

**Entwicklung einer Vorgehensweise zur automatisierten Erkennung  
eines Bedarfs an klinisch-pharmazeutischer Betreuung aus  
GKV-Routinedaten mittels Data-Mining**

Dissertation  
zur Erlangung des akademischen Grades  
Dr. Public Health (Dr. P. H.)

**Universität Bremen  
Zentrum für Sozialpolitik (ZeS)**

vorgelegt von Kerstin Boldt

Bremen, im April 2015

Gutachter:

1. Prof. Dr. rer. nat. Gerd Glaeske, Universität Bremen
2. Prof. Dr. med. Dirk O. Stichtenoth, Medizinische Hochschule Hannover

Kolloquium am: 05.02.2015

Die Arbeit hat dem Promotionsausschuss als Dissertation vorgelegen. Die Promotion wurde am 12.03.2015 vom Promotionsausschuss Dr. Public Health bestätigt.

## **Danksagung**

Ich danke Herrn Prof. Glaeske dafür, diese Arbeit von Beginn an unterstützt und mir auch über die Entfernung hinweg stets schnell und hilfreich zur Seite gestanden zu haben. Insbesondere danke ich für die freundlichen und aufmunternden Worte während der Phasen mit besonderen Herausforderungen. Ich danke Herrn Prof. Stichtenoth, mit seinem Hintergrundwissen über diese Arbeit und das darauf folgende Projekt zum AMTS-Gesundheitscoaching, als Gutachter fungiert und seine medizinische Perspektive eingebracht zu haben. Ich danke dem Team im ZES, insbesondere Frau Stück, Frau Trittin und Frau Höfel, für Ihre stetige Hilfe und Geduld. Den Mitdoktoranden danke ich für den Erfahrungsaustausch und die Hilfe bei den ersten Schritten. Ich danke dem Datenlieferanten für die Bereitstellung der Daten und namentlich Frau Anke Schäfer für die Datenbereitstellung, sowie Herrn Emin Asgarov und Herrn Vaske für den Erfahrungsaustausch in Bezug auf die Modellierungen.

Ein besonderer Dank gilt auch Freunden und Verwandten, ohne deren Unterstützung diese Arbeit nicht möglich gewesen wäre. Mein ganz spezieller Dank geht dabei an Anke Rogon und Petra Tipp für ihr Engagement beim Korrekturlesen und der liebevollen Betreuung von Julian. Des Weiteren danke ich Tobias Bansen, Dr. Susanne Kossow, Dr. Cornelia Heitmann-Gerdau, Dr. Christoph Vauth, Dr. Britta Kragert, Christian Boldt, Katrin Pieper und Dr. Dawid Pieper für Ihre vielfältige Unterstützung. Für die wertvolle Kinderbetreuung danke ich zudem Silvia, Sabine, Jana und Frau Hahn. Ich danke meinen Eltern und Ullu, für die kontinuierliche Unterstützung und den tollen Empfang nach vollbrachter Tat. Ich danke Julian für geduldiges Aushalten und keinen Tag zu früh Krabbeln lernen.

In Gedenken an Blume († 07.02.2012)

Hinweis:

Aus Gründen der besseren Lesbarkeit wird auf geschlechtsspezifische Formulierungen verzichtet. Sämtliche Rollenbezeichnungen gelten im Sinne der Gleichbehandlung für beide Geschlechter. Wird nur ein Geschlecht erwähnt, beziehen sich die Aussagen explizit nur auf dieses.

## Inhalt

Abkürzungsverzeichnis .....	5
Verzeichnis der Abbildungen .....	7
Verzeichnis der Tabellen .....	9
<b>1 Einführung .....</b>	<b>11</b>
1.1 Wissenschaftlicher Hintergrund .....	11
1.2 Nutzen aus der Perspektive einer Gesetzlichen Krankenkasse.....	16
1.3 Zielsetzung und Vorgehensweise .....	17
1.4 Erwartete Limitationen .....	19
<b>2 Fachliche Vorüberlegungen (Business Understanding).....</b>	<b>20</b>
2.1 Grundverständnis einer klinisch-pharmazeutischen Betreuung .....	20
2.2 Auswahl des Zielmerkmals .....	21
2.3 Zielgruppe und ihre Charakteristika .....	24
2.4 Vorauswahl an GKV-Routinedaten .....	28
2.5 Public Health Aspekte .....	30
<b>3 Methodische Vorgehensweise .....</b>	<b>31</b>
3.1 Vor- und Nachteile von GKV-Routinedaten .....	31
3.2 Data-Mining-Prozess und Methode.....	34
3.3 Qualitätssicherung .....	40
3.4 Analyseumfang .....	43
3.5 Ergebnissicherung .....	44
3.6 Ethik und Datenschutz .....	45
<b>4 Voranalysen (Data Understanding).....</b>	<b>46</b>
4.1 Technische Definition der Polypharmazie .....	47
4.2 Alters- und Geschlechtsverteilung bei Polypharmazie .....	66
4.3 Charakteristische Arzneistoffe .....	68
4.4 Charakteristische Diagnosen .....	75
4.5 Prüfung des Zielmerkmals auf Signifikanz .....	83
<b>5 Datenzusammenstellung (Data Preparation) .....</b>	<b>86</b>
5.1 Ausgewählte Variablen .....	86
5.2 Datenermittlung am Beispiel des Modellierungsdatensatzes.....	86

<b>6</b>	<b>Modellentwicklung (Modeling)</b> .....	<b>89</b>
6.1	Konkretisierung der Data-Mining Ziele (Hypothesen) .....	89
6.2	Prüfung des Ausgangsdatensatzes .....	90
6.3	Optimierung und Streichung von Variablen .....	93
6.4	Prüfung der Variablen .....	102
6.5	Modellierung .....	107
6.6	Ergebnis.....	115
<b>7</b>	<b>Ergebnisprüfung und Zielerreichung (Evaluation)</b> .....	<b>120</b>
7.1	Beurteilung der Prädiktoren .....	120
7.2	Beurteilung des Modells.....	122
7.3	Ausprägung wichtiger Variablen .....	125
7.4	Klassifikationsmatrix (Zielerreichung) .....	132
<b>8</b>	<b>Schlussfolgerung (Deployment)</b> .....	<b>136</b>
8.1	Diskussion .....	136
8.2	Fazit und Ausblick.....	150
<b>9</b>	<b>Literatur</b> .....	<b>152</b>
<b>10</b>	<b>ANLAGEN</b> .....	<b>163</b>
10.1	Analyseplan.....	163
10.2	UAW-Triggersymptome .....	164
10.3	UAW-Triggersymptome nach ATC .....	167
10.4	Analyse – Ausschlussdiagnosen .....	169
10.5	Ergebnistabellen zu den Voranalysen .....	170
10.6	Feldbeschreibung für die Variablen.....	174
10.7	Übersicht der Variablen für Datensatz, Modellierung und finales Modell .	180
10.8	Gruppierte ATC-Kombinationen (Matrix) .....	184
10.9	Dokumentation zur Prüfung der Variablen .....	185
10.10	Dokumentation zur Modellbildung .....	191
10.11	Validierungsdatensatz.....	198
10.12	Dokumentation zur Modellvalidierung .....	199
10.13	Dokumentation zur Ergebnisprüfung (Variablen).....	203

**Abkürzungsverzeichnis**

ABH	Arzneimittelbedingte Hospitalisation(en)
ALU	Arbeitslose
AMTS	Arzneimitteltherapiesicherheit
ANOVA	Analysis of Variance (Varianzanalyse)
ANZ	Anzahl
ARN	Arbeitnehmer
ARZN	Arzneimittel
ASS	Acetylsalicylsäure
ATC	Anatomisch-therapeutisch-chemische Klassifikation (für Wirkstoffe)
AUC	Area under curve
BG	Berufsgruppe
COPD	Chronisch Obstruktive Lungenerkrankung
CRISP	Cross Industry Standard Process (for Data Mining)
DDD	Defined daily dose(s)
ESCP	European Society of Clinical Pharmacy
Exp(B)	Effektkoeffizient (Exponentialfunktion an der Stelle B)
FAVE	Familienversicherte
GKV	Gesetzliche Krankenversicherung
GRP	Gruppe
HI	Herzinsuffizienz
ICD	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
ID	Identifikationsnummer
intervall.	intervallskaliert
i. o.	intra-okulär
J	Jahr
k. A.	Keine Angaben (keine Daten)
KHK	Koronare Herzkrankheit
KN	Kennzeichen
Ln	natürlicher Logarithmus
M	männlich
M 1	Modell 1

Min	Mindestanzahl
NSAR	nicht-steroidale Antirheumatika
OTC	Over The Counter (apothekenpflichtige, nicht verschreibungspflichtige Arzneien)
p-Wert	Wahrscheinlichkeit
PPI	Protonenpumpenhemmer
PPV	positive predictive value
PRE	Vorhersagewahrscheinlichkeit (nach dem log. Regressionsmodell)
Q	Quartal
REN	Rentner
ROC	Receiver Operating Characteristics
R-Quadrat	Bestimmtheitsmaß
SELB	Selbständige
SGB-V	fünftes Sozialgesetzbuch
SONST	Sonstige
SOZ	Sozialhilfeempfänger
SPSS	IBM SPSS Statistics (Statistical Package of the Social Sciences)
SQL	Structured Query Language
TAH	Thrombozytenaggregationshemmer
UAE	Unerwünschtes Arzneimittelereignis
UAW	Unerwünschte Arzneimittelwirkung
W	weiblich
WW	Wechselwirkungen
VO	Verordnung
vs.	versus

## Verzeichnis der Abbildungen

Abbildung 1: Phasen des CRISP (Chapmann et al., 2000, S.10-12, modifiziert) .....	37
Abbildung 2: Darstellung der Polypharmazie-Ermittlung bei variierendem Stichtag (Voranalysen I-III) .....	48
Abbildung 3: Ablauf der Datenselektion aus dem Gesamtpool (Voranalysen I-III) .....	49
Abbildung 4: Anzahl Patienten nach Polypharmazie-Gruppe in 2005-2007 (Voranalysen I-III) .....	50
Abbildung 5: Prozentuale Verteilung der Polypharmazie-Gruppen in 2005-2007 (Voranalysen I-III) .....	51
Abbildung 6: Prozentuale Verteilung der ermittelten Stichtage in 2005-2007 (Voranalysen I-III) .....	52
Abbildung 7: Mittlere Anzahl Krankenhauseinweisungen in 2008 nach Polypharmazie-Gruppe (Voranalysen I-III) .....	53
Abbildung 8: Anteil Patienten mit und ohne Krankenhauseinweisungen in 2008 nach Polypharmazie- Gruppe (Voranalysen I-III) .....	54
Abbildung 9: Anteil Patienten nach Art der Krankenhauseinweisung in 2008 und nach Polypharmazie-Gruppe (Voranalysen I-III) .....	55
Abbildung 10: Ablauf der Datenselektion aus dem Gesamtpool (Voranalysen IV-V) .....	60
Abbildung 11: Anzahl Patienten nach Polypharmazie-Gruppe bei variierenden Betrachtungszeiträumen und ATC/Quartal in 2005-2007 (Voranalysen I-V) .....	61
Abbildung 12: Mittlere Anzahl Krankenhauseinweisungen in 2008 nach Polypharmazie-Gruppe bei variierenden Beobachtungszeiträumen (Voranalysen IV-V) .....	62
Abbildung 13: Anteil Patienten mit und ohne Krankenhauseinweisung in 2008 nach Polypharmazie- Gruppe bei variierenden Beobachtungszeiträumen (Voranalysen IV-V) .....	63
Abbildung 14: Anteil Patienten nach Art der Krankenhauseinweisungen in 2008 bei variierenden Beobachtungszeiträumen (Voranalysen IV-V) .....	64
Abbildung 15: Anzahl Patienten mit Polypharmazie nach Alter und Geschlecht in 2007 (Voranalyse VI) .....	67
Abbildung 16: Anzahl ATC nach Verordnungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhauseinweisung in 2008 (Voranalyse VII) .....	69
Abbildung 17: Auffälligste ATC nach Verordnungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhauseinweisung in 2008 (Voranalyse VII) .....	70
Abbildung 18: Anzahl ICD nach Codierungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhauseinweisung in 2008 (Voranalyse VIII) .....	76
Abbildung 19: Auffälligste ICD nach Codierungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhausaufenthalt in 2008 (Voranalyse VIII) .....	77

Abbildung 20: Darstellung der Polypharmazie-Ermittlung mit einheitlichem Stichtag 31.12.2007 (Datenbasis 2007).....	84
Abbildung 21: Ablauf der Datenzusammenstellung (Entwicklungsdatensatz: 2005-2008).....	87
Abbildung 22: Mittelwert des Zielmerkmals in 2008 nach Altersgruppe zum 31.12.2007.....	94
Abbildung 23: Mittelwert des Zielmerkmals in 2008 nach Bundesland zum 31.12.2007 .....	95
Abbildung 24: Mittelwert des Zielmerkmals in 2008 nach Wohnort West/Ost zum 31.12.2007 .....	96
Abbildung 25: Mittelwert des Zielmerkmals in 2008 nach Anzahl FAVE zum 31.12.2007.....	97
Abbildung 26: Mittelwert des Zielmerkmals in 2008 nach Berufsgruppe zum 31.12.2007 .....	98
Abbildung 27: Mittelwert des Zielmerkmals in 2008 nach Polypharmazie-Gruppe in 2007 .....	100
Abbildung 28: Mittelwert des Zielmerkmals in 2008 nach DDD-Gruppe in 2007 .....	101
Abbildung 29: Allgemeine Formel einer logistischen Funktion (Garbade, 2009, S. 41).....	117
Abbildung 30: Regressionsformel zum finalen Modell (Datenbasis: 2005-2008) .....	118
Abbildung 31: Finaler Algorithmus zur Patientenidentifizierung .....	119
Abbildung 32: ROC-Kurve und AUC des Modells (Datenbasis 2007-2010).....	124
Abbildung 33: Vorhersagewahrscheinlichkeit für 2010 nach Geschlecht .....	126
Abbildung 34: Vorhersagewahrscheinlichkeit für 2010 nach Alter zum 31.12.2009 .....	127
Abbildung 35: Vorhersagewahrscheinlichkeit für 2010 nach Anzahl ATC (2009).....	128
Abbildung 36: Vorhersagewahrscheinlichkeit für 2010 nach Polypharmazie-Gruppe (2009).....	129
Abbildung 37: Vergleich von Vorhersage und tatsächlich eingetretenem Zielmerkmal in 2010 .....	130
Abbildung 38: Häufigkeitsprozent der Vorhersagewahrscheinlichkeiten für 2010 (Datenbasis Validierungsdatensatz: 2007-2010) .....	132
Abbildung 39: Kumulierte Häufigkeitsprozent der Vorhersagewahrscheinlichkeiten für 2010 (Datenbasis Validierungsdatensatz: 2007-2010).....	133



## Verzeichnis der Tabellen

Tabelle 1: Vergleich von Definitionen und Merkmalen der Polypharmazie (Mertens, 2009, S. 31, modifiziert) .....	25
Tabelle 2: Auffällige Arzneistoffe im Zusammenhang mit ABH in der Literatur .....	28
Tabelle 3: Auffällige Diagnosen im Zusammenhang mit ABH in der Literatur .....	29
Tabelle 4: Vergleich von Primär- versus Sekundärdaten (Anderson & Garbe, 2008) .....	32
Tabelle 5: Einteilung Strukturen prüfender Verfahren (Backhaus, 2008) .....	38
Tabelle 6: Gebildete ATC-Merkmale für Patienten mit Polypharmazie und Krankenhauseinweisungen (Voranalyse VII) .....	72
Tabelle 7: Vergleich auffälliger ATC aus der Literatur und in Voranalyse VII .....	73
Tabelle 8: Gebildete ICD-Merkmale für Patienten mit Polypharmazie und Krankenhausaufenthalten (Voranalyse VIII) .....	79
Tabelle 9: Vergleich auffälliger ICD aus der Literatur und in Voranalyse VIII .....	80
Tabelle 10: Ausprägung der Krankenhauseinweisungen in den Gruppe A und B (Voranalyse IX) .....	84
Tabelle 11: Ergebnis der Prüfung auf Signifikanz (Voranalyse IX) .....	85
Tabelle 12: Ausgewählte GKV-Routinedaten für die Modellierung .....	86
Tabelle 13: Verteilung der potenziellen Prädiktoren im Entwicklungsdatensatz (Datenbasis 2005-2008) .....	92
Tabelle 14: Vergleich von Voranalysen und Entwicklungsdatensatz (Datenbasis 2005-2008) .....	93
Tabelle 15: Korrelationen zwischen metrischen Merkmalen und Zielmerkmal (Datenbasis 2005-2008: Entwicklungsdatensatz) .....	103
Tabelle 16: Korrelationen zwischen kategorialen Merkmalen und Zielmerkmal (Datenbasis 2005-2008: Entwicklungsdatensatz) .....	105
Tabelle 17: Übersicht über die Variablen für die Modellierung am Entwicklungsdatensatz .....	109
Tabelle 18: Übersicht über erstellte und getestete Modelle (Datenbasis 2005-2008) .....	111
Tabelle 19: Korrelationskoeffizienten fachlicher relevanter ATC-Kombinationen (Datenbasis 2005-2008: Entwicklungsdatensatz) .....	113
Tabelle 20: Prüfung auf Korrelationen zwischen schwer interpretierbaren Variablen (Datenbasis 2005-2008: Entwicklungsdatensatz) .....	114
Tabelle 21: Übersicht der Variablen im finalen Modell (Datenbasis 2005-2008: Entwicklungsdatensatz) .....	116
Tabelle 22: Klassifikationsmatrix für den Schwellenwert $PRE > 0,50$ und Modellanwendung am Validierungsdatensatz (2007-2010) .....	134

Tabelle 23: Klassifikationsmatrix für den Schwellenwert  $PRE > 0,69$  und Modellanwendung am Validierungsdatensatz (2007-2010) ..... 135

Tabelle 24: Klassifikationsmatrix für den Schwellenwert  $> 13$  ATC/Quartal in 2007 und Ermittlung des Zielmerkmals in 2010 (Datenbasis Validierungsdatensatz: 2007-2010)..... 135

# 1 Einführung

## 1.1 Wissenschaftlicher Hintergrund

Arzneimittel zählen „*bei adäquater Anwendung zu den wirksamsten und effizientesten Instrumenten ärztlicher Hilfe*“ (SVR, 2009; S. 43). Gleichzeitig bestehen bei jeder Arzneimittelanwendung Risiken durch **unerwünschte Arzneimittelereignisse** (UAE). Dabei kann zwischen arzneimittel- und anwenderbedingten Ereignissen unterschieden werden (Edwards & Aronson, 2000; Pintor-Marmol et al., 2012). Das heißt, nicht nur durch unerwünschte Arzneimittelwirkungen (UAW), sondern auch durch Medikationsfehler des Anwenders kann es zu UAE kommen. Tendenziell steigt das Risiko allein dadurch, dass Patienten durchschnittlich immer älter werden. **Ältere Menschen** sind anfälliger für UAE (Frölich, 2000). Sie haben nicht nur häufiger Anwendungs- und Verständnisprobleme beim Umgang mit Arzneistoffen, sondern weisen auch abnehmende Stoffwechsel- und Organleistungen auf. Insbesondere Leber und Nieren sind von zentraler Bedeutung bei der Metabolisierung von Arzneistoffen. Aufnahme, Einbau, Umbau und Elimination von Wirkstoffen im Körper sind beim alten Menschen verändert (Wehling & Peiter, 2003; Mangoni & Jackson, 2003; Turnheim, 2004). Durch Muskelabbau und Gewichtsverlust im Alter verändert sich zudem das Verhältnis von hydrophilen und lipophilen Kompartimenten und damit die Verteilung von Arzneistoffen im Körper. Das hat Einfluss auf die Arzneimittelwirksamkeit und macht diese schwerer kalkulierbar. Hinzu kommt, dass Kompensationsmechanismen im Alter schlechter funktionieren und auftretende UAE oft nicht mehr ausgleichen können. So ist das Risiko für UAW bei 70- bis 80-jährigen Patienten 4- bis 5-mal höher als bei jüngeren Menschen (Frölich, 2000).

Mit zunehmendem Alter nimmt Multimorbidität zu und führt häufig zu **Polypharmazie** (Beyer et al., 2007; Glaeske et al., 2008). Beides sind neben einem höheren Alter weitere wichtige Risikofaktoren für arzneimittelbedingte UAE (Mühlberg et al., 1999; Mühlberg, 2002; BMFSFJ, 2010). Polypharmazie kann, muss aber keine Folge von Multimorbidität sein (Glaeske, 2011). Studien aus deutschen Altenheimen zeigen, dass Polypharmazie für das Auftreten von Nebenwirkungen relevant ist (Thürmann et al., 2007; Schmiedl et al., 2010; Jaehde & Thürmann, 2012). Als Polypharmazie beziehungsweise Multimedikation, wird die Therapie mit mehreren Wirkstoffen gleich-

zeitig bezeichnet, wobei die existierenden Definitionen uneinheitlich sind. In der Literatur variiert die Anzahl der Wirkstoffe, ab der von einer Polypharmazie gesprochen wird, beispielsweise zwischen 2 (Veehof et al., 2000) und 13 Wirkstoffen (Foster et al., 2005). Bekannte **Studien zur Polypharmazie** betrachten überwiegend die sogenannte „*kumulative Polypharmazie*“ (Mukhtar, 2010). Dabei wird Polypharmazie über einen definierten, längeren Zeitraum betrachtet. Studien zu „*simultaner Polypharmazie*“ mit Stichtagsbetrachtung oder zu einer „*kontinuierlichen Polypharmazie*“, mit Betrachtung von mindestens zwei Zeiträumen, sind seltener. Die meisten Studien zur Polypharmazie werden in Skandinavien durchgeführt (36%), gefolgt von Nordamerika (22%). Lediglich etwa 8% der Studien kommen aus Deutschland, wobei ein Drittel davon Routinedaten der Gesetzlichen Krankenversicherungen (GKV) als Datenbasis nutzen. Insgesamt handelt es sich bei den Studien zu 80% um Querschnittsstudien und zu 20% um Kohortenstudien (Mukhtar, 2010). Neben Routinedaten werden Primärdaten, die beispielsweise in Patienteninterviews oder mittels Fragebögen erhoben werden, untersucht. Im Mittel (Median) werden auf diese Weise Studienpopulationen in der Größenordnung von 600 bis 8.000 Patienten erreicht. Größere Studien erfolgen in der Regel unter Verwendung von Routinedaten (Mukhtar, 2010). Endpunkte für solche Studien sind häufig der Umfang der Leistungsanspruchnahme oder Morbidität und Mortalität. Untersuchungen mit Routinedaten ergeben in der Gesamtbevölkerung eine **Prävalenz** für Polypharmazie von 10%. Diese steigt im stationären Bereich auf 27%, bei Einschränkung auf Ältere auf 48% und bei Älteren im Krankenhaus auf 59%. Die ermittelte Prävalenz für Polypharmazie aus Primärdaten liegt höher, im Mittel bei 33% (Mukhtar, 2010). Wegen der uneinheitlichen Definitionen und Bezugsgrößen schwanken die Angaben zur Prävalenz der Polypharmazie. Polypharmazie tritt überwiegend bei Patienten höheren Alters auf (Mertens, 2009). Etwa jeder vierte Patient über 65 Jahre in Deutschland erhält 5 oder mehr Arzneistoffe gleichzeitig (WIdO, 2012). Bei 42% der Patienten über 65 Jahre besteht eine kumulative Polypharmazie über 3 Monate (Thürmann et al., 2012; Jaehde & Thürmann, 2012). **Ursachen** für die Entstehung von Polypharmazie sind beispielsweise die Behandlung durch verschiedene Therapeuten bei unzureichender Kommunikation und fehlendem Gesamtkonzept (Boyd et al., 2005; Glaeske, 2011). Dieses sind auch Folgen eines dezentralen Gesundheitssystems, wie es in Deutschland existiert und das beispielsweise „Ärzte-Hopping“ zulässt. Für das deutsche Gesundheitswesen sind vielfältige Koordinationsdefizite beschrieben (SVR, 2009). Bei schlechter Thera-

piekontrolle verstärkt der Leidensdruck der Patienten oft das Bedürfnis nach weiteren Verordnungen durch den Arzt zusätzlich, was Polypharmazie steigern kann. Unter Älteren ist zudem Selbstmedikation mit „*over the counter*“-Arzneimitteln (OTC) weit verbreitet und verläuft weitestgehend unkontrolliert (Albert et al., 2014). Durch auftretende UAE kann eine **Verschreibungskaskade** ausgelöst werden (Rochon & Gurwitz, 1997). Polypharmazie an sich muss nicht zwangsläufig ein Risiko darstellen, kann sich jedoch oft in vielfältiger Weise negativ auswirken. Mit steigender Anzahl von gleichzeitig angewendeten Arzneistoffen wird von einem Anstieg an UAE ausgegangen (Iniesta-Navalon et al., 2011). Die Zunahme von UAE lässt sich bereits durch den mathematischen Zusammenhang zwischen der Anzahl theoretisch möglicher Wechselwirkungen (*WW*) und der Anzahl (*n*) an Arzneistoffen beschreiben:  $Anzahl\ WW = ((n^2 - n) / 2)$ . Mit zunehmender Zahl an Arzneistoffen steigt die Zahl der möglichen Wirkstoffpaare, unter denen es zu einer Wechselwirkung kommen kann, exponentiell an. So sind bei 5, 10 beziehungsweise 15 gleichzeitig verwendeten Arzneistoffen 10, 45 beziehungsweise 105 mögliche Wirkstoffpaare denkbar, zwischen denen es zu Wechselwirkungen kommen kann. In der Literatur variiert, welche Anzahl an gleichzeitig eingenommenen Wirkstoffen eine Polypharmazie definiert und mit welchen Folgen dieses verbunden ist (Mertens, 2009; Schneeweiss et al., 2002; Defort, 2011; Foster et al., 2005; Glaeske & Janhsen, 2006).

Auftretende UAE scheinen überwiegend wirkstoffbedingt zu sein (Schneeweiss et al., 2002). Es gibt einzelne oder Gruppen von Arzneistoffen, für die häufige oder schwerwiegende UAW beobachtet werden und die man unter diesem Aspekt als **Risikoarzneistoffe** bezeichnen kann. Zu nennen wären hier beispielsweise Insulin und Antithrombotika. Untersuchungen, beispielsweise auf Basis von Pharmakovigilanz-Daten, zeigen, welche Arzneistoffe davon besonders oft in Zusammenhang mit arzneimittelbedingten Krankenhauseinweisungen standen (Keller, 2006; Schneeweiss et al., 2002). Weitere Untersuchungen benennen Risikoarzneistoffe im ambulanten Bereich (Foster et al., 2005; SVR, 2009; BMFSFJ, 2010), in höheren Altersgruppen (Fick et al., 2003; Beers, 1997; SVR, 2009 S. 474; Holt et al., 2010) und in Kombinationen (Picksak, 2009). Welche und wie viele dieser Risikoarzneistoffe in Verbindung mit Polypharmazie konkret auftreten, scheint nicht systematisch untersucht. Zu vermuten wäre jedoch, dass mit steigender Anzahl an Wirkstoffen und zunehmender Multimorbidität auch der Einsatz der genannten Risikoarzneistoffe zunimmt.

Mit zunehmender Komplexität der Arzneimitteltherapie leiden die Übersichtlichkeit

und die korrekte Umsetzung der Therapie durch den Patienten (Gallagher & O'Mahony, 2011). Die Wahrscheinlichkeit von **Medikationsfehlern** steigt. Als Medikationsfehler gelten UAE, die „*vermeidbar sind und zu einer unangemessenen Medikationsverwendung führen oder diese verursachen oder zu einer Schädigung des Patienten führen*“ (NCC MERP, 2013; übersetzt aus dem Englischen). Eine sorgfältige Arzneimittelberatung durch den behandelnden Arzt oder anderer Fachberufe im Gesundheitswesen ist essentiell und von zunehmender Bedeutung. Gleichzeitig ist in Deutschland die Kontaktzeit zwischen Arzt und Patient mit durchschnittlich weniger als 15 Minuten sehr kurz für eine angemessene Patientenbetreuung und die Vorbeugung von UAE durch eine entsprechende Beratung (Yarnall et al., 2003; SVR, 2009). Erschwerend kommt hinzu, dass Arztanweisungen, wenn sie erfolgen, von Patienten oftmals nicht eingehalten werden. Die Arzneimittelcompliance beziehungsweise –adhärenz wird mit bis zu 50% als sehr niedrig angenommen (Haynes et al., 2001; Sackett et al., 2003; SVR, 2002). Das Bewusstsein für Arzneimittelrisiken in der Bevölkerung ist demnach gering. Bei Älteren ist ein geradezu sorgloses Umgehen mit Arzneimitteln zu erkennen (BMFSFJ, 2010). So halten sich von den älteren Patienten, die von Polypharmazie und UAE besonders betroffen sind, nur 20% für gefährdet und 60% sorgen sich nicht um Nebenwirkungen, obwohl ein steigendes Risiko mit steigender Zahl gleichzeitig angewendeter Arzneistoffe bekannt ist (WIdO, 2012). Folglich überrascht nicht, dass die Prävalenz von UAE hoch ist. Sowohl im ambulanten, als auch im stationären Bereich kommt es in zirka 5% aller Fälle zu UAE (Gurwitz et al., 2003; Schnurrer & Frölich, 2003). Etwa ebenso häufig sind mit etwa 3 bis 6% **arzneimittelbedingte Hospitalisationen (ABH)** (Muehlberger et al., 1997; Roughead et al., 1998; Pirmohamed et al., 2004). Der „*Patient Safety Report*“ der WHO spricht sogar von bis zu 10% UAE im Rahmen von Krankenhauseinweisungen, insbesondere in weniger entwickelten Ländern (WHO, 2009). Für einen durchschnittlichen geriatrischen Patienten liegen die Schätzungen für UAE in der Krankenhausumgebung mit 12% etwa doppelt so hoch wie für die Gesamtbevölkerung und sogar tödliche Arzneimittelereignisse kommen mit etwa 0,2% bis 1% bedenklich häufig vor (Lazarou et al., 1998; Ebbesen et al., 2001; Schneeweiss et al., 2002). Diese und andere Untersuchungen lassen dabei erkennen, dass ein höheres Alter, Multimorbidität und Polypharmazie mit Krankenhauseinweisungen häufiger in Zusammenhang stehen (Varallo et al., 2014). Die Tatsache, dass es für Deutschland hochgerechnet im Vergleich zu jährlich etwa 5.000 Toten durch Verkehrsunfälle, schät-

zungsweise mindestens 10.000 bis 60.000 Tote durch UAE gibt, verdeutlicht die Relevanz der Thematik (Schnurrer & Frölich, 2003). Bedeutsam ist, dass unabhängig vom Setting, ambulant oder stationär, ein hoher Anteil der UAE, etwa ein Drittel bis die Hälfte, vermeidbar scheint (Lazarou et al., 1998; Gurwitz et al., 2003; Roughead et al., 1998).

Eine **klinisch-pharmazeutische Betreuung** kann die Einhaltung von ärztlichen Anweisungen und die Patientenmotivation verbessern. Es können UAE vermieden und ABH reduziert werden (Rothschild et al., 2010). Durch eine klinisch-pharmazeutische Betreuung bei geriatrischen Patienten in einem Krankenhaus sanken innerhalb eines Jahres weitere Einweisungen insgesamt um 16% und ABH um 80% (Gillespie et al., 2009). Auch in Deutschland werden Klinische Pharmazeuten als kompetente Arzneimittelexperten betrachtet, die bei der Arzneimitteltherapie stärker in die Verantwortung einbezogen werden sollten (SVR, 2009). Demzufolge liegt es nahe, UAE und damit ABH oder gar Todesfälle durch gezielte klinisch-pharmazeutische Betreuung von Risikopatienten, wie Patienten mit Polypharmazie, entgegenzuwirken. Wirksame Strategien zur Verbesserung der Arzneimitteltherapiesicherheit (AMTS) bei Patienten mit Polypharmazie sind nötig. Die Leitliniengruppe Hessen führt in ihrer „*Leitlinien Multimedikation*“ unter anderem an, arzneimittelbedingte Probleme bei Einnahme von mehr als 5 Wirkstoffen und bei bestimmten Risikoarzneistoffen zu erwarten (LGH et al., 2013). Patienten mit Polypharmazie weisen neben der komplexen Arzneimitteltherapie mit einem Risiko für UAE und Hospitalisierungen insgesamt auch eine intensive Leistungsanspruchnahme und daher vermutlich Wirtschaftlichkeitsreserven auf. Damit ist Polypharmazie auch in pharmakoökonomischer Hinsicht ein Indikator für eine Patientengruppe mit einem potenziellen Bedarf an klinisch-pharmazeutischer Betreuung. Patienten mit Polypharmazie sind daher die Zielgruppe dieser Untersuchung. Eine weiterführende, differenziertere Identifizierung von Patienten mit Polypharmazie und einem Betreuungsbedarfs wird im Rahmen dieses Promotionsvorhabens angestrebt. **Ziele der klinisch-pharmazeutischen Betreuung von Patienten mit Polypharmazie** sollte nach Erkennung eines Betreuungsbedarfs dann die Reduktion von Risikokonstellationen im Zusammenhang mit Risikoarzneistoffen und die daraus potenziell resultierenden ABH sein. Dazu sollten ermittelte Patienten zu ihrer Arzneimitteltherapie informiert und motiviert werden, die korrekte Arzneimittelanwendung üben und mit Hilfe einer schriftlich fixierten Medikamentenliste einen Überblick über die individuelle Nutzen-Risiko-Bewertung

ihrer Arzneimitteltherapie erhalten.

Bisherige Studien zu Risikoreduktion bei Polypharmazie in Deutschland, insbesondere im ambulanten Setting, sind allerdings rar. Ein Grund dafür ist, dass die Zusammenführung der nötigen Informationen für die komplexe Untersuchung von UAE im ambulanten Sektor aufwändig ist. In der GKV stehen zu Abrechnungszwecken routinemäßig große Mengen an patientenbezogenen Daten zu Arzneimitteln, Diagnosen, therapeutischen Prozeduren und anderen Gesundheitsinformationen zur Verfügung. In der vorliegenden Arbeit sollen die Möglichkeiten der GKV-Routinedaten daher bestmöglich genutzt werden, um für betroffene Patienten mit Polypharmazie arzneimittelbezogene Strategien zur Risikoreduktion zu ermitteln. Die häufigere Nutzung von Routinedaten zu Auswertungszwecken wird immer wieder von verschiedenen Autoren gefordert (Hasford et al., 2004; Glaeske & Janhsen, 2006; Hoffmann, 2008; Glaeske et al., 2010). Als inhaltliche Schwerpunkte der zukünftigen Forschung zur AMTS werden die „*Nutzung systematisch erhobener Daten für die AMTS*“ und Untersuchungen zur „*AMTS bei [...] bestimmten Patientengruppen*“ explizit genannt (Koordinierungsgruppe AMTS, 2011, S. 1).

## **1.2 Nutzen aus der Perspektive einer Gesetzlichen Krankenkasse**

Die vorliegende Untersuchung erfolgt aus der Perspektive der GKV. Der Gesetzgeber fordert im §12 des fünften Sozialgesetzbuches (SGB-V) von der GKV die Berücksichtigung einer wirtschaftlichen, ausreichenden, notwendigen und zweckmäßigen Versorgung. In Deutschland wird neben Fehlversorgung jedoch auch Über- und Unterversorgung mit Arzneimitteln vermutet (SVR, 2002). Zahlreiche weitere Gesetze wurden daher erlassen, um Verbesserungen in der Gesundheitsversorgung weiter voranzutreiben (GMG, 2003; AVWG, 2006; GKV-WSG, 2007). Die Gesetzlichen Krankenversicherungen sind dadurch stetig dazu angehalten, Maßnahmen zur Versorgungsoptimierung zu entwickeln. Neben den Qualitäts- und Sicherheitsaspekten zum Wohl der Patienten ist dieses für eine Krankenversicherung auch aus ökonomischen Gründen und für die Positionierung im Wettbewerb von Nöten.

Ein Bedarf an einer Verbesserung der AMTS in Deutschland findet auch Ausdruck in den seit 2007 regelmäßig aufgestellten Aktionsplänen der Bundesregierung (BMG, 2007; BMG, 2010; BMG 2013). Bereits 2003 und 2007 hatte der Sachverständigen-



rat zur Begutachtung der Entwicklung im Gesundheitswesen dieses Thema aufgegriffen und Maßnahmen gefordert (SVR, 2003; SVR 2007). Anlass dazu gaben Berichte zur AMTS-Problematik aus dem Ausland (IOM, 1999; DOH, 2000). Die Aktionspläne 2010-12 und 2013-2015 zur Verbesserung der AMTS in Deutschland nennen als eine Maßnahme explizit die Sensibilisierung von Patienten für die Risiken der Arzneimitteltherapie (BMG, 2010; BMG, 2013). Gleichzeitig verstehen sich viele Gesetzliche Krankenversicherungen zunehmend als „Anwälte der Patienten“ und streben danach, vom „Payer zum Player“ zu werden. In diesem Sinne bietet sich für Krankenversicherungen ein neues Betätigungsfeld mit dem Ziel, die Verbesserung der AMTS zu fördern. Einige Krankenversicherungen sind bereits zu diesem Thema aktiv geworden und bieten ihren Versicherten in verschiedener Form entsprechende Arzneimittelinformationen. In manchen Fällen wird auch der Zugang zu einer Arzneimittelberatung ermöglicht. Für die bedarfsgerechte Gestaltung solcher Projekte ist das Wissen über prioritäre Arzneimittelrisiken und gefährdete Zielgruppen wichtig.

Versicherte mit Polypharmazie stehen aus mehreren Gründen besonders im Fokus solcher Maßnahmen. Diese Versicherten bekommen viele und oftmals auch unnötige Arzneimittel verordnet. Gleichzeitig stellen die Arzneimittelkosten in der GKV neben den Ausgaben für Krankenhausversorgung und Arzthonorare einen der größten Ausgabenblöcke dar. Zusätzlich kommen bei Patienten mit Polypharmazie neben den hohen Kosten auch hohe Risiken wegen der Komplexität der Therapie und damit der Gefahr von UAE zusammen.

### **1.3 Zielsetzung und Vorgehensweise**

Das Promotionsvorhaben findet im Umfeld einer Gesetzlichen Krankenversicherung statt, die seit Jahren ein telefonisches Gesundheitscoaching für chronisch Kranke anbietet. Im Rahmen einer Datenauswertung sind Versicherte mit Polypharmazie aufgefallen, für die ein Gesundheitscoaching zum Umgang mit Arzneimitteln aus klinisch-pharmazeutischer Perspektive heraus aufgebaut werden soll. Ziel des Promotionsvorhabens ist es, eine Vorgehensweise zu entwickeln, auf Basis von GKV-Routinedaten Patienten mit Polypharmazie und einem Bedarf an klinisch-pharmazeutischer Betreuung automatisiert zu erkennen. Diese Patienten sollen anschließend daran einem Arzneimittelexperten zur Einschätzung der Medikation zuge-

leitet werden und neben allgemeingültigen auch zu individuellen Risiken betreut werden. Für die Untersuchung sind folgende fachliche und methodische Fragen zu klären:

- I. Lassen sich Patienten mit Polypharmazie und einem erhöhtem Bedarf an klinisch-pharmazeutischer Betreuung aus GKV-Routinedaten erkennen?
  - wie ist Polypharmazie zu definieren?
  - welche Aspekte kennzeichnen einen Betreuungsbedarf?
  
- II. Wie kann dieser Bedarf erkannt werden?
  - sind GKV-Routinedaten für die Erkennung ausreichend?
  - sind Data-Mining-Methoden geeignet?
  - wie zuverlässig gelingt die Erkennung?
  
- III. Welche Einflussfaktoren und welche Wirkstoffe spielen eine Rolle?

Zur Beantwortung dieser Fragen werden Sekundärdaten der betroffenen, bundesweit agierenden Ersatzkasse zugrunde gelegt. Da große Datenmengen zu untersuchen sind, soll die Untersuchung methodisch einem Data-Mining-Prozess folgen. Für die Analyse wird ein etablierter Data-Mining Prozess herangezogen (Chapmann et al., 2000). Die vorliegende Arbeit gliedert sich in Anlehnung daran in die Phasen: 1. Business Understanding, 2. Data Understanding, 3. Data Preparation, 4. Modeling, 5. Evaluation und 6. Deployment:

Einleitend werden in der *Business Understanding-Phase* die zugrunde gelegten Begrifflichkeiten und das fachlich zu bestimmende Zielmerkmal zur Kennzeichnung eines klinisch-pharmazeutischen Betreuungsbedarfs erläutert. Charakteristika der GKV-Routinedaten werden diskutiert, und die Wahl der logistischen Regressionsanalyse wird begründet. Die Durchführung der Untersuchung beginnt mit der *Data Understanding-Phase*, in der die ausgewählten Daten in Voranalysen getestet werden. Im Abschnitt *Data Preparation* wird die Zusammenstellung des Entwicklungssatzes erläutert, der letztendlich für die Modellbildung zum Einsatz kommen soll. Beim *Modeling* erfolgt die eigentliche Modellbildung zur Erkennung eines Bedarfs an klinisch-pharmazeutischer Betreuung durch Testung verschiedener Variablen-Sets. Das Ergebnis der Modellierung wird in der *Evaluations-Phase* an einem Validie-

rungsdatensatz geprüft und der Analyseprozess beurteilt. Abschließend wird das Gesamtergebnis der Arbeit zusammengefasst und unter Einbeziehung der Fragestellungen ein Fazit gezogen (*Deployment*). Außerdem wird ein Ausblick über notwendige Maßnahmen zur Implementierung und möglichen Verbesserung des entwickelten Modells gegeben.

