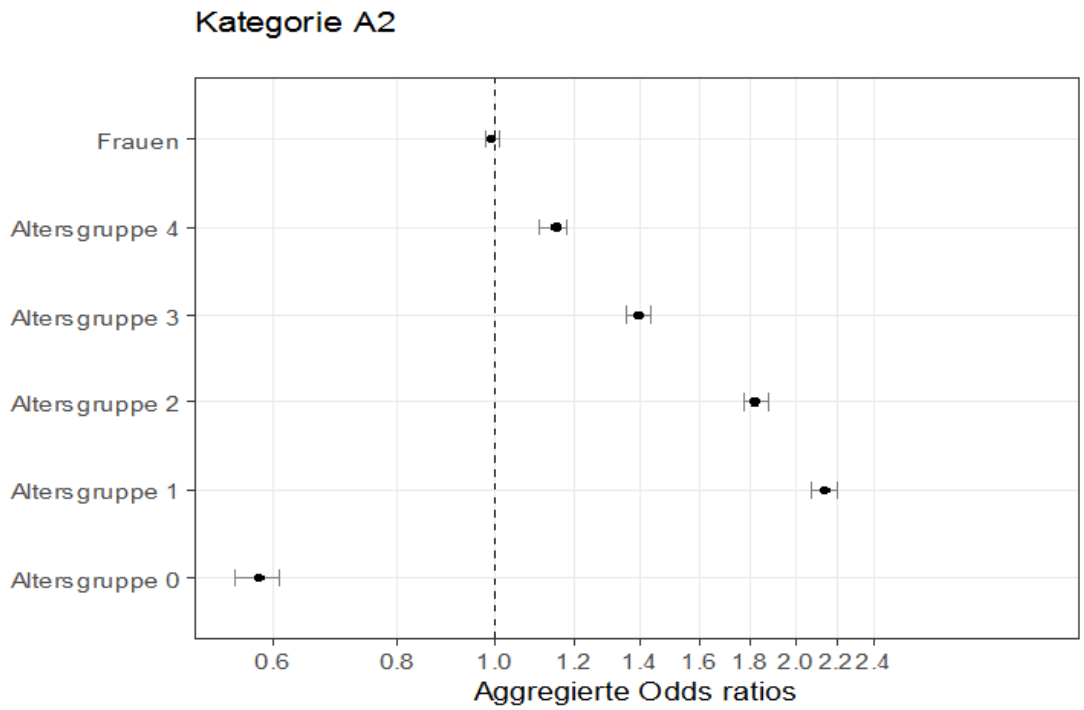

Abbildung 12 Einfluss der demografischen Faktoren bei der Kategorie A2

Bei der Kategorie B1 hatten vor allen die Patienten der Altersgruppen 0 (OR=0,3; KI (0,26-0,34)), 3 (OR=0,35; KI (0,32-0,38)) und 4 (OR=0,27; KI (0,24-0,3)) ein niedrigeres Risiko einer UAE-Diagnose als die Referenzgruppe. Das aggregierte Odd-Ratio für Frauen betrug OR=1,38 mit KI (1,31-1,45). (Christoph Rinner, 2015, S. 58,60)

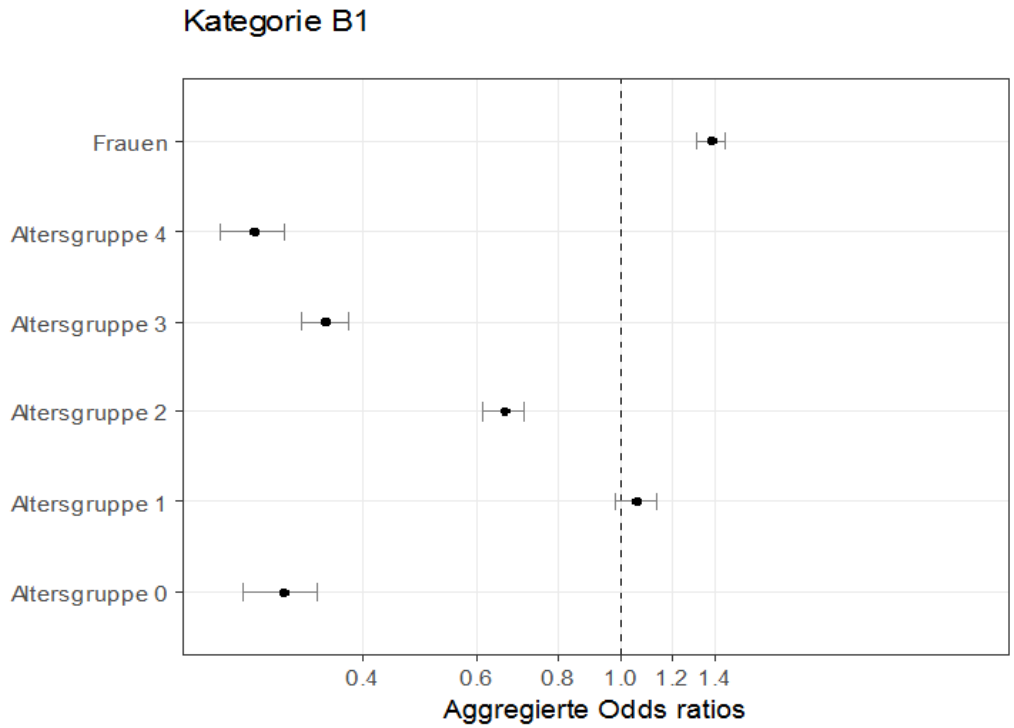

Abbildung 13 Einfluss der demografischen Faktoren bei der Kategorie B1

Vergiftung oder schädlicher Gebrauch durch Arzneimittel oder andere Ursachen ist die Definition der Kategorie B2. Hier hatten vor allen die Patienten der Altersgruppe 1 (OR=3,59; KI (3,22-4,02)) und der Gruppe 2 (OR=2,22; KI (1,99-2,48)) ein erhöhtes Risiko für diese Diagnose.