Nicht Gegenstand der Untersuchung ist die Wirtschaftlichkeit von Arzneimitteln bei Patienten mit Polypharmazie im Sinne einer detaillierten Kosten-Nutzen Analyse. Es ist des Weiteren nicht Ziel des Promotionsvorhabens, alternative Strategien für eine Arzneimittelberatung durch den behandelnden Arzt zu suchen, sondern eine Methode zu entwickeln, die die Arzneimittelberatung sinnvoll ergänzen kann. Der Ausdruck „*klinisch-pharmazeutische Betreuung*“ wird oft quasi gleichbedeutend mit „Beratung zur *Arzneimitteltherapie*“ verwendet. Damit soll nicht ausgedrückt werden, dass eine Beratung zur Arzneimitteltherapie nur durch Klinischen Apotheker erfolgen kann, sondern der Fokus der Beratung soll dadurch exakter definiert werden (siehe Kap. 1.1.). Die Arzneimittelberatung durch den behandelnden Arzt ist grundlegend und wird vorausgesetzt. Eine ergänzende Beratung zu komplexen Arzneimitteltherapien bieten verschiedene Spezialisten im Gesundheitswesen an, insbesondere auch Klinische Pharmakologen und Geriater.

#### **1.4 Erwartete Limitationen**

GKV-Routinedaten sind Sekundärdaten und bilden vermutlich viele der zu betrachtenden Aspekte zur Bestimmung eines klinisch-pharmazeutischen Betreuungsbedarfs nicht ab. Dieses sind vielmehr Informationen, die für den primären Zweck von GKV-Routinedaten, der Abrechnung von Behandlungsleistungen, nicht benötigt und abgefragt werden. Dadurch werden Abstriche an der vollständigen Erkennung von klinisch-pharmazeutischem Betreuungsbedarf gemacht werden müssen. Zusätzlich sind Teile der Leistungsdaten von schlechterer Qualität oder werden beeinflusst von vielfältigen Störeffekten, wie beispielsweise Veränderungen im allgemeinen Versorgungsgeschehen. Die Vor- und Nachteile von GKV-Routinedaten werden im nachfolgenden Kapitel genauer beschrieben. Insgesamt kann die Verwendung von GKV-Routinedaten dazu führen, dass wegen der Nachteile der Datengrundlage falsche Schlussfolgerungen aus den Analyseergebnissen gezogen und Patienten selektiert werden, die in der Realität kein Optimierungspotenzial bei ihrer Arzneimitteltherapie

aufweisen.

Für Data-Mining werden große Datenmengen benötigt. Es wäre möglich, dass für interpretierbare Analyseergebnisse die Zielgruppe fachlich so eng definiert werden muss, dass eine zu geringe Anzahl an Patienten für eine Data-Mining Analyse resultiert. Gegebenenfalls gelingt es nicht, eine Zielgruppe oder charakteristische Merkmale dieser Gruppe so zu definieren, dass sich in der Data-Mining Analyse deutliche und fachlich erklärbare Zusammenhänge zum Zielmerkmal zeigen lassen. Zu viele andere Begleitumstände könnten die erwarteten Effekte zudem überlagern und die Erkennung erschweren. Gegebenenfalls kann weder die Erkennung von Versicherten noch die Konkretisierung von Handlungsempfehlungen ausreichend genau erfolgen, sodass aus der Analyse nur ein geringer praktischer Nutzen resultiert. Die erhaltene Gruppe an Patienten, für die ein Betreuungsbedarf schließlich erkannt werden kann, könnte zu klein sein für den Aufbau eines umfangreichen Betreuungskonzepts.

## 2 Fachliche Vorüberlegungen (Business Understanding)



### 2.1 Grundverständnis einer klinisch-pharmazeutischen Betreuung

Die Klinische Pharmazie ist ein Fachgebiet, das sich in Anlehnung an das amerikanische Vorbild aus den 60er Jahren in Deutschland gebildet und erst 2005 fest etabliert hat (Jaehde & Ammon, 1999; Jaehde, 2009). Wesentliches Element ist die starke Patientenzentrierung mit dem Ziel „aufbauend auf pharmazeutisch-naturwissenschaftlichen Kenntnissen die Optimierung der Arzneimittelanwendung am und durch den Patienten“ zu erreichen (Jaehde & Ammon, 1999; Beney, 2010; ACCP, 2008). Von der Klinischen Pharmazie abzugrenzen ist der jüngere Begriff der „klinisch-pharmazeutischen Betreuung“ (pharmaceutical care). Beide Begriffe werden uneinheitlich verwendet, wobei die Klinische Pharmazie eher das Tätigkeitsfeld an sich, die klinisch-pharmazeutische Betreuung eher die Handlungen bezogen auf den Patienten darstellt (Jorgensen, 2010). Die Handlungen werden umschrieben als Maßnahmen zum Erreichen von „patient outcomes“ (Hepler & Strand, 1990; ACCP,

2008) beziehungsweise zur Maximierung von Arzneimittelleffekten und Minimierung von Risiken und Kosten (Beney, 2010; Fachkommission Klinische Pharmazie der Bundesapothekerkammer, 2005). Die Handlung besteht in der Regel darin, dass ein Arzneimittelexperte eine Arzneimittelanamnese durchführt, eine patientenindividuelle Einschätzung der Arzneimittel und Risiken vornimmt sowie Vorsichtsmaßnahmen oder konkrete Verordnungsänderungen empfiehlt. Die einzelnen Aspekte der klinisch-pharmazeutischen Betreuung scheinen in unterschiedlichen Nationen unterschiedlich betont zu werden. Die europäische und deutsche Definition der klinisch-pharmazeutischen Betreuung wirkt beispielsweise durch die verwendeten Begrifflichkeiten stärker arzneimittel- statt patientenorientiert. Im Unterschied zum Wirtschaftlichkeitsgebot im §12 des SGB-V, in dem auch der Gesetzgeber eine Balance zwischen Kosten und Nutzen fordert, kommt bei der klinisch-pharmazeutischen Betreuung der Aspekt der Arzneimittelsicherheit hinzu. Seit der in den letzten Jahren international und national lauter werdenden Diskussion um Fehler in der Medizin scheint insgesamt ein stärkeres Bewusstsein für Arzneimittelfehler gewachsen zu sein (BMG, 2007; BMG, 2010). Bei der Bestimmung eines Bedarfs an klinisch-pharmazeutischer Betreuung sollen im Rahmen dieser Arbeit die Aspekte Wirksamkeit, Sicherheit und Wirtschaftlichkeit von Arzneimitteln möglichst gemeinsam auf Basis der GKV-Routinedaten betrachtet werden.

## **2.2 Auswahl des Zielmerkmals**

Für die angestrebte Untersuchung ist zur Erkennung eines Bedarfs an klinisch-pharmazeutischer Betreuung ein Zielmerkmal zu definieren, das näherungsweise als Indikator für einen Bedarf fungieren kann. Hiermit sollen Einflussfaktoren ermittelt werden, die einen Zusammenhang zum Zielmerkmal erkennen lassen und für eine Vorhersage eines Betreuungsbedarfs genutzt werden können.

Obwohl verschiedene Prüfalgorithmen, wie beispielsweise der *Medication Appropriateness Index* nach Hanlon (Hanlon, 1992; Brengnhoj, 2005; Cousins, 2009) und der Algorithmus nach Garfinkel (Garfinkel, 2007), zur Prüfung auf eine optimale Arzneimitteltherapie entwickelt wurden, würde ihre Anwendung durch die Vielzahl an Prüfschritten pro Arzneistoff einen hohen Komplexitätsgrad erreichen. Für die Beurteilung der therapeutischen Wirksamkeit eines Arzneimittels auf Patientenebene fehlen darüber hinaus die notwendigen klinischen Daten, wie objektive Laborparameter zum

Gesundheitszustand, in der verfügbaren Datengrundlage.

Aus dem umgekehrten Blickwinkel betrachtet sind Merkmale, die eine nicht ausreichende Wirksamkeit von Arzneimitteln oder gar ein Arzneimittelrisiko erkennen lassen, leichter zu identifizieren. Dazu bieten GKV-Routinedaten Anhaltspunkte (Foster et al., 2005; SVR, 2009; BMFSFJ, 2010). Bei mangelnder therapeutischer Wirksamkeit von Arzneimitteln ist beispielsweise mit einer höheren Leistungsanspruchnahme zu rechnen, die sich in verschiedenen Parametern ausdrücken kann. Ein Arzneimittelrisiko ist erkennbar bei offensichtlichem, nicht bestimmungsgemäßen Gebrauch von Arzneimitteln, also bei potenziellen Medikationsfehlern. Eine Prüfung auf grobe Abweichungen von Erwartungswerten, wie beispielsweise übermäßige Mengen an Arzneistoffen, gemessen in „*daily defined doses*“ (DDD), ist näherungsweise aus GKV-Daten möglich. UAW können zu Erkrankungen führen, die als Diagnosen wiederum in den Routinedaten codiert sind. Es wird zwar eine mangelhafte Dokumentation von UAE angenommen, aber vorhandene Daten sollten dennoch bestmöglich genutzt werden. Folgende Parameter aus den GKV-Leistungsdaten wären als Zielmerkmal denkbar:

Mögliche Indikatoren für eine mangelnde Wirksamkeit von Arzneimitteln:

- Anzahl der Arztkontakte je Quartal
- Anzahl verschiedener verordnender Ärzte
- Anzahl Krankenhauseinweisungen: allgemeine, notfallmäßige, internistische
- Krankenhausliegezeiten
- Höhe der Arzneimittel-, Krankenhaus- beziehungsweise Gesamtkosten
- Todesfall

Mögliche Indikatoren für ein Arzneimittelrisiko:

- Anzahl an Arzneistoffen (nach ATC), Verordnungen, Rezepten
- DDD/ATC pro Jahr deutlich über 365
- Doppelverordnung
- arzneimittelbedingte Erkrankungen (UAE)
- konkrete risikoreiche Arzneistoffkombinationen
- konkrete Risikoarzneistoffe
- konkrete Arzneistoffe mit vorliegender Kontraindikation

- arzneimittelbedingte Hospitalisationen (ABH)

Von den genannten Möglichkeiten soll das auszuwählende Zielmerkmal ein möglichst eindeutig zu bestimmender Parameter sein, der in guter Datenqualität vorliegt und ausreichend häufig im Datensatz codiert ist. Vorteilhaft wäre ein eigenständiger Patienten Outcome-Parameter, der für eine mangelnde Arzneimittelwirkung, ein Arzneimittelrisiko oder beides steht und zugleich kostenrelevant ist. Zudem sollte der Parameter ein Problem mit ausreichender klinischer Relevanz darstellen, der den Aufwand einer klinisch-pharmazeutischen Betreuung rechtfertigt.

Von den Indikatoren *Arztbesuche*, *Hospitalisationen* und *Todesfälle* sind vergleichend betrachtet die Arztbesuche schwer zu ermitteln und schätzungsweise Ereignisse mit geringerer klinischer Relevanz als Krankenhauseinweisungen der Fall. Todesfälle als Zielmerkmal dagegen bergen die Gefahr, Patienten zu finden, für die eine Beratung erst in einem schweren und gegebenenfalls zu späten Krankheitsstadium stattfindet. Daher erscheint die Hospitalisation als Zielmerkmal die sinnvollste Wahl zu sein.

Die potenziell arzneimittelbedingten Erkrankungen sind zwar fachlich gesehen die erste Wahl, lassen sich in den Routinedaten jedoch vermutlich nicht in ausreichender Zahl finden. Doppelverordnungen sind zwar theoretisch gut zu ermitteln, erweisen sich aber praktisch oft als weniger relevant. Vermeintliche Doppelverordnungen können sich als zeitlich versetzt eingenommene Arzneimittel herausstellen, beispielsweise morgens als Mono- und abends als Kombinationspräparat eingesetzt, um die Anzahl der einzunehmenden Medikamente pro Einnahmezeit zu verringern. Auch ist eine Unterscheidung zwischen Dauer- und Bedarfsmedikation als „top up“ nicht anhand der Daten möglich, beispielsweise bei Schmerzmedikamenten, und würde fälschlicherweise eine Doppelverordnung suggerieren. Konkrete einzelne Risikoarzneistoffe oder risikoreiche Arzneistoffkombinationen sind nicht eindeutig definiert und eignen sich daher eher als beschreibende Variable statt als Zielmerkmal. Ebenso stellt die Anzahl an Arzneistoffen allein nicht zwangsläufig ein Risiko dar. Eine Polypharmazie kann ohne Komplikationen verlaufen, während auch eine Monotherapie zu UAEs führen kann. Die Definition der Polypharmazie ist wiederum nicht eindeutig, sondern wäre festzulegen. Allerdings ist für Polypharmazie ein Zusammenhang zu erhöhter Morbidität und Krankenhauseinweisungen gezeigt worden, so dass es auch ein mögliches Zielmerkmal wäre. Eine Reduktion von Arzneistoffen ist laut Literatur

durch eine Betreuung möglich, jedoch nicht signifikant. Letzteres ist jedoch für die Anzahl DDDs der Fall (Pitkala et al., 2001).

Krankenhauseinweisungen als Zielmerkmal für die Analyse haben den Vorteil, dass sie direkt ermittelbar und eindeutig definiert sind sowie Arzneimittelrisiko beziehungsweise -therapieversagen und Kostenrelevanz widerspiegeln. Daher werden Krankenhauseinweisungen als Zielmerkmal ausgewählt. Voranalysen sollen klären, inwiefern sich potenziell arzneimittelbedingte oder generelle Krankenhauseinweisungen als Zielmerkmal besser eignen. Ein weiterer Vorteil ist, dass durch die anvisierte Betreuung durch Arzneimittelexperten Krankenhauseinweisungen leicht reduziert werden könnten (Krska et al., 2001).

### **2.3 Zielgruppe und ihre Charakteristika**

Zu einer ersten Eingrenzung der Zielgruppe mit einem Bedarf an klinisch-pharmazeutischer Betreuung werden Patienten mit Polypharmazie ausgewählt. Für die Datenanalyse ist der Begriff Polypharmazie eindeutig zu beschreiben. Dazu werden Definitionen verschiedener Autoren betrachtet (Tabelle 1). Die Anzahl der verschiedenen, gleichzeitig eingenommenen Wirkstoffe, ab der von einer Polypharmazie gesprochen wird, variiert und liegt beispielsweise bei 2 (Veehof et al., 2000; Knopf & Melchert, 2003), bei 5 (Thomas et al., 1999; Junius-Walker et al., 2006), bei 8 (Defort 2011) oder bei 13 Wirkstoffen (Foster et al., 2005). Manche Autoren unterscheiden zusätzlich zwischen „*minor*“, „*moderate*“ und „*major*“ Polypharmazie bei 2-3, 4-5 und >5 Arzneistoffen (Veehof et al., 2000) oder grenzen Polypharmazie ab 10 Wirkstoffen als „*high-risk*“ Polypharmazie ab. Mit 44% findet der Schwellenwert ab 5 Arzneistoffen die häufigste Verwendung, gefolgt von 16% mit einem Schwellenwert bei 6 Wirkstoffen (Mukhtar, 2010). In der nachfolgenden Tabelle sind verschiedene Definitionen für eine Polypharmazie aufgeführt und um Patientenmerkmale ergänzt, die Patienten mit Polypharmazie bei den jeweiligen Untersuchungen charakterisieren (Tabelle 1).



Nr.	Definition der Polypharmazie	Vermutlich relevante Patientenvariablen	Quelle
1	≥2 verordnete Wirkstoffe (Langzeitmedikation)	Alter, Geschlecht, Neuerkrankungen und Zustandsverschlechterung	Veehof et al., 2000
2	≥2 verordnete Wirkstoffe (inklusive Selbstmedikation)	Alter, Geschlecht, SÖS*, subjektive Gesundheitseinschätzung	Knopf & Melchert, 2003
3	≥3 verordnete Wirkstoffe		WHO, 2006;
4	≥3 verordnete Wirkstoffe (inklusive Selbstmedikation)		Brekke et al., 2006
5	≥5 verordnete Wirkstoffe (inklusive Selbstmedikation)	Alter, männliches Geschlecht, SÖS*, subjektive Gesundheitseinschätzung	Thomas et al., 1999
6	≥5 verordnete, regelmäßige, tägliche, unterschiedliche Wirkstoffe	Alter, Geschlecht, Haushaltsgröße/Familienstand, SÖS*, subjektive Gesundheitseinschätzung, Lebensqualität (IADL)	Junius-Walker et al., 2006
7	≥5 simultan gebrauchte Wirkstoffe	Alter, Geschlecht, Haushaltsgröße/Familienstand, SÖS*, subjektive Gesundheitseinschätzung, Lebensqualität (IADL)	Linjakumpu et al., 2002
8	≥8 verschiedene Wirkstoffe		Defort, 2011
9	≥13 verschiedene Wirkstoffe gleichzeitig		Foster et al., 2005

\* SÖS=sozioökonomischer Status

**Tabelle 1: Vergleich von Definitionen und Merkmalen der Polypharmazie (Mertens, 2009, S. 31, modifiziert)**

Des Weiteren bleibt oft unklar, welche Arzneistoffe bei der Zählung berücksichtigt werden. Die Betrachtung der regelmäßig einzunehmenden Dauermedikation von Patienten ist die Regel. In den überwiegenden Untersuchungen wird auch die zeitlich begrenzt einzunehmende Akutmedikation einbezogen (Mukhtar, 2010). Unklar ist, ob eine unregelmäßig benötigte Bedarfsmedikation darin inbegriffen ist. Manche Autoren schließen Arzneistoffe aus der **Selbstmedikation** explizit in ihre Polypharmazie-Definition ein (Thomas et al., 1999; Knopf & Melchert, 2003; Brekke et al., 2006). Bei Studien ist das jedoch nur etwa bei einem Drittel der Fall (Mukhtar, 2010). Selbstmedikation wird definiert als die „*Verwendung rezeptfreier Medikamente durch Men-*

*schen aufgrund ihrer eigenen Initiative“* (FIP & WSMI, 1999, S. 1). Verschreibungspflichtige Arzneistoffe, die auf Wunsch des Patienten vom Arzt auf Privat- oder dem sogenannten grünen Rezept verordnet werden und somit nicht zur klassischen Selbstmedikation zählen, fallen jedoch auch hier aus der Betrachtung heraus. Ein Beispiel dafür sind **PRISCUS-Medikamente** (Holt et al., 2010), wie Benzodiazepine, die daher häufig in GKV-Routinedaten nicht vollständig erfasst sind.

Dieses bedingt, dass die Definition der Polypharmazie in Abhängigkeit von der verfügbaren Datenbasis beziehungsweise der **Methodik** der Datenerhebung bei Untersuchungen entsprechend variiert. Doch auch innerhalb vergleichbarer Datensätze, wie beispielsweise bei den GKV-Routinedaten, gibt es keine standardisierte Definition, und es kommen verschiedene Vorgehensweisen zur Ermittlung von Polypharmazie zum Einsatz (Mukhtar, 2010). Bei Verwendung von **GKV-Routinedaten** wird häufig zur Vereinfachung die Anzahl an verschiedenen, gleichzeitig dokumentierten Arzneistoffen nach ATC pro Jahr aus den Rezeptdaten bestimmt (Defort, 2011). Hierbei ist fraglich, ob bei den in einem Jahr ermittelten Arzneistoffen von einer dauerhaften, gleichzeitigen Einnahme ausgegangen werden kann und ob der gewählte Zeitraum von einem Jahr zweckmäßig ist. Die Wahl des **Betrachtungszeitraumes** für die Ermittlung einer Polypharmazie variiert. Bei etwa der Hälfte der Studien werden 3 Monate, oft auch 12 Monate gewählt (Mukhtar, 2010). Polypharmazie kann zu einem Stichtag, über einen oder auch über zwei definierte Zeiträume betrachtet werden, was auch als simultane, kumulative beziehungsweise kontinuierliche Polypharmazie bezeichnet wird (Mukhtar, 2010). Oft wird dabei nicht unterschieden in eine neu entstandene oder eine bei Betrachtung bereits bestehende Polypharmazie, was die exakte Bestimmung von **Inzidenzen** für Polypharmazie erschwert.

Die Ermittlung einer Polypharmazie aus GKV-Routinedaten soll daher für die geplante Untersuchung auf einer eigenen, zu entwickelnden Definition basieren. Zur Festlegung einer geeigneten Mindestanzahl an Arzneistoffen in Bezug auf einen definierten Zeitraum werden Voranalysen durchgeführt. Die GKV-Routinedaten enthalten keine Daten zur Selbstmedikation, weswegen diese bei der Bestimmung einer Polypharmazie auf Basis von Routinedaten vernachlässigt werden muss. Eine Einbeziehung wäre aber grundsätzlich wünschenswert, da die Selbstmedikation mit bis zu 30% bis 40% einen großen Anteil der Medikation ausmacht (SVR, 2009 S. 466; Knopf & Melchert, 2003; Schulz, 2012; Albert et al., 2014) und tendenziell weiter zunimmt.

### *Charakteristika der Zielgruppe*

Patienten mit Polypharmazie werden durch unterschiedliche Autoren ähnliche charakteristische Merkmale zugeordnet. Ein fortgeschrittenes Alter wird durchweg als kennzeichnend genannt, teilweise konnte ein signifikanter Zusammenhang nachgewiesen werden (Knopf & Melchert, 2003). Als weitere Merkmale sind das weibliche Geschlecht, das Alleinleben und der Rentnerstatus beschrieben. Allerdings gibt es Untersuchungen, bei denen davon abweichend das männliche Geschlecht bei den untersuchten Patienten mit Polypharmazie überwog (Thomas et al., 1999). Der Einfluss der Aspekte Bildung und Einkommen wird in der Literatur uneinheitlich bewertet und insgesamt als gering eingeschätzt (Flores & Mengue, 2005; Knopf & Melchert, 2003; Mertens, 2009). Stadt- und Landbevölkerung werden ebenfalls im Zusammenhang mit Polypharmazie nicht als relevantes Unterscheidungsmerkmal bewertet (Mertens, 2009).

Die kennzeichnenden Merkmale höheres Alter, weibliches Geschlecht, allein lebend und Rentnerstatus decken sich mit Merkmalen für Patienten, bei denen vermehrt UAW und ABH beobachtet werden. Das Alter wird dabei mit 50-85 Jahre noch genauer eingegrenzt und ein Verhältnis von weiblichen zu männlichen Geschlecht von zirka 60:40 angegeben (Keller, 2006; Mertens, 2009; Schneeweiss et al., 2002).

Genauer in den Voranalysen zu untersuchen bleibt:

- Ab wie vielen Arzneistoffen kann von einer Polypharmazie mit einem erkennbar erhöhten, arzneimittelbedingten Risiko gesprochen werden?
- Welcher Betrachtungszeitraum sollte zugrunde gelegt werden?
- Welche Merkmale weisen Patienten mit Polypharmazie im Untersuchungsdatensatz auf?
- Lässt sich ein Zusammenhang zwischen einer Polypharmazie und der Häufigkeit von Krankenhauseinweisungen, dem Zielmerkmal, bestätigen?

## 2.4 Vorauswahl an GKV-Routinedaten

### Arzneimitteldaten

Aus der Rezeptabrechnung stehen in GKV-Routinedaten große Datenmengen zur Arzneimittelversorgung zur Verfügung. Arzneistoffe sind in diesen Daten nach dem Anatomisch-Therapeutisch-Chemischen Index, der sogenannten ATC-Systematik erfasst (WIdO, 2010). Rezeptdaten sollen bevorzugt für die Untersuchung genutzt werden. In der Tabelle sind relevante Arzneistoffe mit Bezug zu ABH und damit mögliche Indikatoren für Betreuungsbedarf zusammengestellt (Tabelle 2).

Nr.	ATC	Arzneistoff	ABH (Keller, 2006)	ABH (Schneeweiss et al., 2002)
1.	A10A	Insulin	X	X
2.	B01AA04	Phenprocoumon	X	(X)
3.	verschiedene	Antithrombotika	X	X
4.	C03	Diuretika	X	X
5.	C07AA07	Sotalol	X	
6.	B01AB, -C	TAH	X	(X)
7.	C01AA	Digitalisglykoside (Digoxin, Digitoxin)	X	X
8.	C07	Betablocker	X	X
9.	A10B	Orale Antidiabetika	X	X
10.	M01A	NSAR	X	X
11.	C08	Calcium-Antagonisten		X
12.	C09A+C	ACE-Hemmer/ Sartane		X
13.	C01DA	Nitrate		X
14.	H02	Corticosteroide, systemisch		X

**Tabelle 2: Auffällige Arzneistoffe im Zusammenhang mit ABH in der Literatur**

### Diagnosedaten

In den GKV-Routinedaten aus Arztpraxen und Krankenhäusern werden Diagnosen nach der „*Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme – German Modifikation*“, kurz der ICD-Systematik, codiert (z.B. DIMDI, 2010). Im Folgenden sind Diagnosen aufgeführt, die gehäuft im Zusammenhang mit UAW und ABH aufzutreten scheinen und Merkmale zur Erkennung eines Bedarfs an klinisch-pharmazeutischer Betreuung darstellen könnten (Tabelle 3).

Nr.	ICD	Diagnose (verkürzt)	ABH (Schneeweiss et al., 2002)	ABH (Keller, 2006)	ABH (Stausberg & Hast., 2010)
1.	I1	Hypertonie		X	
2.	I4	Arrhythmien	X	X	
3.	I2-5	Herzerkrankungen (KHK, HI)		X	
4.	I6-9	<b>Gefäßerkrankungen</b>	X	X	X
5.	K5	<b>Gastrointestinale Reaktionen</b>	X	X	X
6.	D	<b>Blutbildstörungen</b>	X	X	X
7.	T	Allergische Reaktionen		X	
8.	N	Nierenfunktionsstörungen		X	
9.	K	Leberfunktionsstörungen	X	X	
10.	F	psychische Reaktionen		X	X
11.	E	Stoffwechselstörungen	X	X	
12.	H	Sonstiges: Hör-/ Sehstörungen		X	
13.	J	Lungenerkrankungen		X	
14.	R	Allgemeinsymptome	X	X	
15.	T	Vergiftung		X	X
16.	A04	Enterokolitis (Clostridium difficile)			X
17.	L27	Dermatitis durch Arzneimittel		X	X
18.	Y57	Unerwünschte Nebenwirkung			X

**Tabelle 3: Auffällige Diagnosen im Zusammenhang mit ABH in der Literatur**

## 2.5 Public Health Aspekte

Die Deutsche Gesellschaft für Public Health (DGPH) definiert Public Health als „*die Wissenschaft und die Praxis zur Verhinderung von Krankheiten, zur Verlängerung des Lebens und zur Förderung von physischer und psychischer Gesundheit unter Berücksichtigung einer gerechten Verteilung und einer effizienten Nutzung der vorhandenen Ressourcen*“ (Winslow, 1920). Auf die Arzneimittelversorgung bezogen wird dieses interpretiert als Entwicklung und Umsetzung von praktischen Maßnahmen zur Verbesserung der Arzneimittelversorgung der Bevölkerung hinsichtlich der Arzneimittelwirksamkeit und Wirtschaftlichkeit. Hinzu kommt der Aspekt einer geeigneten Allokation von Ressourcen. In diesen Bestrebungen decken sich Ziele der klinisch-pharmazeutische Betreuung mit der Intention von Public Health. So findet sich neben den patientenbezogenen Dienstleistungen in der Definition der klinisch-pharmazeutische Betreuung auch die allgemeine Aufforderung zur „*Mitwirkung an Maßnahmen, die eine optimale Arzneimitteltherapie gewährleisten*“ (Fachkommission Klinische Pharmazie der Bundesapothekerkammer, 2005, S. 2).

Nach einer älteren Definition ist Public Health „*Theorie und Praxis der auf Gruppen beziehungsweise Bevölkerungen bezogenen Maßnahmen und Strategien zur Verminderung von Erkrankungs- und Sterbewahrscheinlichkeiten durch überwiegend nicht medizinische Interventionen zur Senkung von Gesundheitsbelastungen und zur Stärkung beziehungsweise Vermehrung von Gesundheitsressourcen*“ (Flick, 2002, S. 21). Das Ziel der vorliegenden Arbeit ist die frühzeitige Erkennung einer Patientengruppe mit arzneimittelbezogenen Problemen, um präventive Maßnahmen einleiten zu können. Als konkrete Handlungsempfehlung aus dem Ergebnis der Analyse wäre beispielsweise eine Antwort denkbar auf die Frage, um wie viele und welche Arzneistoffe eine Polypharmazie reduziert werden sollte, um eine ABH zu vermeiden. Damit will die geplante Untersuchung einen praktischen Beitrag zur Verbesserung der Arzneimittelversorgung in einer großen Gruppe der Bevölkerung leisten, und folgt einer Public Health Intention.

### 3 Methodische Vorgehensweise

Orientierend an verfügbaren Leitfäden für die Auswertung von Sekundärdaten, wie dem „*Memorandum III: Methoden für die Versorgungsforschung - Teil 1*“ (Pfaff et al., 2009) und „*Gute Praxis Sekundärdatenanalyse - GPS*“ (AGENS & Arbeitsgruppe Epidemiologische Methoden, 2008) wurde für die Untersuchung ein Analyseplan erstellt (Anlage 10.1). In dem Analyseplan sind die eingangs erläuterten Fragestellungen, Intention und Motivation zu der vorliegenden Untersuchung dargestellt sowie das ausgewählte Zielmerkmal angegeben. Das Untersuchungsdesign, die Datengrundlage, die ausgewählte Untersuchungsmethodik und geplanten Maßnahmen zur Qualitäts- und Ergebnissicherung werden nachfolgend erläutert und sind ebenfalls im Analyseplan zusammengefasst.

#### 3.1 Vor- und Nachteile von GKV-Routinedaten

Für die Erkennung eines Bedarfs an klinisch-pharmazeutischer Beratung sollen GKV-Routinedaten untersucht werden. Da die Routinedaten zum Zeitpunkt der Analyse bereits erfasst worden sind, stellt die vorliegende Untersuchung eine **retrospektive Datenbankanalyse mit Sekundärdaten** dar. Im Gegensatz zu Primärdaten sind Sekundärdaten (hier: Abrechnungsdaten) nicht gezielt für die Beantwortung einer bestimmten Fragestellung erhoben worden (AGENS & Arbeitsgruppe Epidemiologische Methoden, 2008), sondern werden nachträglich für einen anderen Zweck genutzt. Daraus ergeben sich für die Bestimmung eines klinisch-pharmazeutischen Betreuungsbedarfs Vor- und Nachteile. Die nachfolgende Tabelle zeigt beispielhaft die unterschiedlichen Inhalte von Primär- und Sekundärdaten (Anderson & Garbe, 2008) (Tabelle 4).

Nr.	Charakteristikum	Primärdaten (Feldstudie)	Sekundärdaten (Datenbankstudie)
1	Anzahl der eingeschlossenen Individuen	Eher niedrig	Hoch
2	Bestimmung der Inzidenzen sehr seltener Ereignisse möglich	Nein	Ja
3	Arzneimittelinformationen auch von schwerkranken Patienten verfügbar	Meist nein, evtl. vom behandelnden Arzt	Ja
4	Informationen über die Einnahme nicht verschreibungspflichtiger Arzneimittel	Ja	Nein
5	Recall-Bias möglich	Ja	Nein
6	Interviewer-Bias möglich	Ja	Nein
7	Selektionsbias durch Nonresponse (Verweigerung der Studienteilnahme)	Ja	Nein
8	Zeitbedarf vom Beginn der Studie bis Auswertung der Ergebnisse	Hoch	Gering
9	Kosten	Hoch	Gering
10	Bevölkerung-Repräsentativität	Gelegentlich	Häufig
11	Zustimmung der Patienten erforderlich	Ja	Nein
12	Zugang zu Originalunterlagen (Krankenakten, etc.)	Möglich, abhängig vom Studien-Design	Nein
13	Angabe zur Erkrankungsschwere	Ja	Selten
14	Informationen zu Rauchen, Alkoholggebrauch und anderen Lebensgewohnheiten	Ja	Nein
15	Laborwerte oder andere klinische Messgrößen	Ja	Nein
16	Informationen zur Patienten-Compliance	Prinzipiell ja	Nein
17	Verschreibungen im Krankenhaus	Möglich, abhängig vom Studien-Design	Meist nein

**Tabelle 4: Vergleich von Primär- versus Sekundärdaten (Anderson & Garbe, 2008)**



Ein Nachteil von GKV-Routinedaten ist, dass ambulante Diagnosedaten zwar mittlerweile intensiver bei der Vergütung von Leistungen genutzt werden und daher besser dokumentiert sind, die Datenqualität jedoch nunmehr durch eine exzessive Dokumentation weiterhin problematisch ist. Eine eingeschränkte Datenqualität ist bei der Datennutzung zu berücksichtigen, da es sonst zu Fehlinterpretationen kommen kann (Hoffmann, 2008). Durch angehäuften, nicht aktualisierte Codierungen von Diagnosen kann es beispielsweise zur Überschätzung oder durch fehlende oder fehlerhafte Codierung zur Unterschätzung der Krankheitsschwere kommen (Pfaff et al., 2009). Bei Auswertungen von Datensätzen über mehrere Jahre hinweg ist zu berücksichtigen, dass sich wesentliche Änderungen an den generellen Erfassungs- oder Abrechnungsprozessen in den Daten abbilden (Pfaff et al., 2009). Dieses kann die Interpretation von Auswertungsergebnissen erschweren. Es kann sich dabei um „regulatorische, gesetzliche, strukturelle und organisatorische Einflüsse“ handeln (Pfaff et al., 2009). Wo eine routinemäßige Datenerfassung nicht stattfindet, können keine Aussagen getroffen werden (Pfaff et al., 2009). Das ist beispielsweise der Fall bei der Arzneimittelversorgung im Rahmen der Selbstmedikation und bei pauschal vergüteten Krankenhausbehandlungen. Dieses ist nicht unerheblich, da insbesondere die Selbstmedikation tendenziell zunimmt. Für die Prüfung auf eine optimale Arzneimitteltherapie beziehungsweise eine klinisch-pharmazeutische Betreuung generell sind daher fallbezogene, klinische Informationen wichtig. Diese fehlen in den Routinedaten. Dazu gehören objektive Messwerte, wie Körpergröße, Körpergewicht und zur Bewertung von Arzneimittelwirkungen spezifische Laborparameter. Hinweise zum Gesundheitsverhalten der Patienten, wie beispielsweise der Arzneimittel-Compliance beziehungsweise der -Adhärenz, das tägliche Bewegungspensum, die Ernährungsgewohnheiten und der Umgang mit Alkohol- oder Nikotin sind ebenfalls nicht in den Routinedaten abgebildet. Oft bieten seltene, außergewöhnliche klinische Veränderungen oder deren zeitliche Abfolge im Zusammenhang mit einem Arzneimitteleinsatz die entscheidenden Informationen für die Interpretation von UAW (vgl. Naranjo Algorithmus zur Prüfung auf UAW, (Naranjo, 1981). Solche sind kaum in Routinedaten erfasst oder nachträglich erkennbar, sondern selbst bei kontinuierlicher, direkter Patientenbetreuung und gezielter Erfassungsabsicht schwierig zu identifizieren.

Vorteile der routinemäßig erfassten GKV-Leistungsdaten sind, dass sie für hohe Patientenzahlen und in großem Umfang vorliegen, leicht zugänglich sind und Informationen aus mehreren, verschiedenen Versorgungssektoren enthalten (Pfaff et al.,

2009). Letzteres erlaubt sektorenübergreifende Datenanalysen. Die leichte Zugänglichkeit spart Zeit und Bearbeitungskosten bei der Auswertung. Da Daten für eine sehr große Anzahl an Patienten erfasst werden, ist die Untersuchung seltener Ereignisse möglich (Anderson & Garbe, 2008). Des Weiteren enthalten GKV-Daten auch Informationen über Schwerkranke, wie beispielsweise Intensivpatienten, die aus praktischen oder ethischen Gründen meist schwer ermittelt werden können und daher in Primärdaten häufig fehlen (Anderson & Garbe, 2008). GKV-Routinedaten sind systematisch erfasst und liegen oft für mehrere Jahre zurückliegend in gleicher Struktur vor (Pfaff et al., 2009). Dieses eröffnet die Möglichkeit, Vergleiche zwischen den Datensätzen anzustellen und Zusammenhänge über längere Zeiträume hinweg zu untersuchen. Die Routinedaten werden nicht zu Untersuchungszwecken erfasst und sind daher unbeeinflusst von nachträglich definierten Fragestellungen (Pfaff et al., 2009). Die Gefahr von Verzerrungen (Bias) durch den Forscher ist damit geringer als bei Primärdaten. Routinedaten bilden daher die Versorgungspraxis direkt und realistisch ab (Hoffmann, 2008). Wo die Routinedaten unmittelbar Abrechnungszwecken dienen, beispielsweise bei den Rezeptdaten, erfolgt die Erfassung durch den Leistungserbringer in der Regel vollständig. Dadurch ist die Datenqualität von GKV-Routinedaten in diesen Bereichen hoch (Hoffmann et al., 2007). Bei der angestrebten Untersuchung sollen GKV-Routinedaten trotz mancher bereits beschriebener Nachteile genutzt werden.

### **3.2 Data-Mining-Prozess und Methode**

#### *Eignung von Data-Mining*

Data-Mining bezeichnet, frei ins Deutsche übersetzt, das „Graben“ beziehungsweise „Schürfen“ nach Wissen aus Daten. Bedingt durch verschiedene Anwendungsfelder für Data-Mining existieren zahlreiche, unterschiedliche Definitionen. Nach den Amerikanern Berry und Linoff wird unter Data-Mining die *„automatisierte oder halbautomatisierte Erkundung und Analyse von großen Datenmengen zur Aufdeckung von aussagekräftigen Mustern und Regeln“* verstanden (Berry & Linoff, 1997 nach Wiedmann & Buckler, 2003 S. 24; übersetzt aus dem Englischen). Fast gleichlautend ist die deutsche Definition nach Ferber, der ergänzt, dass durch Data-Mining Bezie-

lungsmuster in eine „für Nutzende verständliche Form gebracht werden“ (Ferber, 2003, S.14). Ashby & Simms fügen hinzu, dass Data-Mining dadurch gekennzeichnet ist, dass „sowohl Technologien zur Erkennung als auch statistische und mathematische Methoden“ eingesetzt werden (Ashby & Simms, 1998 nach Wiedmann & Buckler, 2003, S. 24; übersetzt aus dem Englischen). Eine entscheidende Voraussetzung für Data-Mining ist, dass die zu untersuchenden großen Datenmengen systematisch erfasst worden sind (Wiedmann & Buckler, 2003). Das Ergebnis von Data-Mining Analysen können deskriptive Modelle oder Vorhersagemodelle sein oder auch beides. (Wiedmann & Buckler, 2003). Es sollen „gültige, neue oder möglicherweise nützliche und verständliche Muster identifiziert“ werden (Wiedmann & Buckler, 2003, S. 22).

Bei den GKV-Routinedaten handelt es sich um Massendaten mit begrenzter Datenqualität und -aussagekraft. Sie werden kontrolliert, also systematisch erfasst, und sind damit für die Untersuchung mit Methoden des Data-Mining grundsätzlich geeignet. Die Erkennung eines Bedarfs an klinisch-pharmazeutischer Betreuung ist allerdings wegen der vielfältigen zu prüfenden Aspekte komplex. Sie sind daher auf eine überschaubare Prüffregel (Algorithmus) einzugrenzen. Bei der Masse der Daten sind dabei rasch zeitliche und verarbeitungstechnische Grenzen zu erwarten. Eine Herangehensweise, die die Datenauswahl halb manuell und halb automatisiert einengt, erscheint geeignet und mittels Data-Mining möglich. Da die Ergebnisse einer Data-Mining Analyse einen Algorithmus ergeben, erscheint dieses für die praktische Verwendung im Sinne der Fragestellung nützlich. Zusätzlich besteht die Möglichkeit, durch die sich ergebende Auswahl und statistische Gewichtung von Merkmalen im Algorithmus, Hinweise für praktisch nützliche Handlungsempfehlungen zu erhalten.

Für die Anwendung von Methoden des Data-Mining spielt es keine Rolle, welche inhaltliche Fragestellung bearbeitet werden soll (Wiedmann & Buckler, 2003). Data-Mining besitzt eine hohe Flexibilität und ist damit auch auf die zu untersuchende Fragestellung anwendbar. Klassische Beispiele für Data-Mining sind die Fehlererkennung in Herstellungsprozessen und Zukunftssimulationen im strategischen Management (Wiedmann & Buckler, 2003). Gemäß der Literatur hat Data-Mining auch bei medizinischen Themen Anwendung gefunden. Beispielsweise gibt es Untersuchungen zu Patienten mit hoher Mortalität nach Lungentransplantationen (Rivo et al., 2012; Varela et al., 2003; Oztekin et al., 2009) und zur Vorhersage von Überlebenschancen bei Brustkrebspatientinnen (Delen et al., 2005). Im Zusammen-

hang mit Arzneimitteln finden Methoden des Data-Mining offenbar häufig zur Untersuchung der Pharmakovigilanz Anwendung. So wird Data-Mining beispielsweise zum Aufspüren von UAW in großen Datensätzen von Meldesystemen eingesetzt (Harpaz et al., 2012). Die Entwicklung eines Vorhersagemodells für UAW findet sich bei Graziano et al. und kommt der hier angestrebten Untersuchung nahe (Graziano, 2010). Obwohl es verschiedene Algorithmen zur Erkennung einer notwendigen Arzneimitte-optimierung gibt (Kap. 2), ist keine Untersuchung bekannt, mit der Patienten bezüglich dieses Aspekts automatisiert erkannt werden können. Data-Mining als methodische Herangehensweise erscheint für die Beantwortung der Fragestellung daher sinnvoll.

### *Cross-Industry Standard Process (CRISP)*

Für die Durchführung der Data-Mining Analyse gilt es, systematisch vorzugehen. In der Literatur sind verschiedenen Data-Mining Prozesse beschrieben, die meist 5, 6 oder mehr Stufen aufweisen (Wiedmann & Buckler, 2003). Der Versuch eines Expertenkonsortiums seit 1996 einen systematischen, standardisierten Data-Mining Prozess zu entwickeln, hat im Rahmen eines EU-geförderten Projekts zur Formulierung des „*Cross Industry Standard Process for Data Mining*“ (CRISP-DM 1.0) geführt (Chapmann et al., 2000). Da der CRISP als praxisorientierter und von vielen Experten akzeptierter Leitfaden gilt, wird dieser der geplanten Untersuchung zugrunde gelegt.

Der CRISP-DM 1.0 besteht aus 6 unterschiedlichen Prozessschritten, die aufeinander aufbauen und nacheinander durchlaufen werden. Dieses soll eine systematische Herangehensweise bei der Klärung der Fragestellung gewährleisten. Erweisen sich während des Prozesses gefundene Erkenntnisse als nicht nützlich, werden diese verworfen und die Untersuchung ein oder mehrere Prozessschritte zurück erneut angesetzt. Dieses kann im Rahmen einer Data-Mining Analyse mehrfach und über alle Prozessschritte hinweg erfolgen. Data-Mining wird als kontinuierlicher Prozess verstanden, in den neue Erkenntnisse und Erfahrungen immer wieder zur Optimierung der Untersuchung in den Prozess einfließen (Wiedmann & Buckler, 2003). Die 6 Phasen sind in der folgenden Abbildung dargestellt (Chapmann et al., 2000; Chou & Voit, 2009) (Abbildung 1).

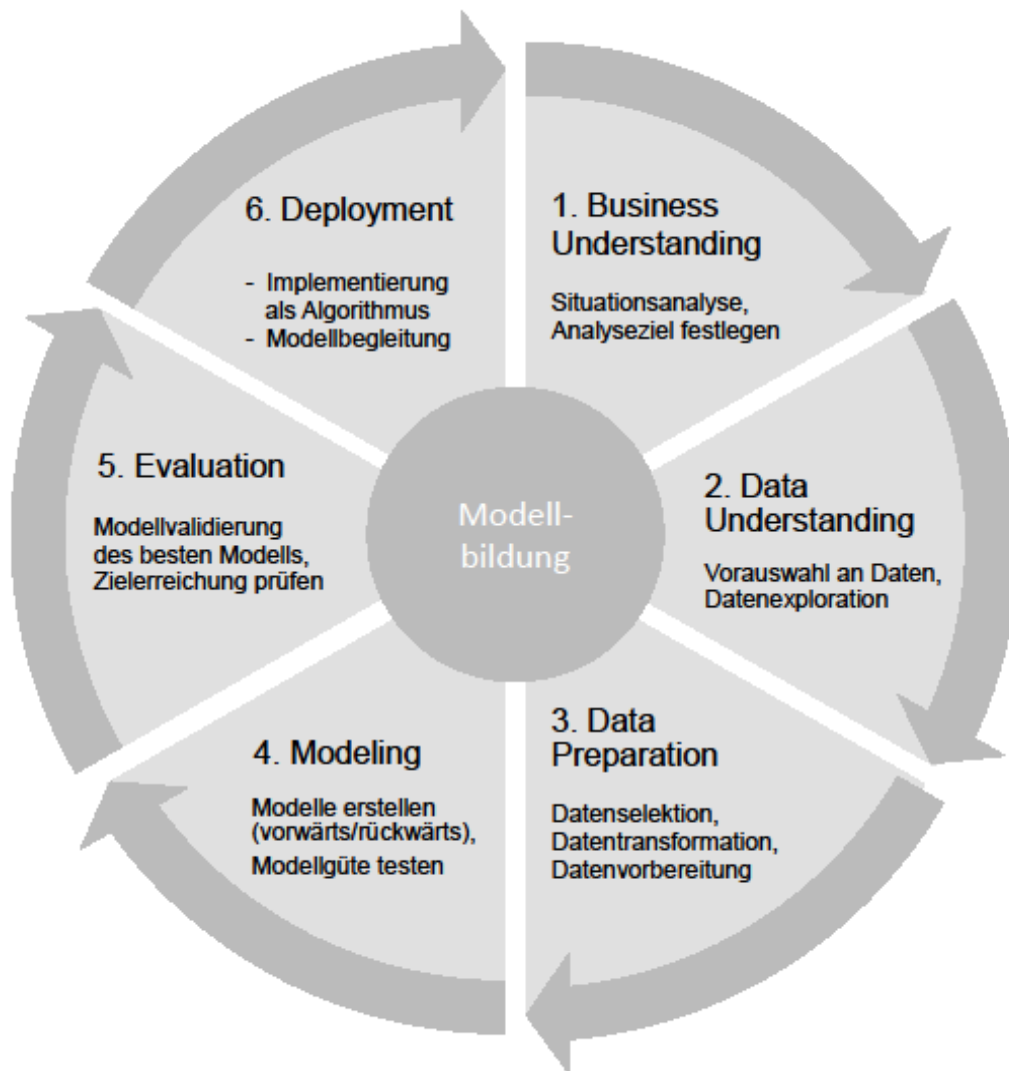


Abbildung 1: Phasen des CRISP (Chapmann et al., 2000, S.10-12, modifiziert)

### Auswahl der Data-Mining Methode

Für den Hauptteil der Untersuchung, nämlich der Prüfung von Abhängigkeiten zwischen dem Zielmerkmal und erklärenden Merkmalen, wird ein beschreibendes Verfahren benötigt. Mit diesem sollen nicht nur ein Merkmal (univariate Analyse), sondern mehrere erklärende Patientenmerkmale gleichzeitig betrachtet werden. Für solche simultanen, mehrdimensionalen Datenanalysen sind Methoden der **multivariaten Statistik** geeignet (Timischl, 2000; Backhaus, 2008). Multivariate Analysemethoden werden in der Literatur unterschiedlich eingeteilt. Manche Autoren unterscheiden zwischen Interdependenz- und Dependenzanalysen. Mit **Dependenzanalysen** können Abhängigkeiten (Dependenzen) zwischen Variablen nach Art und Richtung un-

tersucht werden. Da Abhängigkeiten zwischen Patientenmerkmalen und dem Zielmerkmal angenommen werden, kommt ein Verfahren vom Typ der Dependenzanalyse für die Untersuchung in Frage. Es handelt sich weniger um eine Strukturen entdeckende als vielmehr um eine **Strukturen prüfende Datenanalyse** (Backhaus, 2008 S. 245). Dieses spielt eine Rolle für die Auswahl einer Analysemethode und für die spätere Interpretation der Ergebnisse, da gefundene „*Korrelationen nicht mit Kausalitäten gleichgesetzt und nicht mit jeder Methode gleichermaßen untersucht werden können*“ (Reinboth, 2006).

Für die Wahl einer geeigneten Dependenzanalyse wiederum ist entscheidend, wie die Skalierung beziehungsweise das Messniveau der zu betrachtenden abhängigen und unabhängigen Variablen aussehen (Pepels, 2008) (Tabelle 5).

Abhängige Variable	unabhängige Variable	
	metrisch	nominal
metrisch	Regression Zeitreihenanalyse	Varianzanalyse Regression mit Dummies
nominal	Diskriminanzanalyse logistische Regression*	Kontingenzanalyse auswahlbasierte Conjoint-Analyse

\*=unabhängige Variablen können metrisch oder nominal sein.

**Tabelle 5: Einteilung Strukturen prüfender Verfahren (Backhaus, 2008)**

Das gewählte Zielmerkmal „Krankenhauseinweisung“ kann als abhängige Variable interpretiert werden, das in seiner Ausprägung von den erklärenden Patientenmerkmalen, den unabhängigen Variablen, beeinflusst wird. Für die statistische Analyse solcher Beziehungen werden **Regressionsanalysen** wegen ihrer hohen Flexibilität am häufigsten eingesetzt (Backhaus, 2008). Da unter den unabhängigen Merkmalen metrische Parameter, wie etwa Alter oder Arzneimittelkosten, zu erwarten sind, kommen die Verfahren aus der linken Spalte der oben dargestellten Tabelle in Frage. Da als Zielmerkmal für die Untersuchung ein zu bestimmender Krankenhausaufenthalt mit der Ausprägung „ja“ oder „nein“ vorgesehen ist, liegt ein sogenanntes binäres Merkmal vor. Es handelt sich damit um eine nominale und damit diskrete Variable, bei der **binäre logistische Regressionsanalysen** angewendet werden können (Backhaus, 2008). Im Gegensatz zur ebenfalls geeigneten Diskriminanzanalyse liegt der Vorteil der logistischen Regressionsanalyse darin, dass diese robuster ist und

beispielsweise keine normalverteilten Variablen vorliegen müssen (Backhaus, 2008 S. 244). Ebenso wie bei einer klassischen, linearen Regressionsanalyse werden mit einer logistischen Regressionsanalyse Gewichte ermittelt, mit denen die erklärenden Merkmale die Ausprägung des Zielmerkmals beeinflussen (Backhaus, 2008). Wegen des binären Zielmerkmals ist das Ergebnis einer logistischen Regressionsanalyse jedoch eine Wahrscheinlichkeit für das Eintreten des Zielmerkmals zwischen 0 und 1 (Backhaus, 2008). Die logistische Regression stellt somit ein geeignetes statistisches Verfahren dar, um ein **Vorhersagemodell** für das Eintreten des Zielmerkmals zu ermitteln und damit indirekt einen Betreuungsbedarf abzuschätzen.

Da die Gewichtung anfangs darüber unklar ist, welche Variablen am meisten zu einer treffenden Vorhersage beitragen, soll gemäß den Prinzipien einer Data Mining-Analyse bei der Entwicklung eines Vorhersagemodells eine **schrittweise logistische Regression** durchgeführt werden. Das schrittweise Vorgehen wird mit Hilfe der Statistiksoftware **SPSS 19.0** (IBM Corp., 2010) automatisiert durchgeführt werden. Zur testweisen Bildung mehrerer Modelle sollen folgende verfügbare Modellierungsalgorithmen verwendet werden:

- **Vorwärtsselektion:**  
Schrittweise Variablenauswahl mit einem Test auf Aufnahme und einem Test auf Ausschluss. Der Test auf Aufnahme beruht auf der Signifikanz der Werte-Statistik. Der Test auf Ausschluss beruht bei der Methode „nach Wald“ auf der Wald-Statistik, bei der Methode „bedingt“ auf der Wahrscheinlichkeit einer Likelihood-Quotienten-Statistik (IBM Corp., 2010).
- **Rückwärtselimination:**  
Eine Methode zur Variablenauswahl, bei der alle Variablen in die Gleichung aufgenommen und anschließend schrittweise ausgeschlossen werden. Die Variable mit der kleinsten Teilkorrelation zur abhängigen Variablen wird als erste für den Ausschluss in Betracht gezogen. Das Verfahren wird beendet, wenn keine Variablen mehr zur Verfügung stehen, die die Ausschlusskriterien erfüllen. Für den Test auf Ausschluss stehen verschiedene Methoden zur Auswahl, verwendet wird die Methode „nach Wald“ (IBM Corp., 2010).

- **Einschluss:**  
Eine Prozedur für die Variablenauswahl, bei der alle Variablen in einem einzigen Schritt aufgenommen werden. Bei der Einschluss-Methode findet keine Gewichtung oder ein Ausschluss der Variablen nach einer statistisch ermittelten Relevanz statt (IBM Corp., 2010). Diese Methode kann daher verwendet werden, um die Reihenfolge und den Einschluss von Variablen manuell und nach fachlichen Gesichtspunkten zu bestimmen und gezielt zu testen (Fromm, 2005).

Für die Data-Mining Analyse sind aus den Kernfragen der Untersuchung Hypothesen für das Data-Mining zu bilden. Unter Berücksichtigung des ausgewählten Zielmerkmals als Proxy für einen Bedarf an klinisch-pharmazeutischer Betreuung lauten diese:

1. Mit zunehmender Polypharmazie steigt die Wahrscheinlichkeit für eine (arzneimittelbedingte) Hospitalisation.
2. Mit Data-Mining Methoden lassen sich Patienten mit Polypharmazie, die innerhalb von einem Jahr (arzneimittelbedingt) in ein Krankenhaus eingewiesen werden, aus GKV-Routinedaten vorhersagen.
3. Mit Data-Mining lassen sich bei Polypharmazie-Patienten Merkmale identifizieren, die eine (arzneimittelbedingte) Hospitalisation wahrscheinlicher machen.

### **3.3 Qualitätssicherung**

Orientiert an den Qualitätsanforderungen für Sekundärdatenanalysen werden nachfolgend Vorüberlegungen und Maßnahmen dargestellt, die eine gute Auswertungsqualität der Untersuchung gewährleisten sollen (AGENS & Arbeitsgruppe Epidemiologische Methoden, 2008; Pfaff et al., 2009).

**Objektivität:**

Bei der schrittweisen Annäherung an die Definition der Zielgruppe und bei der Auswahl von Zielmerkmal und erklärenden Variablen, besonders zu Beginn des Data-



Mining Prozesses, werden viele fachlich motivierte Entscheidungen getroffen. Auch wenn dieser Prozess durch Voranalysen unterstützt werden soll, wird die Entscheidungsfindung teilweise einen experimentellen und durch die Autorin subjektiv geprägten Prozess darstellen. Dieses beeinflusst die Analysegestaltung und damit die Ergebnisse. Zur Datenreduktion ist die Eingrenzung der für die Modellbildung zu testenden Merkmale notwendig. Gleichzeitig sollen bei der Sondierung keine Merkmale, die das zu untersuchende Geschehen maßgeblich beeinflussen, für die Analyse verloren gehen. Verzerrungen der Ergebnisse sind denkbar durch:

- die Art der Zielgruppendefinition
- die Auswahl und Ausgrenzung von erklärenden Merkmalen
- die Wahl des Zielmerkmals
- die Bestimmung von Ein- und Ausschlusskriterien
- die Begrenzung des Auswertungszeitraums
- nicht zugängliche Daten und Analysemethoden

Die getroffenen Regelungen im Analyseplan, wie beispielsweise die vorab formulierten Fragestellungen, sollen jedoch ein objektives Vorgehen weitgehend gewährleisten.

Reliabilität der Ergebnisse:

Zur Reproduzierbarkeit der Ergebnisse und für eine größtmögliche Transparenz bezüglich der Voranalysen und der Modellentwicklung wird für jede Teilanalyse Datenbasis und Durchführung beschrieben. Die Modellierung ist determiniert durch die eingesetzten Variablen, das Zielmerkmal und den verwendeten Modellierungsalgorithmus. Wo aus fachlichen Gründen in die Modellierung manuell eingegriffen wird, wird dieses beschrieben und begründet. Es ist nicht auszuschließen, dass mit anderen Data-Mining Verfahren, anderen Merkmalen oder Merkmalskombinationen bei Nutzung von GKV-Routinedaten Modelle mit einer besseren Modellgüte entwickelt werden können. Dazu sind zu viele verschiedene Varianten möglich. Die Modellierung wiederum ist ein experimenteller Prozess, der je nach den Modellierungserfahrungen des Durchführenden unterschiedlich ausfallen kann. Die Auswahl an Merkmalen, die zu guten Modellen führen, sollte sich allerdings ähnlich darstellen und an vergleichbaren Datensätzen reproduzierbar werden können. Das finale Vorhersage-

modell an sich wird durch die logistische Regressionsformel eindeutig beschrieben. Bei Anwendung des Modells auf ein und denselben Datensatz wird dieses daher unabhängig vom Durchführenden zu stets identischen Vorhersageergebnissen führen.

#### Validität:

Die Validität des Vorhersagemodells ist durch die begrenzte Datenqualität der GKV-Routinedaten eingeschränkt. Diese wurde beschrieben. Auswirkungen auf das Auswertungsergebnis sind bei der Interpretation zu berücksichtigen. Die Eignung von Data-Mining als Methode wurde hergeleitet und wird als gegeben betrachtet. Eine systematische Vorgehensweise wird durch die Orientierung an einem etablierten Data-Mining Prozess (CRISP-DM 1.0) unterstützt. Weitere Maßnahmen zur Sicherstellung der Validität des Vorhersagemodells sind bei den Ausführungen zur Ergebnissicherung dargestellt.

#### Statistische Qualitätssicherung:

Zur Beschreibung der Analyseergebnisse werden im Verlauf der Untersuchung Methoden der **deskriptiven Statistik** angewendet und absolute, relative oder prozentuale Häufigkeiten, Lagemaße, meist Mittelwert oder Median, und gegebenenfalls Streuungsmaße ermittelt. Wo es wichtig ist, für die Entscheidungsfindung oder die Ergebnissicherung werden Methoden der **analytischen Statistik** genutzt, Parameter-, Verteilungs- und Unabhängigkeitstests durchgeführt und Konfidenzintervalle beziehungsweise Überschreitungswahrscheinlichkeiten (p-Werte) geschätzt. Weitere Maßnahmen zur Sicherstellung der statistischen Qualität finden sich in den Ausführungen zum Analyseumfang und zur Ergebnissicherung.

#### Praktische Aspekte:

Bei der Zusammenstellung des Entwicklungs- und Validierungsdatsatzes werden die Datentabellen vor der Modellierung mittels SPSS 19.0 auf unterschiedliche Weise qualitätsgeprüft. Dieses betrifft insbesondere Maßnahmen zur Erkennung von Datenlücken und doppelten Einträgen, sowie eine inhaltliche Plausibilitätsprüfung. Die Vorgehensweise wird im Abschnitt *Data-Preparation* detailliert beschrieben.

### 3.4 Analyseumfang

Bei der Analyseplanung ist zu prüfen, welche Anforderungen an die statistische Aussagekraft der Untersuchung gestellt werden (AGENS & Arbeitsgruppe Epidemiologische Methoden, 2008). Diese werden dadurch bestimmt, dass untersucht wird, mit welcher Häufigkeit die zu untersuchenden Merkmale im Untersuchungsdatensatz auftreten und welcher Datenumfang notwendig ist, um auch seltene Merkmale identifizieren und Aussagen über diese mit einer ausreichenden statistischen Aussagekraft treffen zu können. Da Patienten mit Polypharmazie aus einem Routinedatensatz einer mittelgroßen, bundesweit agierenden Krankenkasse identifiziert werden sollen, ist nicht mit geringen Fallzahlen zu rechnen. Durch die Voranalysen werden die erwarteten Patientenzahlen in der zugrunde gelegten Datenbasis im Vorfeld der Modellbildung untersucht und durch eine entsprechende Definition von Zielgruppe, Zielmerkmal und erklärenden Merkmalen ausreichende Fallzahlen sichergestellt. Zusätzlich soll vor der Modellierung die Merkmalsverteilung in den Analysedatensätzen ermittelt und geprüft werden. Sogenannte „Fehler 2. Art“ sind unter diesen Voraussetzungen unwahrscheinlich, daher wird auf die Durchführung einer Machbarkeitsanalyse verzichtet.

#### *Ein- und Ausschlusskriterien*

Für die ausschließliche Betrachtung von Patienten mit Polypharmazie wird die entwickelte Definition dafür als Einschlusskriterium festgelegt. Damit noch ein möglichst aktueller Bezug zwischen Polypharmazie und Krankenhauseinweisung angenommen werden kann, werden nur Patienten eingeschlossen, die ein Quartal vor dem Betrachtungszeitraum für die Bestimmung keine Polypharmazie aufgewiesen haben. Um Verzerrungen des Analyseergebnisses durch fehlende Daten auszuschließen, werden nur vollständige Datensätze eingeschlossen, das heißt nur Daten von Patienten, die durchgängig im Auswertungszeitraum versichert waren. Aus praktischen Erwägungen im Hinblick auf eine mögliche spätere klinisch-pharmazeutischen Betreuung wird bei der Datenauswahl für die Analyse das Alter der einzubeziehenden Patienten auf 18-85 Jahren begrenzt. In Anlehnung an Schneeweiss et al. werden zusätzlich Diagnosen, die auf eine hohe Morbidität oder Erkrankungen mit spezieller Beratungsthematik hindeuten, ausgeschlossen (Schneeweiss et al., 2002).

### *Auswertungszeitraum*

Für die Bestimmung der Häufigkeiten der beschreibenden Merkmale im Datensatz wird rückblickend, von einem Auswertungstichtag betrachtet, ein Auswertungszeitraum von 3 Jahren festgelegt. Dieser ist so gewählt, dass die vermeintlich charakteristischen Merkmale für Patienten mit Polypharmazie und einem Betreuungsbedarf ausreichend häufig im Datensatz vertreten sind und Spielraum für die Merkmalsbildung vorhanden ist. Das soll die Bildung eines Vorhersagemodells mit guter Modellgüte ermöglichen. Für die Ereignisprognose, die Detektion des Zielmerkmals, wird ein weiterer Auswertungszeitraum von einem Jahr nach dem Auswertungstichtag definiert. Es werden ein Datensatz zur Modellbildung (Entwicklungsdatensatz) und ein zweiter Datensatz zur Modellvalidierung (Validierungsdatensatz) erstellt:

- Entwicklungsdatensatz: 2005-2007 mit Ereignisprognose für 2008
- Validierungsdatensatz: 2007-2009 mit Ereignisprognose für 2010

### **3.5 Ergebnissicherung**

Zur Sicherung der internen Validität des Vorhersagemodells ist vorgesehen, in SPSS 19.0 für jeden Prädiktor einen Hypothesentest zum Einfluss auf das Zielmerkmal anzufordern. Zusätzlich soll für jede eingeführte neue Variable während der schrittweisen Modellierung die Signifikanz der Chi-Quadrat-Veränderung mit dem „*Omnibus-Test der Modellkoeffizienten*“ geprüft werden. Eine signifikante Chi-Quadrat Veränderung kennzeichnet eine Modellverbesserung. Damit soll sichergestellt werden, dass alle im finalen Vorhersagemodell eingesetzten Merkmale einen signifikanten Einfluss auf das Zielmerkmal haben und gut ausgewählt wurden. Für die Effektkoeffizienten ist geplant, mit SPSS 19.0 95-prozentige Vertrauensintervalle zu berechnen, um die erkennbaren Aussagen zur Effektstärke abzusichern. Um zu gewährleisten, dass das entwickelte Modell nur geringfügig durch Ausreißer verzerrt ist, sollen die Patientenfälle ermittelt werden, die durch das Modell unrichtig eingeordnet wurden.

Um Zwischenergebnisse der Modellierung zu prüfen und die beste Auswahl an Variablen für das finale Modell zu gewährleisten, soll anhand des Pseudo-R-Quadrats nach Nagelkerke und dem prozentualen Anteil der richtigen Klassifizierung die Modellgüte abgeschätzt werden. Für das finale Modell ist geplant, eine Klassifikations-

matrix für einen definierten Wertebereich an Vorhersagewahrscheinlichkeiten zu erstellen und die Modellgüte unter diesen Rahmenbedingungen durch Bestimmung von Spezifität, Sensitivität, Trefferquote und PPV (engl.: positive predictive value) genauer darzustellen.

Zur Prüfung der externen Validität (Generalisierbarkeit) des Vorhersagemodells ist die exaktere Bestimmung der Modellgüte vorgesehen. Dazu soll das finale Modell an einer Hälfte des gesplitteten Entwicklungsdatensatzes angewendet und an der anderen Hälfte der AUC-Wert des Modells bestimmt werden. Zur Prüfung der Modellstabilität über einen längeren Zeitraum wird das entwickelte Vorhersagemodell an einem Validierungsdatensatz mit zwei Jahre aktuelleren Daten angewendet und die Güte der Prädiktoren und des Modells wie für die Sicherstellung der internen Validität überprüft.

### **3.6 Ethik und Datenschutz**

Gemäß den Untersuchungsleitfäden muss im Rahmen von Sekundärdatenanalysen keine Ethikkommission für die Genehmigung solcher Untersuchungen eingeschaltet werden, solange der Datenschutz eingehalten wird. Wenn keine personenbezogenen Daten verwendet werden oder wenn eine Pseudonymisierung der Daten vorgenommen wird, sodass keine Rückschlüsse auf Einzelpersonen möglich sind, ist dieses ausreichend (AGENS & Arbeitsgruppe Epidemiologische Methoden, 2008). In diesem Fall ist auch eine Genehmigung für die Datennutzung durch Patienten nicht notwendig, was die Umsetzung von Sekundärdatenanalysen erleichtert. Zur Sicherstellung der angemessenen Berücksichtigung datenschutzrechtlicher Aspekte bei der praktischen Durchführung der Analysen und der späteren, unbeeinflussten Veröffentlichung der Ergebnisse wurden folgende Maßnahmen ergriffen:

- Pseudonymisierung der zugrunde gelegten Daten (Kap. 3.6)
- Einholung einer schriftlichen Schweigepflichtserklärung der Beteiligten, soweit nicht abgedeckt durch den Dateneigner

- Abstimmung und Freigabe der Untersuchungsergebnisse mit dem Datenschutzbeauftragten des Dateneigners
- Schriftliche Vereinbarung mit dem Dateneigner über die Erlaubnis zur Durchführung der Analysen und einer zeitnahen, unbehinderten und vollständigen Darstellung der Ergebnisse im Rahmen einer Veröffentlichung

## 4 Voralysen (Data Understanding)



Bei den Vorüberlegungen zum Zielmerkmal und den Patienten mit Polypharmazie in der *Business Understanding*-Phase haben sich Fragen ergeben, die in Voralysen geklärt werden sollen. So sollen die Definitionen von Zielmerkmal (Kap. 2.2) und Zielgruppe (Kap. 2.3) konkreter bestimmt werden und die Vorauswahl der erklärenden Merkmale (Kap. 2.4), der Prädiktoren, für die Modellentwicklung näher auf ihre Eignung überprüft werden.

### *Methodisches Vorgehen bei den Voralysen*

Entsprechend den Prinzipien des Data-Mining werden verschiedene Voralysen mit variierenden Rahmenbedingungen durchgeführt. Da für die Umsetzung der Analysen auf einen großen Datenpool aus unterschiedlichen Datenquellen zugegriffen wird, erfolgt die Datenzusammenstellung mittels **SQL-Abfragen** (Structured Query Language). Unter Nutzung des Analysis Services des SQL Servers 2005 erfolgen die Datenselektionen im Data Warehouse des Datenlieferanten nach den fachlichen Vorgaben der Autorin. Die erhaltenen Datenpakete werden für die weiteren Analysen zur Verfügung gestellt und für Detailanalysen mit **SPSS 19.0 Statistics 19.0** durch die Autorin weiterbearbeitet.

#### **4.1 Technische Definition der Polypharmazie**

Die vielfältigen Definitionen der Polypharmazie in der Literatur ist in den fachlichen Vorüberlegungen dargestellt worden. Diese gilt es für die Analysen in eine geeignete, technisch umsetzbare Definition zu übersetzen, die der fachlichen möglichst nahe kommt. In den Voranalysen soll geklärt werden, welcher Schwellenwert für die Mindestanzahl und welcher Beobachtungszeitraum für die Prüfung auf gleichzeitig eingenommene Arzneistoffe sinnvollerweise zu wählen ist.

##### *Voranalysen I-III*

Betrachtung von mehr als 5, 9, 13 Wirkstoffen über 4 Quartale:

In der Literatur wird überwiegend ab 5 oder mehr Wirkstoffen von einer Polypharmazie gesprochen. Daher wird zunächst überprüft, ob ein Schwellenwert bei **mindestens 5** gleichzeitig verordneten Arzneistoffen für die Definition der Polypharmazie sinnvoll ist oder die Festlegung des Schwellenwertes auf 9 oder 13 Arzneistoffe geeigneter erscheint (Voranalysen I-III).

Um die gleichzeitige Einnahme von Arzneistoffen als Teil der Definition der Polypharmazie nachzubilden, ist zusätzlich ein zu überprüfender Zeitraum festzulegen. Dazu bietet sich die tägliche, die wöchentliche, die monatliche oder die quartalsweise Prüfung der gleichzeitigen Einnahme von Arzneistoffen an. Da die zu verwendende Datenbasis zur Prüfung auf Polypharmazie Rezeptdaten sind und diese in der Regel Verordnungen von Arzneistoffen für ein Quartal oder länger enthalten, wird für die erste Voranalyse die **quartalsweise** Betrachtung gewählt.

Eine exaktere Bestimmung der zeitgleichen Arzneistoffeinnahme wäre durch die Berechnung von Reichweiten möglich. Dazu könnte der Zeitraum berechnet werden, für den die verordnete Arzneistoffmenge bei zugrunde legen der pro Arzneistoff definierten durchschnittlichen Tagesdosis (DDD: defined daily dose) ausreichen würde. Diese Herangehensweise wird verkompliziert dadurch, dass die ermittelten Reichweiten der einzelnen Arzneistoffe eines Patienten unterschiedlich lang sein können und die Prüfung auf das gleichzeitige Vorliegen von Arzneistoffen pro Patient an zahlreichen verschiedenen Stichtagen, beispielsweise täglich oder wöchentlich, erfolgen müsste. Zur Vereinfachung erfolgt daher die quartalsweise Betrachtung ohne Berechnung

einer Reichweite.

Um eine Unterscheidung von Patienten mit einer längerfristigen von denen mit einer kurzzeitigen Polypharmazie vornehmen zu können, wird die Anzahl an Arzneistoffen pro Quartal über mehrere Quartale hinweg betrachtet. In den ersten Voranalysen wird auf die Verordnung von mindestens 5, 9 beziehungsweise 13 Arzneistoffen pro Quartal **über mindestens 4 aufeinander folgende Quartale**, also über ein Jahr, geprüft (Abbildung 2).

Fall	2005				2006				2007				2008				
	Nr.	Q1*	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
I.		0	0	0	0	0	6	9	9	10	10	13	13				
II.		0	0	5	3	7	5	7	8	8	4	4	5				
III.		5	0	5	5	0	5	5	5	5	5	0	5				
IV.		5	5	7	8	8	5	4	4	5	7	4	7				

\*=Q=Quartal; Zahlenwerte=Anzahl ATC pro Quartal; hellgrau=berücksichtigtes Quartal mit  $\geq 5$  ATC (Polypharmazie definiert als  $\geq 5$  ATC/Quartal über die ersten 4 zusammenhängenden Quartale ab 31.12.2007); dunkelgrau=Auswertungszeitraum bezüglich Zielmerkmal (Krankenhausaufenthalt)

**Abbildung 2: Darstellung der Polypharmazie-Ermittlung bei variierendem Stichtag (Voranalysen I-III)**

#### *Durchführung (Voranalysen I-III)*

Als Datenbasis für alle Voranalysen zur genaueren Bestimmung der technischen Definition der Polypharmazie werden Rezeptdaten zugrunde gelegt. Arzneistoffe sind in den GKV-Routinedaten nach der Anatomisch-therapeutisch-chemischen Klassifikation eindeutig beschrieben (WIdO, 2010). Dabei stellt der ATC-Code in der letzten Ebene als 7-stelliger Code den Wirkstoff eindeutig dar und wird für die Prüfung auf die Anzahl verschiedener Arzneistoffe herangezogen. In Anlehnung an die geplanten Auswertungszeiträume und die Einschlusskriterien wird folgende Datenbasis verwendet (Abbildung 3):

- Auswertungstichtag beziehungsweise Stand der Daten: 31.12.2007
- Patient ist durchgängig versichert von 2005-2007
- (keine Altersbeschränkung)



- Rezeptdaten von 2005-2007 zur Ermittlung von Patienten mit mindestens 5 verschiedenen verordneten Wirkstoffe in mindestens einem der 12 Quartale zur groben Datenreduktion
- Krankenhausdaten von 2008 zur Ermittlung der Krankenhausaufenthalte
- Datenpool mit n=416.140 Patienten

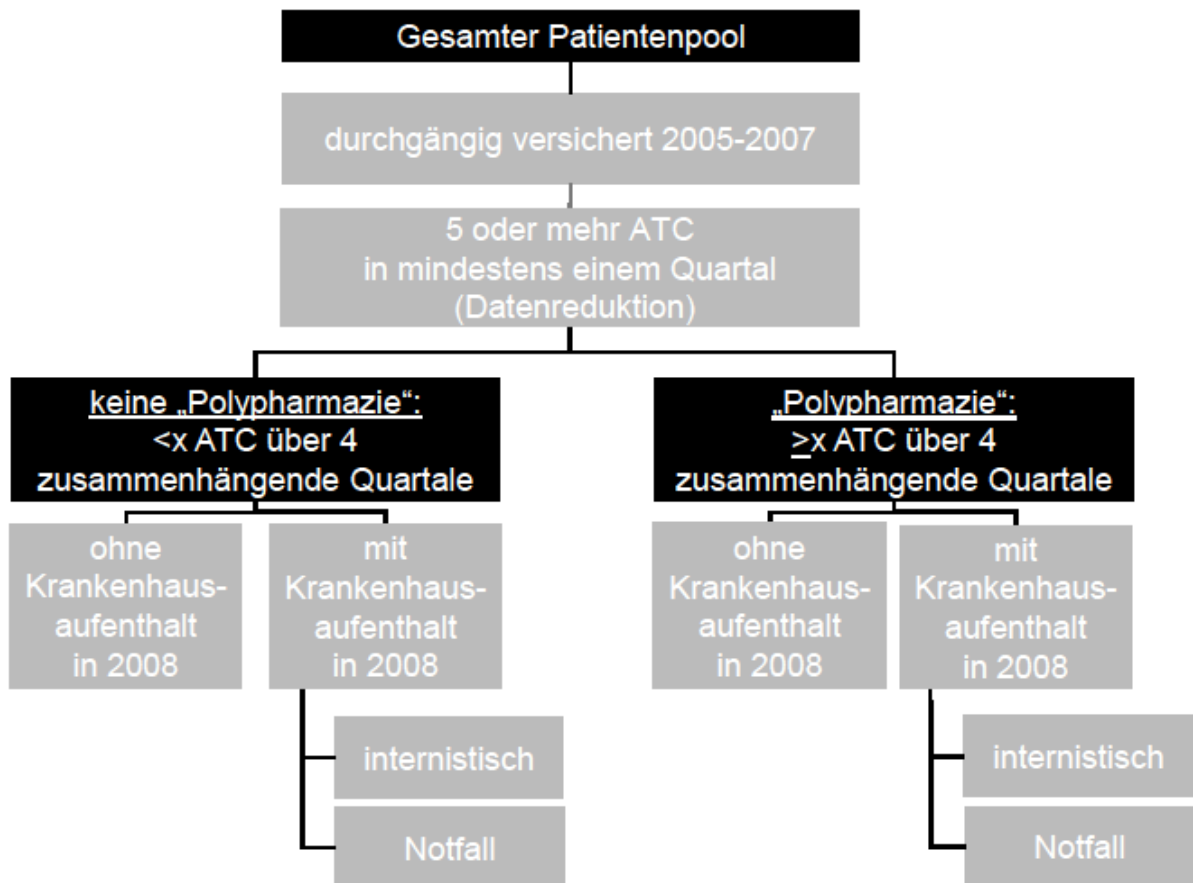


Abbildung 3: Ablauf der Datenselektion aus dem Gesamtpool (Voranalysen I-III)

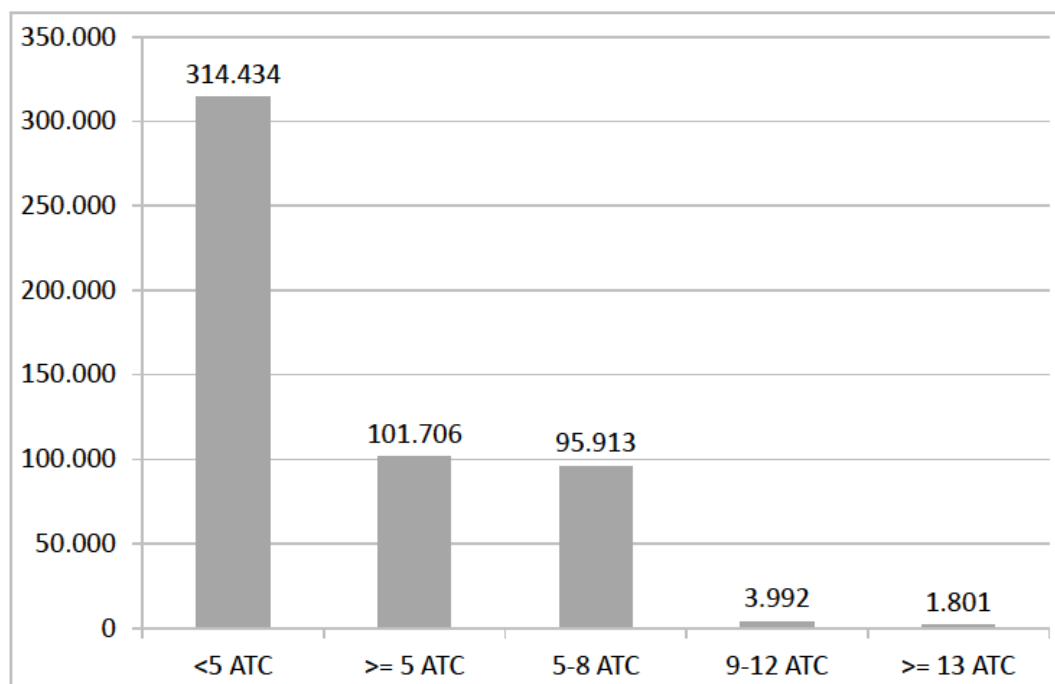
Die Durchführung der Voranalysen I-III erfolgt schrittweise mit Hilfe von direkten SQL-Abfragen aus dem beschriebenen Datenpool im Data Warehouse des Datenlieferanten nach Vorgabe der Autorin mit den nachfolgend genannten Zielen.

- Ermittlung der Polypharmazie durch Prüfung der letzten 4 aufeinanderfolgenden Quartale in 2005-2007 auf
  - **5 oder mehr** Arzneistoffe (Voranalyse I)
  - **9 bis 12** Arzneistoffe (Voranalyse II)
  - **13 oder mehr** Arzneistoffe (Voranalyse III)

- Ermittlung der Anzahl an Krankenhauseinweisungen im den folgenden 12 Monaten ab dem letzten Tag des ermittelten Zeitraums mit Polypharmazie. (Es ergeben sich für die betrachteten Patientenfälle uneinheitliche Stichtage.)
- Ermittlung der Anzahl an Krankenhauseinweisungen differenziert nach internistischen Einweisungen und Notfalleinweisungen.

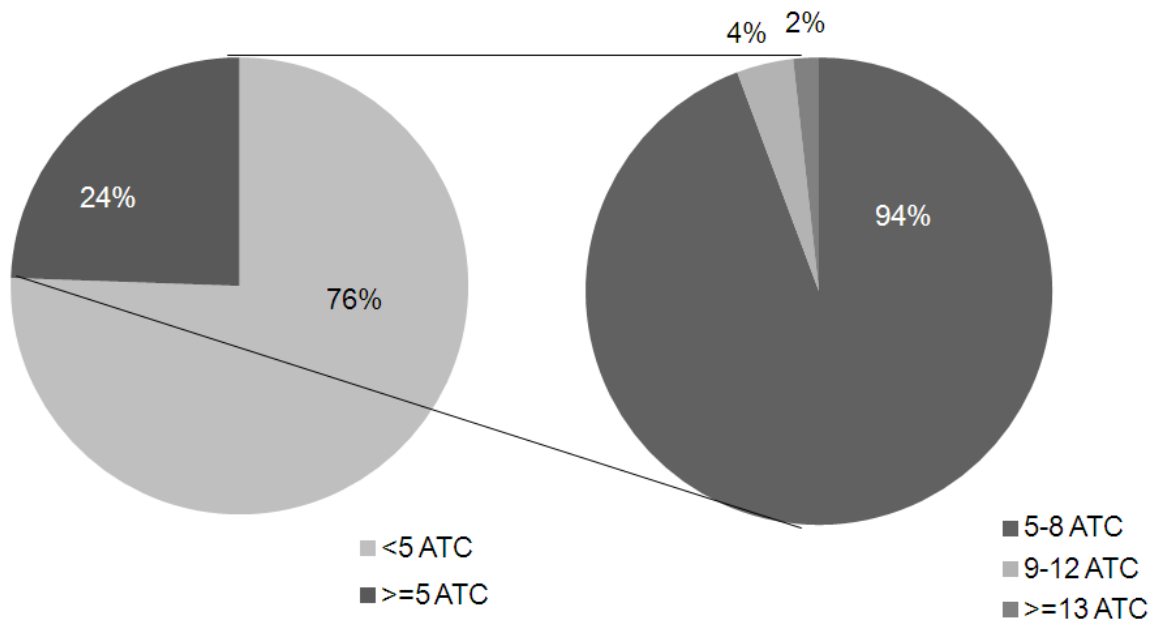
### *Ergebnis (Voranalysen I-III)*

Die nachfolgende Abbildung (Abbildung 4) zeigt die Häufigkeitsverteilung für die verschiedenen gebildeten „Polypharmazie“-Gruppen (siehe auch Anhang 10.5).