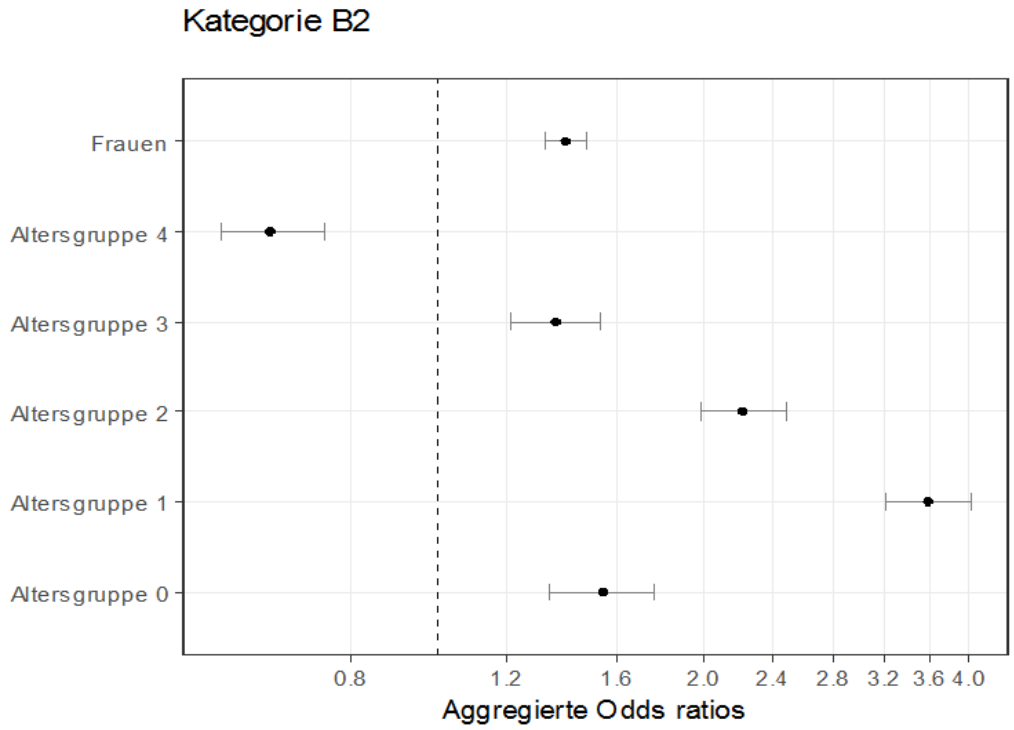

Abbildung 14 Einfluss der demografischen Faktoren bei der Kategorie B2

5.4 Diagnosen Kombinationen bei Aufenthalten mit unerwünschten Arzneimittelereignissen.

Krankenhausaufenthalte die eine UAE-Diagnose als Haupt- oder Nebendiagnose haben und bei denen zusätzlich eine weitere Diagnose gestellt wurde, wurden mithilfe der Assoziationsanalyse analysiert. Um die UAE-Diagnosen zu finden, die für Bildung der Assoziationsregeln verwendet werden, ist die Bedingung eines Supports von 5% an allen Aufenthalten vordefiniert worden. Somit konnten vier Diagnosen gefunden werden, für die Assoziation mit anderen Erkrankungen analysiert worden ist. An der Stelle sei es vermerkt, dass die vollständigen Spezifikationen der 4 Diagnosen die hier analysiert werden im Anhang angeführt sind. Die 4 gefundenen Diagnosen stammen aus der Kategorie A2 der UAE-Diagnosen, die Kombinationen dieser Diagnosen werden in folgenden Abschnitten beschrieben[7].

[7] Die Eigennamen der Diagnosen in dieser Arbeit sind der freizugänglichen Homepage mit dem Vermerk „ICD-10-GM-2017 Systematik online lesen“, mit dem Suchkatalog für ICD-10 Codes unter der Adresse http://www.icd-code.de entnommen worden

Die Katalogisierung der Diagnosen in dem freizugänglichen online-Katalog erfolgte auf der Basis von (Graubner, 2013).

5.4.1 Kombinationen mit T88.7

T88.7 definiert als: Nicht näher bezeichnete unerwünschte Nebenwirkung eines Arzneimittels oder einer Droge

Diese Diagnose kommt in allen 11 Jahren am häufigsten vor. Der durchschnittliche *Support* zwischen den Jahren 2001-2011 betrug 23,7%. Für diese UAE-Diagnose konnten die meisten Assoziationsregeln gefunden werden.

Zwischen den Jahren 2001 und 2011 werden durchgehend 7 Regeln gebildet, vier davon mit den ICD10-Codes I10, I25.9, I25.1 und I48, die Herzkrankheiten beziehungsweise Vorhofflimmern und Vorhofflattern (I48) charakterisieren. Die Durchschnittslifts dieser Regeln:

{I10 → T88.7} = 1,6;

{I25.9 → T88.7} = 1,58;

{I25.1 → T88.7} = 1,99;

{I48 → T88.7} = 1,2

Somit konnte festgestellt werden, dass die unerwünschten Nebenwirkungen eines Arzneimittels besonders oft bei Patienten mit einem Herzleiden vorkommen.

Zwischen 2008-2011 kommt zusätzliche Assoziationsregel {I73.9 → T88.7} mit einem mittleren *Lift* von 2, was bedeutet, dass Patienten mit Peripheren Gefäßkrankheiten ebenfalls übermäßig häufig Nebenwirkungen eines Medikamentes verspüren.

Eine weitere, wichtige Regel {E11.9 → T88.7} mit einem mittleren *Lift* von 1,4 (über die Jahre 2001-2011), die festgestellt werden konnte, betrifft die Erkrankung an Diabetes mellitus, Typ 2, ohne Komplikationen. Hier sind die Patienten ebenfalls öfter von Nebenwirkungen eines Arzneimittels betroffen. Diesbezüglich tritt im Jahr 2004 zusätzliche Assoziationsregel mit Adipositas {E66.9 → T88.7} auf, mit einem *Lift* von 1,75.

Weiter führt die medikamentöse Behandlung der Patienten mit Störungen des Lipoproteinstoffwechsels und sonstigen Lipidämien deutlich vermehrt zu unerwünschten Nebenwirkungen {E78 → T88.7} *Lift* =1,9. Zusätzlich kommen hier in den Jahren 2004, 2005 sowie 2008-2011 weitere Assoziationsregeln, mit den Untercodes E78.2 und E78.5. Diese haben ebenfalls einen hohen *Lift*. Die letzte Regel die über den gesamten Zeitraum von 11 Jahren vorkommt ist {N39 → T88.7} (N39 – Krankheiten des Harnsystems), mit einem mittleren *Lift* von 1,3.

Zwischen den Jahren 2002-2011 wird eine weitere wichtige Assoziationsregel {J44.9 → T88.7} gebildet mit durchschnittlichem *Lift* von 1,4. Diese Regel deutet darauf hin, dass J44.9 – chronische obstruktive Lungenkrankheit eine positive Korrelation mit der Häufigkeit des Auftretens der Nebenwirkungen eines Arzneimittels aufweist. In den Jahren 2001, 2003-2006 treten unerwünschte Nebenwirkungen vermehrt bei Patienten die Hyperurikämie ohne Zeichen von entzündlicher Arthritis oder tophischer Gicht aufweisen. Zwischen 2005-2011 wurden unerwünschte Nebenwirkungen eines Arzneimittels übermäßig häufig bei Patienten mit Postmenopausale Osteoporose {M81.9 → T88.7}, *Lift* = 2 festgestellt.

In den Jahren 2008-2011 werden weitere 2 Regeln mit hohen *Lift* gebildet: die erste dieser Regel {Z98.8 → T88.7} mit *Lift*=2.8 weist darauf hin, dass Patienten mit Zuständen nach chirurgischen Eingriffen ebenfalls unerwünschte Nebenwirkungen hatten. Weiter treten die Nebenwirkungen häufiger bei Patienten mit Divertikulose des Dickdarmes {K57.3 → T88.7}, *Lift* = 2,1. Zuletzt wird zwischen 2007-2011 wird eine Assoziationsregel die zwei UAE-Diagnosen enthaltet {N18.9 → T88.7} mit *Lift* = 1,6 gebildet. Somit kommen chronische Nierenkrankheit und nicht näher bezeichnete unerwünschte Nebenwirkung eines Arzneimittels oder einer Droge häufig zusammen vor.

5.4.2 Kombinationen mit T78.4

T78.4 definiert als: Allergie, nicht näher bezeichnet

Für jedes Jahr zwischen 2001-2011 konnte die Assoziationsregel {I10 → T78.4} gefunden werden, diese weist aber einen *Lift* von 1 auf, da der Support beider Diagnosen hoch ist. Diagnose die mit T78.4 kodiert ist, wurde im Jahr 2001 bei 11,4% aller

UAE-Aufenthalte gestellt, bis zum Jahr 2011 stieg der Anteil dieser Diagnose auf 16,3%. Es konnten aber keine weiteren Assoziationsregeln ermittelt werden.

5.4.3 Kombinationen mit F19.2

F19.2 definiert als: Psychische- und Verhaltensstörungen durch multiplen Substanzgebrauch und Konsum anderer psychotroper Substanzen: Abhängigkeitssyndrom

Für die F19.2 Diagnose die durchschnittlich in 6,6% aller Aufenthalte vorkommt wurde für die Jahre 2003 und 2005-2007 eine Assoziationsregel gefunden und zwar mit der Diagnose B18.2 – Chronische Virushepatitis C. In 52% der Aufenthalte die B18.2 Diagnose enthalten kommt ebenfalls die Diagnose F19.2 vor. Diese Regel ergibt in den 4 Jahren einen durchschnittlichen *Lift* von 6,4. Das heißt, dass die psychischen- und Verhaltensstörungen aufgrund von Konsum psychotroper Substanzen 6,4-mal öfters bei Patienten die an Hepatitis erkrankt als bei anderen Patienten vorkommen. Diese Assoziationsregel liegt einen Zusammenhang zwischen Drogenkonsum und Hepatitis Erkrankung nahe.

5.4.4 Kombinationen mit F13.2

F13.2 definiert als: Psychische- und Verhaltensstörungen durch Sedativa (Beruhigungsmitteln) oder Hypnotika (Schlafmitteln): Abhängigkeitssyndrom.

Diese Diagnose überschreitet die 5% Marke bei den UAE-Aufenthalten erst im Jahr 2006, dafür stieg ihre Häufigkeit von Jahr zu Jahr (7,6% im Jahr 2011). Es konnten drei Assoziationsregeln zu anderen Diagnosen gefunden werden.

Ab dem Jahr 2006 bis 2011, liegt eine Assoziation mit der Diagnose F10.2 - Psychische und Verhaltensstörungen durch Alkohol: Abhängigkeitssyndrom mit einem durchschnittlichen *Lift* von 5,8. Das bedeutet eine starke Korrelation zwischen Auftreten der psychischen- und Verhaltungsstörungen bei Patienten die Beruhigungs- oder Schlafmittel einnehmen (im gewissen Grade abhängig sind) und Patienten die diese Störungen aufgrund von Alkoholkonsum (ebenfalls Abhängigkeitssyndrom) aufweisen.

Ab Jahr 2008 besteht ein Zusammenhang zwischen dem Auftreten der Diagnose F13.2 mit einer anderen A2 Diagnose F11.2 - Psychische und Verhaltensstörungen durch Opi-

oide (Schmerzmittel auf Opiumbasis): Abhängigkeitssyndrom, der durchschnittlicher Lift betrug 4,6.

Ab dem Jahr 2007 konnte eine dritte Assoziationsregel mit der essentiellen Hypertonie ermittelt werden. Der mittlere Lift dieser Regel betrug 0,69 was bedeutet, dass bei Patienten die Schlaf- oder Beruhigungsmittel einnehmen, Patienten die an Hypertonie leiden unterrepräsentiert sind.

5.5 Unerwünschte Arzneimittelereignisse und Begleiterkrankungen

Die Auswahl des Patientenkollektivs für die weitere Analyse, basiert auf den Erkenntnissen, die aus der Assoziationsanalyse gewonnen wurden. Studien auf dem Gebiet der Herzkrankheiten bestätigen ebenfalls, dass das Kollektiv der hier ausgewählten Patienten sinnvoll ist. Demzufolge die an arterielle Hypertonie oder koronaren Herzerkrankung erkrankten Patienten häufig ebenfalls an COPD erkrankt sind (S Blum, 2014). Da die kardiovaskulären Erkrankungen oft mit der Verordnung von Betablockern einhergehen, entsteht automatisch eine gewisse Wechselwirkung dieses Arzneistoffes mit COPD (Mihaela S Stefan, 2012). Auch Diabetes mellitus wird in Verbindung mit der chronisch obstruktiver Lungenerkrankung gebracht (Johanna R Feary, 2010; 65).

Das Patientenkollektiv, das für die weitere Analyse aufgrund der Ergebnisse der Assoziationsanalyse gebildet wurde, beinhaltet ausschließlich Patienten mit der Grunderkrankung koronarer Herzkrankheit. Für diese Patienten werden ebenfalls die Informationen über mögliche Begleiterkrankungen wie chronische Lungenkrankheit und Diabetes mellitus analysiert. Für die Analyse bezüglich der unerwünschten Arzneimittelereignisse steht die mögliche Einnahme von Betablockern als der Aspekt einer medikamentösen Behandlung in Fokus. In diesem Abschnitt werden die deskriptive Analyse, univariate Testungen sowie die Ereigniszeitanalyse für die Daten aus den Jahren 2006 und 2007 beschrieben.

Im Jahr 2006 wurden 35.516 Patienten mit der koronaren Herzkrankheit stationär aufgenommen. Bei 76,3% dieser Patienten ist eine medikamentöse Behandlung mit Betablockern dokumentiert worden. 46,5% der Patienten litten zusätzlich an der COPD-Erkrankung, 24,8% hatten Diabetes mellitus.

Der Anteil weiblicher Patienten betrug 37,6%, 73% waren zwischen 61 - 90 Jahre alt. 8.461 Patienten die 2006 mit einer koronaren Herzkrankheit aufgenommen worden sind, starben.

Im Jahr 2007 sind 31.328 Patienten mit koronarer Herzkrankheit stationär aufgenommen worden. Betablocker Einnahme wurde bei 74,1% Patienten dokumentiert, 44,8% der Patienten hatten zusätzlich eine COPD-Erkrankung und 23,9% waren an Diabetes mellitus erkrank.

Die Verteilung bezüglich der demografischen Faktoren war ähnlich wie im Vorjahr (37,2% der Patienten waren weiblich, 73,5% waren zwischen 61 – 90 Jahre alt). 6.065 Patienten die 2007 aufgenommen worden sind, sind verstorben.

5.5.1 Univariate Tests

In diesem Abschnitt ist durch die Analyse der Häufigkeitsverteilungen der UAE-Diagnosen getestet worden, ob ein signifikanter Zusammenhang zwischen: den Komorbiditäten, den demografischen Faktoren, medikamentöser Behandlung mit Betablocker, der Aufenthaltsdauer und schließlich dem Tod und der UAE-Diagnose besteht.

Im Jahr 2006 gab es einen signifikanten Zusammenhang zwischen dem Auftreten einer UAE-Diagnose und: der COPD Erkrankung ($\chi^2 = 38{,}54$, p-Wert <0,001) und der Erkrankung an Diabetes mellitus ($\chi^2 = 22{,}68$, p-Wert <0,001). Die demografischen Faktoren Geschlecht und Altersgruppe wiesen ebenfalls einen signifikanten Zusammenhang mit dem Auftreten einer UAE-Diagnose auf (Geschlecht: $\chi^2 = 33{,}15$, p-Wert <0,001; Altersgruppe: $\chi^2 = 49{,}65$, p-Wert <0,001). Es konnte kein signifikanter Zusammenhang für die Einnahme der Betablocker bezüglich der Häufigkeit der UAE-Diagnosen festgestellt werden. Ein hochsignifikanter Zusammenhang bestand zwischen dem Vorhandensein einer UAE-Diagnose und der Anzahl der Sterbefälle ($\chi^2 = 107{,}86$, p-Wert <0,001). Bezüglich der Aufenthaltsdauer bestand ebenfalls ein signifikanter Unterschied (p-Wert <0,001). Patienten, bei denen eine Diagnose die auf unerwünschte Arzneimittelereignisse hinweist festgestellt worden ist, waren durchschnittlich 60 Tage (Standardabweichung = 47 Tage) in stationärer Behandlung, wohingegen die Patienten ohne einer UAE-Diagnose verbrachten durchschnittlich 25 Tage (Standardabweichung = 29 Tage) in stationärer Behandlung.

Im Jahr 2007 gab es einen Zusammenhang zwischen dem Auftreten einer UAE-Diagnose und: der COPD Erkrankung ($\chi^2 = 13{,}38$, p-Wert <0,001). Im Vergleich zu 2006 gibt es keinen signifikanten Zusammenhang mehr, zwischen UAE-Diagnose und der Erkrankung an Diabetes mellitus. Die demografischen Faktoren Geschlecht und Altersgruppe wiesen nach wie vor einen signifikanten Zusammenhang mit dem Auftreten einer UAE-Diagnose auf (Geschlecht: $\chi^2 = 5{,}43$, p -Wert = 0,02; Altersgruppe: p-Wert <0,001). Es besteht weiterhin kein signifikanter Zusammenhang für die Einnahme der Betablocker bezüglich der Häufigkeit der UAE-Diagnose. Zwischen dem Vorhandensein einer UAE-Diagnose und der Anzahl der Sterbefälle konnte erneut ein hochsignifikanter Zusammenhang festgestellt werden ($\chi^2 = 83{,}88$, p-Wert <0,001). Patienten mit einer UAE-Diagnose waren durchschnittlich 56 Tage (Standardabweichung = 42 Tage) in stationärer Behandlung, wohingegen die Patienten ohne einer UAE-Diagnose verbrachten durchschnittlich 21 Tage (Standardabweichung = 26 Tage) in stationärer Behandlung (p-Wert <0,001).