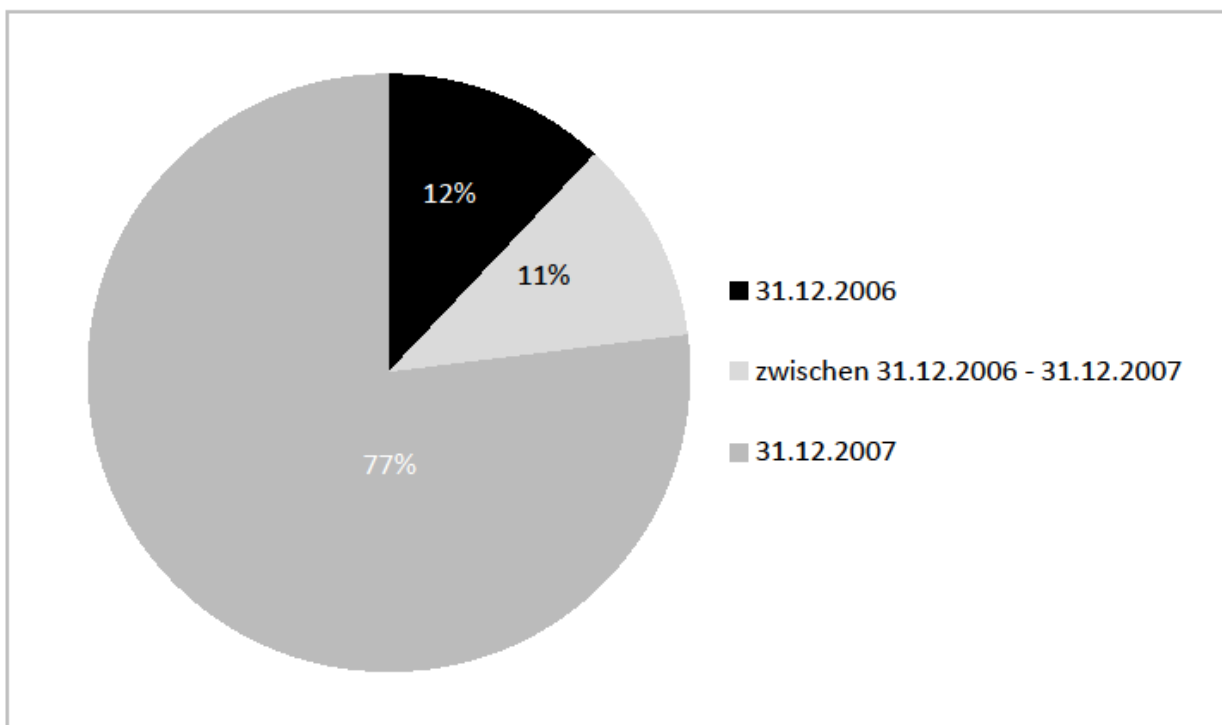
**Abbildung 4: Anzahl Patienten nach Polypharmazie-Gruppe in 2005-2007 (Voranalysen I-III)**

Die Betrachtung der absoluten Patientenzahlen in den jeweiligen Polypharmazie-Gruppen zeigt eine insgesamt hohe Anzahl von Patienten mit 5 oder mehr verschiedenen verordneten Arzneistoffen über 4 zusammenhängende Quartale in dem betrachteten Patientenpool. Bei den gebildeten Untergruppen wird die Dominanz der Polypharmazie-Gruppe mit Patienten, die zwischen 5 und 8 Arzneistoffen pro Quartal über 4 Quartale verordnet bekommen haben, deutlich.



**Abbildung 5: Prozentuale Verteilung der Polypharmazie-Gruppen in 2005-2007 (Voranalysen I-III)**

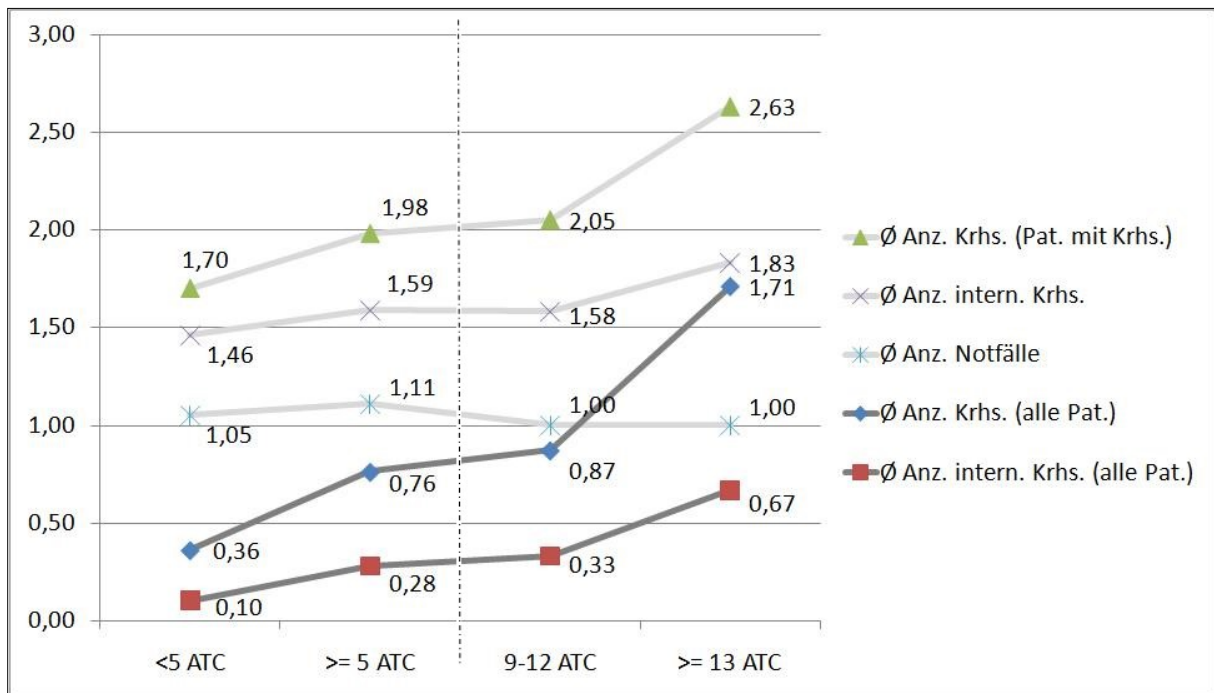
Prozentual ausgedrückt liegt der Anteil der Patienten mit 5 oder mehr verschiedenen verordneten Arzneistoffen in 4 aufeinander folgenden Quartalen mit 24% bei etwa einem Viertel des betrachteten Patientenpools (Abbildung 5). Innerhalb dieser Gruppe ist der Anteil an Patienten mit 9-12 beziehungsweise 13 oder mehr verschiedenen Arzneistoffen mit 4% und 2% vergleichsweise gering.



**Abbildung 6: Prozentuale Verteilung der ermittelten Stichtage in 2005-2007 (Voranalysen I-III)**

Zur Bestimmung der Polypharmazie werden die aktuellsten 4 zusammenhängenden Quartale mit 5 oder mehr Arzneistoffen pro Quartal im Betrachtungszeitraum von 2005 bis 2007 ermittelt (Abbildung 6). Bei 77% der Patienten zeigen die letzten 4 Quartale in 2007 jeweils 5 oder mehr Arzneistoffe und erhalten den Auswertungstichtag 31.12.2007. Ab diesem Stichtag wird die Anzahl der Krankenhauseinweisungen in den nächsten 12 Monaten (Zielmerkmal) beginnend ab dem 01.01.2008 betrachtet. Bei 12% der Patienten werden die ersten 4 Quartale des Datensatzes, die Quartale 1 bis 4 in 2005, und damit der Auswertungstichtag 31.12.01.2006 ermittelt. Bei 11% der Patienten ergeben sich Auswertungstichtage zwischen dem 31.12.2006 und dem 31.12.2008.

### Betrachtung der Krankenhauseinweisungen



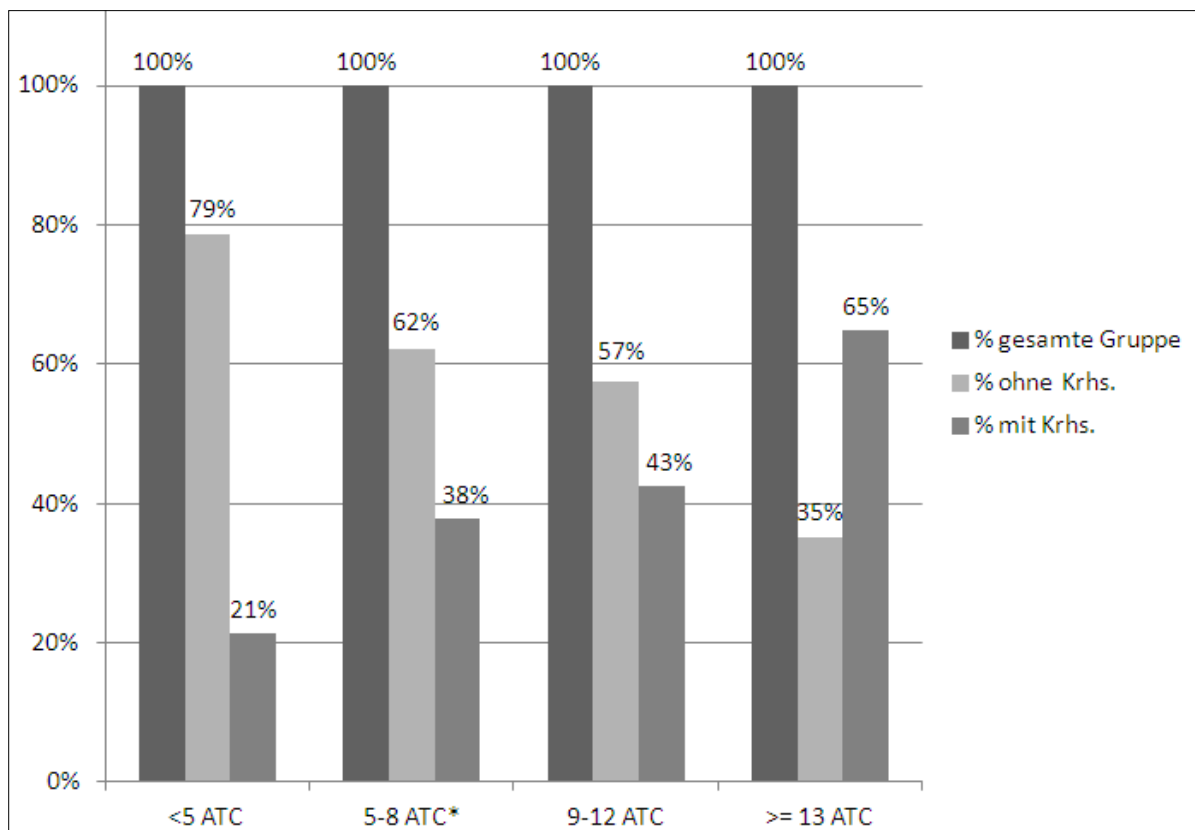
Anz.=Anzahl; intern.=internistische; Krhs.=Krankenhauseinweisungen; Notfälle=Notfalleinweisung

**Abbildung 7: Mittlere Anzahl Krankenhauseinweisungen in 2008 nach Polypharmazie-Gruppe (Voranalysen I-III)**

Die Polypharmazie-Gruppen 9-12 ATC und  $\geq 13$  ATC sind Teilmengen der Gruppe  $\geq 5$  ATC und daher streng genommen nicht direkt miteinander vergleichbar. Dieses wird in der Abbildung durch einen Trennstrich angedeutet (Abbildung 7). Die Gruppe 5-8 ATC ist nicht dargestellt, da die Anzahl der Krankenhauseinweisungen für diese Gruppe nicht separat ermittelt wurde. Es wird davon ausgegangen, dass die durchschnittliche Anzahl an Krankenhauseinweisungen für diese Gruppe unterhalb des Durchschnittswerts der Gruppe  $\geq 5$  ATC liegt. Das Ergebnis der Voranalyse zeigt mit zunehmender Zahl an Arzneistoffen pro Quartal ansteigende durchschnittliche Häufigkeiten an **Krankenhauseinweisungen generell** (0,36 vs. 0,76 vs. 0,87 vs. 1,71; entsprechend +111% vs. +14% vs. +96%). Insgesamt zeigt sich, dass es ab 5, 9 beziehungsweise 13 Arzneistoffen zu einem 2,1-, 2,4- beziehungsweise 4,8-fachen Anstieg der durchschnittlichen Krankenhauseinweisungen kommt (0,76/0,36; 0,87/0,36; 1,71/0,36). Ebenso wie für die Krankenhauseinweisungen generell steigt die durchschnittliche Anzahl der **internistischen Krankenhauseinweisungen** (0,10 vs. 0,28 vs. 0,33 vs. 0,67; entsprechend + 180% vs. +22% vs. +103%). Es kommt damit zu

einem 2,8-, 3,3- beziehungsweise 6,7-fachem Anstieg (0,28/0,10; 0,33/0,10; 0,67/0,10).

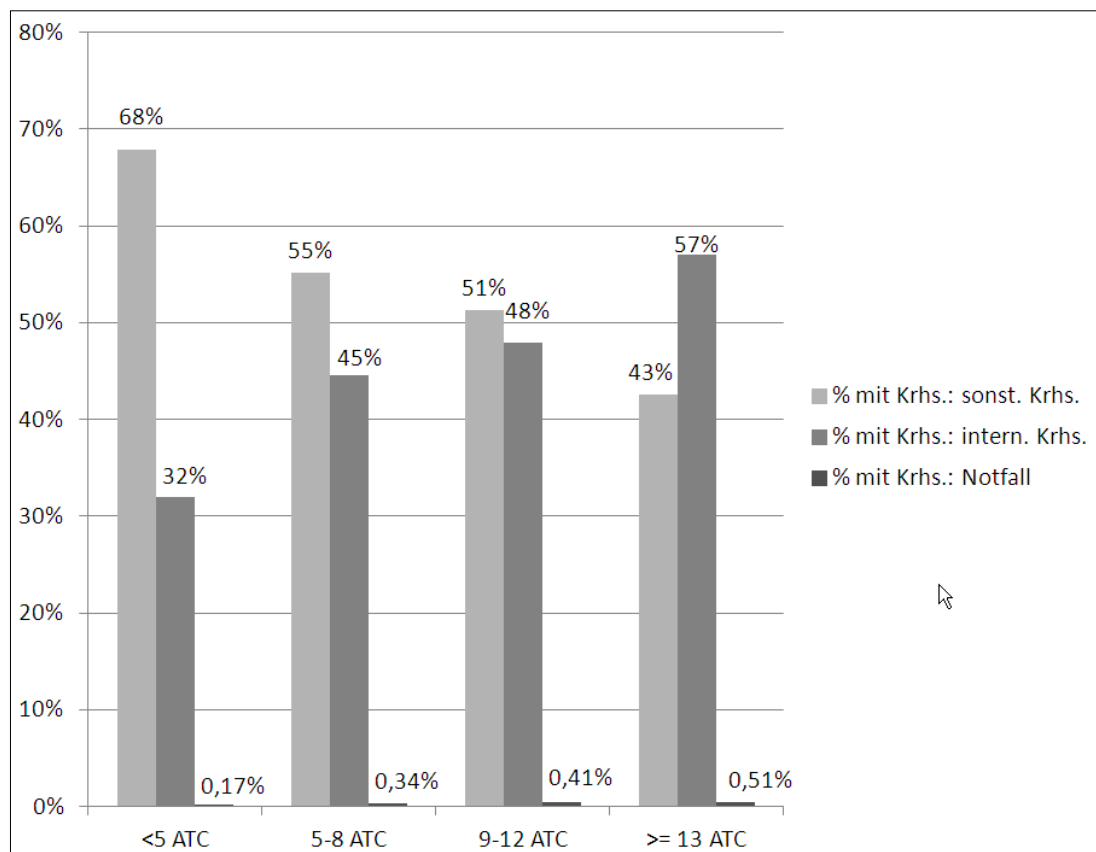
Bei der ausschließlichen Betrachtung der **Patienten mit mindestens einer Krankenhauseinweisung** und bei Unterscheidung zwischen generellen und internistischen Krankenhausaufenthalten innerhalb des ausgewerteten Jahres zeigt sich bei den Durchschnittswerten über die Polypharmazie-Gruppen hinweg das gleiche Bild ansteigender Werte (Abbildung 8). Die Durchschnittswerte liegen auf einem vergleichsweise höheren Niveau. In der Gruppe  $\geq 13$  ATC erscheint der Anstieg der Durchschnittswerte dabei ausgeprägter als in den anderen Polypharmazie-Gruppen (1,70 vs. 1,98 vs. 2,05 vs. 2,63; mit +16% vs. +4% vs. +28%), insbesondere bei den internistischen Krankenhauseinweisungen (1,46 vs. 1,59 vs. 1,58 vs. 1,83 mit +9% vs. -1% vs. +16%). Die durchschnittliche Anzahl an **Notfalleinweisungen** ins Krankenhaus stellt sich dagegen über alle Polypharmazie-Gruppen hinweg insgesamt niedrig und in etwa gleichbleibend dar.



Krhs.=Krankenhauseinweisungen

**Abbildung 8: Anteil Patienten mit und ohne Krankenhauseinweisungen in 2008 nach Polypharmazie-Gruppe (Voranalysen I-III)**

Bei der vergleichenden Gegenüberstellung der prozentualen Anteile an Krankenhauseinweisungen generell über die Polypharmazie-Gruppen hinweg, zeigen sich mit zunehmender Anzahl an Arzneistoffen pro Quartal steigende Anteile von Patienten mit und abnehmende Anteile von Patienten ohne eine Krankenhauseinweisung (Abbildung 9).



intern.=internistische; Krhs.=Krankenhauseinweisungen; sonst.=sonstige

**Abbildung 9: Anteil Patienten nach Art der Krankenhauseinweisung in 2008 und nach Polypharmazie-Gruppe (Voranalysen I-III)**

In den verschiedenen Polypharmazie-Gruppen mit mehr als 5, 9 beziehungsweise 13 gleichzeitig verordneten Arzneistoffen pro Quartal über 4 aufeinanderfolgende Quartale, verteilen sich die verschiedenen Arten an Krankenhauseinweisungen prozentual ähnlich, aber tendenziell ansteigend. Internistische Krankenhauseinweisungen stellen in den ATC-Gruppen 5-8 ATC beziehungsweise 9-12 ATC mit 45% und 48% grob die Hälfte der Krankenhauseinweisungen dar, neben 55% und 51% Krankenhauseinweisungen, die weder internistische noch Notfalleinweisungen sind. In der Polypharmazie-Gruppe  $\geq 13$  ATC dreht sich das Verhältnis um. Hier überwiegen mit 57%

die internistischen Einweisungen vor den Krankenhauseinweisungen generell mit 43%. In einem sehr viel geringeren Ausmaß, aber ebenfalls über die Polypharmazie-Gruppen hinweg erkennbar ansteigend, treten Notfalleinweisungen mit 0,35%, 0,41% beziehungsweise 0,51% auf.

### *Diskussion (Voranalysen I-III)*

Ausreichende Patientenzahlen und Verteilung der ATC-Gruppen:

Bei der Betrachtung von Patienten mit 5 oder mehr verschiedenen, verordneten Arzneistoffen über 4 zusammenhängende Quartale bilden Patienten mit 5-8 ATC mit 94% den größten beziehungsweise fast ausschließlichen Anteil im betrachteten Patientenpool. Nur bei Einbeziehung dieser Gruppe findet sich damit eine ausreichende Anzahl an Patienten mit so definierter Polypharmazie für eine Modellbildung.

Eignung der dynamischen Betrachtung für die Bestimmung der Polypharmazie:

Die dynamische Ermittlung der letzten 4 aufeinanderfolgenden Quartalen mit mindestens 5 verschiedenen Wirkstoffen pro Quartal im Zeitraum 2005-2007 führt zu uneinheitlichen Stichtagen pro Patient, von denen aus Krankenhauseinweisungen innerhalb eines Jahres als Zielmerkmal bestimmt werden sollen. Die Verteilung der Stichtage zeigt, dass ein kleinerer Teil der Patienten eine Polypharmazie in 2005 oder 2006 aufwies, aber nicht mehr in den aktuelleren Daten von 2007. Diese Patienten sind für die Zielgruppe weniger relevant, da für eine Beratung eine aktuelle Polypharmazie vorliegen soll. Aus diesem Grund werden im Folgenden ausschließlich die letzten, aktuellen Quartale zur Bestimmung einer Polypharmazie ab einem einheitlichen Stichtag vorgenommen. Wie viele Quartale dabei optimaler Weise betrachtet werden sollten, wird in den nachfolgenden Analysen untersucht.

Festlegung einer Anzahl an Arzneistoffen für die Bestimmung der Polypharmazie:

Bei der gleichzeitigen Verordnung von 5 oder mehr verschiedenen Arzneistoffen pro Quartal über 4 aufeinander folgende Quartale ( $\geq 5$  ATC) scheint sich die Anzahl an Krankenhausfällen generell innerhalb von 12 Monaten im Gegensatz zu der Gruppe  $< 5$  ATC etwa zu verdoppeln (0,36 vs. 0,76; entsprechend dem 2,1-fachen). Bezogen auf internistische Krankenhauseinweisungen zeigt sich etwa eine Verdreifachung (0,10 vs. 0,28; entsprechend dem 2,8-fachen). Es wird daher als sinnvoll betrachtet,



die fachliche und technische Definition für Polypharmazie bereits ab einer Anzahl von 5 gleichzeitig verordneten Arzneistoffen weiter zu verfolgen.

Die Anzahl der Krankenhausfälle bei Patienten mit einer Verordnung von weniger als 5 Arzneistoffen wurde daher nicht weiter, beispielsweise differenziert in 1-3 ATC und 3-5 ATC, untersucht. Die Analyseergebnisse für die Patienten mit mehr als 5 verordneten Arzneistoffen zugleich zeigen in der Gruppe 9-12 ATC zwar eine weitere Zunahme der Krankenhausfälle innerhalb eines Jahres (0,76 versus 0,87; + 15%), diese wird jedoch als zu geringfügig angesehen, um bei der technischen Definition der Polypharmazie von der auch in der Literatur am häufigsten zugrunde gelegten Anzahl von 5 Arzneistoffen abzuweichen. Theoretisch wäre zu erwarten, dass mit jedem zusätzlichen Arzneistoff die Anzahl an möglichen UAE und damit wiederum die Anzahl an Krankenhausfällen exponentiell steigt. Dieses ist zwischen der Gruppe  $\geq 5$  ATC und der Untergruppe mit 9-12 ATC nicht ohne weiteres erkennbar.

Die Anzahl der generellen und internistischen Krankenhauseinweisungen, von denen letztere eher mit arzneimittelbedingten Einweisungen in Verbindung gebracht werden (Schneeweiss et al., 2002), scheint ab 13 gleichzeitig verordneten Arzneistoffen pro Quartal im Vergleich zur Gruppe 9-12 ATC deutlich anzusteigen (+28% beziehungsweise +15%). Damit wäre neben 5 auch die Verwendung von 13 verschiedenen Arzneistoffen als Mindestanzahl für die Definition der Polypharmazie denkbar. Allerdings sinkt die Anzahl der Patienten mit identifizierbaren, internistischen Krankenhauseinweisungen in dieser Polypharmazie-Gruppe stark, wodurch die Datengrundlage für die Entwicklung eines Vorhersagemodells als zu gering eingeschätzt wird. Daher werden 5 Arzneistoffe pro Quartal als geeigneter Grenzwert für die Definition und die Bestimmung einer Polypharmazie im Rahmen dieser Untersuchung herangezogen.

Die Betrachtung der Krankenhauseinweisungen im Zusammenhang mit einem Notfall trägt für die große Masse der Patienten wenig für die Festlegung auf eine Mindestanzahl an Arzneistoffen zur Bestimmung einer Polypharmazie bei. Die Darstellung der prozentualen Verteilung der verschiedenen Krankentypen zeigt zwar mit zunehmender Anzahl an Arzneistoffen den erwarteten Anstieg, die Werte erscheinen aber unzuverlässig und aufgrund der sehr geringen, zum Teil nur noch einstelligen Patientenzahlen in den Polypharmazie-Gruppen nicht für die Modellierung nutzbar.

### Bewertung möglicher Zielmerkmale:

Neben den Krankenhauseinweisungen generell wurden internistische Krankenhaus- und Notfalleinweisungen betrachtet, um ihre Eignung als mögliche alternative Zielmerkmale zu prüfen. Die durchschnittlichen Häufigkeiten für generelle sowie auch für internistische Krankenhauseinweisungen und damit für eine potenzielle ABH, steigt bei internistischen Einweisungen über die Polypharmazie-Gruppen hinweg stärker (2,8-, 3,3- beziehungsweise 6,7-facher Anstieg) als bei der generellen, undifferenzierten Betrachtung von Krankenhauseinweisungen (2,1- , 2,4 beziehungsweise 4,8-facher Anstieg ). Eine extremere Ausprägung ist für die Verwendung als Zielmerkmal günstiger, weil zweckdienlich für die Erkennung von Patienten mit klinisch-pharmazeutischem Betreuungsbedarf. Aufgrund der niedrigeren Patientenzahlen mit internistischen Krankenhauseinweisungen, insbesondere in den Gruppen mit zunehmenden Arzneistoffen pro Quartal, wird jedoch mit Problemen bei der Modellerstellung gerechnet und von der Verwendung als Zielmerkmal abgesehen. Geringe Patientenzahlen zeigen sich noch extremer bei den Notfalleinweisungen. Der Einsatz von Notfalleinweisungen als Zielmerkmal wird daher ebenfalls nicht weiter erwogen.

Bei der Bewertung der Analyseergebnisse, besonders in den Gruppen mit nur geringen Patientenzahlen und nur geringen Differenzen zwischen den Ergebnissen, wäre die statistische Signifikanz zu überprüfen. Da es sich um Voranalysen zur groben Einschätzung des Patientenpools handelt, wird an dieser Stelle darauf verzichtet. Eine Signifikanzprüfung erfolgt nach endgültiger Festlegung der Zielgruppen-Definition.

### *Fazit (Voranalysen I-III):*

Für die technische Definition der Polypharmazie werden 5 verschiedene, verordnete Arzneistoffe pro Quartal als Mindestanzahl festgelegt. Die Betrachtung der Quartale wird rückblickend ab einem einheitlichen Auswertungstichtag vorgenommen. Die optimale Anzahl der insgesamt zu betrachtenden Quartale ist zu ermitteln.

Als Zielmerkmal eignet sich die Bestimmung von Krankenhausweinweisungen innerhalb eines Jahres. Internistische oder Notfalleinweisungen eignen sich aufgrund der geringen Häufigkeiten im Datenpool weniger für die Verwendung als Zielmerkmal bei der Modellbildung.

*Voranalysen IV-V:*

Betrachtung von 5 Wirkstoffen über 8 und 12 Quartale:

Um zu prüfen, ob mit zunehmender Dauer einer Polypharmazie die Anzahl an Krankenhauseinweisung weiter zunimmt, soll im Folgenden die Betrachtung von 5 oder mehr gleichzeitig verordneten Arzneistoffen über 4 Quartale hinaus erfolgen. Dadurch sollen der zugrunde gelegte Datensatz mit GKV-Routinedaten von 2005-2008 besser ausgenutzt und gegebenenfalls Auswirkungen eines größeren Betrachtungszeitraumes erkannt werden.

Durchführung (Voranalysen IV-V)

Die Datenbasis, die Eingrenzung des Datenpools und die Art der Durchführung erfolgt wie in den Voranalysen I-III. Die Mindestanzahl der verordneten Arzneistoffe pro Quartal wird für die Bestimmung einer Polypharmazie auf 5 oder mehr fixiert (Abbildung 10). Die Länge des Betrachtungszeitraumes wird von 4 Quartalen erhöht auf:

- **8 Quartale**, das heißt von 2006 bis 2007 (*Voranalyse VI*),
- **12 Quartale**, das heißt von 2005 bis 2007 (*Voranalyse V*).

Die Prüfung auf Polypharmazie erfolgt vom Stichtag 31.12.2007 rückblickend betrachtet, die Prüfung auf Krankenhausaufenthalte erfolgt vorausschauend für 2008.

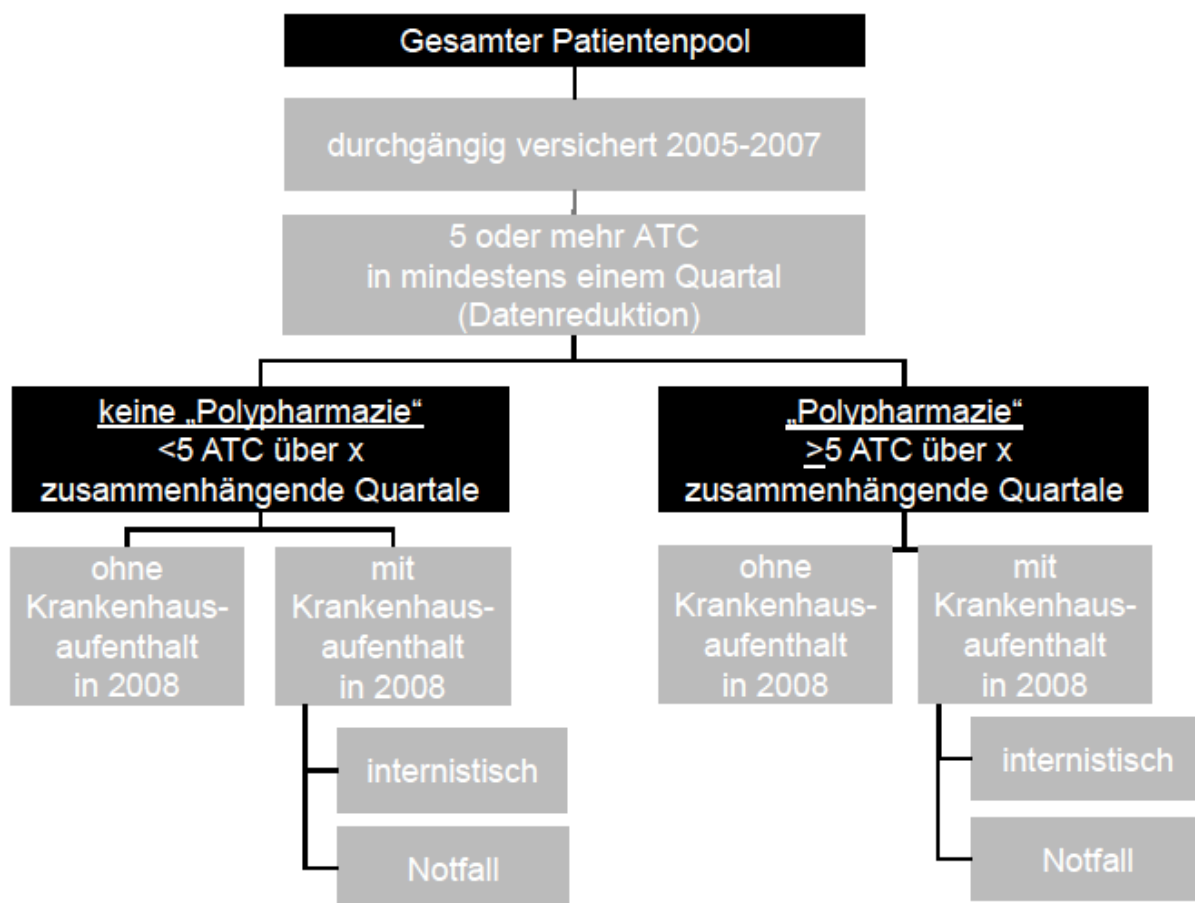
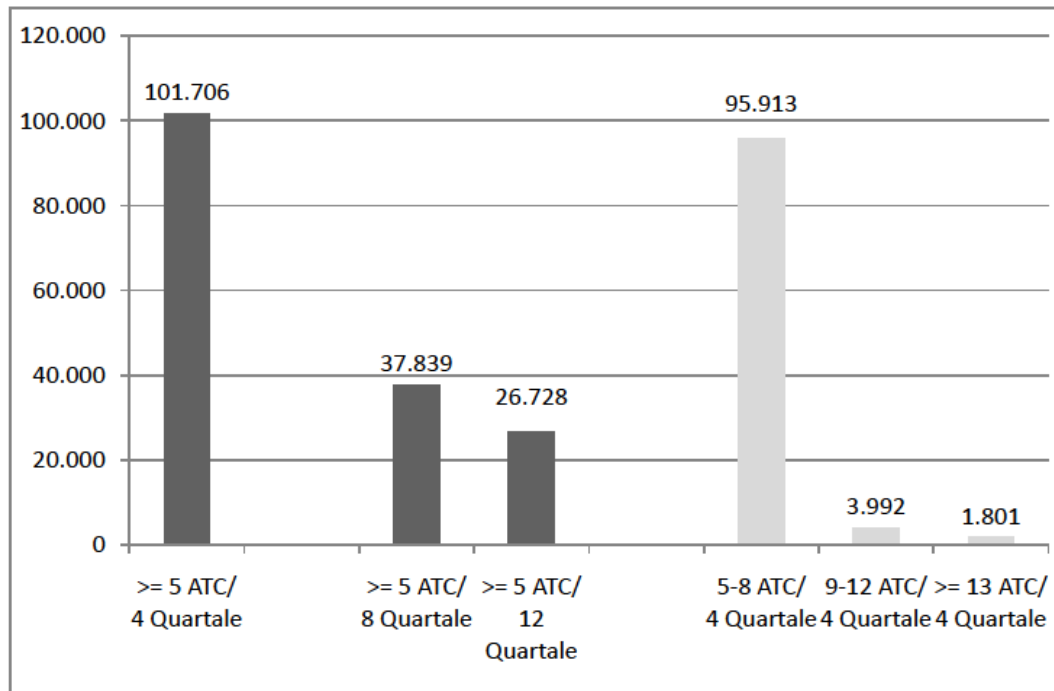


Abbildung 10: Ablauf der Datenselektion aus dem Gesamtpool (Voranalysen IV-V)

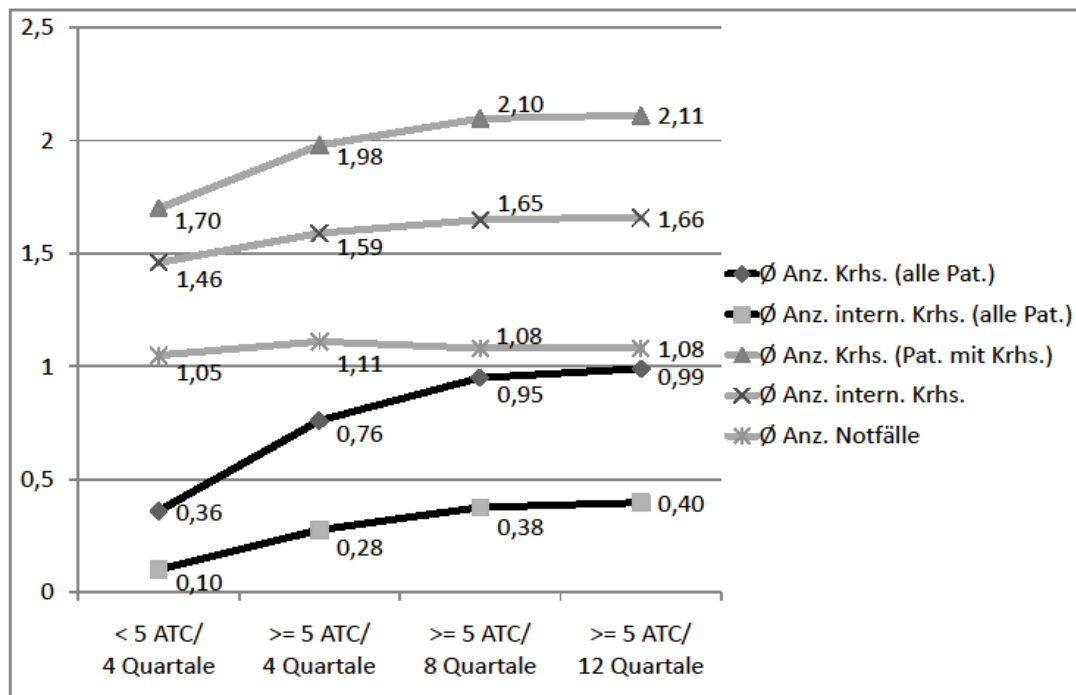
### Ergebnisse (Voranalysen IV-V)

Nachfolgend sind die wichtigsten Ergebnisse dargestellt (Abbildung 11). Die vollständige Datentabelle mit einer Übersicht über die Ergebnisse der Voranalysen findet sich im Anhang (Anhang 10.5).



**Abbildung 11: Anzahl Patienten nach Polypharmazie-Gruppe bei variierenden Betrachtungszeiträumen und ATC/Quartal in 2005-2007 (Voranalysen I-V)**

Bei vergleichender Betrachtung der Polypharmazie-Gruppe mit 5 oder mehr gleichzeitig verordneten Arzneistoffen pro Quartal (Abbildung 11) über 4 Quartale gegenüber der Betrachtung von 8 beziehungsweise 12 Quartalen sinken die Patientenzahlen um 63% (von 101.706 auf 37.839) beziehungsweise um 74% (von 101.706 auf 26.728). Insgesamt liegen die Patientenzahlen in allen Gruppen im 5-stelligen Bereich. Daneben sind zum Vergleich die absoluten Patientenzahlen bei Variation der Mindestanzahl an gleichzeitig verordneten Arzneistoffen dargestellt. Bei höherer Mindestanzahl an Arzneistoffen sinken die Patientenzahlen um 96% (von 101.706 auf 3.992) beziehungsweise um 98% (von 101.706 auf 1.801) von 5- auf 4-stellige Zahlenwerte und damit vergleichsweise deutlich stärker.



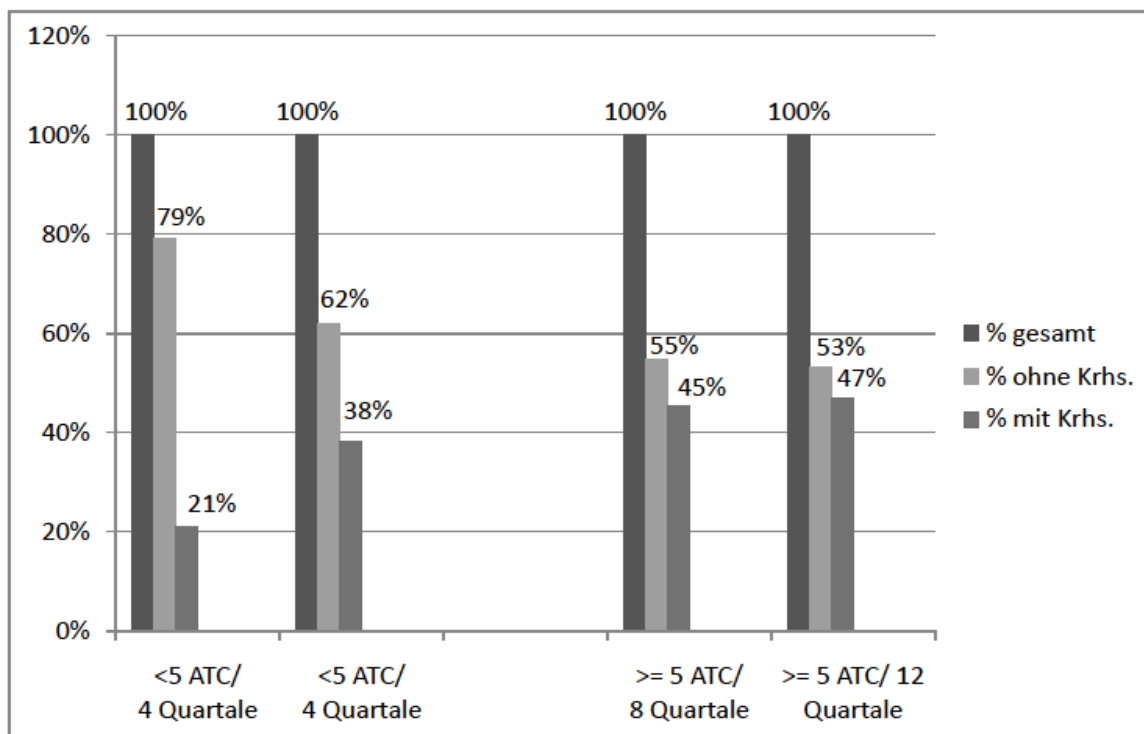
**Abbildung 12: Mittlere Anzahl Krankenhauseinweisungen in 2008 nach Polypharmazie-Gruppe bei variierenden Beobachtungszeiträumen (Voranalysen IV-V)**

Die beiden Polypharmazie-Gruppen, die über 4 Quartale hinweg betrachtet wurden, sind nicht direkt mit den anderen beiden Gruppen vergleichbar, da die Datenbasis 4 Quartale in einem Zeitraum von 2005-2007 darstellte und nicht 2005 allein. Für das Erkennen von tendenziellen Entwicklungen werden die Werte jedoch vergleichend nebeneinander dargestellt (Abbildung 12).