Die Analyse der Daten aus den beiden Jahren ergab ähnliche Ergebnisse, was ebenfalls bedeutet, dass die Erkenntnisse bestätigt werden konnten. In beiden Jahren 2006 und 2007 besteht zwischen den demographischen Faktoren Alter und Geschlecht, Aufenthaltsdauer sowie der Erkrankung an COPD ein Zusammenhang in Bezug auf Häufigkeit einer UAE-Diagnose. Es konnte kein signifikanter Zusammenhang zwischen der Einnahme der Betablocker und dem Auftreten einer UAE-Diagnose festgestellt werden. Lediglich bei der Analyse der Häufigkeiten bezüglich der Erkrankung an Diabetes mellitus unterscheiden sich die Ergebnisse. Im Jahr 2006 konnte ein signifikanter Zusammenhang festgestellt werden, 2007 ist das Ergebnis knapp nicht mehr signifikant. Die univariaten Tests sind in folgender Tabelle dargestellt.

Tabelle 10 Univariate Tests für die Einflussfaktoren für die Jahre 2006 und 2007

	Jahr 2006 n = 35.516	Jahr 2007 n = 31.328
	p-Wert	
Geschlecht	<0,001†	0,020†
Altersgruppe	<0,001†	<0,001†
Diabetes	<0,001†	0,062†
Betablocker	0,617†	0,188†
COPD	<0,001†	<0,001†
Aufenthaltsdauer	<0,001‡	<0,001‡
Tod	<0,001†	<0,001†

† Chi 2 – Test

‡ Mann-Whitney-U-Test

5.5.2 Ereigniszeitanalyse

Die Ereigniszeitanalyse ist für drei verschiedene Zeitpunkte durchgeführt worden. In die Analyse sind alle Patienten eingeschlossen worden, die zwischen 01.01.2006 und 31.07.2007 stationär aufgenommen wurden. Insgesamt waren es 54.734 Patienten, bei 354 Patienten wurde eine UAE-Diagnose dokumentiert.

5.5.2.1 Ereignisanalyse 15 Tage nach der Aufnahme

Innerhalb der ersten 15 Tage nach der stationären Aufnahme ist bei 84 Patienten eine UAE-Diagnose gestellt worden, das entspricht einem Prozentanteil von 23,7% aller gestellten UAE-Diagnosen. Mittels univariater Cox-Regressionen ist das Risiko einer UAE-Diagnose quantifiziert worden, dabei sind die Hazard-Ratios (HR) und deren 95% Konfidenzintervalle ermittelt worden.

Tabelle 11 Univariate Cox-Regressionen (15 Tage nach der stationären Aufnahme)

	15 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	2,04	(1,32 – 3,15)	0,001
Alter			
0 – 45 Jahre	0,54	(0,07 – 4,19)	0,55
61 -75 Jahre	1,09	(0,52 – 2,31)	0,82
76 - 90 Jahre	2,38	(1,19 – 4,78)	0,015
über 90 Jahre	10,95	(4,80 – 24,97)	< 0,001
Diabetes ja	1,69	(1,09 – 2,65)	0,02
Betablocker ja	0,44	(0,28 – 0,68)	<0,001
COPD ja	2,33	(1,48 – 3,66)	<0,001

Referenzgruppen: Geschlecht – männlich

Altersgruppe 2 (46-60 Jahre)

Demnach war das Risiko einer UAE-Diagnose bei Frauen doppelt so hoch als bei Männer (HR = 2,04, p-Wert = 0,001). Patienten der Altersgruppen 4 (76 - 90 Jahre) und Altersgruppe 5 (über 90 Jahre) hatten ein höheres Risiko einer UAE-Diagnose als die Referenzgruppe 2 (46 - 60 Jahre) (Altersgruppe 4: HR = 2,38, p-Wert = 0,015; Altersgruppe 5: HR = 10,95, p-Wert <0,001). Die Begleiterkrankungen COPD und Diabetes mellitus erhöhten ebenso das Risiko einer UAE-Diagnose (COPD: HR = 2,33, p-Wert <0,001; Diabetes mellitus: HR = 1,69, p-Wert = 0,023). Bei Patienten die Betablocker eingenommen haben war das Risiko einer UAE-Diagnose um den Faktor 0,44 niedriger als bei Patienten, die keine Betablocker einnahmen (HR = 0,44, p-Wert <0,001).

Im multivariaten Modell ist das Risiko bei Frauen nicht mehr signifikant größer als bei den Männern (HR: 1,49, KI (0,95 – 2,35), p-Wert = 0,083), das Risiko für Patienten in der Altersgruppe 4 ist ebenfalls nicht mehr signifikant grösser als das der Referenzgruppe (Altersgruppe 4: HR = 1,87, KI (0,92 – 3,82), p-Wert = 0,084). Das Risiko einer

UAE-Diagnose wird durch die Erkrankung an Diabetes mellitus nicht mehr signifikant beeinflusst (Diabetes mellitus: HR = 1,32, KI (0,79 – 2,20), p-Wert = 0,288).

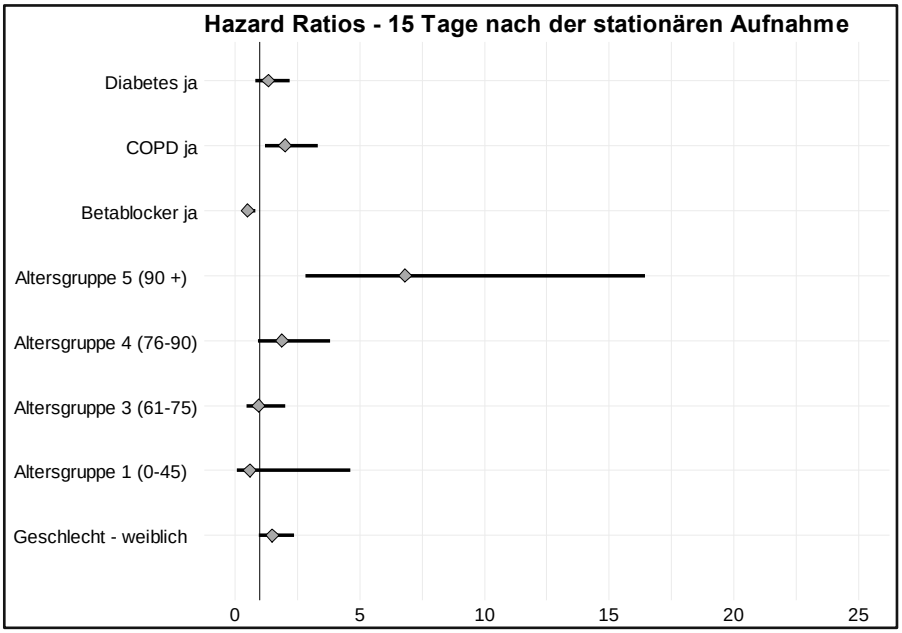

Abbildung 15 Multivariates Cox-Regressionsmodel - 15 Tage nach der stationären Aufnahme

5.5.2.2 Ereignisanalyse 60 Tage nach der Aufnahme

In den Zeitraum von 60 Tagen ab der stationären Aufnahme des Patienten, ist bei 127 Patienten die UAE-Diagnose dokumentiert worden, ein Prozentanteil von 35,9% aller gestellten UAE-Diagnosen.

Univariate Auswertungen haben ergeben, dass das Risiko einer UAE-Diagnose erneut für Frauen und für die Altersgruppen 4 und 5 signifikant höher ist als für die Patienten der Referenzgruppen. Im Vergleich zu der Analyse zum Zeitpunkt 15 Tage nach der stationären Aufnahme, ist das Risiko für diese Faktoren gestiegen (Geschlecht weiblich: HR = 2,43, p-Wert <0,001, Altersgruppe 4: HR = 2,84, p-Wert <0,001; Altersgruppe 5: HR = 11,04, p-Wert <0,001).

Patienten die an COPD oder Diabetes mellitus erkrankt sind haben ebenfalls ein erhöhtes Risiko einer UAE-Diagnose (COPD: HR = 1,91, p-Wert <0,001; Diabetes mellitus:

HR = 1,68, p-Wert = 0,005). Die Einnahme der Betablocker reduziert im univariaten Modell das Risiko um den Faktor 0,61 (HR = 0,61, p-Wert = 0,009).

Tabelle 12 Univariate Cox-Regressionen (60 Tage nach der stationären Aufnahme)

	60 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	2,43	(1,70 – 3,47)	<0,001
Alter			
0 – 45 Jahre	0,41	(0,05 – 3,15)	0,39
61 -75 Jahre	1,41	(0,75 – 2,66)	0,28
76 - 90 Jahre	2,84	(1,56 – 5,19)	<0,001
über 90 Jahre	11,04	(5,36 – 22,72)	<0,001
Diabetes ja	1,68	(1,17 – 2,41)	0,005
Betablocker ja	0,61	(0,42 – 0,88)	0,009
COPD ja	1,91	(1,34 – 2,74)	<0,001

Referenzgruppen: Geschlecht – männlich

Altersgruppe 2 (46-60 Jahre)

In dem multivariaten Modell sind Geschlecht und Alter weiterhin signifikante Risikofaktoren (Geschlecht weiblich: HR = 1,82, 95 % KI (1,26 – 2,64), p-Wert = 0,001; Altersgruppe 4: HR = 2,22, KI (1,20 – 4,10), p-Wert = 0,011; Altersgruppe 5: HR = 7,26, KI (3,38 – 15,58), p-Wert <0,001), ebenso die Erkrankung an COPD (HR = 1,58, KI (1,04 – 2,38), p-Wert = 0,030) erhöht das Risiko einer UAE-Diagnose. Ein signifikanter Einfluss auf das Risiko bei Patienten mit Diabetes und bei Patienten, die Betablocker verabreicht bekamen, ist nicht mehr gegeben.

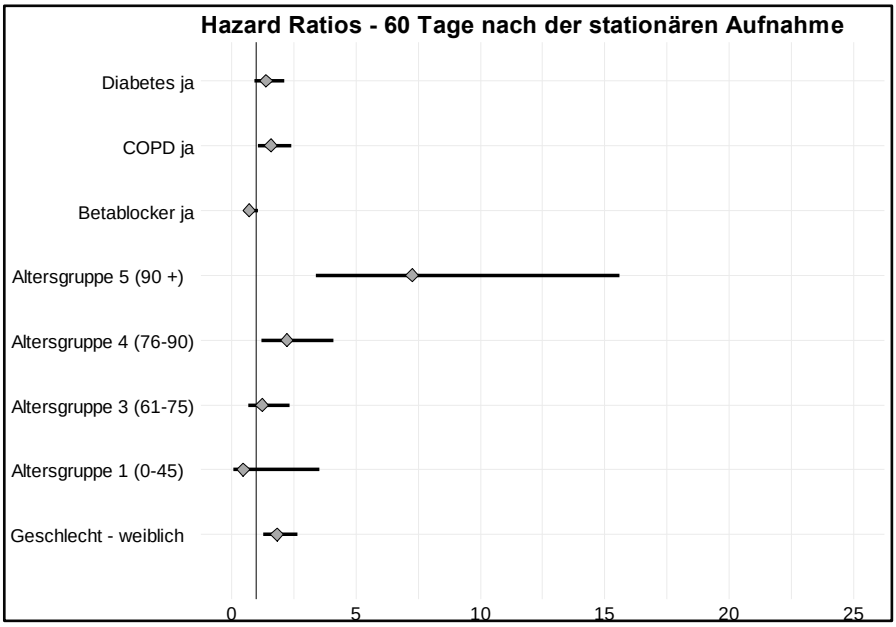

Abbildung 16 Multivariates Cox-Regressionsmodel - 60 Tage nach der stationären Aufnahme

5.5.2.3 Ereignisanalyse 5 Monate nach der Aufnahme

Der letzte Zeitpunkt für den die Ereigniszeitanalyse durchgeführt worden ist, sind 5 Monate ab der Aufnahme eines Patienten in einem Krankenhaus. Bis zu diesem Zeitpunkt ist die Anzahl der festgestellten UAE-Diagnosen auf 203 gestiegen.

Die univariaten Modelle ergaben signifikante Risikounterschiede für alle getesteten Faktoren. Demnach haben weibliche Patienten und Patienten aus den Altersgruppen 4 und 5 ein signifikant erhöhtes Risiko einer Diagnose die auf unerwünschte Arzneimittelereignisse hindeutet (Geschlecht weiblich: HR = 2,20, p-Wert <0,001; Altersgruppe 4: HR = 2,79, p-Wert <0,001; Altersgruppe 5: HR = 8,85, p-Wert <0,001).

COPD und Diabetes mellitus erhöhen ebenfalls dieses Risiko (COPD: HR = 2,03, p-Wert <0,001; Diabetes mellitus: HR = 1,41, p-Wert = 0,023). Die Einnahme der Betablocker reduziert das Risiko einer UAE-Diagnose um den Faktor 0,7 (HR = 0,70, p-Wert = 0,016).

Tabelle 13 Univariate Cox-Regressionen (5 Monate nach der stationären Aufnahme)

	5 Monate		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	2,20	(1,66 - 2,90)	<0,001
Alter			
0 – 45 Jahre	0,73	(0,22 – 2,44)	0,61
61 -75 Jahre	1,28	(0,78 – 2,11)	0,31
76 - 90 Jahre	2,79	(1,76 - 4,44)	<0,001
über 90 Jahre	8,85	(4,93 - 15,87)	<0,001
Diabetes ja	1,41	(1,05 - 1,89)	0,023
Betablocker ja	0,70	(0,52 - 0,93)	0,016
COPD ja	2,03	(1,52 - 2,7)	<0,001

Referenzgruppen: Geschlecht – männlich

Altersgruppe 2 (46-60 Jahre)

Im multivariaten Modell sind die demografischen Faktoren Geschlecht und Altersgruppe 4 und 5 die signifikanten Einflussfaktoren (Geschlecht weiblich: HR = 1,68, p-Wert <0,001; Altersgruppe 4: HR = 2,25, p-Wert <0,001; Altersgruppe 5: HR = 6,27, p-Wert <0,001). Die COPD Erkrankung erhöht ebenfalls das Risiko einer UAE-Diagnose (HR = 1,91, p-Wert <0,001). Bei den Altersgruppen 1 und 3, Diabetes mellitus und die Einnahme von Betablockern konnte keine signifikante Änderung des Risikos für eine UAE-Diagnose festgestellt werden.

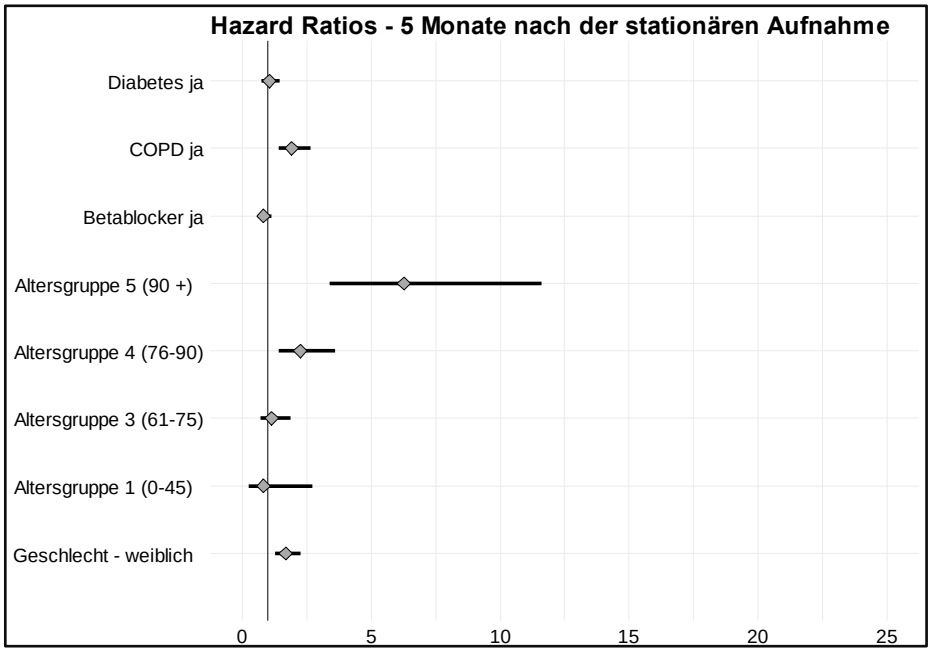

Abbildung 17 Multivariates Cox-Regressionsmodel - 5 Monate nach der stationären Aufnahme