Das Ergebnis der Voranalysen zeigt über alle Polypharmazie-Gruppen hinweg betrachtet mit zunehmend längerem Betrachtungszeitraum ansteigende durchschnittliche **Krankenhauseinweisungen** (0,36 vs. 0,76 vs. 0,95 vs. 0,99; entsprechend +111% vs. +25% vs. +4%). Insgesamt zeigt sich, dass es damit bei 4, 8 beziehungsweise 12 Quartalen zu einem 2,1-, 2,6 beziehungsweise 2,8-fachen Anstieg der durchschnittlichen Krankenhauseinweisungen kommt (0,76/0,36; 0,95/0,36; 0,99/0,36).

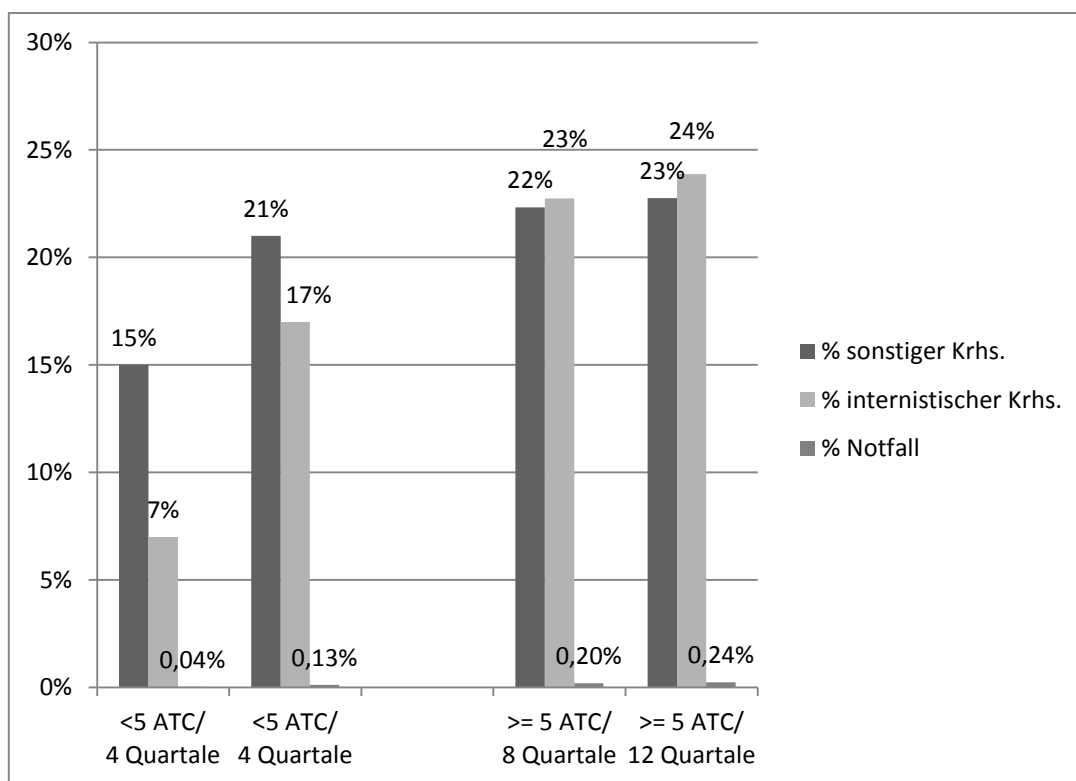
Ebenso zeigt auch die Betrachtung differenziert nach **internistischen Krankenhauseinweisungen** über alle Polypharmazie-Gruppen hinweg einen kontinuierlichen Anstieg (0,10 vs. 0,28 vs. 0,38 vs. 0,40; entsprechend + 180% vs. +36% vs. +5%). Dieses entspricht einem 2,8-, 3,8- beziehungsweise 4,0-fachem Anstieg (0,28/0,10; 0,38/0,10; 0,40/0,10) der durchschnittlichen Häufigkeiten an internistischen Krankenhauseinweisungen

Bei der differenzierteren Betrachtung der **Patienten mit mindestens einer Krankenhauseinweisung** und Unterscheidung zwischen generellen und internistischen Krankenhausaufenthalten innerhalb des ausgewerteten Jahres zeigt sich bei den Durchschnittswerten über alle betrachteten Polypharmazie-Gruppen hinweg das gleiche Bild ansteigender Werte (Abbildung 13). Die Durchschnittswerte liegen auf einem erwartungsgemäß höheren Niveau. Bei der Betrachtung von 12 Quartalen erscheint der Anstieg der Durchschnittswerte der Krankenhauseinweisungen generell gering (1,70 vs. 1,98 vs. 2,10 vs. 2,11; mit +16% vs. +6% vs. +0,5%). Dieses ist auch der Fall bei den internistischen Krankenhauseinweisungen (1,46 vs. 1,59 vs. 1,65 vs. 1,66 mit +9% vs. +4% vs. +0,6%). Die durchschnittliche Anzahl an Notfalleinweisungen ins Krankenhaus stellt sich über alle betrachteten Polypharmazie-Gruppen hinweg als insgesamt gleich dar.



**Abbildung 13: Anteil Patienten mit und ohne Krankenhauseinweisung in 2008 nach Polypharmazie-Gruppe bei variierenden Beobachtungszeiträumen (Voranalysen IV-V)**

Bei den Patientenanteilen mit und ohne Krankenhauseinweisung in einem Jahr zeigt sich ein ähnliches Bild wie bei der Variation der Mindestanzahl an Arzneistoffen pro Quartal. Mit zunehmender Länge des Auswertungszeitraumes steigt der Anteil an Patienten mit einer Krankenhauseinweisung.



**Abbildung 14: Anteil Patienten nach Art der Krankenhauseinweisungen in 2008 bei variierenden Beobachtungszeiträumen (Voranalysen IV-V)**

Bei Betrachtung nach Art der Krankenhauseinweisungen (Abbildung 14) bleibt der Anteil an Krankenhauseinweisungen pro Jahr bei Ausweitung des Beobachtungszeitraumes von 4 auf 8 und 12 Quartale in etwa gleich. Mit Ausweitung des Beobachtungszeitraumes von 4 auf 8 Quartale steigt dagegen der Anteil an internistischen Krankenhaus- und Notfalleinweisungen, bei Betrachtung über 12 Quartale ergibt sich kaum eine Änderung. Die Gruppenbelegungen sind allerdings bei 8 und 12 Quartalen nur noch 4- beziehungsweise 2-stellig.



### *Diskussion der Ergebnisse (Voranalysen IV-V):*

#### Ausreichende Patientenzahlen:

Die Ergebnisse der Voranalyse zeigen, dass bei Ausweitung der Betrachtungszeiträume zur Bestimmung einer Polypharmazie die Patientenzahlen deutlich sinken, aber nicht in dem Ausmaß wie bei der Erhöhung der Mindestanzahl von gleichzeitig verordneten Arzneistoffen. Damit bleibt mehr Spielraum für die Entscheidung zur Festlegung auf einen der getesteten Auswertungszeiträume.

#### Auswahl eines Mindestzeitraums für die Bestimmung der Polypharmazie:

Mit der Erhöhung des Betrachtungszeitraumes von 4 auf 8 Quartale lassen sich häufigere Krankenhauseinweisungen vermuten, bei weiterer Ausdehnung des Betrachtungszeitraumes auf 12 Quartale steigt die Häufigkeit jedoch kaum noch. Prinzipiell wäre denkbar, die Prüfung auf Vorliegen einer Polypharmazie über 8 Quartale statt 5 Quartale anzustellen, um eine Patientengruppe mit möglichst vielen Krankenhausfällen für die Erstellung des Vorhersagemodells zur Verfügung zu haben. Wegen der geringeren Patientenzahlen bei der Betrachtung von 4 statt 8 beziehungsweise 12 Quartalen, soll jedoch die Betrachtung von 4 Quartalen zur Bestimmung einer Polypharmazie im weiteren Verlauf beibehalten werden. Dieses stellt einen Kompromiss dar, um ausreichend hohe Patientenzahlen für die Durchführung einer Data-Mining Analyse und die Erstellung eines Vorhersagemodells mit einer akzeptablen Güte zur Verfügung zu haben.

#### Bewertung möglicher Zielmerkmale:

Bei Ausweitung des Betrachtungszeitraumes für die Bestimmung einer Polypharmazie von 4 auf 8 und 12 Quartale sind keine nennenswerten Zunahmen von absoluten oder prozentualen Häufigkeiten der Krankenhauseinweisungen, insbesondere der internistischen Krankenhaus- und Notfalleinweisungen, zu erkennen. Zudem sinken die Patientenzahlen deutlich. Bei der Betrachtung von 4 Quartalen im Vergleich zu 8 Quartalen verringert sich die Patientengruppe bereits um 63% (von 101.706 auf 37.839), bei den Patienten mit einer generellen beziehungsweise internistischen Krankenhauseinweisungen nochmals mit 56% beziehungsweise 51% um etwa die Hälfte (39.064 vs. 17.133 beziehungsweise 17.611 vs. 8.608). Ebenso wie bei den

Voranalysen I-III erscheinen die Patientenzahlen trotz Ausweitung des Beobachtungszeitraumes zu gering, um internistische Krankenhaus- oder Notfalleinweisungen als mögliche Zielmerkmale zu erwägen.

#### *Fazit (Voranalysen I-V)*

Polypharmazie wird für die folgenden Untersuchungen technisch definiert als die gleichzeitige Verordnung von 5 oder mehr verschiedenen Arzneistoffen pro Quartal über mindestens 4 zusammenhängenden Quartale ab einem zu definierenden Auswertungstichtag.

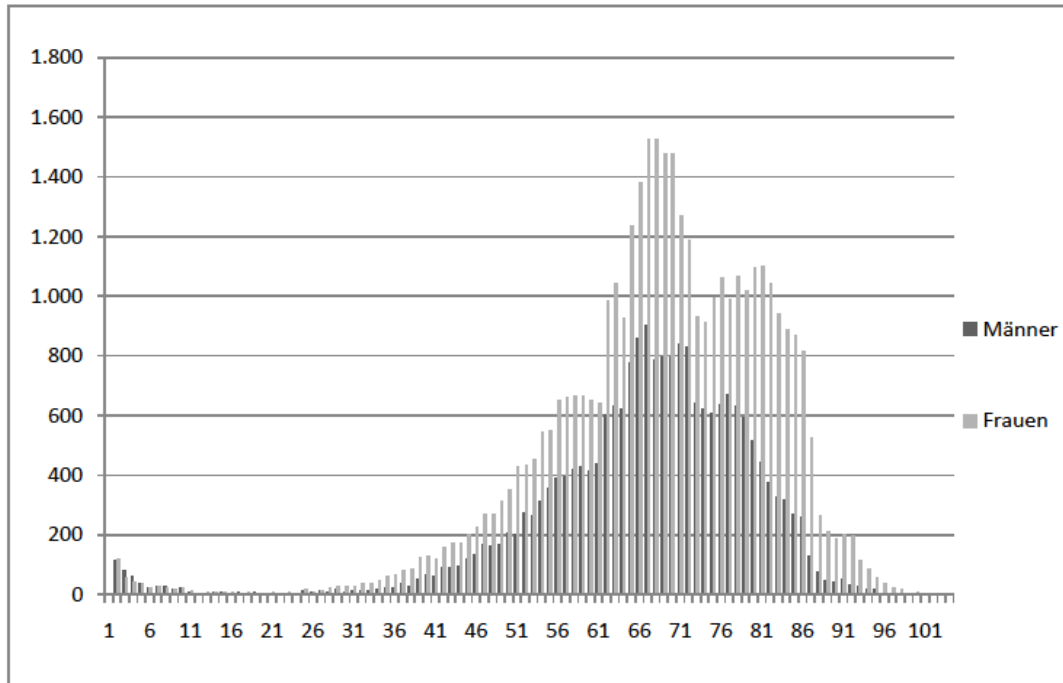
#### **4.2 Alters- und Geschlechtsverteilung bei Polypharmazie**

Zur Überprüfung, ob als Einschlusskriterium ein begrenzter Altersbereich ausgewählt werden sollte, wird nach Festlegung der Definition für Polypharmazie mittels der Voranalysen I-V nun die Altersverteilung in der Zielgruppe überprüft.

#### *Durchführung (Voranalyse VI)*

Als Datengrundlage dient folgender Datenauszug:

- Auswertungstichtag 31.12.2007
- durchgängig versicherte Patienten in 2005-2007
- (keine Alterseinschränkung)
- Polypharmazie im Sinne von mindestens 5 verschiedene ATC in 4 zusammenhängenden Quartalen in 2007 (rückblickend ab dem Stichtag 31.12.2007)
- n=61.352

*Ergebnis (Voranalyse VI)*

**Abbildung 15: Anzahl Patienten mit Polypharmazie nach Alter und Geschlecht in 2007 (Voranalyse VI)**

Patienten mit Polypharmazie finden sich in dem untersuchten Patientenpool aus den Jahren 2005-2007 in jedem Alter zwischen 2 bis 105 Jahren und haben ein Durchschnittsalter von 69 Jahren (Abbildung 15). Mit 72% (44.107 von 61.352) stellt die Altersgruppe von 60-85 Jahren den überwiegenden Teil der Patienten mit Polypharmazie dar. In der deutschen Gesamtbevölkerung stellt diese Altersgruppe zum Vergleich zirka 25% dar (20 Mio. vs. 82 Mio.; Bundesamt, 2009).

Differenziert nach Männern und Frauen zeigt sich jeweils eine ähnliche Verteilungskurve, wobei der Anteil der Frauen bei Patienten mit Polypharmazie in fast allen Altersgruppen überwiegt. Dieses trifft insbesondere für die höheren Altersstufen zu. Unter den Patienten mit Polypharmazie insgesamt findet sich ein Verhältnis zwischen Frauen und Männern von  $w:m=1,87$  (39.543 vs. 21.809). Im Altersbereich 60-85 Jahre liegt dieses bei  $w:m=1,74$  (28.017 vs. 16.090). Im Vergleich dazu liegt das Verhältnis in der deutschen Gesamtbevölkerung insgesamt bei  $w:m=1,04$  beziehungsweise bei 1,13 für 70-jährige und 1,51 für 80-jährige (Bundesamt, 2009).

*Fazit: (Voranalyse VI)*

Übereinstimmend mit den Hinweisen aus der Literatur zeigt sich eine Häufung von Patienten mit Polypharmazie in der Altersgruppe von 60-85 Jahren, die überwiegend weiblich sind. Da dennoch möglichst auch junge Erwachsene in die Analyse einbezogen werden sollen, wird die untere Altersgrenze bei 18 Jahren definiert. Nach oben wird das Alter auf 85 Jahre begrenzt, um möglichst viele von Polypharmazie betroffene Patienten einzuschließen. Das ist möglich, da Schwerkranke durch entsprechende Ausschlusskriterien bereits unberücksichtigt gelassen werden.

**4.3 Charakteristische Arzneistoffe**

Durch weitere Voranalysen sollen einzelne Arzneistoffe beziehungsweise Arzneistoffgruppen nach ATC ermittelt werden, die für Patienten mit Polypharmazie im Zusammenhang mit Krankenhauseinweisungen charakteristisch sind und in die Modellentwicklung eingehen sollen. Die Ergebnisse der Voranalysen sollen die in der *Business Understanding*-Phase dargestellten Erwartungen zu relevanten Arzneistoffen bestätigen.

*Durchführung (Voranalyse VII)*

Die Datenbasis wird unter Verwendung der entwickelten Definition für Polypharmazie folgendermaßen gebildet:

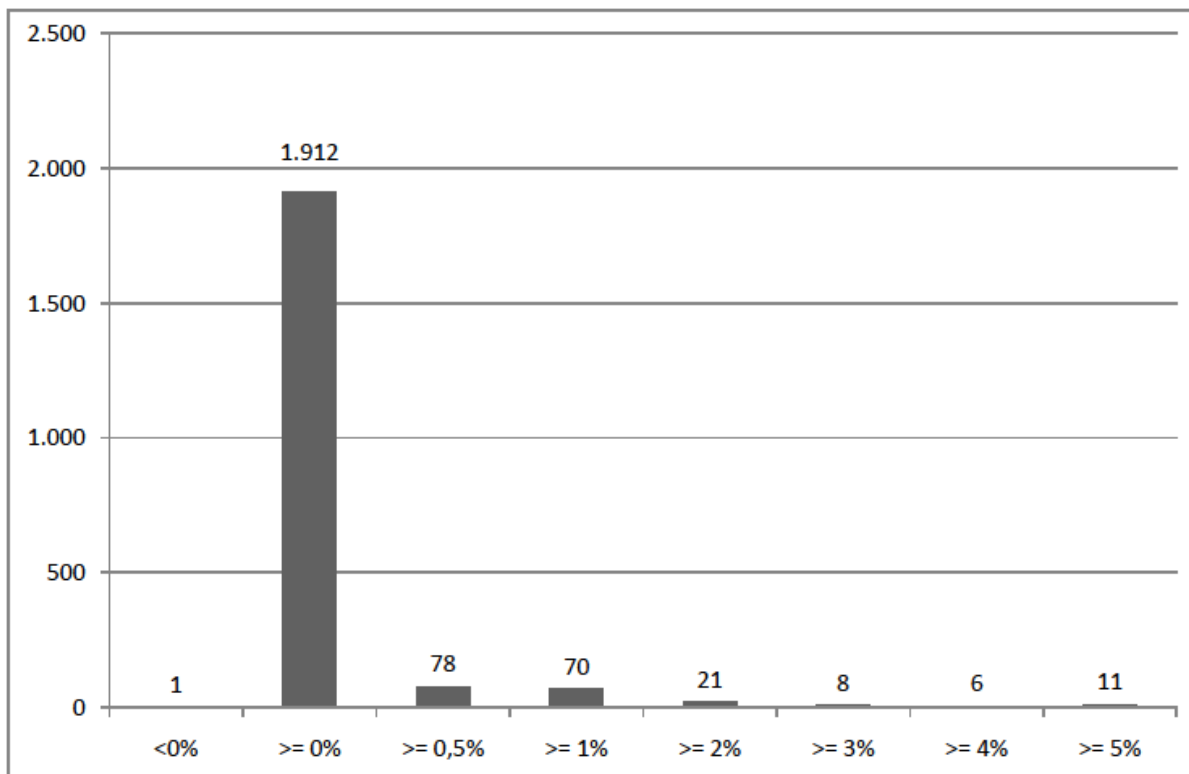
- Auswertungstichtag und Stand der Daten: 31.12.2007
- Patient ist durchgängig versichert in 2005-2007
- (keine Altersbeschränkung)
- Ermittlung einer Polypharmazie, das heißt von Patienten mit mindestens 5 verschiedenen Wirkstoffen pro Quartal in allen 4 Quartalen in 2007
- Rezeptdaten von 2007 zur Ermittlung der verordneten ATC
- Krankenhausdaten von 2008 zur Ermittlung der Krankenhausaufenthalte

Der Auswertungszeitraum für die Betrachtung der ATC wird für diese Voranalyse von 3 Jahren (2005-2007) auf ein Jahr (2007) verkürzt, um bei Betrachtung der Krankenhausaufenthalte im Folgejahr (2008) auf lediglich die im Vorjahr angewendeten Arz-

neistoffe zu fokussieren. Für die Patienten mit Polypharmazie werden alle verordneten Arzneistoffe nach ATC insgesamt und differenziert nach Patienten mit und ohne Krankenhausaufenthalt selektiert. Dieses geschieht mittels SQL-Abfrage im Data Warehouse des Datenlieferanten nach den oben genannten Vorgaben. Die sich ergebenden ATC werden durch die Autorin manuell gruppiert nach auffälligen absoluten Häufigkeiten und gleichzeitig hohen prozentualen Differenzen im Vergleich zwischen den beiden Patientengruppen. Dabei werden die gemäß Fachliteratur als relevant erwarteten ATC bevorzugt betrachtet und ausgewählt.

### *Ergebnisse (Voranalyse VII)*

Es wurden innerhalb des Auswertungszeitraumes, dem Jahr 2007, bei den betrachteten 61.352 Patienten 2.107 unterschiedliche ATC ermittelt.

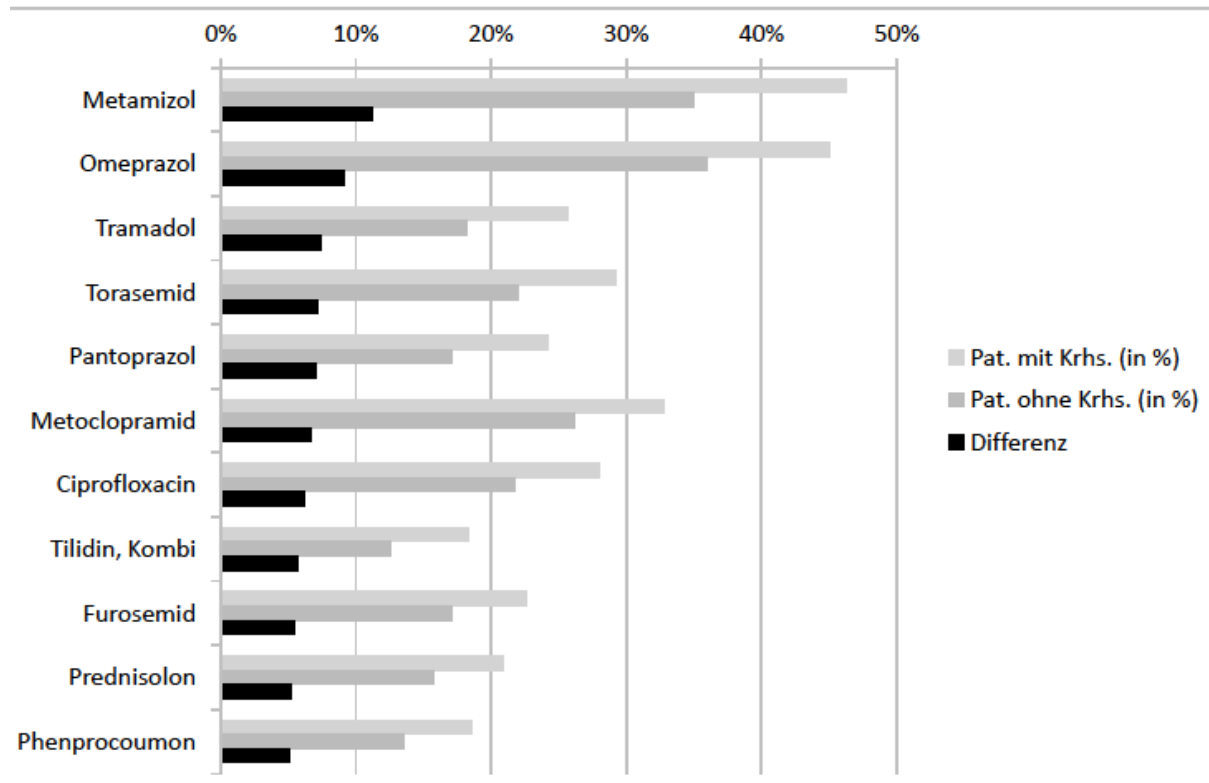


**Abbildung 16: Anzahl ATC nach Verordnungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhauseinweisung in 2008 (Voranalyse VII)**

Bei 90% der ATC (1.912 von 2.107) zeigen sich mit 0 bis 0,5% vergleichsweise geringe prozentuale Abweichungen zwischen den Verordnungen in 2007 für Patienten

mit und ohne Krankenhausaufenthalt in 2008 (Abbildung 16).

Vergleichend dargestellt sind im Folgenden die ATC mit den größten prozentualen Differenzen und damit möglicherweise einem Zusammenhang zu Krankenseinweisungen bei Patienten mit Polypharmazie (Abbildung 17).



Pat.=Patienten; Krhs.=Krankenhausaufenthalt

**Abbildung 17: Auffälligste ATC nach Verordnungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenseinweisung in 2008 (Voranalyse VII)**

Der Wirkstoff Metamizol zeigt beim Vergleich von Polypharmaziepatienten mit und ohne Krankenhausaufenthalt einen Unterschied von +11% (46% vs. 35%) und damit die größte Differenz. Den kleinsten Unterschied (nicht abgebildet) zeigt sich mit -5% (21% vs. 26%) für den Wirkstoff Metformin.

## Zusammenstellung von ATC-Merkmalen:

Von den 2.107 verordneten Arzneistoffen in der betrachteten Patientengruppe werden 52 Arzneistoffe als auffällig im Zusammenhang mit einem Krankenhausaufenthalt betrachtet und daraus **26 ATC-Merkmale**, 16 aus einzelnen und 10 aus kombinierten ATC, gebildet (Tabelle 6):

Nr.	Variable	Arzneistoff (Differenz in %) **	Differenz in % (Gruppe) **
1.	ATC_GRP_N02BB02	Metamizol (11%)	11%
2.	ATC_GRP_A02BC01	Omeprazol (9%) + Pantoprazol (7%)	7-9%
3.	ATC_GRP_A03FA01	Metoclopramid (7%)	7%
4.	ATC_GRP_N02A_SPEZ	Tramadol (7%), Tilidin-Kombinationen (6%) oder Fentanyl (3%)	3-7%
5.	ATC_GRP_C03CA*	Furosemid (5%) oder Torasemid (7%)	5-7%
6.	ATC_GRP_B01AA04*	Phenprocoumon (5%)	5%
7.	ATC_GRP_B01AC04*	Clopidogrel (4%)	4%
8.	ATC_GRP_C03DA01	Spironolacton (3%)	3%
9.	ATC_GRP_C09AA05*	Ramipril (3%)	3%
10.	ATC_GRP_J01EE01	Sulfamethoxazol+ Trimethoprim (3%)	3%
11.	ATC_GRP_NSAR_SPEZ*	Diclofenac oder Ibuprofen (3%) oder ASS (2%)	2-3%
12.	ATC_GRP_N_SPEZ	Gabapentin oder Pregabalin (3%) oder L- Dopa + Benserazid (2%)	2-3%
13.	ATC_GRP_C01AA04*	Digitoxin (2%)	2%
14.	ATC_GRP_C01DX12	Molsidomin (2%)	2%
15.	ATC_GRP_J01_SPEZ	Clindamycin oder Levofloxacin oder Moxiflo- xacin (2%)	2%
16.	ATC_GRP_M04AA01	Allopurinol (2%)	2%
17.	ATC_GRP_N02BG07	Flupirtin (Analgetikum, 2%)	2%
18.	ATC_GRP_R05CB01	Acetylcystein (2%)	2%
19.	ATC_GRP_N05_SPEZ	Zopiclon (2%) oder Zolpidem oder O-	1-2%

		xazepam oder Lorazepam oder Bromazepam oder Promethazin (1%)	
20.	ATC_GRP_N06_SPEZ	Amitriptylin oder Citalopram oder Mirtazapin (2%) oder Doxepin oder Pipofezin (1%)	1-2%
21.	ATC_GRP_R03_SPEZ	Tiotropium (2%) oder Salmeterol-Kombi oder Formoterol(-Kombi) oder Fenoterol oder Formoterol + Budesonid (1%)	1-2%
22.	ATC_GRP_C01BD01	Amiodaron (1%)	1%
23.	ATC_GRP_R03DA04	Theophyllin (1%)	1%
24.	ATC_GRP_S01_SPEZ*	Prednisolon oder Dexamethason + Antiinfektiva (als Ophthalmika, 1%)	1%
25.	ATC_GRP_B01AC05*	Ticlopidin (0%)	0%
26.	ATC_GRP_A10BA02*	Metformin (-5%)	- 5%

\*=Auffällige Arzneistoffe bei Krankenhauseinweisungen gemäß Literatur (Keller, 2006; Schneeweiss et al., 2002);

\*\*=%: ermittelte prozentuale Differenz zwischen Polypharmaziepatienten mit bzw. ohne Krankenhausaufenthalt in 2008

**Tabelle 6: Gebildete ATC-Merkmale für Patienten mit Polypharmazie und Krankenhauseinweisungen (Voranalyse VII)**

#### *Diskussion (Voranalyse VII):*

Die Ergebnisse der Voranalyse lassen zum Teil die erwarteten statistischen Auffälligkeiten zwischen einzelnen Arzneistoffen und häufigen Krankenhauseinweisungen sichtbar werden. Bei 9 von 17 in der herangezogenen Literatur genannte ATC können bereits in der Voranalyse Auffälligkeiten identifiziert werden (Tabelle 7). Dazu zählen insbesondere **Phenprocoumon** (5%), **Schleifendiuretika**, wie Torasemid (7%) und Furosemid (6%), und **Clopidogrel** (4%). Allerdings fällt auf, dass **Insulin** mit 0% Differenz zwischen den Gruppen mit und ohne Krankenhausaufenthalt zahlenmäßig unauffällig bleibt, während für Insulin in der Literatur eine gehäufte Anwendung im Zusammenhang mit ABH beschrieben wird (Keller, 2006).



Nr.	ATC	Arzneistoff	Keller, 2006 *	Schneeweiss et al., 2002 **	Vor- analyse
1.	A10A	Insulin	X	X	
2.	B01AA04	Phenprocoumon	X	(X)	X
3.	diverse	Antithrombotika	X	X	
4.	C03	Diuretika	X	X	X
5.	C07AA07	Sotalol	X		
6.	B01AB/AC	TAH (inkl. Clopidogrel)	X	(X)	X
7.	C01AA	Digoxin, Digitoxin	X	X	(X)
8.	C07	Betablocker	X	X	
9.	A10B	Orale Antidiabetika	X	X	
10.	M01A	NSAR	X	X	(X)
11.	C08C/D	Calcium-Antagonisten		X	
12.	C09A	ACE-Hemmer / Sartane		X	(X)
13.	C01DA	Nitrate		X	
14.	H02A	Corticosteroide, systemisch		X	
15.	N02BB02	Metamizol			X
16.	A02BC	PPI			X
17.	N02A_Spez	Opioide			X

\*=Keller, 2006 (S. 91); \*\*=Schneeweiss et al., 2002 (S. 285-291); X=Auffälligkeit; (X)=schwach ausgeprägte Auffälligkeit

**Tabelle 7: Vergleich auffälliger ATC aus der Literatur und in Voranalyse VII**

Ebenfalls abweichend von den Hinweisen in der Literatur finden sich mit den **Protonenpumpenhemmern (PPI)** Omeprazol (9%) und Pantoprazol (7%) vermeintlich sichere, nebenwirkungsarme Wirkstoffe in den oberen Rängen der statistisch auffälligen ATC (Tabelle 7). Als risikoreicher einzustufende NSAR zeigen sich dagegen mit einer Differenz von 2% nur mäßig auffällig in der Voranalyse (Tabelle 6). **Metamizol** (11%) stellt sich im Analyseergebnis unter den auffälligen ATC als Spitzenreiter dar. Die herangezogene Literatur (Keller, 2006; Stausberg & Hasford, 2010, Schneeweiss et al., 2010) erwähnt Metamizol jedoch nicht als statistisch auffällig im Zusammenhang mit Krankenhauseinweisungen. Zu ABH bei Anwendung von **Opioiden**, wie Tramadol (7%) und Tilidin (6%), finden sich Hinweise, insbesondere im Rahmen ei-

ner Medikamentenabhängigkeit (Stausberg & Hasford, 2010). Bei Tramadol und Tildin handelt es sich zwar um schwächere Opioide, allerdings mit bekanntem Missbrauchs- und Gefährdungspotenzial. Das statistisch am wenigsten hervorstechende **Metformin** (-5%) gilt zwar beispielsweise in Bezug auf Hypoglykämien als sicherer in der Anwendung bei Diabtes als Insulin (0%), jedoch überrascht die Position am Ende der Tabelle (Tabelle 6). Insulin dagegen ist als potenter Wirkstoff sicherlich risikoreicher als physiologische Kochsalzlösung (+2%). Letztere zeigt im Analyseergebnis jedoch größere Auffälligkeiten.

Die erhaltenen statistischen Verhältnisse zwischen vermeintlich sicheren und unsicheren Arzneistoffen (Abbildung 17) stimmen im Analyseergebnis der Voranalyse insgesamt nicht durchgängig mit den Erwartungen überein. Unter Umständen ist der gewählte Auswertungszeitraum von einem Jahr für die Ermittlung der ATC zu weit gefasst, um im Analyseergebnis klare Signale der erwarteten Risikoarzneistoffe im Zusammenhang mit Krankenhauseinweisungen erkennen zu können. Die ATC werden daher zwar vorrangig nach auffälligen prozentualen Differenzen aber auch nach den Erwartungen aus der Literatur gruppiert. Die absolute Häufigkeit einer ATC wird weniger gewichtet. Inwiefern die definierten ATC-Merkmale zur Bildung eines guten Modells beitragen können, wird sich bei der Testung verschiedener Modelle zeigen und in diesem Rahmen detaillierter überprüft. Mit 8 von 26 ATC-Merkmalen werden zirka ein Drittel der gebildeten Merkmale übereinstimmend mit den Hinweisen aus der Literatur ausgewählt (siehe \* in Tabelle 6). **Metformin** (-5%) wird wegen der niedrigsten Differenz als zu testendes Merkmal aufgenommen. **Ciprofloxacin** war auffällig, wurde aber nicht als ATC-Merkmal ausgewählt, da der Einsatz als Antibiotikum vermutlich kurzzeitig erfolgt und eine Intervention den Patienten zu spät erreichen wird. Daher soll auf typische Dauermedikation fokussiert werden. Um die gebildeten ATC-Gruppen bei der Modellierung bei Bedarf spontan weiter aggregieren zu können, wurde durch die Autorin eine **Matrix** erstellt, die eine Übersicht über fachlich sinnvoll erscheinende Variationsmöglichkeiten bietet (Anhang 10.8).

### *Fazit (Voranalyse VII):*

Auf Grundlage der Literaturrecherche und der Voranalyse werden ATC-Merkmale für Patienten mit Polypharmazie gebildet und Kombinationsmöglichkeiten in einer Matrix zusammengestellt. Diese sollen bei der Modellierung auf ihren Nutzen bei der Vorhersage von Krankenhauseinweisungen in der Zielgruppe untersucht werden.

#### **4.4 Charakteristische Diagnosen**

Ebenso wie für die Arzneistoffe werden Diagnosen, für die ein Zusammenhang mit ABH vermutet wird, in einer Voranalyse näher untersucht und gruppiert.

#### *Durchführung (Voranalyse VIII)*

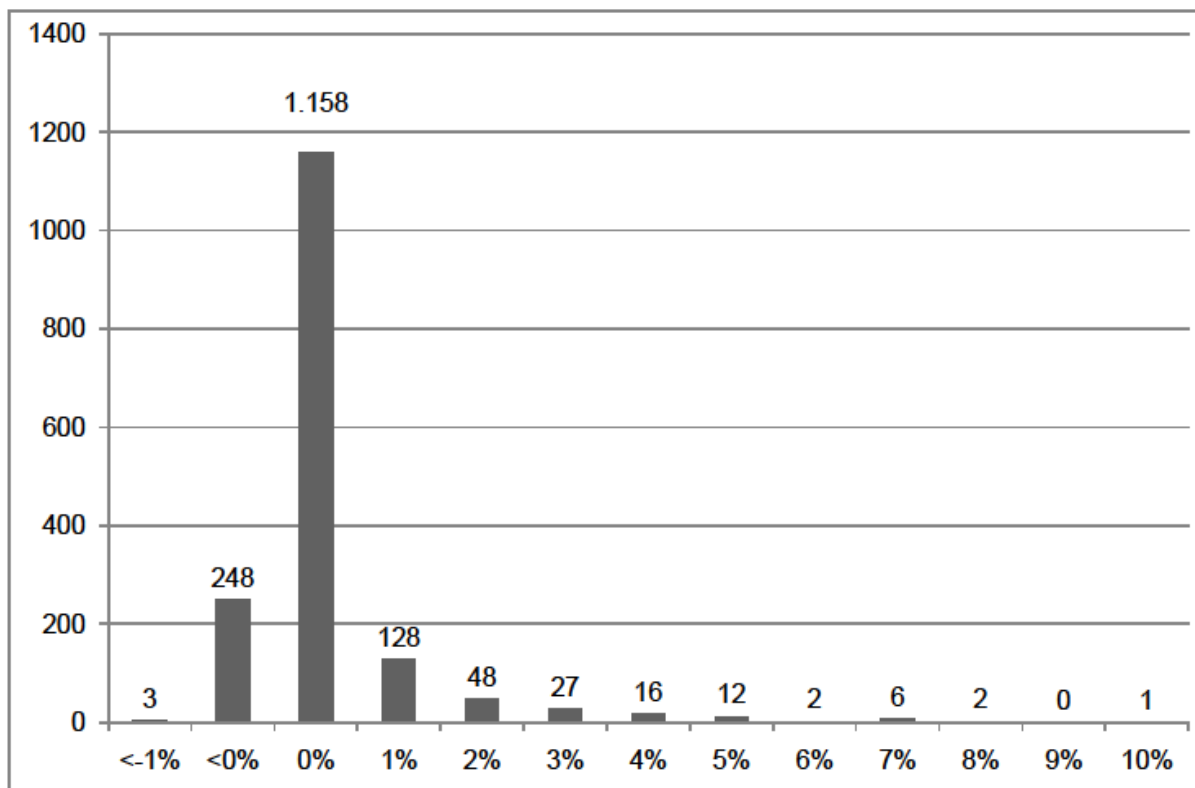
Die Datenbasis wird unter Verwendung der entwickelten Definition für Polypharmazie folgendermaßen gebildet:

- Auswertungstichtag und Stand der Daten: 31.12.2007
- Patient ist durchgängig versichert in 2005-2007
- (keine Altersbeschränkung)
- Rezeptdaten von 2007 zur Ermittlung einer Polypharmazie, das heißt von Patienten mit  $\geq 5$  verschiedenen Wirkstoffen pro Quartal in allen 4 Quartalen
- Ambulante und stationäre Diagnosedaten aus 2007 zur Ermittlung der codierten Diagnosen nach ICD
- Krankenhausdaten von 2008 zur Ermittlung der Krankenhausaufenthalte

Für die betrachtete Patientengruppe werden insgesamt, sowie differenziert nach Patienten mit und ohne Krankenhausaufenthalt, alle ambulant und stationär codierten Diagnosen nach ICD und die zugehörige Anzahl der Patienten mittels SQL-Abfrage im Data Warehouse des Datenlieferanten gemäß der oben genannten Vorgaben selektiert. Die sich ergebenden ICD werden durch die Autorin nach auffälligen Differenzen der prozentualen Häufigkeiten beim Vergleich zwischen den Patienten mit und ohne Krankenhausaufenthalt ausgewählt und gruppiert. Dabei werden die gemäß Fachliteratur erwarteten ICD bevorzugt berücksichtigt.

*Ergebnisse (Voranalyse VIII)*

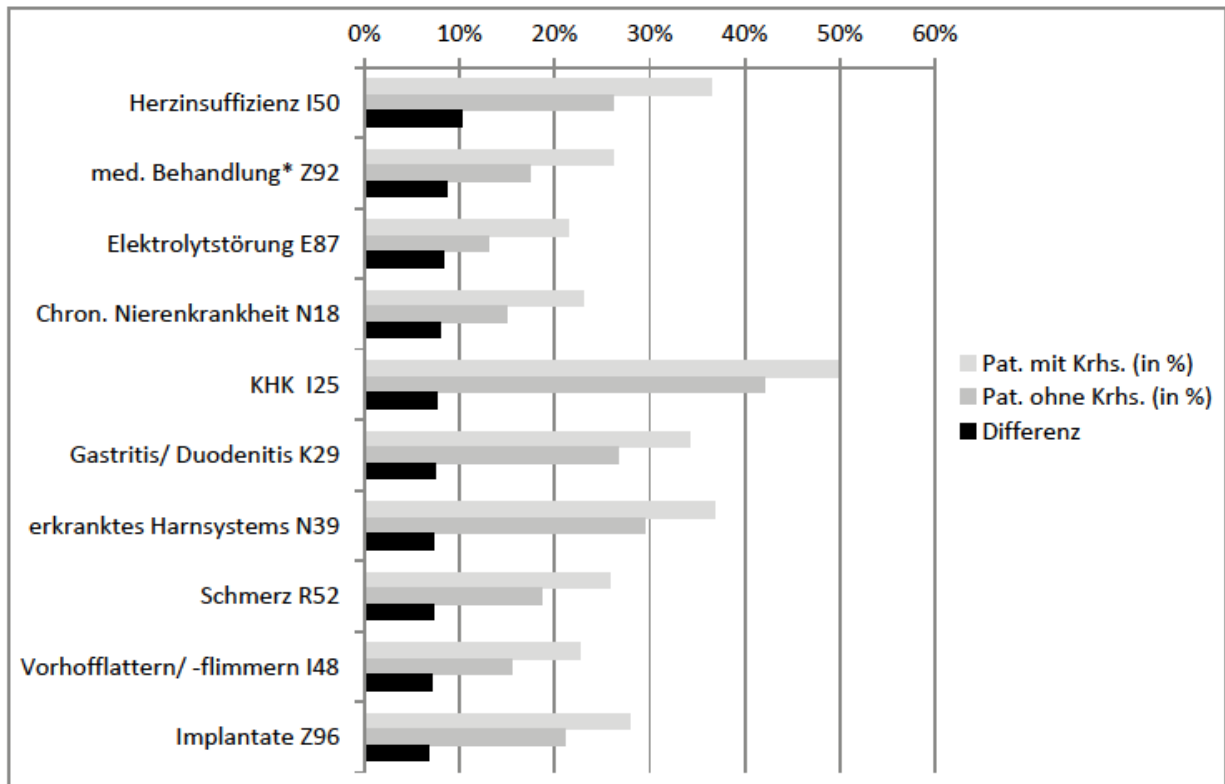
Es werden innerhalb des Auswertungszeitraumes und der betrachteten 61.352 Patienten insgesamt 1.651 unterschiedliche 3-stellige ICD ermittelt.