5.5.2.4 Ereigniszeitanalyse mit imputierten Daten

Nachdem die fehlenden Werte bei den Faktoren Geschlecht und Altersgruppe ersetzt worden sind, wurde erneut die Ereigniszeitanalyse durchgeführt. Die Ergebnisse untermauern die Aussagen der ersten Auswertung, es konnten jedoch bei dem univariaten Modell für Geschlecht 1.080 Patienten mehr in die Analyse eingeschlossen werden. Bei dem Modell bezüglich der Altersgruppen konnten 13 Patienten mehr in die Auswertung einfließen. Für das multivariate Modell konnten 1.085 Patienten mehr analysiert werden. Im multivariaten Modell zum Zeitpunkt 15 Tage nach der stationären Aufnahme erhöhten die Altersgruppe 5 (90+ Jahre alt) sowie die COPD-Erkrankung das Risiko einer UAE-Diagnose (Altersgruppe 5: HR = 7,60, KI (3,20 – 18,05), p-Wert <0,001; COPD: HR = 2,08, KI (1,25 – 3,46), p-Wert = 0,004). Die Einnahme der Betablocker reduzierte das Risiko signifikant (HR = 0,53, KI (0,34 – 0,83), p-Wert = 0,005). Zum Zeitpunkten 60 Tage und 5 Monate nach der stationären Aufnahme, waren die Ergebnisse der multivariaten Modelle nach der Imputation, mit den Ergebnissen der Modelle, die mit dem Rohdatensatz berechnet worden sind, einstimmig. Detaillierte Ergebnisse sind im Anhang ersichtlich.

6 Diskussion

Die inhaltlichen Erkenntnisse die aus Analyse der Daten in dieser Arbeit gewonnen wurden können bezüglich drei Aspekte zusammengefasst werden.

Der erste Aspekt betrifft die Auswertung der MBDS-Daten bezüglich demografischen Faktoren. Was das Geschlecht betrifft sind Frauen durchgehend öfter von einer AUE-Diagnose betroffen, bis auf die Diagnosen aus der Kategorie A2 hatten die Frauen höheres Risiko einer Diagnose die auf unerwünschte Arzneimittelereignisse deutet. Bezüglich des Alters war der Anteil und auch das Risiko einer Diagnose aus der Kategorie A2 - Induzierung durch Arzneimittel oder andere Ursachen und der Kategorie B2 - Vergiftung oder schädlicher Gebrauch durch Arzneimittel oder andere Ursachen, besonders hoch bei den Altersgruppen 0, 1 und 2. Dieses Ergebnis ist eindeutig dem unbedachten beziehungsweise zweckentfremdeten Arzneimittelkonsum der Jugendlichen und jungen Menschen zuzuordnen.

Die Assoziationsanalyse hat wiederrum eine Grunderkrankung sichtbar gemacht, die eng mit den unerwünschten Arzneimittelereignissen in Verbindung gebrachte werden konnte. Demnach sind vor allem Patienten mit koronarer Herzkrankheit und Herzinfarkt oft von einer AUE- Diagnose betroffen.

Ebenfalls durch die Assoziationsanalyse konnte ein beunruhigender Trend beobachtet werden. In diesem Fall waren die Assoziationsregeln ein Indiz dafür, dass der Gebrauch von Schlaf- und Beruhigungsmitteln mit Kombination mit Alkohol und schließlich Schmerzmitteln immer häufiger und deutlicher wird.

Abschließend können Verbesserungsvorschläge beziehungsweise kritische Anmerkungen nicht aus der Acht gelassen werden. Für die Alters- und Geschlechtsstandardisierung verwendete Standardbevölkerung stammt aus dem Jahr 1990. Um die exakte Anpassung der Bevölkerungszahlen Österreichs war keine andere Standardbevölkerung vorhanden, die gleichermaßen das Geschlecht und Alter berücksichtigt. Das Bestreben der WHO neue, aktualisierte und flexibel gestaltete Version der Standardbevölkerungen herauszugeben wäre wünschenswert.

Weitere Aspekt der angesprochen wird, sind die fehlenden Werte in Geschlecht und Altersgruppe der Patienten in den Verrechnungsdaten. Das nicht Vorhandensein zweier

der wichtigsten Merkmale der Patienten die stationär aufgenommen werden, ist bei Verrechnungsdaten durch nichts zu begründen. Ein anderer Aspekt der Datenqualität bei den MBDS-Daten betrifft den fehlenden Patientenbezug. Mit einer einfachen Maßnahme, der Vergabe einer Identifikationsnummer für den aufgenommenen Patienten, könnte man, mit so gewonnener Zusatzinformation, den gesamten Verlauf der Krankengeschichte eines Patienten aufzeichnen und verfolgen. Dieser Aspekt ist im Zeitalter von elektronischer Gesundheitsakte (ELGA) unbedingt zu beachten und zu berücksichtigen. Eine umfangreiche, strategische Planung einer Datenbank ist in erster Linie essentiell für die Koordination der medizinischen Versorgung der Patienten, aber auch für wissenschaftliche und ebenso für wirtschaftliche Analysen, ist sie von großer Bedeutung. Ein einheitliches Konzept sowie ein zentrales Register mit strikten Anweisungen der Dateneingabe, würden eine bessere Verknüpfung von Daten ermöglichen, und somit die Möglichkeit bieten diese optimal auszuwerten.

7 Tabellenverzeichnis

8 Abbildungsverzeichnis

9 Literaturverzeichnis

1. Andreas Behr, U. P. (2014). *Einführung in die Statistik mit R.* Vahlen. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=rkpLBAAAQBAJ&printsec=frontcover&dq=Einführung+in+die+Statistik+mit+R&hl=de&sa=X&ved=0ahUKEwiuraeA_Z3bAhXDKJoKHacABjgQ6AEIJzAA#v=onepage&q=Einführung%20in%20die%20Statistik%20mit%20R&f=false

2. Christoph Rinner, P. R.-Z. (09. 12 2015). *Drug safety 3 – Analysis of Adverse Drug Events in Austrian hospital discharge diagnoses between 2001 and 2011.* Abgerufen am 23. 05 2018 von https://cemsiis.meduniwien.ac.at/mim/wf/projekte/drug-safety3/: http://www.meduniwien.ac.at/msi/mias/papers/ADE3-Endbericht-2015-12-04.pdf

3. David W. Hosmer, J. S. (2004). *Applied Logistic Regression.* John Wiley & Sons. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=Po0RLQ7USIMC&pg=PA66&hl=de&source=gbs_selected_pages&cad=2#v=onepage&q&f=false

4. Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences, 9th Edition.* Cengage Learning. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=zzV-BAAAQBAJ&pg=PA619&dq=chi+2+test&hl=de&sa=X&ved=0ahUKEwidlZiP05LbAhWMZFAKHduwBys4RhDoAQhrMAk#v=onepage&q=chi%202%20test&f=false

5. Graubner, B. (2013). *ICD-10-GM 2014 Systematisches Verzeichnis - Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme* . Deutscher Ärzteverlag.