Hinweis: Die Zuordnung der ICD zu einer der angegebenen Differenzen in % erfolgt unter Vernachlässigung der Nachkommastellen. Beispiel: ICD mit einer Differenz von 1,6% ist unter 1% subsumiert.

**Abbildung 18: Anzahl ICD nach Codierungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhauseinweisung in 2008 (Voranalyse VIII)**

Bei 70% der vorkommenden ICD (1.158 von 1.651) zeigen sich mit 0 bis 0,9% nur geringe Unterschiede zwischen den prozentualen Häufigkeiten der Patienten mit und ohne Krankenhausaufenthalt (Abbildung 18). Eine kleine Gruppe von 1% der ICD (11 von 1.651) zeigen jedoch vergleichsweise große Differenzen zwischen 6% und 10%.



Pat.=Patienten; Krhs.=Krankenhausaufenthalt, \* = in Eigenanamnese

**Abbildung 19: Auffälligste ICD nach Codierungsunterschieden in 2007 bei Polypharmaziepatienten mit und ohne Krankenhausaufenthalt in 2008 (Voranalyse VIII)**

Von den 1.651 beobachteten ICD in der betrachteten Patientengruppe werden 116 ICD als auffällig im Zusammenhang mit einem Krankenhausaufenthalt betrachtet. Abgebildet ist die Top 10 (Abbildung 19). Aus den 116 verschiedenen ICD werden 20 ICD-Merkmale, 3 aus einzelnen und 17 aus kombinierten ICD, gebildet (Tabelle 8).

Nr.	Variable	Feldbeschreibung	Differenz in %**
1.	ICD Grp I	<b>Herzinsuffizienz I50</b> (10%), <b>KHK I25</b> (8%), Hypertensive Herzerkrankung I11 oder Angina pectoris I20 (6%) oder weitere Herzerkrankungen: I27, I34, I35, I51 (2-3%)	10%
2.	ICD Grp Z	<b>medizinische Behandlung in der Eigenanamnese Z92</b> , inkl. Arzneimittel wie ASS (9%), <b>Implantate Z95/Z96</b> (7%) oder weitere: Z03, Z51, Z74, Z85, Z90, unterstützende Apparate Z99 (2-3%)	9%
3.	ICD Grp E8*	<b>Elektrolytstörungen E87</b> (8%) oder Volumenmangel E86* (4%)	8%
4.	ICD Grp N*	<b>Chronische Nierenerkrankung N18*</b> (8%), <b>erkranktes Harnsystems N39</b> , inkl. unklare Harnwegsinfektion und Inkontinenz (7%) oder weitere, wie Glomeruläre Krankheiten N08, unklare Niereninsuffizienz N19, Zystitis N30, oder Prostatahyperplasie N40 (2-4%)	8%
5.	ICD Grp K*	<b>Gastritis K29</b> , inkl. akute hämorrhagische Gastritis (7%) oder <b>funktionelle Darmstörung K59</b> , inkl. Obstipation und Diarrhö (5%), sonstige nichtinfektiöse Gastroenteritis und Kolitis K52* und weitere K57, sonstige, inkl. Magenblutungen K92* (2-3%)	7%
6.	ICD Grp R*	Unspezifische Symptome: <b>Schmerz R52*</b> (7%), Kurzatmigkeit R06* (6%) oder weitere: Herz-/Brustschmerzen R00/R07, Bauchschmerzen/Übelkeit R10/R11*, Stuhl-/Harnstörung R15/R32/R39, Schwindel R26/R42*, Unwohlsein/Kollaps R53/R55, Ödem R60 (2-4%)	7%
7.	ICD Grp I4	<b>Vorhofflattern I48</b> (7%) oder kardiale Arrhythmien I49 (6%) oder AV-Block I44 (2%)	7%
8.	ICD Grp F*	Akute Depression F32 (6%) oder weitere: rezidivierende Depression F33, Belastungs-/ Somatoforme Störungen F43/F45 oder organische psychische Störungen F03, F06 (2-3%)	6%
9.	ICD Grp I6-I9*	Atherosklerose I70 (6%) oder sonstige periphere Gefäßerkrankung I73 (5%) oder weitere: I65, I67, I69, I83, I87, I89, Hypotonie I95* (2-3%)	6%
10.	ICD Grp J	COPD J44 (6%) oder respiratorische Insuffizienz J96, Pneumonie J18, Bronchitis J20 (2-3%)	6%

11.	ICD Grp M	Spondylose M47 (6%), Osteoporose M81 (5%) oder weitere: M06, 13, 15-17, 19, 25, 42, 48, 51, 53, 54, 62, 75, 79, 79, 99 (2-4%)	6%
12.	ICD Grp G	Schlafstörungen G47 (5%) oder Polyneuropathien G62-63 (3-4%) oder G20, G25, G40, G45, G55, G56, G81, G93 (1%)	5%
13.	ICD Grp D*	Anämien D50, D62, D64, D68 (3-4%)	4%
14.	ICD Grp A*	Infektiöse Erkrankungen: Gastroenteritis und Kolitis A09 (3%)	3%
15.	ICD Grp E7	Purin-Pyrimidinstoffwechselstörung E79 (3%)	3%
16.	ICD Grp H	Augenerkrankungen beziehungsweise Katarakt H04, H25, H26, H35, H52, H53 (2-3%) oder Hörschwäche H61, H90, H91 (2%)	3%
17.	ICD Grp T*	Verletzungen T14 oder Komplikationen nach Eingriffen T81, T82, T88* (2-3%)	3%
18.	ICD Grp B	Kandidosen (B37) oder Streptokokken oder Staphylokokken-Infektion (B95) (2%)	2%
19.	ICD Grp E	Sonstige Hypothyreosen E03 (2%)	2%
20.	ICD Grp L*	Dermatitis L23 oder Dekubitus L89 oder weitere: Pruritus L29, Ekzeme L30, Ulcus/Unterhauterkrankung L97/L98 (1-2%)	2%

\*=Auffällige Arzneistoffe bei Krankenhauseinweisungen gemäß Literatur (Stausberg & Hasford,2010; Schneeweiss et al., 2002)

\*\*=Differenz der prozentualen Häufigkeit der ICD in 2007 bei Patienten mit beziehungsweise ohne Krankenhausaufenthalt in 2008

**Tabelle 8: Gebildete ICD-Merkmale für Patienten mit Polypharmazie und Krankenhausaufenthalten (Voranalyse VIII)**

*Diskussion (Voranalyse VIII):*

Nr.	ICD	Diagnose (verkürzt)	1)	2)	3)	Analyse
1.	A04*	Enterokolitis (Clostridium difficile)		X		(X)
2.	D	<b>Blutbildstörungen</b>	X	X	X	X
3.	E	Stoffwechsel-/Elektrolytstörungen	X		X	X
4.	F	psychische Reaktionen	X	X		X
5.	G	Nervenerkrankungen	X			X
6.	H	Sonstiges: Hör/Sehstörungen	X			(X)
7.	I1	Hypertonie	X			
8.	I2-5	Herzerkrankungen (KHK, HI)	X			X
9.	I4	Arrhythmien	X		X	X
10.	I6-9	<b>Gefäßerkrankungen**</b>	X	X	X	X
11.	J	Lungenerkrankungen	X			X
12.	K2/5	<b>Gastrointestinale Reaktionen</b>	X	X	X	X
13.	K7	Leberfunktionsstörungen	X		X	X
14.	L27	Dermatitis durch Arzneimittel	X	X		(X)
15.	M	Muskel-/Skeletterkrankungen				X
16.	N	Nierenfunktionsstörungen	X			X
17.	R	Allgemeinsymptome	X		X	X
18.	T	Vergiftungen	X	X		(X)
19.	T	Allergische Reaktionen	X			
20.	Y57	Unerwünschte Nebenwirkungen		X		
21.	Z92	Behandlung (Eigenanamnese)				X
22.	Z96	Vorhandensein eines Implantats				X

1) Keller, 2006; 2) Stausberg & Hasford, 2010; 3) Schneeweiss et al., 2002

\*=Vergleiche mit K5 „pseudomembranöse Kolitis“; \*\*=inklusive Hypotonie, sowie zerebrovaskuläre, arterielle und venöse Gefäßerkrankungen; (X)=schwach ausgeprägte Auffälligkeit

**Tabelle 9: Vergleich auffälliger ICD aus der Literatur und in Voranalyse VIII**



Unter den auffälligsten ICD bei Patienten mit Polypharmazie und Krankenhausaufenthalten sind Erkrankungen, die gemäß Literatur mit ABH in Zusammenhang gebracht werden (Tabelle 9). Dieses sind *kardiovaskuläre Störungen*, *Elektrolystörungen*, *Nierenfunktionsstörungen* und *gastrointestinale Störungen* (Schneeweiss et al., 2002; Stausberg & Hasford, 2010). Denkbar wäre bei diesen ein kausaler Zusammenhang mit Diuretika beziehungsweise NSAR, die derartige Erkrankungen als UAW auslösen können und ebenfalls auffällig häufig bei Patienten mit Polypharmazie und Krankenhausaufenthalten in den Voranalysen aufgetreten sind. Mit *N18 chronische Niereninsuffizienz* ist eine weitere ICD-Analyse auffällig, die im Hinblick auf ABH gut erklärbar ist. Bei eingeschränkter Nierenfunktion besteht eine erhöhte Arzneimittelempfindlichkeit gegenüber renal eliminierten Wirkstoffen und damit eine höhere Wahrscheinlichkeit für UAW und damit auch ABH. Mit *I95 Hypotonie* und *K29 Gastritis* beziehungsweise *K92 Magenblutungen* werden ICD ermittelt, die mit den von Stausberg & Hasford genannten ICD *I95 Hypotonie* und *K52 Gastroenteritis und Kolitis* nahezu in Deckung gebracht werden können. Bei genauerer Überprüfung der ICD-Systematik bezüglich der Codierung einer Magenblutung, ergibt sich, dass mehrere Möglichkeiten zur Codierung bestehen. Eine Gastritis kann unter anderem als *K29 akute hämorrhagische Gastritis*, *K92.2 nicht näher bezeichnete gastrointestinale Blutung* oder *K52.1 toxische Gastroenteritis* codiert werden. Es ist daher davon auszugehen, dass beispielsweise arzneimittelbedingte Magenblutungen uneinheitlich codiert werden und dadurch zusätzlich schwer zu detektieren sind. Damit wäre erklärbar, dass für die ermittelten auffälligen ICD wenig exakte, jedoch viele übereinstimmende, sinnverwandte Diagnosen in der Literatur gefunden werden.

Unerwartet ist die Auffälligkeit der ICD *Z92 medizinische Behandlung in Eigenanamnese*. Der Code schließt sehr unterschiedliche medizinische Behandlungen ein, was die Interpretation erschwert. Es fällt allerdings die „*Dauertherapie mit Acetylsalicylsäure*“ darunter. Fraglich ist, ob die Selbstmedikation mit ASS auf diese Weise codiert wird und durch die Auffälligkeit der ICD eine erhöhte Wahrscheinlichkeit für ABH abgeleitet werden kann. Weitere Analysen wären notwendig, um diese Vermutung zu bestätigen.

Vor dem Hintergrund der Ergebnisse aus den Voranalysen zu den ATC werden einige ICD vermisst. So steht Metamizol, das durch eine seltene, aber unter Umständen schwerwiegende Agranulozytose zu Krankenseinweisungen führen kann, in der ATC-Analyse auf dem ersten Rang der auffälligen ATC. In der ICD-Analyse ist je-

doch *D70 Agranulozytose* mit 0,7% Differenz zwischen den Patientengruppen nicht auffällig. Codierungsfehler sind wegen der eindeutigeren ICD-Auswahl unwahrscheinlicher. Eine auffällige ATC stellt außerdem Ciprofloxacin dar, dass eine pseudomembranöse Kolitis auslösen und als *A07.4 Enterokolitis durch Clostridium difficile* codiert werden kann. Bei der Analyse zur ICD-Gruppierung waren zwar die *A09 Infektiöse Erkrankungen: Gastroenteritis und Kolitis* und die *K29 Gastritis*, jedoch nicht die *A07* auffällig. Hier könnten erneut uneinheitliche Codierungen eine Rolle spielen.

Einige ICD zeigen negative Differenzen und scheinen damit nicht im Zusammenhang mit Krankenseinweisungen zu stehen. Bei genauerer Betrachtung dieser ICD handelt es sich entweder um keine oder nur geringfügige Erkrankungen, wie *Z00 Allgemeinuntersuchung* und *J00 Erkältungsschnupfen* oder um präventive Maßnahmen wie *Z24, Z27 Impfungen gegen verschiedene Infektionskrankheiten*. Die Ergebnisse erscheinen damit plausibel.

Zu den ICD, die zur Identifizierung von ABH eine wichtige Rolle spielen sollten, zählen *Y57 Unerwünschte Nebenwirkung bei der therapeutischen Anwendung von Arzneimitteln* und *T36-T50 Vergiftungen mit Arzneimitteln*. Mit diesen ist die Verschlüsselung eindeutiger arzneimittelbedingter Probleme möglich. Dementsprechende Hinweise finden sich in der Literatur (Stausberg & Hasford, 2010). Beide ICD sind bei der Voranalyse zu den ICD unauffällig. Allerdings sind im Unterschied zu den Angaben aus der herangezogenen Literatur in die Voranalyse nicht nur stationäre, sondern auch ambulante Diagnosen eingeflossen. Diese könnten die erwarteten ICD-Auffälligkeiten überlagern. Denkbar wäre auch, dass der kausale Zusammenhang von Erkrankungen mit Arzneimitteln nicht erkannt und entsprechend nicht mit *Y57* oder *T36-T50* codiert wird.

Ebenso wie bei der Voranalyse zur Untersuchung charakteristischer ATC fällt auf, dass sich für häufige ICD von Volkskrankheiten, wie Herzinsuffizienz, KHK, Lungenerkrankungen und Osteoporose, auch auffällig hohe Differenzen zwischen den Patientengruppen mit und ohne Krankenhausaufenthalt finden. Dieses könnte ein Ausdruck höherer Morbidität sein und die gesuchten arzneimittelbedingten Erkrankungen störend überlagert haben.

### *Fazit (Voranalyse VIII):*

Die auf Grundlage der Literaturrecherche und der Voranalyse gebildeten ICD-Merkmale für Patienten mit Polypharmazie und einem Krankenhausaufenthalt sollen bei der Modellierung auf ihren Nutzen untersucht werden. Gegenüber den ICD sollen ATC-Merkmale bei der Modellbildung jedoch vorrangig zum Einsatz kommen, um einen stärkeren Arzneimittelbezug zu gewährleisten.

### **4.5 Prüfung des Zielmerkmals auf Signifikanz**

Nachdem mittels der Voranalysen eine Definition für eine Polypharmazie festgelegt worden ist, gilt es vor Beginn der Modellbildung abzuschätzen, ob das ausgewählte Zielmerkmal (Krankenhausaufenthalt im Folgejahr) bei Patienten mit Polypharmazie generell signifikant häufiger eintritt als bei Patienten ohne Polypharmazie. Ein signifikanter Unterschied beim Vorkommen von Krankenhauseinweisungen würde zum Ausdruck bringen, dass das gewählte Zielmerkmal ein relevantes Unterscheidungsmerkmal darstellt und als Zielmerkmal zur Entwicklung eines Vorhersagemodells geeignet ist. Als Zielmerkmal wird das Auftreten einer *Krankenhauseinweisung innerhalb von einem Jahr ab Auswertungstichtag (ja /nein)* festgelegt und nachfolgend überprüft.

### *Durchführung (Voranalyse IX):*

Patienten mit Polypharmazie werden von einem einheitlichen Stichtag aus rückblickend ermittelt. Vom diesem Stichtag aus ein Jahr vorausschauend wird das Vorkommen von Krankenhauseinweisungen bestimmt (Abbildung 20).

### Datengrundlage:

- Stichtag und Stand der Daten: 31.12.2007
- alle durchgängig versicherten Patienten in 2007
- (keine Alterseinschränkung)

- Rezeptdaten von 2007 zur Ermittlung einer Polypharmazie, das heißt von Patienten mit  $\geq 5$  verschiedenen Wirkstoffen pro Quartal in allen 4 Quartalen (Gruppe A: Polypharmazie; Gruppe B: keine Polypharmazie)
- Krankenhausdaten von 2008 zur Ermittlung der Krankenhausaufenthalte

Fall	2005				2006				2007				2008			
	Nr.	Q1*	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3
I.	0	0	0	0	0	6	9	9	10	10	13	13	keine Polypharmazie			
II.	0	0	5	3	7	5	7	8	8	4	4	5				
III.	5	0	5	5	0	5	5	5	5	5	0	5				
IV.	5	5	7	8	8	5	4	4	5	7	5	7				

\*=Q=Quartal; Zahlenwerte=Anzahl ATC pro Quartal; hellgrau=berücksichtigtes Quartal mit  $\geq 5$  ATC (Polypharmazie definiert als  $\geq 5$ ATC/Quartal über 4 zusammenhängenden Quartale ab 31.12.2007); dunkelgrau=Auswertungszeitraum bezüglich Zielmerkmal (Krankenhausaufenthalt)

**Abbildung 20: Darstellung der Polypharmazie-Ermittlung mit einheitlichem Stichtag 31.12.2007 (Datenbasis 2007)**

*Ergebnis (Voranalyse IX):*

	Gruppe A	Gruppe B	Gesamt
Anzahl Patienten	52.160 (100%)	1.021.560 (100%)	1.073.720 (100%)
Polypharmazie in 2007(>5 ATC/ Q1-4)	ja	nein	-
Krankenhauseinweisungen: Ja	21.294 (41%)	149.236 (15%)	170.530 (16%)
Krankenhauseinweisungen: Nein	30.866 (59%)	872.324 (85%)	903.190 (84%)

\*=Patienten haben möglicherweise mehr als einen Krankenhausfall im Jahr

**Tabelle 10: Ausprägung der Krankenhauseinweisungen in den Gruppe A und B (Voranalyse IX)**

In einer mittels SPSS 19.0 erstellten Kreuztabelle (Tabelle 10) stellt sich das relative Risiko für eine Krankenhauseinweisung bei Patienten mit Polypharmazie fast 3-fach höher dar ( $0,41/0,15=2,7$ ) als bei Patienten ohne Polypharmazie. Dieses entspricht den Erwartungen aus den Voranalysen I-III.

Test	Wert	Freiheits- grad	Asymptotische Signifikanz*	Exakte Signi- fikanz
Chi-Quadrat nach Pearson	25529,195	1	0,000	
Exakter Test nach Fisher				0,000

\*=2-seitig; \*\*=1-seitig

**Tabelle 11: Ergebnis der Prüfung auf Signifikanz (Voranalyse IX)**

Das Merkmal Krankenhauseinweisung ist binär (ja/nein), womit von einer Binominalverteilung auszugehen ist. Die Patientengruppen A und B sind unabhängig voneinander. Zur Signifikanzprüfung wird daher der „*Chi-Quadrat*“- Test und der „*Exakte Test nach Fisher*“ angewendet, da die Voraussetzungen für die Anwendung beider Tests erfüllt sind. Das Ergebnis des Chi-Quadrat-Tests ist wegen Verwendung einer Approximation an eine Normalverteilung bei der vorliegenden Binominalverteilung allerdings weniger genau.

Beide Tests zeigen im Ergebnis, dass sich die Patientengruppe mit Polypharmazie (Gruppe A) von der Patientengruppe ohne Polypharmazie (Gruppe B) in Bezug auf das Vorkommen von Krankenhauseinweisungen höchst signifikant ( $p < 0,001$ ) unterscheidet (Tabelle 11).

#### *Diskussion (Voranalyse IX):*

Mit der Signifikanzprüfung wird kein kausaler Zusammenhang nachgewiesen. Allerdings zeigt das Ergebnis, dass mit dem Zielmerkmal für die Zielgruppe ein gutes Differenzierungsmerkmal zum übrigen Patientenpool identifiziert werden konnte.

#### *Fazit (Voranalyse IX):*

Das Vorkommen von Krankenhauseinweisungen ist in der Zielgruppe im Vergleich zum übrigen Patientenpool höchst signifikant verschieden ( $p < 0,001$ ). Damit ist es als Zielmerkmal geeignet und wird als solches für die Modellbildung verwendet.

## 5 Datenzusammenstellung (Data Preparation)



### 5.1 Ausgewählte Variablen

Aufgrund der Ergebnisse aus den Voranalysen werden 64 Variablen aus den GKV-Routinedaten ausgewählt, die für die Modellbildung getestet werden sollen. Eine Übersicht über alle für den Ausgangsdatensatz festgelegten Variablen mit ausführlicher Feldbeschreibung, möglichen Ausprägungen und zugrundeliegenden Auswertungszeiträume findet sich im Anhang (Anhang 10.6). Die Nutzung der stationären Daten wird auf das Zielmerkmal beschränkt, das heißt, internistische Krankenhauseinweisungen und Notfalleinweisungen werden gestrichen (Tabelle 12).

Patientenstammdaten	Rezeptdaten	Stationäre Krankenhausdaten	Ambulante Arztdaten
Patientennummer	Anzahl DDD*	Krankenhauseinweisung	ICD-Merkmale (20x)
Mitgliedsstatus	Anzahl Verordnungen*		
Alter	Anzahl Rezepte*		
Geschlecht	Arzneimittelkosten*		
Familienmitglieder	Anzahl ATC*		
Bundesland	ATC-Merkmale (26x)		
Berufsgruppe			
Anzahl: 7	Anzahl: 36	Anzahl: 1	Anzahl: 20

Intern. Krhs. =internistische Krankenhauseinweisung; \*=Merkmal wird über ein Jahr und über 3 Jahre ermittelt

**Tabelle 12: Ausgewählte GKV-Routinedaten für die Modellierung**

### 5.2 Datenermittlung am Beispiel des Modellierungsdatensatzes

Wie in den Erläuterungen zur Methodik dargestellt, werden zwei Datensätze für die Untersuchung erstellt. Ein Datensatz dient zur Merkmalsauswahl und Modellbildung (Entwicklungsdatensatz), ein zweiter zur Validierung und Stabilitätsprüfung des Modells (Validierungsdatensatz). Die Datenermittlung und -zusammenstellung verläuft für beide Datensätze identisch, mit dem Unterschied, dass dem Entwicklungsdatensatz GKV-Routinedaten aus den Jahren 2005-2008 und dem Validierungsdatensatz Daten aus den Jahren 2007-2010 zugrunde liegen. Die Datenzusammenstellung wird im Folgenden am Beispiel des Ausgangsdatensatzes beschrieben.

### Durchführung der Datenzusammenstellung

Für die Datenzusammenstellung ist der Zugriff auf viele verschiedene Datentabellen aus verschiedenen Datenbanken notwendig. Daher erfolgt die Datenzusammenstellung durch Mitarbeiter im Data Warehouse des Datenlieferanten mittels SQL-Abfragen nach Vorgaben der Autorin. Für die Datenzusammenstellung sind mehrere Teilschritte erforderlich. Die folgende Abbildung zeigt, aus welchen Datenquellen der GKV-Routinedaten die Daten nach Filterung beziehungsweise durch Berechnungen ermittelt werden (Abbildung 21).

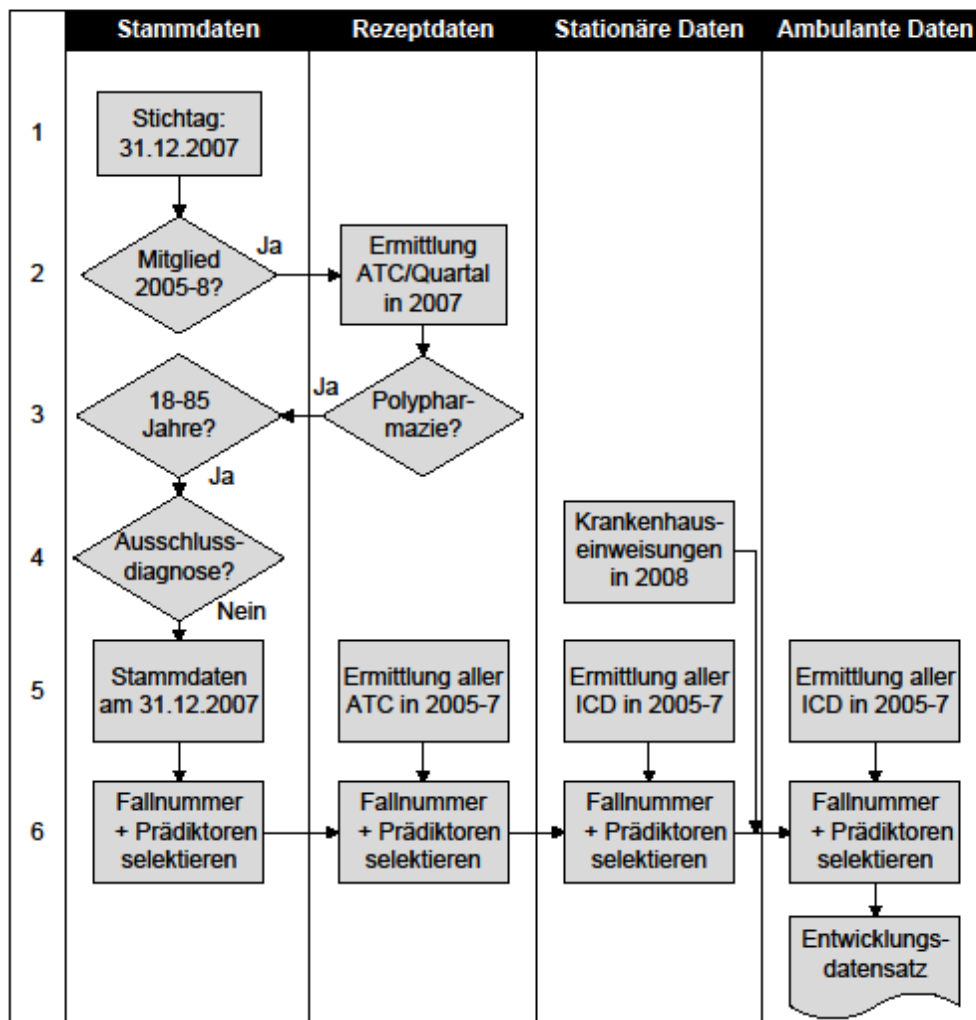


Abbildung 21: Ablauf der Datenzusammenstellung (Entwicklungsdatensatz: 2005-2008)

Alle für die Modellierung benötigten Datensätze mit den gewählten, vermeintlichen Prädiktoren aus verschiedenen Datenquellen der GKV-Routinedaten werden fallbezogen in einer Datentabelle zusammengeführt. Dabei werden folgende Bearbei-

tungsschritte bereits im Data Warehouse durchgeführt:

- Datenselektion und -reduktion:

Die Datenselektion erfolgt gemäß der in den Voranalysen entwickelten und durch die Autorin vorgegebenen Selektionsbedingungen für die ausgewählten Variablen. Dazu sind die Selektionsbedingungen in SQL-Abfragen zu übersetzen. Durch die Zielgruppendefinition, die festgelegten Ein- und Ausschlusskriterien, die ausgewählten Variablen und das definierte Zielmerkmal wird der Datenumfang für die Analysedatensätze bereits deutlich eingegrenzt.

- Datenformatierung:

Die separat ermittelten Falldaten zu den Variablen werden für die Zusammenführung in eine einzige Datentabelle einheitlich formatiert. Die Zusammenführung der verschiedenen Datenauszüge erfolgt über pseudonymisierte Patientennummern. Im Ausgangsdatensatz für die Modellierung werden alle Variablen in Spalten nebeneinander und die Patientendaten in Zeilen untereinander angeordnet, um eine spätere Weiterbearbeitung in SPSS 19.0 zu erleichtern.

- Datencodierung:

Die ursprünglichen technischen Variablennamen aus den GKV-Routinedaten werden umbenannt in eine geeignete Kurzform der Variablennamen aus der entworfenen Variablenübersicht. Die Variablenbeschreibung wird, wo technisch nötig, neu abgestimmt und bezüglich Feldbeschreibungen, verwendeten Datenquellen, den möglichen Ausprägungen und dem betreffenden Auswertungszeitraum des Datenauszugs angepasst.

Nach Überlassung der Datenauszüge durch den Datenlieferanten erfolgt die Weiterbearbeitung durch die Autorin:

- Datenüberführung und -einrichtung

Die Datentabellen für Entwicklungs- und Validierungsdatensatz werden nach SPSS 19.0 importiert. Die Variableneinstellungen, wie Spaltenbreite und Wertelabel werden überprüft und ein geeigneter Datentyp sowie ein geeignetes Messniveau definiert, um die Voraussetzung für die korrekte Durchführung von weiterführenden Berechnungen zu schaffen.



- Datenbereinigung

Die erhaltene Datentabelle wird mittels SPSS 19.0 auf leere Datenfelder und doppelte Einträge überprüft, um Datenlücken und fehlerhafte doppelte Einträge auszuschließen. Es wird eine grobe fachliche Plausibilitätsprüfung vorgenommen, indem ausgewählte Parameter wie Patientenzahl, Altersdurchschnitt, Geschlechtsverteilung und das Vorkommen von Krankenhauseinweisungen ermittelt und mit den Erwartungen aus den Voranalysen abgeglichen werden.

## 6 Modellentwicklung (Modeling)



### 6.1 Konkretisierung der Data-Mining Ziele (Hypothesen)

Für die Modellierung werden nach Abschluss der Voranalysen und unter Berücksichtigung der Ergebnisse daraus die Hypothesen für das Data-Mining konkretisiert:

1. Polypharmazie im Sinne von  $\geq 5$  ATC/Quartal über ein Jahr steht im Zusammenhang mit (arzneimittelbedingten) Krankenhauseinweisungen.

Mit zunehmender Polypharmazie nimmt die Wahrscheinlichkeit für (arzneimittelbedingte) Krankenhauseinweisungen zu.

2. Patienten mit Polypharmazie ( $\geq 5$  ATC/Quartal über ein Jahr), die innerhalb von einem Jahr (arzneimittelbedingt) in ein Krankenhaus eingewiesen werden, lassen sich aus GKV-Routinedaten vorhersagen.
3. Krankenhauseinweisungen bei Patienten mit Polypharmazie stehen im Zusammenhang mit bestimmten Patientenmerkmalen und Arzneistoffen, woraus sich konkrete Handlungsempfehlungen für die pharmazeutische Betreuung ableiten lassen.

## **6.2 Prüfung des Ausgangsdatensatzes**

Die Datenermittlung und -zusammenstellung erfolgt wie im vorherigen Kapitel dargestellt. Der Entwicklungsdatensatz stellt den Ausgangsdatensatz für die Modellierung dar.

### *Datengrundlage (Entwicklungsdatensatz)*

- Patienten sind durchgängig versichert in 2005-2008
- Altersbeschränkung auf 18-85 Jahre
- Patienten mit Polypharmazie, das heißt mit 5 oder mehr verschiedenen ATC pro Quartal über die Quartale 1-4 in 2007 (und ohne Polypharmazie im 4. Quartal 2006 als quasi „run-in“ Periode)
- Patienten ohne Ausschlussdiagnosen
- n=44.108

### *Datensatzbeschreibung (Entwicklungsdatensatz)*

Der Ausgangsdatensatz für die Modellierung enthält Daten von 44.108 Patienten mit Polypharmazie und den etwa 50 für die Modellierung ausgewählten Variablen, für die ein Zusammenhang mit ABH vermutet wird (Anlage 10.2 und 10.3). Eine Übersicht über die Verteilung der wichtigsten Merkmale im Entwicklungsdatensatz ist in der nachfolgenden Tabelle differenziert nach Patienten mit und ohne Krankenhausaufenthalt dargestellt (Tabelle 13).

Variable (2005-2007)	Mess- niveau	Polypharmazie- Patient mit Krhs. (2008)	Polypharmazie- Patient ohne Krhs. (2008)	p-Wert
Patienten-Merkmale				
n (Gesamt: 44.108)		17.019 (39%)	27.089 (61%)	
Alter (18-85 Jahren)	metrisch	Ø 69,2	Ø 67,5	0,000
Geschlecht: weiblich*	kategorial	10.721 (63,0%)	17.461 (64,5%)	0,002
Bundesland: Westen*	kategorial	12.987 (76,3%)	20.282 (74,9%)	0,001
Berufsgruppe: Rentner*	kategorial	14.168 (83,2%)	21.067 (77,8%)	0,000
Polypharmazie-Merkmale				
ATC pro Jahr	metrisch	Ø 14,6	Ø 12,6	0,000
kleinste Anzahl ATC pro Quartal über Quartal 1-4*	metrisch	Ø 6,6	Ø 6,2	0,000
ATC 5-8 in Quartal 1-4*	kategorial	14.548 (85,5%)	24.773 (91,5%)	0,000
ATC 9-11 in Quartal 1-4*	kategorial	2.211 (13,0%)	2.137 (7,9%)	0,000
ATC $\geq$ 12 in Quartal 1-4*	kategorial	260 (1,5%)	179 (0,7%)	0,000
Rezept-Merkmale				
DDD pro Jahr	metrisch	Ø 3.024	Ø 2.830	0,000
DDD pro ATC pro Jahr*	kategorial	Ø 220,2	Ø 239,9	0,000
Rezepte pro Jahr	metrisch	Ø 30,0	Ø 26,1	0,000
Arzneimittelkosten/ Jahr	metrisch	Ø 2.771	Ø 2.234	0,000
ATC-Merkmale (Auszug)				
Schleifendiuretika*	kategorial	7.242 (42,6%)	9.252 (34,2%)	0,000
Phenprocoumon	kategorial	3.327 (19,5%)	3.811 (14,1%)	0,000
Risikoarzneistoffe*,**	kategorial	15.742 (92,5%)	23.880 (88,2%)	0,000
ICD-Merkmale (Auszug)				
Volumenmangel (E8)	kategorial	3.264 (19,2%)	3.132 (11,6%)	0,000
Schlafstörungen (G)	kategorial	9.166 (53,9%)	12.131 (44,8%)	0,000
Atherosklerose (I6-9)	kategorial	10.662 (62,6%)	14.501 (53,5%)	0,000
COPD (J)	kategorial	7.380 (43,4%)	9.875 (36,5%)	0,000
Gastritis (K)	kategorial	8.584 (50,4%)	11.124 (41,1%)	0,000

Nierenerkrankung (N)	kategorial	9.722 (57,1%)	12.795 (47,2%)	0,000
Übelkeit, u.a. (R)	kategorial	12.420 (73,0%)	16.930 (62,5%)	0,000
Eigenanamnese ... (Z)	kategorial	10.305 (60,5%)	12.673 (46,8%)	0,000

Krhs.=Krankenhausaufenthalt; \*=berechnete Variablen; \*\*=Allopurinol, Clopidogrel, Digitoxin, Phenprocoumon, Spironolacton, Theophyllin oder Ticlopidin

**Tabelle 13: Verteilung der potenziellen Prädiktoren im Entwicklungsdatensatz (Datenbasis 2005-2008)**

### *Signifikanztest der Variablen*

Die Tabelle enthält zusätzlich Angaben zur Abschätzung, ob die Unterschiede der Variablen zwischen den Gruppen mit und ohne Krankenhausaufenthalt im Entwicklungsdatensatz und voraussichtlich in der Grundgesamtheit signifikant sind (p-Wert). Für die metrischen Variablen wurde die Prüfung mittels des *Mann-Whitney-U-Tests für unabhängige Stichproben* durchgeführt, da die betreffenden Variablen bei vorheriger Testung nach *Kolmogorov-Smirnov* keine Normalverteilung zeigten. Für kategoriale Variablen mit nur zwei Ausprägungen (binär) wurde eine Kreuztabelle erstellt und ein Chi-Quadrat-Test durchgeführt. Für die kategorialen Merkmale mit mehr als zwei Ausprägungen (Fallgruppen) erfolgt die Prüfung mittels des *einfaktoriellen ANOVA-Tests* (Varianzanalyse). Geprüft wird, ob die Mittelwerte der verschiedenen Fallgruppen für das Zielmerkmal in der Grundgesamtheit alle gleich oder verschieden sind. Für diesen Test wird das dichotome Zielmerkmal mit zwei Ausprägungen als metrisch interpretiert. In Verbindung mit einem *Tamhane-T2-Mehrfachvergleichstest* kann konkretisiert werden, welche Mittelwerte der Fallgruppen untereinander signifikant verschieden sind. Zu den Variablen mit mehr als zwei Ausprägungen gehören beispielsweise die Merkmale *BG\_Status*, *Anzahl\_FAVE*, *Bundesland\_ID* und das Polypharmazie-Merkmal *kleinste ATC pro Quartal über Quartale 1-4*, entsprechend *min\_ATC\_Q* (Anhang 10.6). Für alle betrachteten Merkmale kann mit  $p < 0,001$  von höchst signifikanten beziehungsweise mit  $p < 0,002$  für die Geschlechtsvariable von sehr signifikanten Unterschieden ausgegangen werden. Das ist für die Verwendung der Merkmale zur Modellbildung von Vorteil.

### *Abgleich mit den Voranalysen*

Die Verteilungen der für die Modellbildung zusammengestellten Variablen im Entwicklungsdatensatz werden für ausgewählte Merkmale mit denen in den Voranalysen verglichen. Insgesamt zeigen sich trotz leicht unterschiedlicher Datengrundlage ähnliche Häufigkeiten und Mittelwerte. Beispielsweise sind Durchschnittsalter, Geschlechtsverteilung (siehe Voranalyse VI) und die prozentuale Häufigkeit von Krankenhauseinweisungen und Polypharmazie-Gruppen (siehe Voranalysen I-III) im Entwicklungsdatensatz ähnlich wie in den Voranalysen (Tabelle 14).

<b>Merkmal</b>	<b>Voranalyse</b>	<b>Entwicklungsdatensatz</b>
Altersdurchschnitt	Ø 69,0	Ø 68,1
Geschlechtsverteilung w:m	1,8	1,7 (64% vs. 36%)
Polypharmazie-Patient mit Krankenhausaufenthalt innerhalb eines Jahres ab Stichtag	38,4%	38,5%
Polypharmazie (5-8 ATC/Quartal)	94%	89%
Polypharmazie (9-12 ATC/Quartal)	4%	10%
Polypharmazie ( $\geq 13$ ATC/Quartal)	2%	1%

**Tabelle 14: Vergleich von Voranalysen und Entwicklungsdatensatz (Datenbasis 2005-2008)**

### **6.3 Optimierung und Streichung von Variablen**

Zur Signifikanzprüfung wurden kategoriale Variablen mit mehr als zwei Ausprägungen mittels des *einfaktoriellen ANOVA-Tests* in Verbindung mit einem *Tamhane-T2-Mehrfachvergleichstest* untersucht. Mit Hilfe der Testergebnisse lassen sich die betroffenen Variablen zusätzlich so optimieren, dass Fallgruppen entstehen, die in der Grundgesamtheit schätzungsweise signifikant verschiedene Mittelwerte aufweisen. Dadurch wäre zu erwarten, dass bessere Unterscheidungsmerkmale (Prädiktoren) zur Erkennung der Zielgruppe mit stärkeren Korrelationen zum Zielmerkmal entstehen. Zeigen Fallgruppen keine signifikant verschiedenen Mittelwerte, sollen diese oder die Variable insgesamt gestrichen werden.

### Altersvariable

Das Alter ist gemäß den definierten Einschlusskriterien begrenzt auf 18-85 Jahre. Die Variable *Alter\_ID* ist metrisch und bildet jedes auftretende Alter ab. Nach der Varianzanalyse ergeben sich 4 Altersgruppen mit signifikant unterschiedlichen Mittelwerten: 18-60, 61-70, 71-80, 81-90 Jahre, beziehungsweise 81-85 Jahre (Abbildung 22).

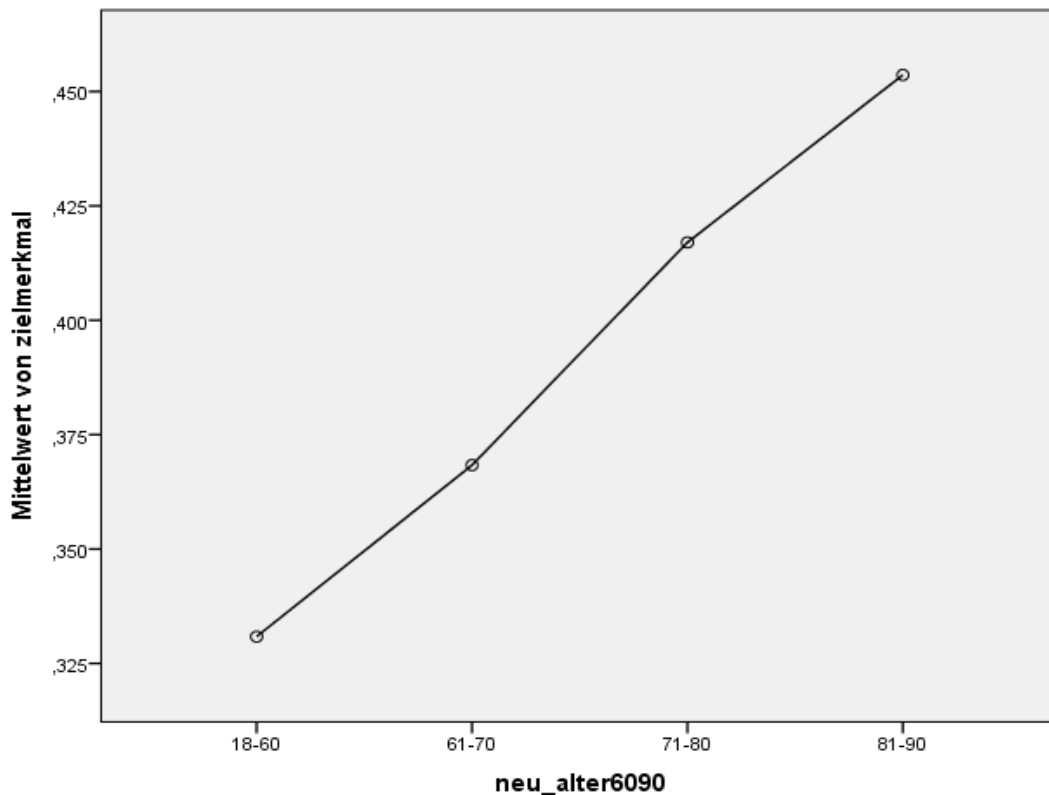
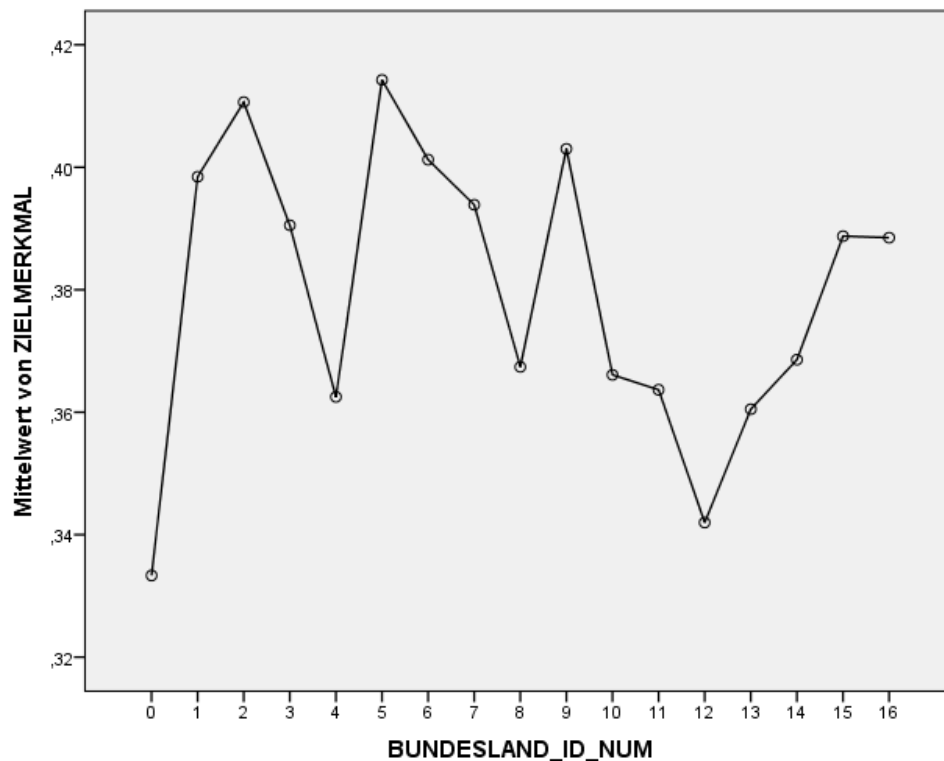


Abbildung 22: Mittelwert des Zielmerkmals in 2008 nach Altersgruppe zum 31.12.2007

Aufgrund des Ergebnisses wird die Altersvariable *Alter\_60708085* mit entsprechenden Fallgruppen gebildet. Der Altersschwerpunkt bei Polypharmazie und bei UAW liegt zudem laut Literatur bei 60-85 Jahren. Diese Altersgruppe macht im Datensatz mit 72% den größten Anteil aus, weswegen zusätzlich die neue Variable *ist\_Alter ≥ 60* gebildet wird. Da ansonsten überwiegend dichotome Variablen in die Modellbildung eingehen werden, das Einbringen von metrischen Variablen die Modellgüte jedoch verbessern kann, wird neben den beiden neuen Merkmalen *Alter\_60708085* und *ist\_Alter ≥ 60* die metrische Altersvariable *Alter\_ID* beibehalten und in die nachfolgende Prüfung der Variablen einbezogen.

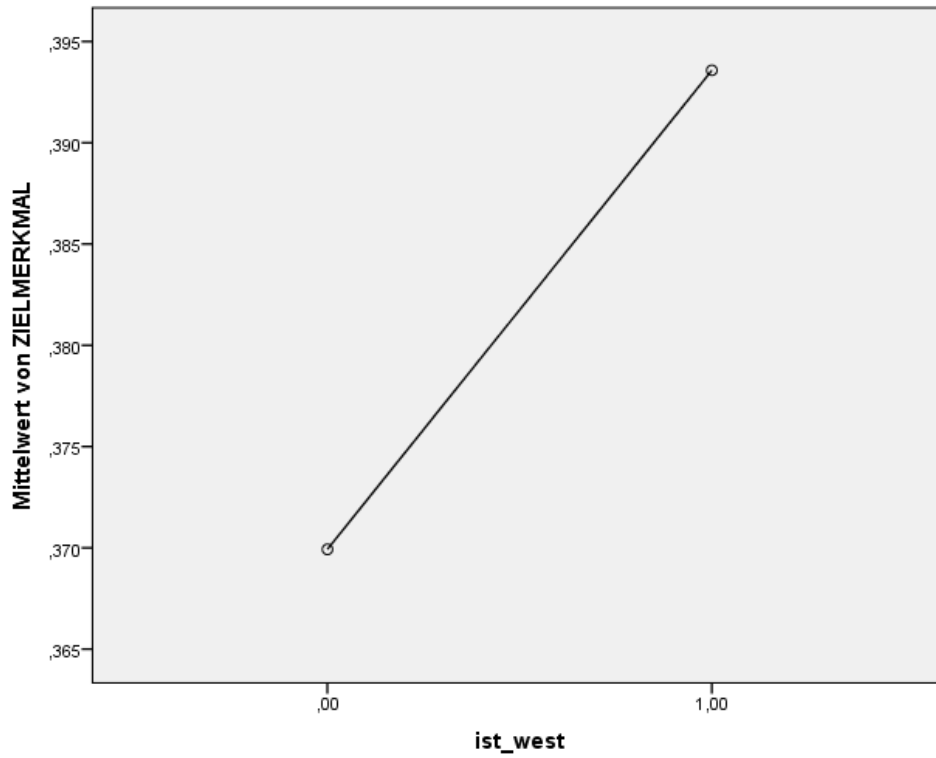
### Bundesland des Wohnorts

Das Merkmal *Bundesland\_ID*, mit der Codierung 1-10 für westliche Bundesländer und 11-16 für östliche Bundesländer jeweils von Nord nach Süd, lässt in der Varianzanalyse mit *Mehrvvergleichstest nach Tamhane* keinen einheitlichen Zusammenhang zum Zielmerkmal erkennen (Abbildung 23).



**Abbildung 23: Mittelwert des Zielmerkmals in 2008 nach Bundesland zum 31.12.2007**

Insgesamt leben zirka 75% aller im Entwicklungsdatensatz betrachteten Patienten in einem westlichen Bundesland. Es wird daher eine Transformation in das Merkmal *ist\_west* (1,0) vorgenommen und mittels des *einfaktoriellen ANOVA* erneut eine Varianzanalyse ausgeführt.



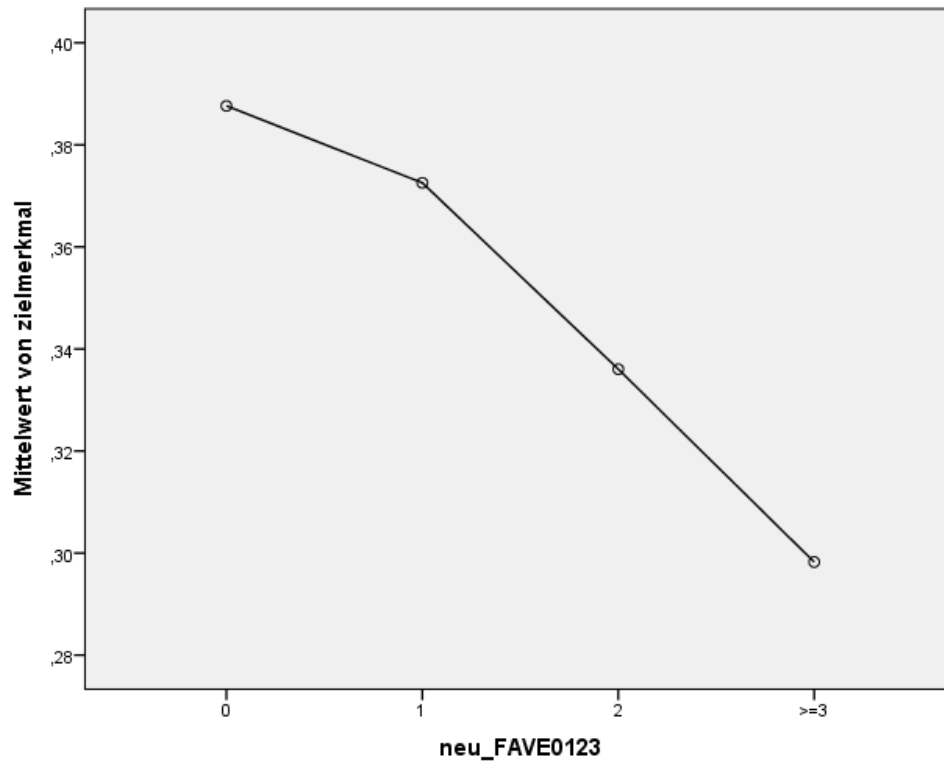
**Abbildung 24: Mittelwert des Zielmerkmals in 2008 nach Wohnort West/Ost zum 31.12.2007**

Das Ergebnis der Varianzanalyse (Abbildung 24) zeigt für *ist\_west* einen signifikanten Unterschied bezogen auf das Zielmerkmal. Die Variable *ist\_west* wird daher zusammen mit der Variablen *Bundesland\_ID* weiter überprüft.



### Anzahl Familienmitglieder

Die Anzahl der mitversicherten Familienmitglieder, in der Regel Ehefrau und Kinder, werden codiert im Merkmal *Anz\_FAVE*. Die Prüfung mittels Varianzanalyse zeigt zunächst an, dass eine zunehmende Anzahl Familienmitglieder mit weniger Krankenhauseinweisungen einherzugehen scheint (Abbildung 25).

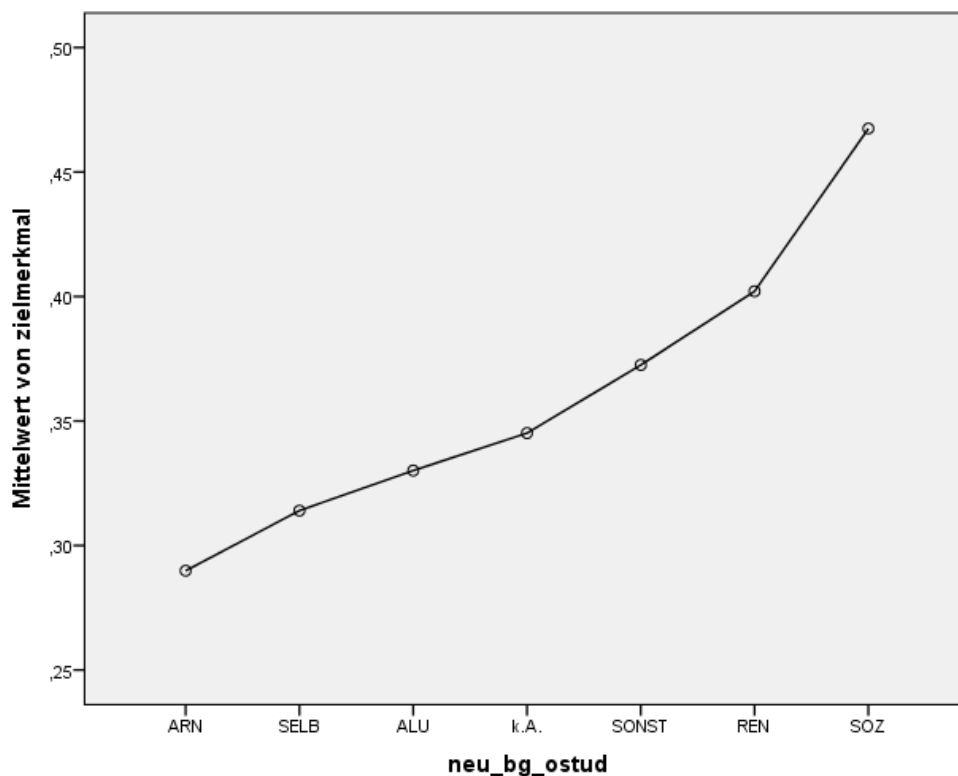


**Abbildung 25: Mittelwert des Zielmerkmals in 2008 nach Anzahl FAVE zum 31.12.2007**

Die Unterschiede der Mittelwerte sind mit  $p=0,000$  im Entwicklungsdatensatz zwar signifikant, jedoch mit  $p>0,05$  schätzungsweise nicht in der Grundgesamtheit der Zielgruppe. Eine Optimierung der Variablen in *Anz\_FAVE* (0, 1, 2,  $\geq 3$ ) ergibt keine Verbesserung. Das Merkmal *Anz\_FAVE* wird daher für die Modellbildung nicht weiter berücksichtigt.

### Berufsgruppen-Status

Das Merkmal *BG\_Status* enthält Informationen über den Beschäftigungsstatus der betrachteten Patienten. Die Gruppe der Rentner macht mit 80% den größten Anteil aus. Der Mittelwertvergleichstest (Abbildung 26) lässt zudem erkennen, dass die Gruppe der Rentner mit 0,42 den zweithöchsten Mittelwert unter den Fallgruppen und damit die größte Tendenz für eine Krankenhauseinweisung hat.



**Abbildung 26: Mittelwert des Zielmerkmals in 2008 nach Berufsgruppe zum 31.12.2007**

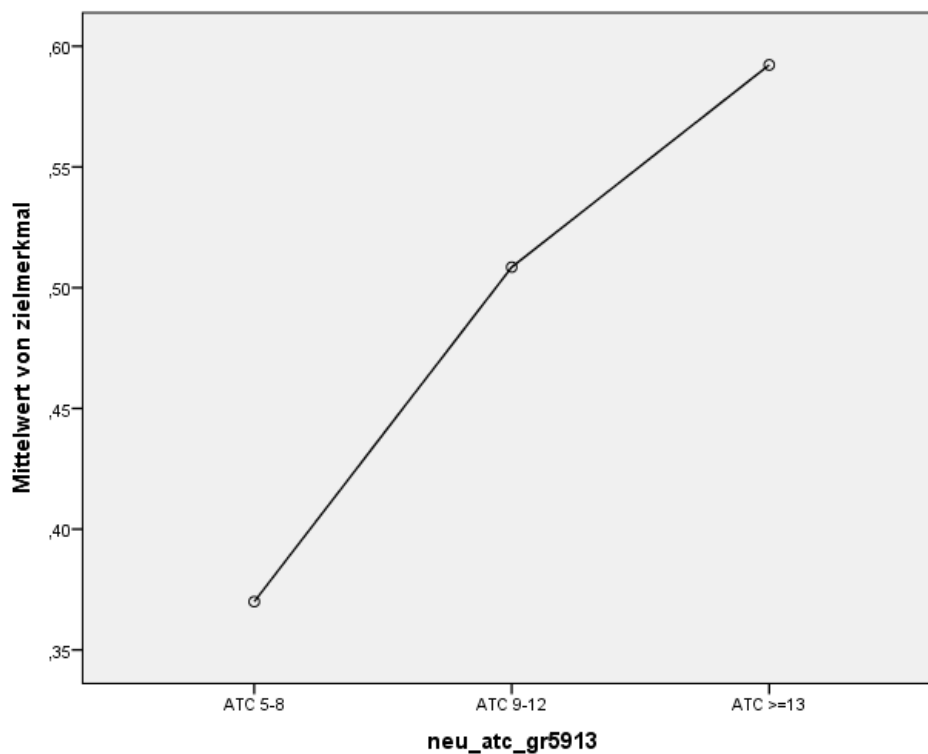
Aufgrund des Ergebnisses der Analyse wird ein neues Merkmal *ist\_Rentner* (1,0) gebildet. Gemäß der Varianzanalyse wäre auch ein klassiertes Merkmal mit 3 Fallgruppen denkbar aus Arbeitnehmern (1.), Selbstständigen, Arbeitslosen und Patienten mit unbekanntem Status (2.) und Patienten mit sonstigem Status, Rentnerstatus und Sozialhilfeempfängern (3.). Allerdings entstehen dabei 3 heterogene Patientengruppen, weswegen von einer Gruppierung abgesehen wird. Die Variable *BG\_Status* mit 7 Fallgruppen, sowie die optimierte Variable *ist\_Rentner* sollen beide weiterhin überprüft werden, da der Einfluss der Variablen im Vergleich unklar bleibt.

### *Arzneimittelkosten*

Die Arzneimittelkosten werden durch zwei Merkmale, *Arzn\_Kosten\_1J* und *Arzn\_Kosten\_3*, beschrieben. Diesen sind Daten aus 1 beziehungsweise 3 Jahren zugrunde gelegt. Beide Variablen zeigen eine Verteilung, die nicht der Normalverteilung entspricht. Durch Logarithmieren lässt sich die Verteilung annähernd in eine solche überführen. Die Merkmale *arzn\_kosten\_1J* und *arzn\_kosten\_3J* werden gebildet. Die Auswahl zur Verwendung bei der Modellierung wird von der weiteren Überprüfung der Variablen anhand der Korrelationskoeffizienten abhängig gemacht.

### *Mindestanzahl ATC pro Quartal*

Die Mindestanzahl, das heißt die kleinste gemeinsame Zahl an verschiedenen Arzneistoffen (ATC) pro Quartal über 4 Quartale, wird durch das Merkmal *min\_ATC\_Q* beschrieben. Das Merkmal ergibt sich aus der in den Voranalysen entwickelten Definition für eine Polypharmazie. Die Anzahl liegt im Entwicklungsdatensatz zwischen 5 und 25. Ab 15 und mehr ATC sinkt die Anzahl der betroffenen Patienten unter 100. Daher wird diese Variable transformiert in die neue Variable *min\_ATC\_Q14*, die Patienten mit einer Mindestanzahl an ATC pro Quartal zwischen 14 und 25 zu einer Fallgruppe zusammenfasst. Da hierbei die Informationen zum Grad der Polypharmazie verloren gehen, werden beide Variablen im weiteren Verlauf noch näher überprüft. Durch die Betrachtung der Variablen in der Varianzanalyse (Abbildung 27) ergeben sich Hinweise auf eine günstige Klassierung der Variable *min\_ATC\_Q* in die Fallgruppen mit 5-8 ATC/Quartal (1.), 9-12 ATC/Quartal (2.) und  $\geq 13$  ATC/Quartal (3.).

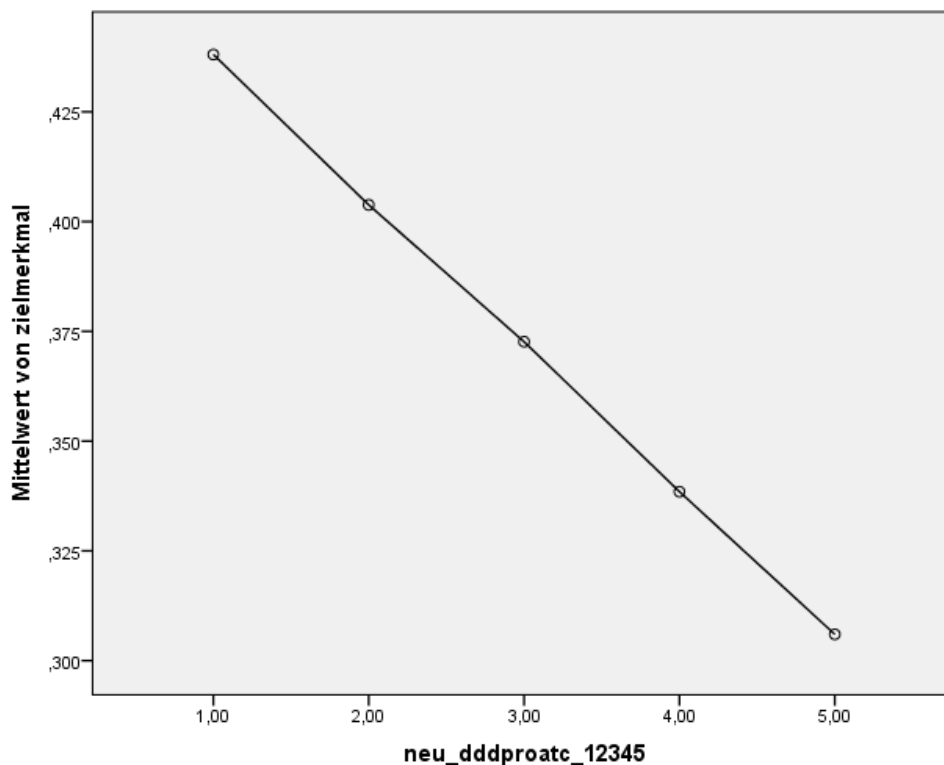


**Abbildung 27: Mittelwert des Zielmerkmals in 2008 nach Polypharmazie-Gruppe in 2007**

Es wird eine neue Variable *Anz\_ATC\_Gr5913* gebildet. Ob die ursprüngliche Variable *min\_ATC\_Q* oder eine der optimierten Variablen *min\_ATC\_Q14* oder *Anz\_ATC\_Gr5913* in die Modellierung eingehen, wird von der weiteren Überprüfung der Variablen anhand der Korrelationskoeffizienten abhängig gemacht. Undeutlich erscheint des Weiteren, ob die Grenze der obersten Fallgruppen bei 12 oder 13 ATC pro Quartal liegen sollte.

*Anzahl DDD (defined daily dose)*

Die Anzahl an DDD (defined daily dose) werden durch zwei Merkmale, *Anz\_DDD\_1J* und *Anz\_DDD\_3J*, beschrieben. Diesen sind Daten aus einem beziehungsweise 3 Jahren zugrunde gelegt. Es wird vermutet, dass steigende DDD, als ein Maß für die Menge der verordneten Arzneistoffe, ähnlich wie eine hohe Anzahl ATC in einem Zusammenhang mit einem Krankenhausaufenthalt stehen. Statt der Betrachtung der DDD insgesamt für einen Patienten ist die Betrachtung der DDD bezogen auf die ATC inhaltlich exakter. Es wird daher ein neues Merkmal *DDDproATC* gebildet. Dieses soll einen genaueren Anhaltspunkt über die angemessene Dosis oder auch Therapiedauer mit den verordneten Arzneistoffen geben. Niedrige Werte können dabei Unterdosierungen oder häufige Wirkstoffwechsel anzeigen. Beides könnte zu Krankenseinweisungen führen. Allerdings entstehen niedrige Werte ebenfalls durch eine Teilung von Darreichungsformen und können irreführend sein. Bei 365 Tagen im Jahr sollten sich zirka 365 DDD pro ATC ergeben. Eine höhere Anzahl an DDD pro ATC kann auf eine Überdosierung hindeuten, was als ein Risiko ebenfalls einen guten Prädiktor für eine ABH darstellen könnte. Mittels einer teils automatischen, teils manuellen Klassierung der Variable in SPSS 19.0 und einer Überprüfung mittels Varianzanalyse (Abbildung 28) werden die günstigsten Fallgruppen ermittelt.



**Abbildung 28: Mittelwert des Zielmerkmals in 2008 nach DDD-Gruppe in 2007**

Die Analyse ergibt 5 Fallgruppen mit signifikant unterschiedlichen Mittelwerten (Abbildung 28). Verschiedene Mittelwerte in Bezug auf das Zielmerkmal werden geschätzt bei einer Anzahl DDD pro ATC von 167, 232, 272, 335 und  $\geq 335$ , entsprechend Gruppe 1 bis 5. Die Zahlenwerte für die DDD in den Gruppen entsprechen dabei zirka 45%, 65%, 75%, 90% und  $>90\%$  der zu erwartenden 365 DDD pro Jahr, wenn von einer DDD pro Tag ausgegangen wird. Allerdings zeigen sich mit zunehmenden DDD pro ATC sinkende Mittelwerte. Dieses entspricht zunächst nicht den angestellten Vermutungen. Welche Rolle die optimierte Variable *DDDproATC* im Vergleich zur ursprünglichen Variable *Anz\_DDD* spielt und welcher Datenumfang, 1 oder 3 Jahre, zugrunde gelegt werden sollte, wird von der weiteren Überprüfung der Variablen anhand der Korrelationskoeffizienten abhängig gemacht. Die optimierte Variable *DDDproATC* sollte jedoch sowohl inhaltlich wie statistisch einen besseren Prädiktor für ABH abgeben als die Variable *Anz\_DDD*.

#### **6.4 Prüfung der Variablen**

Zur Abschätzung, welche Merkmale sich für die Modellbildung besonders eignen, werden auf der Basis des Entwicklungsdatensatzes ( $n=44.108$ ) Korrelationskoeffizienten ermittelt, um Stärke und Richtung des Zusammenhangs zum Zielmerkmal zu überprüfen. Korrelationskoeffizienten können in der Regel Werte zwischen -1 und +1 annehmen. Das Vorzeichen gibt die Richtung und der Betrag die Stärke der Korrelation an. Beträge der Korrelationskoeffizienten kleiner  $\leq 0,2$  werden als sehr geringe,  $>0,5$  als mittlere und  $>0,7$  als hohe Korrelation bewertet (Bühl, 2009).

##### *Korrelationskoeffizienten der metrischen Merkmale*

Zur Beschreibung eines linearen Zusammenhangs zwischen zwei metrischen Merkmalen, die annähernd normalverteilt sind, wird der *Korrelationskoeffizient nach Pearson* eingesetzt. Richtige Ergebnisse werden auch erwartet bei intervallskalierten und dichotomen Daten. Da die metrischen Variablen gegebenenfalls nicht ausreichend normalverteilt sind, wird zusätzlich der *Korrelationskoeffizient nach Kendalls-Tau b* (Rangkoeffizient) bestimmt. Dieser setzt keine linearen Zusammenhänge voraus, ergibt aber tendenziell niedrigere Werte. Der *Korrelationskoeffizient nach Kendalls-Tau* ist für den Einsatz bei einer intervallskalierten und einer ordinalen Variable geeignet. Da andere geeignete Korrelationskoeffizienten in SPSS 19.0 nicht unterstützt

werden, wird das dichotome Zielmerkmal, mit dem korreliert werden soll, zur Vereinfachung als ordinal interpretiert (Bühl, 2009). In die Prüfung eingeschlossen werden neben den ursprünglichen Variablen auch die optimierten Variablen (\*\*). Es wird darüber hinaus differenziert zwischen den über ein Jahr ( $x=1$ ) und den über 3 Jahre ( $x=3$ ) ermittelten Variablen.

Variable	Mess-niveau	Korrelation für $x=1$ (Pearson)	Korrelation für $x=3$ (Pearson)	Korrelation für $x=1$ (Kendalls-Tau b)	Korrelation für $x=3$ (Kendalls-Tau b)	p-Wert*
Patienten-Merkmale						
Alter_ID (in Jahren)	intervall.	0,081	-	0,070	-	0,000
Rezept-Merkmale						
Rezepte pro x Jahr(e)	intervall.	0,162	0,143	0,140	0,120	0,000
Verordnungen pro x Jahr(e)	intervall.	0,142	0,123	0,124	0,103	0,000
Anzahl DDD pro x Jahr(e)	intervall.	0,071	0,064	0,055	0,048	0,000
Arzneikosten pro x Jahr(e)	intervall.	0,070	0,072	0,101	0,092	0,000
Ln_Kosten pro x Jahr **	intervall.	0,122	0,111	0,101	0,092	0,000
Polypharmazie-Merkmale						
Anzahl ATC pro x Jahre	intervall.	0,204	0,207	0,170	0,174	0,000

intervall.=intervallskaliert; \*=die Korrelationen nach Pearson und Kendalls-Tau b sind auf dem Niveau von  $p=0,000$  (2-seitig) signifikant; \*\*=optimierte Ursprungsvariable

**Tabelle 15: Korrelationen zwischen metrischen Merkmalen und Zielmerkmal (Datenbasis 2005-2008: Entwicklungsdatensatz)**

Die bestimmten Korrelationskoeffizienten der metrischen Merkmale (Tabelle 15) zeigen mit  $<0,2$  sehr geringe Korrelationen zum Zielmerkmal. Nur das Merkmal für die Anzahl an ATC pro Jahr hat einen Korrelationskoeffizient knapp über 0,2. Allerdings zeigen die p-Werte an, dass die Korrelationen signifikant sind. Wie erwartet, fallen die *Korrelationskoeffizienten nach Kendalls-Tau b* mit Ausnahme für die Arzneikosten kleiner aus. Insgesamt sind sie ähnlich hoch wie die *Korrelationskoeffizienten nach Pearson*. Dieses kann auf normalverteilte Variablen hindeuten (Bühl, 2009).

Sind bei zwei nicht-dichotomen Variablen die zugrundeliegenden Zusammenhänge

nicht linear, sondern u-förmig, kann der Korrelationskoeffizient als Maß für den Zusammenhang sehr klein werden und versagen (Bühl, 2009). Dieses ist insbesondere bei der Altersvariable zu vermuten. Die versuchsweise Berechnung des für u-förmige Zusammenhänge besser geeigneten *Zusammenhangsmaßes Eta* für die Altersvariable ergibt mit 0,091 jedoch keinen wesentlich höheren Korrelationskoeffizienten.

Bei den Rezept-Merkmalen liegen die Korrelationskoeffizienten im Vergleich zur Altersvariablen höher. Die Anzahl der *ATC pro Jahr* zeigt von den Polypharmazie-Merkmalen den höchsten Korrelationskoeffizienten. Die Korrelation zum Zielmerkmal scheint damit stärker als für die Merkmale *min\_ATC\_Q* und *min\_ATC\_Q14*, die die Polypharmazie exakter beschreiben sollten. Das könnte darauf hindeuten, dass bei der Betrachtung der ATC pro Quartal über 4 Quartale und die Gruppierung aufgrund der durchgängig kleinsten gemeinsamen Anzahl an ATC pro Quartal (*min\_ATC...*) Informationen verloren gehen. Das Merkmal für die *Anzahl DDD pro Jahr* zeigt den niedrigsten Korrelationskoeffizienten, obwohl eine stärkere Korrelation wie bei der *Anzahl ATC pro Jahr* zum Zielmerkmal vermutet worden wäre.

#### *Korrelationskoeffizienten der kategorialen Merkmale*

Die kategorialen Merkmale unterscheiden sich im Messniveau, daher werden verschiedene Korrelationskoeffizienten ermittelt. Der *Korrelationskoeffizient nach Kendalls-Tau* stellt ein Ordinalmaß dar und ist geeignet für ordinale beziehungsweise binominale Variablen, die als ordinal interpretiert werden können (Bühl, 2009). Der *Korrelationskoeffizient nach Cramer V* stellt ein Nominalmaß dar und wird für die nominalen Variablen bestimmt. Das binominale Zielmerkmal, mit dem korreliert werden soll, wird für die Bestimmung als ordinal beziehungsweise nominal interpretiert. In der nachfolgenden Tabelle sind die Ergebnisse aufgeführt (Tabelle 16). Für den *Korrelationskoeffizient nach Kendalls-Tau b* sind die ermittelten p-Werte angegeben. In die Prüfung eingeschlossen werden neben den ursprünglichen Variablen auch die optimierten Variablen (\*\*).



Variable	Kategoriales Messniveau	Korrelation nach Kendalls-Tau b	Korrelation nach Cramer V	p-Wert*
Patienten-Merkmale (Auswahl)				
Alter_60708085**	ordinal	0,075	-	0,000
ist_Alter ≥60**	binominal	0,056	0,056	0,000
ist_weiblich	binominal	-0,015	0,015	0,002
ist_Mann	binominal	0,015	0,015	0,002
BG_Status (1-7)	nominal	-	0,074	0,000
ist_Rentner**	binominal	0,067	0,067	0,000
Bundesland_ID (0-16)	nominal	-	0,043	0,000
ist_west**	binominal	0,016	0,016	0,002
Anzahl_FAVE (0-10)	ordinal	-0,013	-	0,007
Anzahl_FAVE (0,1,2,≥3)**	ordinal	-0,013	-	0,007
Rezept-Merkmale (Auswahl)				
DDDproATC (1-5)**	ordinal	- 0,084	-	0,000
Polypharmazie-Merkmale				
min_ATC_Q14 (1,2,...≥14)**	ordinal	0,098	-	0,000
ATC-Gruppe (5-8,9-11,≥12)**	ordinal	0,138	-	0,000
ATC-Gruppe (5-8,9-12,≥13)**	ordinal	0,093	-	0,000
ATC-Merkmale (Auswahl)				
Diuretika	binominal	0,084	0,084	0,000
Phenprocoumon	binominal	0,072	0,072	0,000
Risikoarzneistoffe	binominal	0,070	0,070	0,000
ICD-Merkmale (Auswahl)				
Volumenmangel (E8)	binominal	0,105	0,105	0,000
Gastritis (K)	binominal	0,092	0,092	0,000

\*=angegeben ist die Signifikanz für die Korrelation nach Kendalls-Tau (2-seitig, auf dem Niveau von 0,01 signifikant);

\*\*=optimierte Ursprungsvariable

**Tabelle 16: Korrelationen zwischen kategorialen Merkmalen und Zielmerkmal (Datenbasis 2005-2008: Entwicklungsdatensatz)**

Alle bestimmten Korrelationskoeffizienten für die überprüften Variablen sind wie bei

den metrischen Variablen  $<0,2$  und drücken damit sehr geringe Korrelationen aus. Allerdings zeigen die p-Werte an, dass die Korrelationen signifikant sind. Insbesondere für die Altersmerkmale, sowie für die Variablen *ist\_weiblich* und *ist\_Rentner* wären gemäß den fachlichen Vorüberlegungen stärkere Korrelationen erwartet worden. Die in 4 Altersgruppen klassierte Variablen *Alter\_60708085* zeigt eine stärkere Korrelation zum Zielmerkmal als die Variable *ist\_Alter  $\geq 60$*  mit zwei Ausprägungen. Die Variable *BG\_Status* mit 7 verschiedenen Fallgruppen scheint ähnlich stark mit dem Zielmerkmal korreliert zu sein wie die dichotome Variable *ist\_Rentner*.

Einige Korrelationskoeffizienten haben negative Vorzeichen. So scheinen weibliches Geschlecht (*ist\_weiblich*), die Anzahl an Familienmitgliedern (*Anzahl\_FAVE*) und das gebildete Merkmal *DDDproATC* negativ mit dem Zielmerkmal korreliert. Das bedeutet bei steigender Anzahl an Familienmitgliedern und steigender Zahl an DDD ein selteneres Vorkommen von Krankenhauseinweisungen. Dieses wäre für die Variable *DDDproATC* im Vergleich zu den inhaltlich ähnlichen Variablen *Anzahl ATC pro Jahr* und *Anzahl DDD pro Jahr*, die positiv mit dem Zielmerkmal korreliert sind, zunächst nicht erwartet worden. Beide Variablen stellen ein Maß für die Menge an eingesetzten Arzneistoffen dar.

Die ICD-Merkmale zeigen mit  $0,09 - 0,10$  die höchsten Korrelationskoeffizienten und erscheinen daher insgesamt stärker mit Krankenhauseinweisungen korreliert als die übrigen Merkmals-Gruppen.

#### Fazit:

Die Aussagekraft der Korrelationskoeffizienten kann insgesamt dadurch eingeschränkt sein, dass die vorliegenden Zusammenhänge durch die gewählten Korrelationskoeffizienten nicht passend dargestellt werden. Die für die Prüfung der Variablen bestimmten Korrelationskoeffizienten lassen jedoch Tendenzen erkennen, die für die weitere Reduktion der Variablenanzahl genutzt werden. Daraus ergibt sich, dass Variablen, die mit *Daten aus 3 Jahren* gebildet werden, nicht weiter für die Modellbildung berücksichtigt werden. Die niedrigen Korrelationskoeffizienten lassen keine bessere Prädiktion im Vergleich zu der Bildung aus Daten von einem Jahr erwarten. Wegen der hohen Korrelationskoeffizienten der *ICD-Merkmale* werden diese für die Modellierung verworfen, da die Effekte der Patienten- und insbesondere der Arzneistoff-Merkmale vorrangig untersucht und nicht überlagert werden sollen. Die Optimierung der Patienten-, Rezept- und Polypharmazie-Merkmale haben größtenteils zu höheren Korrelationskoeffizienten geführt. Für die Modellbildung werden daher die optimierten Variablen vorrangig verwendet. Die Variable *ist\_Alter  $\geq 60$*  wird wegen eines geringen Korrelationskoeffizienten für die Modellbildung nicht berücksichtigt. Bei den undeutlichen Tendenzen wird die günstigste Variable, insbesondere zur Be-

schreibung von Alter, Berufsgruppe, ATC und DDD während der Modellierung getestet. Für die Bildung der ATC-Gruppe als Polypharmazie-Merkmal wird aufgrund des besseren Korrelationskoeffizienten als Grenze für die oberste Fallgruppe *12 ATC pro Quartal* gewählt.

### **6.5 Modellierung**

Für die Modellierung des Vorhersagemodells sind die zu testenden Variablen einzugrenzen. Durch die im vorherigen Kapitel beschriebene Abschätzung seiner Relevanz in Bezug auf eingetretene Krankenhauseinweisungen wurde das ursprüngliche Variablen-Set in der Anzahl weiter reduziert. Einige Variablen wurden transformiert oder neu berechnet, um Variablen mit stärkeren Korrelationen zum Zielmerkmal zu erhalten. Es werden nur die aus den 1-Jahresdaten gebildeten Merkmale weiter verwendet, da die 3-Jahresdaten keine wesentlich besseren Korrelationen zum Zielmerkmal erkennen ließen. Von den ursprünglich etwa 80 Merkmalen (Anlage 10.7) verbleiben etwa 40 Variablen. Zusätzlich werden aus den ATC-Merkmalen gemäß den Hinweisen aus der Literatur fachlich bedeutsame ATC-Kombinationen gebildet und in einer Matrix zusammengestellt (Anhang 10.8). Von diesen werden 11 ATC-Kombinationen nach vermuteter fachlicher Relevanz durch die Autorin ausgewählt und im Entwicklungsdatensatz zur Testung bei der Modellbildung angelegt. Insgesamt gehen damit zirka *50 Variablen* in die Modellierung ein. Die nachfolgende Tabelle zeigt eine Übersicht dieser Variablen (Tabelle 17).