6. Gregory W. Corder, D. I. (2011). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach.* John Wiley & Sons. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=T3qOqdpSz6YC&printsec=frontcover&dq=mann+whitney+U+Test&hl=de&sa=X&ved=0ahUKEwir75uZrqbbAhUCalAKHVkCDgUQ6AEINjAD#v=onepage&q=U%20Test&f=false

7. Guo, S. (2010). *Survival Analysis.* New York, USA: Oxford University Press. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=QR-ye6JJms4C&printsec=frontcover&dq=Survival+Analysis+guo&hl=de&sa=X&ved=0ahUKEwjvkPKpqaPbAhXNyKQKHQuuAr8Q6AEIJzAA#v=onepage&q=partial&f=false

8. Hohl CM, K. A.-W. (12. November 2013). *National Center for Biotechnology Information.* doi:10.1136/amiajnl-2013-002116

9. Howell, D. C. (30. 01 2013). *Graphics in R.* Abgerufen am 23. 05 2018 von https://www.uvm.edu/~dhowell/methods8/Supplements/R-Programs/R-Graphics.html

10. Johanna R Feary, L. C. (2010; 65). Prevalence of major comorbidities in subjects with COPD and incidence of myocardial infarction and stroke: a comprehensive analysis using data from primary care. *Thorax; Volume 65*, S. 956-962. doi:10.1136/thx.2009.128082

11. Jürgen Cleve, U. L. (2016). *Data Mining.* Walter de Gruyter GmbH & Co KG. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=LngDDgAAQBAJ&printsec=frontcover&dq=Data+Mining+Cleve&hl=de&sa=X&ved=0ahUKEwig5KXlrp7bAhUhLZoKHbvbCD8Q6AEIJzAA#v=onepage&q=support&f=false

12. Karsten Lübke, M. V. (2014). *Angewandte Wirtschaftsstatistik: Daten und Zufall.* Wiesbaden: Springer-Verlag. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=GWJ3BQAAQBAJ&printsec=copyright&redir_esc=y#v=onepage&q=lift&f=false

13. Klaus Backhaus, B. E. (2015). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung.* Springer-Verlag. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=3LvYCgAAQBAJ&pg=PA283&dq=interpretation+logistische+regression&hl=de&sa=X&ved=0ahUKEwjq-8iL_6HbAhWJKFAKHci8Apo4ChDoAQgmMAA#v=onepage&q=interpretation%20koeffizienten&f=false

14. Liu, X. (2015). *Applied Ordinal Logistic Regression Using Stata: From Single-Level to Multilevel Modeling.* SAGE Publications. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=CqpiCgAAQBAJ&printsec=frontcover&dq=logistische+regression+model&hl=de&sa=X&ved=0ahUKEwj8mLOu-6HbAhXRy6QKHe2nCF84FBDoAQg2MAI#v=onepage&q=logistische%20regression%20model&f=false

15. Longford, N. T. (2006). *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician.* Springer Science & Business Media. Abgerufen am 18. 06 2018 von https://books.google.at/books?id=YVqpHsQxUu8C&printsec=frontcover&dq=hot+deck+imputation&hl=de&sa=X&ved=0ahUKEwjuhvPa8YTcAhVyyKYKHQWTASo4Ch

DoAQhFMAQ#v=onepage&q=hot%20deck%20imputation&f=false

16. Mara Tableman, J. S. (2003). *Survival Analysis Using S: Analysis of Time-to-Event Data*. CRC Press. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=CgUflE1yBi8C&pg=PA254&dq=kaplan-meier+greenwood&hl=de&sa=X&ved=0ahUKEwjWld2g0abbAhXHJlAKHa4cAtY4HhDoAQhNMAU#v=onepage&q=kaplan-meier%20greenwood&f=false

17. Marco Giesselmann, M. W. (2013). *Regressionsmodelle zur Analyse von Paneldaten.* Springer-Verlag. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=U6ojBAAAQBAJ&printsec=frontcover&dq=logistische+regression&hl=de&sa=X&ved=0ahUKEwj12-yknJ7bAhViIpoKHQAxDHA4ChDoAQgsMAE#v=onepage&q=logistische%20regression&f=false

18. Marwala, T. (2009). *Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques.* Hershey: IGI Global. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=1_O7K9EIl2UC&printsec=frontcover&dq=Computational+Intelligence+for+Missing+Data+Imputation,+Estimation,+and+Management:+Knowledge+Optimization+Techniques&hl=de&sa=X&ved=0ahUKEwjbmJCAhqLbAhWEaFAKHeifAYAQ6AEIJzAA#v=onepage&q

19. Michael Eid, K. S. (2014). *Testtheorie und Testkonstruktion.* Göttingen: Hogrefe Verlag. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=gw_uAwAAQBAJ&printsec=frontcover&dq=konfidenzintervalle+für+wald+test&hl=de&sa=X&ved=0ahUKEwj-zqbDg57bAhWDHpoKHTrqD1M4ChDoAQhMMAc#v=onepage&q=%20wald%20test&f=false

20. Mihaela S Stefan, M. B. (11 2012). Association between β-blocker therapy and outcomes in patients hospitalised with acute exacerbations of chronic obstructive lung disease with underlying ischaemic heart disease, heart failure or hypertension. *Thorax; Volume 67*, S. 977-984. doi:10.1136/thoraxjnl-2012-201945

21. Mills, M. (2011). *Introducing Survival and Event History Analysis.* SAGE. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=NzAdSrHNYr8C&printsec=frontcover&dq=Introducing+Survival+and+Event+History+Analysis&hl=de&sa=X&ved=0ahUKEwjSg4qbp6PbAhVNaFAKHcDRC3MQ6AEIJzAA#v=onepage&q=hazard&f=false

22. O.Jacke, C. (2013). *Evaluation der Versorgungseffektivität eines zertifizierten Brustkrebszentrums im Zeitvergleich (1996-97, 2003-04).* epubli GmbH. Abgerufen am 05 2018 von https://books.google.at/books?id=SE0gAgAAQBAJ&pg=PP1&lpg=PP1&dq=Evaluation+der+Versorgungseffektivität+eines+zertifizierten+Brustkrebszentrums+im+Zeitvergleich+(1996-97,+2003-04)&source=bl&ots=3U4dENAxHT&sig=Tzjy525GKVRXO6lsFIq1nKO_KN8&hl=de&sa=X&ved=0ahU

23. Roger, F. (1981). *The minimum basic data set for hospital statistics in the EEC.* Luxenburg: Publications Office of the European Union, Katalognummer CD-NJ-81-002-EN-C. Abgerufen am 23. 05 2018 von https://publications.europa.eu/en/publication-detail/-/publication/784b148a-776a-495d-9080-4982f64943b6/language-en

24. S Blum, D. J. (2014). Kardiovaskuläre und pneumologische Komorbiditäten steigern das Risiko von COPD-Exazerbationen – Daten der Kohorte Bergmannsheil. *Pneumologie*, S. 68 - V29. doi:10.1055/s-0034-1367739

25. Sarria-Santamera, A. (01. Oktober 2013). The Spanish Minimum Basic Data Set. *European Journal of Public Health, Volume 23*, S. 104. Abgerufen am 23. 05 2018 von https://academic.oup.com/eurpub/article/23/suppl_1/ckt126.256/2838122

26. Schlittgen, R. (2009). *Multivariate Statistik.* München: Oldenbourg Verlag. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=Vt_oBQAAQBAJ&printsec=frontcover&dq=Multivariate+Statistik&hl=de&sa=X&ved=0ahUKEwjnjteA9p3bAhWkA5oKHfpMAtgQ6AEIJzAA#v=onepage&q=Multivariate%20Statistik&f=false

27. Statistisches-Bundesamt. (2006). *Im Informationssystem der GBE zur Altersstandardisierung benutzte Standardbevölkerungen. Gliederungsmerkmale: Alter, Geschlecht, Standardbevölkerung.* Abgerufen am 24. 05 2018 von http://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/dboowasys921.xwdevkit/xwd_init?gbe.isgbetol/xs_start_neu/&p_aid=i&p_aid=21528687&nummer=1000&p_sprache=D&p_indsp=-&p_aid=71996870

28. Stausberg Jürgen, H. J. (22. January 2010). Identification of Adverse Drug Events. *Deutsches Ärzteblatt International*, 23-29. doi:10.3238/arztebl.2010.0023

29. Stausberg, J. (13. März 2014). *BMC Health Services Research.* doi:10.1186/1472-6963-

14-125

30. Sullivan, L. M. (2017). *Essentials of Biostatistics in Public Health.* Jones & Bartlett Learning. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=lwMODgAAQBAJ&pg=PA268&dq=interpretation+hazardratio&hl=de&sa=X&ved=0ahUKEwjbhPmtpabbAhUJKFAKHXBfAx4Q6AEIXDAG#v=onepage&q=interpretation%20hazardratio&f=false

31. Team, R. C. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Abgerufen am 03 2018 von https://www.r-project.org

32. Ton J. Cleophas, A. Z. (2006). *Statistics Applied to Clinical Trials: Self-Assessment Book.* Springer Science & Business Media. Abgerufen am 18. 06 2018 von https://books.google.at/books?id=FcOCccPur1AC&dq=kaplan+meier+plot&hl=de&source=gbs_navlinks_s

33. Trybou J, S. E. (Mai 2014). Costs associated with readmissions in Belgian acute-care hospitals. *Acta Clinica Belgica, Volume 68*, S. 263-267. doi:10.2143/ACB.3263

34. Urban, D. (1993). *Logit-Analyse: statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen.* Lucius & Lucius DE. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=iAxVU9QKUxoC&pg=PA59&dq=konfidenzintervalle+für+wald+test&hl=de&sa=X&ved=0ahUKEwiVq8uCg57bAhUqAZoKHX2YB2QQ6AEIKjAA#v=onepage&q=konfidenzintervalle%20für%20wald%20test&f=false

35. Woodwar, M. (1999). *Epidemiology: Study Design and Data Analysis.* CRC Press. Abgerufen am 23. 05 2018 von https://books.google.at/books?id=OqjB1WBrjjEC&pg=PA168&dq=mantel+haenszel+odds+ratio&hl=de&sa=X&ved=0ahUKEwjnu8vqlpXbAhUSJVAKHcbhCjEQ6AEILTAB#v=onepage&q=mantel%20haenszel%20odds%20ratio&f=false

10 Anhang

Kurzfassung

Der Fokus dieser Arbeit ist die Anwendung statistischer Methoden an reellen Routinedaten hinsichtlich der unerwünschten Arzneimittelereignisse kurz UAE.

Als Datenquellen für die Analyse dienen je nach Fragestellung sowohl die MBDS-Daten für die Jahre 2001-2011 als auch die Verrechnungsdaten für die Jahre 2006 und 2007. Verrechnungsdaten enthalten im Vergleich zur MBDS-Daten zusätzliche Informationen wie zum Beispiel medikamentöse Behandlung des Patienten oder die Informationen über den behandelten Arzt.

Neben der Alters- und Geschlechtsstandardisierung der Aufenthaltsraten sind statistische Methoden wie logistische Regression, Mantel-Haenszel Modell (als eine Methode der Meta-Analyse) für die Aggregation der Odds-Ratios und Berechnung der dazugehörigen Konfidenzintervalle angewendet worden. Dank der Assoziationsanalyse konnten die häufigsten Kombinationen zwischen ICD-10-Codes und den ADE ICD-10-Codes aufgespürt werden. Mit Hilfe der Cox-Regression wurde die Ereigniszeitanalyse durchgeführt.

Abstract

The main focus of this work is to apply statistical methods on real data with regard to adverse (drug) events. Analysis was based on both the MBDS-data from 2001 to 2011 and the billing data from 2006 and 2007. Compared to MBDS data, billing data provide additional information on the patients such as medical treatment or attending doctor.

Besides standardisation of rates of stay with regard to age and gender, further applied statistical methods were logistic regression, Mantel-Haenszel statistics (as a method for meta-analysis) for the aggregation of odd ratios and the calculation of the respective confidence intervals. Using Association rule learning, the most frequent combinations of ICD 10 codes and ADES ICD 10 codes could be determined. Furthermore, survival analysis was conducted by applying Cox regression.

Tabelle 14 Multivariate Cox-Regression

	15 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.49	(0.95 – 2.35)	0.08
Alter			
0 – 45 Jahre	0.59	(0.07 – 4.62)	0.62
61 -75 Jahre	0.94	(0.44 – 2.00)	0.87
76 - 90 Jahre	1.87	(0.92 – 3.82)	0.08
über 90 Jahre	6.81	(2.82 – 16.44)	< 0.001
Diabetes ja	1.32	(0.79 – 2.20)	0.29
Betablocker ja	0.50	(0.32 – 0.79)	0.003
COPD ja	1.99	(1.19 – 3.32)	0.008
	60 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.82	(1.26 – 2.64)	0.001
Alter			
0 – 45 Jahre	0.46	(0.06 – 3.51)	0.45
61 -75 Jahre	1.22	(0.65 – 2.32)	0.53
76 - 90 Jahre	2.22	(1.20 – 4.10)	0.01
über 90 Jahre	7.26	(3.38 – 15.58)	<0.001
Diabetes ja	1.38	(0.90 – 2.10)	0.14
Betablocker ja	0.70	(0.48 – 1.04)	0.07
COPD ja	1.58	(1.04 – 2.38)	0.03
	5 Monate		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.68	(1.26 – 2.25)	<0.001
Alter			

0 – 45 Jahre	0.81	(0.24 – 2.72)	0.74
61 -75 Jahre	1.14	(0.69 – 1.86)	0.61
76 - 90 Jahre	2.25	(1.40 – 3.60)	<0.001
über 90 Jahre	6.27	(3.39 – 11.60)	<0.001
Diabetes ja	1.04	(0.74– 1.45)	0.82
Betablocker ja	0.82	(0.60 – 1.11)	0.20
COPD ja	1.91	(1.39 – 2.63)	<0.001

Tabelle 15 Multivariate Cox-Regression mit imputierten Datensatz

	15 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.42	(0.91 – 2.22)	0.12
Alter			
0 – 45 Jahre	0.60	(0.08 – 4.67)	0.62
61 -75 Jahre	0.98	(0.46 – 2.07)	0.96
76 - 90 Jahre	1.87	(0.92 – 3.81)	0.08
über 90 Jahre	7.60	(3.20 – 18.05)	< 0.001
Diabetes ja	1.32	(0.80 – 2.18)	0.28
Betablocker ja	0.53	(0.34 – 0.83)	0.005
COPD ja	2.08	(1.25 – 3.46)	0.005
	60 Tage		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.76	(1.22 – 2.54)	0.002
Alter			
0 – 45 Jahre	0.46	(0.06 – 3.52)	0.45
61 -75 Jahre	1.26	(0.67 – 2.37)	0.48

76 - 90 Jahre	2.21	(1.20 – 4.08)	0.01
über 90 Jahre	7.84	(3.69 – 16.64)	<0.001
Diabetes ja	1.37	(0.90 – 2.08)	0.14
Betablocker ja	0.73	(0.49 – 1.07)	0.10
COPD ja	1.63	(1.08 – 2.46)	0.02
	5 Monate		
	HR	(95% KI)	p - Wert
Geschlecht weiblich	1.65	(1.24 – 2.20)	<0.001
Alter			
0 – 45 Jahre	0.81	(0.24 – 2.73)	0.74
61 -75 Jahre	1.15	(0.70 – 1.89)	0.57
76 - 90 Jahre	2.22	(1.39 – 3.56)	<0.001
über 90 Jahre	6.58	(3.58 – 12.08)	<0.001
Diabetes ja	1.04	(0.74– 1.45)	0.81
Betablocker ja	0.83	(0.61 – 1.13)	0.25
COPD ja	1.95	(1.42 – 2.68)	<0.001

Abbildung 18 Genaue Spezifikation von Diagnosen T88.7 und T78.4

T88.7- Nicht näher bezeichnete unerwünschte Nebenwirkung eines Arzneimittels oder einer Droge
Inklusive: Allergische Reaktion durch indikationsgerechtes Arzneimittel oder indikationsgerechte Droge bei ordnungsgemäßer Verabreichung, Idiosynkrasie durch indikationsgerechtes Arzneimittel oder indikationsgerechte Droge bei ordnungsgemäßer Verabreichung, Überempfindlichkeit und unerwünschte Nebenwirkung durch indikationsgerechtes Arzneimittel oder indikationsgerechte Droge bei ordnungsgemäßer Verabreichung. Exklusive: Näher bezeichnete unerwünschte Nebenwirkungen von Arzneimitteln und Drogen (A00-R99), (T80-T88.6), (T88.8)
T78.4 – Allergie, nicht näher bezeichnet
Inklusive: Allergische Reaktion, Idiosynkrasie und Überempfindlichkeit ohne nähere Angaben Exklusive: T88.7, näher bezeichnete Formen einer allergischen Reaktion, wie zum Beispiel: allergische Gastroenteritis und Kolitis (K52.2), Dermatitis (L23-L25) und (L27.-) und Heuschnupfen (J30.1).