Variable	Optimierte Variable	Feldbeschreibung	Wert
Zielmerkmal (abhängige Variable)			
ZIELMERKMAL	-	Krankenhausaufenthalt (ja/nein)	1, 0
Patienten-Merkmal			
KN_GESCHLECHT	ist_Mann	Männliches Geschlecht (ja/nein)	1, 0
ALTER_ID	Alter_60708085	Alter (18-60, 61-70, 71-80, 81-85)	1-4
BUNDESLAND_ID	ist_west	Wohnort im Westen (ja/nein)	1, 0
BG_STATUS (1-7)	ist_Rentner	Berufsgruppe (beziehungsweise Rentner)	1-7 (1, 0)
Rezept-Merkmal			
Anz_REZEPTE_1_J	-	Anzahl Rezepte in einem Jahr	0-n
Anz_VO_1_J	-	Verordnungen in einem Jahr	0-n
Anz_DDD_1_J	DDDproATC_12345	Ø DDD pro ATC in einem Jahr	0-5
ARZN_KOSTEN_1_J	InKost1Jplus1	Log. Arzneimittelkosten pro Jahr	0-n
Polypharmazie-Merkmal			
Anz_ATC_1_J	-	Anzahl ATC pro Jahr	0-n
min_ATC_Q	min_ATC_Q14	ATC/Q über 4 Quartale (5-13, ≥14)	5-14
	neu_ATC_Gr5912	ATC/Q über 4 Q. (5-8, 9-11, ≥12)	1-3
Variable	Feldbeschreibung	Variable	Feldbeschreibung
ATC-Merkmale (binär)			
ATC_A02BC01	Omeprazol	ATC_J01_Spez	Clindamycin o. Chinolon
ATC_A03FA01	Metoclopramid	ATC_M04AA01	Allopurinol
ATC_A10BA02	Metformin	ATC_NSAR_SPEZ	NSAR-Auswahl
ATC_B01AA04	Phenprocoumon	ATC_N02A_SPEZ	Opioid-Auswahl
ATC_B01AC04	Clopidogrel	ATC_N02BB02	Metamizol
ATC_B01AC05	Ticlopidin	ATC_N02BG07	Flupirtin
ATC_C01AA04	Digitoxin	ATC_N_SPEZ	Psychopharmaka
ATC_C01BD01	Amiodaron	ATC_N05_SPEZ	Psycholeptika-Auswahl
ATC_C01DX12	Molsidomin	ATC_N06_SPEZ	Antidepressiva-Auswahl
ATC_C03CA	Furo- o. Torasemid	ATC_R03DA04	Theophyllin

ATC_C03DA01	Spironolacton	ATC_R03_SPEZ	Inhalativa-Auswahl
ATC_C09AA05	Ramipril	ATC_R05CB01	Acetylcystein
ATC_J01EE01	„Cotrimoxazol“	ATC_S01_SPEZ	Steroid-Auswahl (i.o.)
Variable	Feldbeschreibung	Variable	Feldbeschreibung
ATC-Kombinationen (binär)			
B01AC04_05	Clopidogrel + ASS	C09AA05_03DA01	Spironolacton + Ramipril
C01_BD01_AA04	Digitoxin + Amiodaron	M04AA01_C09AA05	Allopurinol + Ramipril
C03DA01_CA	Schleifendiuretika + Spironolacton	R03DA04_B01AC04	Theophyllin + Clopidogrel
C09AA05_03CA	Schleifendiuretika + Ramipril	C01BD01_N02Aspz	Amiodaron + Opioide
M04AA01_C03CA	Schleifendiuretika + Allopurinol	ATC_Risiko	Risikoarzneistoffe- Auswahl
R03DA04_M04AA01	Theophyllin + Allopurinol		

o.=oder; i.o.=intra-okulär

**Tabelle 17: Übersicht über die Variablen für die Modellierung am Entwicklungsdatensatz**

### *Vorgehen bei der Modellierung*

Die Modellierung und Testung erfolgt mit SPSS 19.0 unter Anwendung verschiedener Modellierungsalgorithmen (Kap. 3). Für die ersten Modelle wird der Algorithmus „Vorwärts nach Wald“ verwendet. Die Modellierung des Vorhersagemodells erfolgt in 16 Schritten. Bei der Modellierung werden mehrere Modelle mit zunächst allen Variablen und anschließend wechselnden Variablen-Sets gebildet. Nach jeder Modellbildung wird die erhaltene Modellgüte anhand des prozentualen Anteils an richtig klassifizierten Patienten und dem Pseudo-R-Quadrat nach Nagelkerke auf eine Verbesserung hin überprüft. Um über die Variablenauswahl für den nächsten Modellierungsschritt zu entscheiden werden der Effektkoeffizient  $\text{Exp}(B)$ , der Korrelationskoeffizient nach Kendalls-Tau  $b$  oder partielle Korrelationen zur Entscheidungsfindung genutzt. Die Variablen-Sets werden solange variiert, bis ein Modell mit größtmöglicher und konstant bleibender Güte und einer kleinstmöglicher Anzahl an

fachlich interpretierbaren Variablen zusammengestellt ist. Zur Bestätigung des optimalen Variablen-Sets wird von der Methode „Vorwärts nach Wald“ auf „Rückwärts nach Wald“ gewechselt und erneut modelliert. Zum Schluss der Modellierung wird zusätzlich der rechenaufwendigere Algorithmus „Vorwärts bedingt“ getestet. Die verschiedenen Variablen-Sets, die zugehörigen Modelle und Zwischenergebnissen sind in der nachfolgenden Übersicht dargestellt (Tabelle 18).

Modell	Variable(n)	Anzahl Variablen	Richtige Klassifizierung	R2 *	Ergebnis
M 1	alle Variablen (optimierte Variablen ohne DDDproATC_12345 und neu_ATCGr5912)	26	64,3%	9,2%	Ist_west, Anz_DDD_1J und 9 ATC-Merkmale werden ausgeschlossen; ist_weiblich wird umcodiert in ist_Mann
M 2	Variablen wie M 1 (nicht optimiert)	22	63,1%	8,7%	Modellverschlechterung bestätigt optimierte Variablen, diese werden wieder aufgenommen
M 3	M 1 + DDDproATC_12345	27	64,4%	9,2%	DDDproATC_12345 erhöht die Güte und verbleibt
M 4	+ neu_ATCGr5912 (statt min_ATC_Q14)	25	64,4%	9,2%	neu_ATCGr5912 und Theophyllin werden ausgeschlossen, min_ATC_Q14 wird manuell wieder aufgenommen
M 5	M 3 + Alter_ID (statt alter60708085)	27	64,3%	9,2%	Alter_ID verschlechtert die Güte bei niedrigem Exp(B)** , alter_60708085 wird wieder aufgenommen
M 6	+ BG_1234567 (statt ist_Rentner)	26	64,4%	9,2%	BG_234567 wird ausgeschlossen, ist_Rentner wird manuell wieder aufgenommen
M 7	+ ATC-Kombinationen (ohne ATC_Risiko)	28	64,5%	9,3%	2 ATC-Kombinationen verbessern die Güte, 7 ATC-Kombinationen und Ramipril werden ausgeschlossen
M 8	M 7,	29	64,5%	9,3%	M 7 wird bestätigt, eine zu-

	rückwärts nach Wald				sätzliche ATC-Kombination wird dabei aufgenommen
M 9	M 7 + ATC_Risiko, rückwärts nach Wald	27	64,4%	9,3%	ATC_Risiko verschlechtert die Güte, aber wird beibehalten wegen hohem Exp(B), andere Variablen werden dadurch ausgeschlossen
M 10	M 9 + sechs fachlich relevante ATC-Kombinationen	25	64,5%	9,3%	2 ATC-Kombinationen verbessern die Güte, übrige werden ausgeschlossen, ebenso NSAR und Theophyllin , schwer interpretierbare negative Korrelationen vorhanden
M 11	M 10, rückwärts nach Wald	26	64,5%	9,3%	M 10 bestätigt, Digitoxin zusätzlich im Modell
M 12	M 10, vorwärts bedingt	26	64,5%	9,3%	M 11 wird bestätigt
M 13	M 12 ohne Anz_VO_1J und Anz_ATC_1J	24	64,3%	8,8%	Modell verschlechtert sich, min_ATC_Q14 wird ausgeschlossen
M 14	M 12 + Anz_ATC_1J ohne min_ATC_Q14	24	64,5%	9,1%	Modell verbessert durch Anz_ATC_1J, Digitoxin wird ausgeschlossen
M 15	M 14, rückwärts (mit Digitoxin)	25	64,4%	9,1%	M 14 bestätigt, aber bei schlechterer Güte wird Digitoxin eingeschlossen
M 16	M 14, rückwärts (ohne Digitoxin)	24	64,5%	9,1%	M 16 hat mit M 14 identische Variablen und eine gleich hohe Modellgüte=Ende

\*=R-Quadrat nach Nagelgerke als näherungsweise Pseudo-R-Quadrat-Maß

**Tabelle 18: Übersicht über erstellte und getestete Modelle (Datenbasis 2005-2008)**

In das **Modell 1** gehen alle ursprünglich eingeplanten Patientenmerkmale ein. Eine Ausnahme bildet die Anzahl an Familienmitgliedern, von der nach vorheriger Prüfung kein Nutzen für die Modellbildung erwartet wird. Des Weiteren kommen die eingangs in diesem Kapitel tabellarisch dargestellten Rezept-, Polypharmazie- und ATC-Merkmale zum Einsatz. Soweit vorhanden werden für diese die Variablen in optimierter Form genutzt. Die Merkmale DDDproATC\_12345 und ATCGr\_5912 werden wegen des noch schwer einschätzbaren Einflusses ebenso wie die 11 ATC-Kombinationen zu diesem Zeitpunkt noch nicht in das Modell integriert. **Modell 2** enthält alle Variablen in nicht optimierter Form und dient im Vergleich als Test für die Eignung der optimierten Variablen.

In **Modell 3 bis 6** werden nacheinander verschiedene Varianten an optimierten Patienten-, Rezept- und Polypharmazie-Merkmalen einzeln auf Ihre Effekte bezüglich der Modellgüte getestet.

Nach abgeschlossener Testung aller übrigen Variablen-Gruppen werden in **Modell 7** für eine fachlich stärkere Ausrichtung des Modells die 11 gebildeten ATC-Kombinationen integriert. Dieses erfolgt zu diesem Zeitpunkt, damit zuvor Effekte der einzelnen ATC-Merkmale besser beobachtet und die Anzahl der Merkmale reduziert werden kann. Aus dem gleichen Grund wird das Merkmal ATC\_Risiko erst in Modell 9 hinzugenommen. In **Modell 8 und 9** wird zusätzlich die Methode von „Vorwärts nach Wald“ auf „Rückwärts nach Wald“ verändert, um zu testen, ob die automatisch erfolgte Variablenauswahl stabil bleibt.

In das **Modell 10** werden testweise zum Vergleich mit dem Vormodell 6 ATC-Kombinationen nach fachlicher Relevanz und unabhängig von Ein- oder Ausschluss im vorherigen Modell eingebracht. Für die ausgewählten ATC-Kombinationen werden nachträglich Korrelationskoeffizienten bestimmt, um neben der fachlichen die statistische Relevanz erkennen zu können und damit die Auswahl zu unterstützen (Tabelle 19). Die Effektkoeffizienten  $\text{Exp}(B)$  aus dem vorherigen Modell sind dafür nur begrenzt geeignet, da diese sich bei Änderung der Variablen-Sets ebenfalls verändern.



Variable	Einschluss in M 9	Korrelation nach Kendalls-Tau b	Korrelation nach Cramer-V	p-Wert*
Risikoarzneistoffe-Auswahl	Ja	0,070	0,070	0,000
Spironolacton + Ramipril	Nein	0,065	0,065	0,000
Schleifendiuretika + Spironolacton	Nein	0,062	0,062	0,000
Clopidogrel + ASS	Nein	0,053	0,053	0,000
Allopurinol + Ramipril	Ja	0,039	0,039	0,000
Digitoxin + Amiodaron	nein	0,027	0,072	0,000

\*=angegeben ist die Signifikanz für die Korrelation nach Kendalls-Tau b (2-seitig, auf dem Niveau von 0,01 signifikant)

**Tabelle 19: Korrelationskoeffizienten fachlicher relevanter ATC-Kombinationen (Datenbasis 2005-2008: Entwicklungsdatensatz)**

In **Modell 11 und 12** wird zur Bestätigung der Variablenauswahl erneut der Modellierungsalgorithmus auf „Rückwärts nach Wald“ und „Vorwärts bedingt“ gewechselt. Auf die Gegenprüfung nach rein fachlicher Variablenauswahl mittels Einschluss-Methode wird verzichtet, da bei den einzelnen Modellierungsschritten und den Veränderungen der Variablen-Sets bereits fachliche Erwägungen berücksichtigt wurden.

Bei **Modell 13 und 14** werden schwer interpretierbare Variablen getestet. Das finale Modell soll neben einer guten Modellgüte Variablen enthalten, die sich fachlich interpretieren lassen. Die Regressionskoeffizienten im Modell 13 und 14 für die Anzahl von Verordnungen (Anz\_VO\_1J), für die Mindestanzahl an ATC pro Quartal über 4 Quartale (min\_ATC\_Q14) und von DDD pro ATC (DDDproATC\_12345) sind unerwartet negativ (nicht abgebildet). Naheliegend sind inhaltliche Überschneidungen von Variablen, weswegen mit SPSS 19.0 eine Prüfung auf bivariate Korrelationen erfolgt (Tabelle 20).

Bivariate Korrelationen (nach Pearson)						
	MIN_ ATC_ Q14	Anz_ ATC _1J	Anz_ REZEPT E_1J	Anz_ VO _1J	Anz_ DDD _1J	DDR ATC _12345
min_ATC_Q14	1,000	0,595	0,623	0,711	0,557	0,066
Anz_ATC_1J	0,595	1,000	0,626	0,589	0,311	-0,431
Anz_REZEPT E_1J	0,623	0,626	1,000	0,908	0,425	-0,091
Anz_VO_1J	0,711	0,589	0,908	1,000	0,522	0,052
Anz_DDD_1J	0,557	0,311	0,425	0,552	1,000	0,556
DDDproATC_12345	0,066	-0,431	-0,091	0,052	0,556	1,000

\*=Alle Korrelationen sind zweiseitig Signifikanz mit  $p=0,000$

**Tabelle 20: Prüfung auf Korrelationen zwischen schwer interpretierbaren Variablen (Datenbasis 2005-2008: Entwicklungsdatensatz)**

Die Überprüfung lässt mit einem Koeffizienten von 0,908 eine starke Korrelation zwischen Anz\_VO\_1J und Anz\_REZEPT  
E\_1J, sowie mit 0,711 zwischen Anz\_VO\_1J und min\_ATC\_Q14 vermuten. Es lässt sich dagegen zwar mit einem Koeffizienten von 0,595 nur eine mäßige Korrelation zwischen min\_ATC\_Q14 mit Anz\_ATC\_1J zeigen, da beide Variablen jedoch inhaltlich ähnlich sind, werden die Variablen getrennt getestet. Das Merkmal Anz\_DDD\_1J, das mit 0,556 eine mäßige Korrelation zu DDDproATC zeigt, ist bereits aus dem Modell entfernt. Für die Variable DDDproATC\_12345 lässt sich keine weitere störende Korrelation erkennen, weswegen sie trotz negativem Regressionskoeffizienten im **Modell 13** belassen wird. Die Variablen Anz\_VO\_1J und Anz\_ATC\_1J werden dagegen testweise in Modell 13 entfernt. In **Modell 14** wird Anz\_ATC\_1J allerdings wieder eingeschlossen, da min\_ATC\_Q14 keine Modellverbesserung erzielt, sondern durch den Modellierungsalgorithmus verworfen wird.

Da das finale Variablen-Set erreicht scheint, wird in **Modell 15** eine Bestätigung durch Rückwärtsmodellierung nach Wald durchgeführt. Da dabei eine Variable mehr im Modell verbleibt, wird diese in **Modell 16** entfernt und erneut rückwärts modelliert. Modell 16 stellt Modell 14 dar, das lediglich rückwärts statt vorwärts modelliert wurde. Da sich bei gleichem Variablen-Set eine gleichbleibende Modellgüte ergibt und alle eingeschlossenen Merkmale jetzt interpretierbar erscheinen, wird die Modellierung

an dieser Stelle beendet. Modell 14 wird als das finale Modell betrachtet.

## **6.6 Ergebnis**

Grundlage der Modellierung ist der Entwicklungsdatensatz mit 44.108 Patientenfällen. Da alle Datensätze vollständig sind, können alle Fälle in die Modellbildung einbezogen werden.

Die nachfolgende Übersicht (Tabelle 21) zeigt die Variablen, die mittels des Modellierungsalgorithmus und aufgrund fachlicher Entscheidungen ausgewählt wurden und in das finale Modell eingegangen sind (Modell 14). Die Variablen sind in der Reihenfolge ihrer Eingabe in die Statistiksoftware ausgewiesen. Zusätzlich sind der Regressionskoeffizient „B“ und der Effektkoeffizient „Exp(B)“ der einzelnen Prädiktoren aufgeführt (Anhang 10.10).

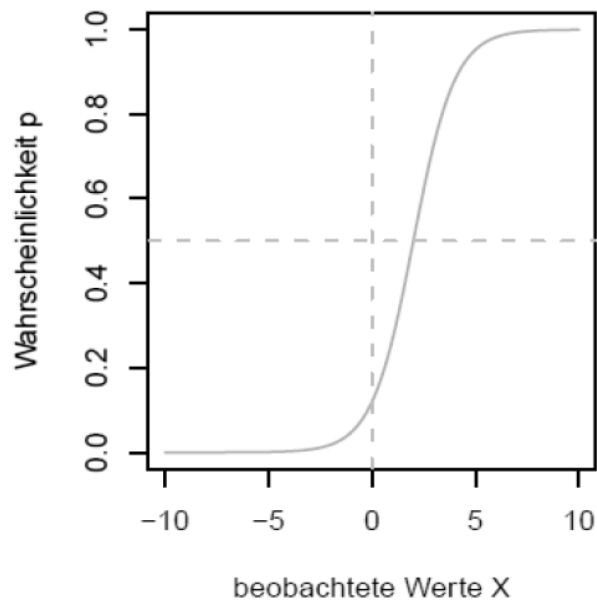
Nr.	Variable im Modell	Variablen- Beschreibung	B *	Signi- fikanz	Exp (B)	95% KI (2-seitig)
1	Anz_ATC_1_J	Anzahl ATC pro Jahr	0,036	,000	1,037	> 1
2	neu_Alter60708085	Alter (18-60,61-70,71-80,81-85 Jahre)	0,128	,000	1,137	> 1
3	neu_InKost1Jplus1	Log. Arzneimittelkosten pro Jahr +1	0,151	,000	1,163	> 1
4	ATC_Grp_n02a	Opiode	0,191	,000	1,210	> 1
5	ATC_Grp_b01aa04	Phenprocoumon	0,251	,000	1,268	> 1
6	ATC_Grp_b01ac04	Clopidogrel	0,238	,000	1,285	> 1
7	ATC_Grp_a02bc	Omeprazol	0,170	,000	1,186	> 1
8	ATC_Grp_a10ba02	Metformin	-0,140	,000	0,869	< 1
9	ATC_Grp_c03da01	Spirolacton	0,193	,000	1,213	> 1
10	ATC_Grp_n	Psychopharmaka	0,144	,000	1,155	> 1
11	neu_ist_Mann	männlich	0,124	,000	1,132	> 1
12	ATC_Grp_n02bb02	Metamizol	0,103	,000	1,109	> 1
13	ATC_Grp_c09aa05_03ca	Furosemid + Ramipril	0,102	,011	1,107	> 1
14	DDDproATC_12345	Ø DDD/ATC pro Jahr	-0,060	,000	0,942	< 1
15	Anz_Rezepte_1_J	Rezepte pro Jahr	0,007	,000	1,007	> 1
16	neu_ist_Rentner	Rentner	0,131	,000	1,140	> 1
17	ATC_Grp_b01ac05	Ticlopidin	0,093	,000	1,097	> 1
18	neu_ATC_Risiko	Risikoarzneistoffe	0,136	,000	1,145	> 1
19	ATC_Grp_j01	Antibiotika (i.v.)	0,080	,000	1,084	> 1
20	ATC_Grp_c01bd01	Amiodaron	0,198	,003	1,219	> 1
21	ATC_Grp_c01dx12	Molsidomin	0,115	,005	1,122	> 1
22	ATC_Grp_n06	Antidepressiva	0,066	,008	1,068	> 1
23	ATC_Grp_m04aa01_c09aa05	Allopurinol + Ramipril	0,096	,032	1,101	> 1
24	ATC_Grp_c03ca	Furosemid, Torasemid	0,063	,011	1,065	> 1
-	Konstante	-	-3,091	,000	0,045	-

\* B=Regressionskoeffizient

**Tabelle 21: Übersicht der Variablen im finalen Modell (Datenbasis 2005-2008: Entwicklungsdatsatz)**

### Regressionsformel

Mit der logistischen Regression werden Wahrscheinlichkeiten (p-Werte) berechnet. Die allgemeine Regressionsformel hat daher die Form einer Wahrscheinlichkeitsformel (Abbildung 29).



$$p(y) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

Abbildung 29: Allgemeine Formel einer logistischen Funktion (Garbade, 2009, S. 41)

### Ermittelte Regressionsformel (finales Modell)

Die Regressionsformel des entwickelten Vorhersagemodells (Abbildung 30) enthält die im Folgenden aufgeführten Parameter. Das  $\alpha$  stellt den Schnittpunkt auf der Y-Achse dar und wird auch als Intercept bezeichnet. Das  $\alpha$  ist negativ, würde aber, da es sich bei der Regressionsformel um eine Exponentialfunktion handelt, einen positiven Wert annehmen, wenn alle X den Wert 0 annehmen.

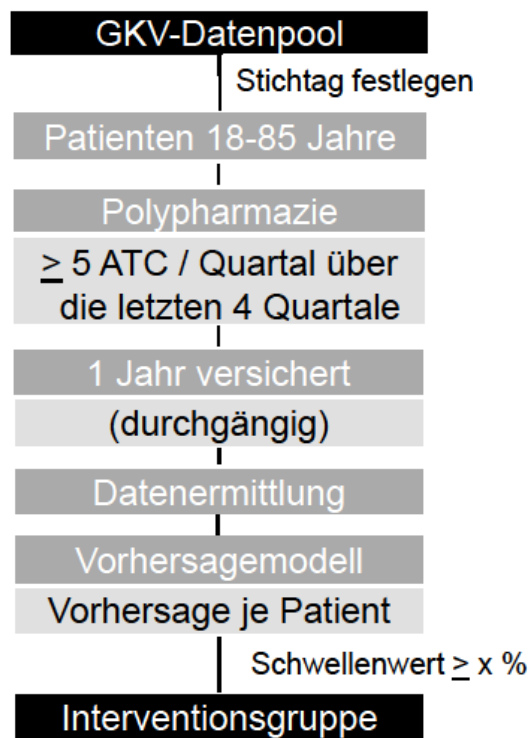
$$\beta_0 = -3,09050946565525$$

$$\begin{aligned} & (\beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_n X_n) = \\ & - 3,09050946565525 \\ & + 0,0363357772963722 \times \text{Anz\_ATC\_1J} \\ & + 0,128076843206269 \times \text{NEU\_ALTER60708085} \\ & + 0,150768813597773 \times \text{NEU\_LNKOST1JPLUS1} \\ & + 0,190900730370938 \times \text{ATC\_N02A\_SPEZ} \\ & + 0,237767289849138 \times \text{ATC\_B01AA04} \\ & + 0,250851613483668 \times \text{ATC\_B01AC04} \\ & + 0,170465780657153 \times \text{ATC\_A02BC} \\ & - 0,139909288403694 \times \text{ATC\_A10BA02} \\ & + 0,193350686844981 \times \text{ATC\_C03DA01} \\ & + 0,143943277724429 \times \text{ATC\_N\_SPEZ} \\ & + 0,124361683476461 \times \text{NEU\_IST\_MANN} \\ & + 0,103203360324656 \times \text{ATC\_N02BB02} \\ & + 0,101905170162405 \times \text{ATC\_C09AA05\_03CA} \\ & - 0,0596426105596153 \times \text{DDDproATC\_12345} \\ & + 0,00695855018996758 \times \text{Anz\_REZEPTTE\_1J} \\ & + 0,131076215497003 \times \text{NEU\_IST\_REN} \\ & + 0,092704848386765 \times \text{ATC\_B01AC05} \\ & + 0,135661870324869 \times \text{NEU\_RISIKO\_MIND1} \\ & + 0,0802860762060424 \times \text{ATC\_J01\_SPEZ} \\ & + 0,198400575592824 \times \text{ATC\_C01BD01} \\ & + 0,114817713938228 \times \text{ATC\_C01DX12} \\ & + 0,0659441855922937 \times \text{ATC\_N06\_SPEZ} \\ & + 0,0629928515588899 \times \text{ATC\_C03CA} \\ & + 0,0964329404598315 \times \text{ATC\_M01AA01\_C09AA05} \end{aligned}$$

**Abbildung 30: Regressionsformel zum finalen Modell (Datenbasis: 2005-2008)**

### Finaler Algorithmus für den Modelleinsatz

Das Vorhersagemodell lässt Patienten mit Polypharmazie und einer vorhergesagten und möglicherweise arzneimittelbedingten Hospitalisation innerhalb von einem Jahr erkennen. Das entwickelte Modell ist für die routinemäßige Anwendung an GKV-Leistungsdaten in einen Algorithmus einzubetten, der nachfolgend dargestellt ist (Abbildung 31). Die auf diese Weise erkannten Patienten sollen eine Interventionsgruppe mit einem vermutlichen Bedarf an klinisch-pharmazeutischer Betreuung ergeben.



**Abbildung 31: Finaler Algorithmus zur Patientenidentifizierung**

Für den routinemäßigen Einsatz des Modells beziehungsweise den Algorithmus sind ein Auswertungstichtag und ein Wiederholungszyklus für die Modellanwendung festzulegen. Für die Bestimmung der Polypharmazie verändern sich mit fortschreitenden Auswertungstichtagen die zugrunde zu legenden Quartale als Auswertungszeitraum. Zur finalen Auswahl der Interventionsgruppe ist zusätzlich nach Anwendung des Vorhersagemodells ein Schwellenwert zu definieren. Die Festlegung des Schwellenwerts ist abhängig von der gewünschten Größe der Interventionsgruppe und der gewünschten Vorhersagegüte des Anwenders (Kap. 7). Unter Umständen

sind „Vorfahrtsregeln“ für andere Identifizierungsalgorithmen und spezielle Ausschlusskriterien durch den Anwender zu definieren.

## 7 Ergebnisprüfung und Zielerreichung (Evaluation)



Die Ergebnisprüfung erfolgt am Validierungsdatensatz und umfasst die Betrachtung der einzelnen Prädiktoren und des Modells insgesamt. Der Validierungsdatensatz wird in gleicher Weise ermittelt wie der zur Modellentwicklung verwendete Entwicklungsdatensatz, jedoch auf Grundlage von aktuelleren GKV-Routinedaten. Die Datenbasis sind Daten aus den Jahren von 2007-2010 statt aus den Jahren 2005-2008. Das jeweils letzte Jahr des Auswertungszeitraumes stellt das Jahr dar, für das eine potenziell arzneimittelbedingte Krankenhauseinweisung vorhergesagt werden soll. Der Validierungsdatensatz enthält Daten zu denselben Variablen wie im Entwicklungsdatensatz. Optimierte und neue Variablen, die für das Modell benötigt werden, werden in dem Datensatz durch die Autorin nachträglich angelegt.

### 7.1 Beurteilung der Prädiktoren

#### *Effektrichtung und Effektgröße*

Das Vorzeichen der in SPSS 19.0 ausgegeben, nicht standardisierten *Regressionskoeffizienten*  $B$  zeigt die Richtung des Variableneffekts auf das Zielmerkmal an. Ein positiver Regressionskoeffizient weist darauf hin, dass mit zunehmendem Wert des Prädiktors die Krankenhauseinweisungswahrscheinlichkeit zunimmt. Aufgrund der Hinweise aus der Literatur (Business Understanding-Phase) und aufgrund der Ergebnisse der Voranalysen (Data Understanding) wurden die Variablen so formuliert, dass sich positive Regressionskoeffizienten ergeben sollten. Mit Ausnahme von zwei Variablen im finalen Modell zeigen alle Prädiktoren eine positive Effektrichtung und entsprechen somit den Erwartungen (Anhang 10.10). Für die zwei Variablen mit ne-



gativer Effektrichtung, *ATC\_Metformin* und *DDDproATC\_12345*, hat sich dieses in den vorherigen Analysen angedeutet und wird als fachlich interpretierbar bewertet. Die Größe des Effekts, den jede einzelne Variable auf die richtige Klassifizierung der Patienten im finalen Modell hat, lässt sich anhand des *Effektkoeffizienten*  $Exp(B)$  abschätzen. Die Effektkoeffizienten liegen zwischen 0,869 für *ATC\_Metformin* und 1,285 für *ATC\_Clopidogrel* und zeigen damit keine extremen Unterschiede, durch die sich einzelne Variablen besonders hervorheben. Für eine vergleichende Betrachtung der Variablen ist die Bestimmung der Variablengüte besser geeignet. Die Effektrichtungen und -größen lassen insgesamt die Erwartungen aus der Literaturrecherche und die beobachteten Tendenzen aus den Voranalysen erkennen und werden als fachlich interpretierbar bewertet (Kap. 8).

### Signifikanz

Der Signifikanztest für die *Regressionskoeffizienten*  $B$  der Prädiktoren zeigt Signifikanz an ( $p$ -Werte  $< 0,05$ ). Die *Effektkoeffizienten*  $Exp(B)$  liegen im beidseitigen 95%-Konfidenzintervall überwiegend über dem Betrag von 1, was gut erklärende Variablen zeigt (Anhang 10.10). Der signifikante Einfluss der Variablen auf das Zielmerkmal wird insbesondere angezeigt durch die *Chi-Quadrat-Zunahme* in jedem Schritt der Modellbildung. Alle Chi-Quadrat-Veränderungen sind mit  $p$ -Wert  $< 0,05$  signifikant. Damit ist davon auszugehen, dass alle Prädiktoren im Modell einen signifikanten Einfluss auf das Zielmerkmal haben.

### Güte

Die Güte der einzelnen Prädiktoren wird optimaler Weise durch die Bestimmung des *AUC-Wertes* (engl.: area under curve) ausgedrückt werden. Dieser kann für eine einzelne Variable durch Bildung eines Ein-Variablen-Modells aus dem Merkmal und dem Zielmerkmal bestimmt werden. Darüber hinaus ist die Bestimmung eines *kumulierten AUC-Wertes* bei schrittweiser Einführung von Merkmalen möglich. Je höher der AUC-Wert einer Variablen, desto höher die Güte. Die Berechnung der AUC-Werte für jede einzelne Variable wird in SPSS 19.0 nicht unterstützt. Für die vergleichende Betrachtung der Variablen wird daher die jeweilige *Chi-Quadrat-Veränderung*

während der schrittweisen Modellierung in SPSS 19.0 angefordert und herangezogen (Anhang 10.10). Die Reihenfolge, in der die Variablen zu einer Chi-Quadrat-Veränderung beitragen, das Modell verbessern und in das Modell eingehen, zeigt die Güte des jeweiligen Prädiktors im Modell an. Das Ergebnis lässt sich mit *Anz\_ATC\_1J* als Variable mit der höchsten Chi-Quadrat-Veränderung im Modell fachlich gut interpretieren und entspricht den Erwartungen.

## 7.2 Beurteilung des Modells

### *Signifikanz*

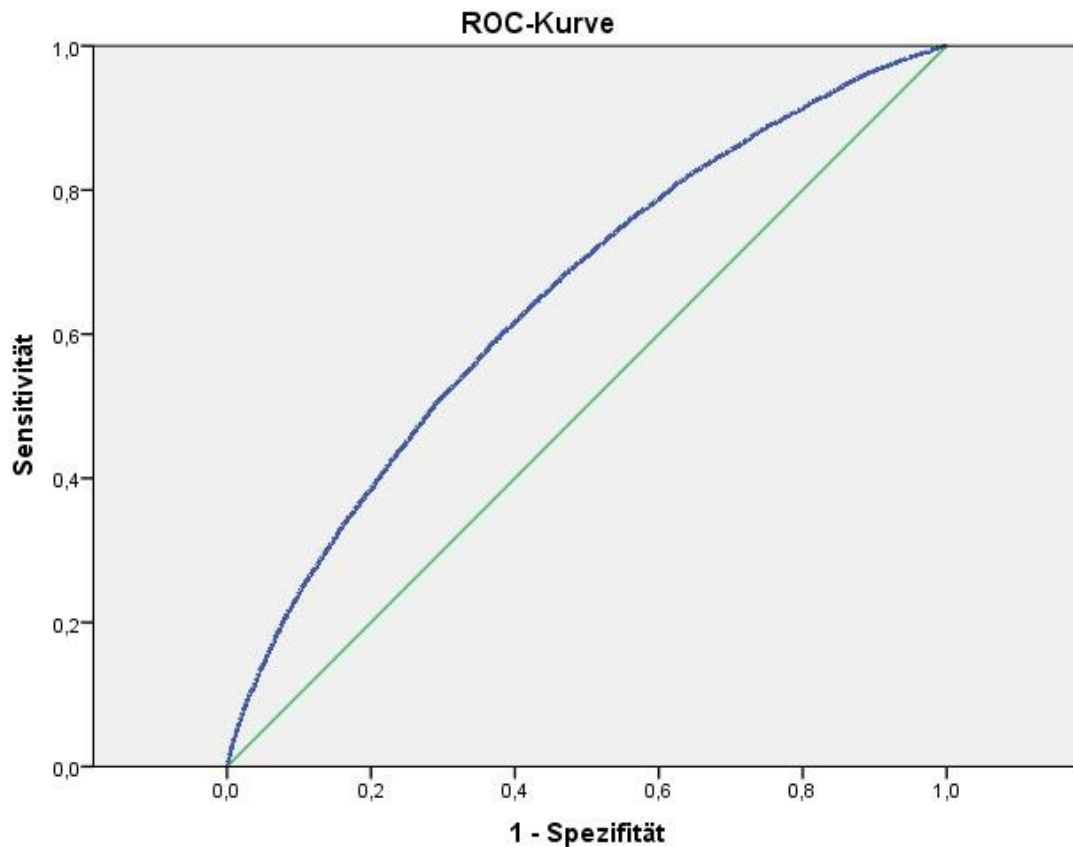
Der *Omnibus-Test der Modellkoeffizienten* zeigt im Gesamtergebnis für das Modell mit  $p=0,000$  eine Senkung des *Chi-Quadrat-Werts* um 3.067. Jede Abnahme stellt eine Verbesserung dar. Das heißt, über alle 24 Schritte der Modellierung hinweg haben alle eingegangenen Variablen zusammen das Modell höchst signifikant verbessert.

### *Güte der Modellanpassung*

Bei linearen Regressionen wird üblicherweise zur Überprüfung der Modellgüte das *Bestimmtheitsmaß R-Quadrat* ermittelt. Bei nicht-linearen Regressionen, wie der logistischen Regression, ist die Ermittlung von R-Quadrat nicht ohne weiteres möglich. Es existieren verschiedene Methoden, ein *Pseudo-R-Quadrat* zu ermitteln. Diese ergeben unterschiedliche Werte, sind nicht direkt miteinander vergleichbar und daher schwerer zu interpretieren. Zur Abschätzung der Modellgüte wird in SPSS 19.0 das *Pseudo-R-Quadrat nach Nagelkerke* ausgegeben. Dieses ist so modifiziert, dass es Werte zwischen 0 und 1 annehmen kann (Bühl, 2009). Ein *Pseudo-R-Quadrat nach Nagelkerke* von 9,1% bedeutet, dass die Varianz des Zielmerkmals nur zu einem kleinen Anteil durch die Gesamtheit der im Modell verwendeten Merkmale erklärt werden kann. Der Anteil sollte möglichst 20% erreichen und ist somit gering. Die ebenfalls in SPSS 19.0 angeforderte Klassifizierungstabelle zeigt eine *richtige Klassifizierung* der Patienten von 64,5%. Im Vergleich zum Ausgangsmodell ohne einen der 24 Prädiktoren steigt der Wert im finalen Modell von 61,4% ausgehend um 3,1%. Der Anteil der 24 erklärenden Variablen an dem Anstieg ist damit verhältnismäßig gering.

Allerdings sind andere Gütemaße bei der logistischen Regressionsanalyse geeigneter als das *Pseudo-R-Quadrat nach Nagelkerke* und der Anteil der *richtigen Klassifizierung*. Diese lassen jedoch während der Modellierung eine grobe Abschätzung der Modellgüte zu.

Für eine genauere Bestimmung der Modellgüte wird der AUC-Wert des Modells ermittelt. Die Bestimmung erfolgt nicht am Entwicklungsdatensatz, sondern an einem Validierungsdatensatz. Dieser beinhaltet zwei Jahre aktuellere GKV-Leistungsdaten, im Vergleich zum Entwicklungsdatensatz ähnlich viele Patientenfälle und eine ähnliche Verteilung des Zielmerkmals (Anhang 10.11). Dieses ist anzustreben, da der AUC-Wert je nach Stichprobe und Stichprobengröße schwanken kann. Die Ermittlung des *AUC-Wertes* erfolgt mittels Anwendung des finalen Modells am Validierungsdatensatz. Auf Grundlage der auf diese Weise für jeden Patienten berechneten Vorhersagewahrscheinlichkeit wird eine *ROC-Kurve* (Receiver Operating Characteristics - Kurve) in SPSS 19.0 angefordert (Abbildung 32). In dieser werden Spezifität (Falsch-Positiv-Rate) und Sensitivität (Richtig-Positiv-Rate) für jeden Vorhersagewert gegeneinander aufgetragen. Die Fläche unter der *ROC-Kurve* ergibt den *AUC-Wert*.



Datenbasis	AUC-Wert	Asymptotische Signifikanz	95% KI Untergrenze	95% KI Obergrenze
Validierungsdaten	65,2%	0,000	64,7%	65,7%

**Abbildung 32: ROC-Kurve und AUC des Modells (Datenbasis 2007-2010)**

Die Diagonale in der Grafik zeigt eine *ROC-Kurve* für die zufällige Zuordnung der Patienten ohne Anwendung eines Modells, das heißt bei einer 50%-prozentigen Wahrscheinlichkeit für eine Krankenhauseinweisung, und hat einen *AUC-Wert* von 50%. Die *ROC-Kurve* für das Modell liegt oberhalb der Diagonalen und hat einen *AUC-Wert* von 65,2%. Damit ist ausgedrückt, in welchem Maße die Zuordnung der Patienten durch das Modell genauer erfolgt als bei zufälliger Zuordnung. *AUC-Werte* von 70% oder größer sind gut und anzustreben. Bei Modellen für komplexe medizinische Fragestellungen wird allerdings oftmals eine geringere Modellgüte erreicht als beispielsweise bei technischen Fragestellungen (Backhaus, 2008). Die ermittelte Güte für das entwickelte Modell wird daher als ausreichend bewertet.

### *Stabilität*

Ein Hinweis auf eine gute Stabilität des Modells ergibt sich aus der gleichbleibenden Variablenauswahl bei vorwärts und rückwärts Modellierung während der Modellbildung am Entwicklungsdatensatz. Genauere Hinweise zur Stabilität ergeben sich durch die Anwendung des Modells am Validierungsdatensatz mit zwei Jahre aktuelleren Daten. Die ermittelte Güte des Modells am Validierungsdatensatz ist, wie dargestellt, an der aktuelleren Datenbasis ausreichend gut und damit über die zwei Jahre stabil geblieben. Um sicherzustellen, dass die Modellgüte auf einem akzeptablen Niveau bleibt, sollte diese bei kontinuierlicher Anwendung des Modells regelmäßig erneut an aktualisierten Datensätzen auf Stabilität überprüft werden.

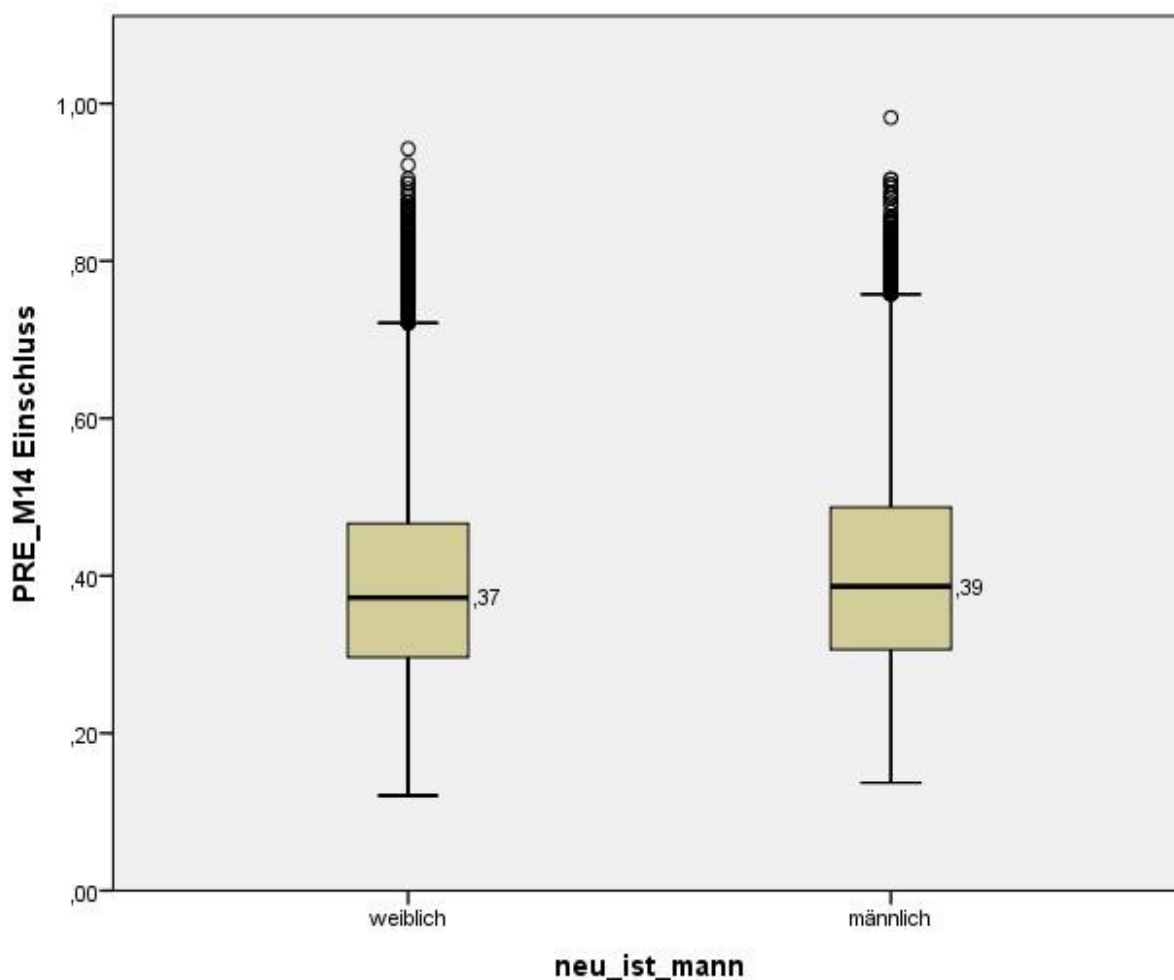
### *Test auf Ausreißer*

Beim Test auf Ausreißer werden 8 von 45.739 Fällen im Validierungsdatensatz als Ausreißer ermittelt. Die fallweise Auflistung enthält alle Patienten, die untypische Fälle darstellen und vom Modell nicht gut beschrieben werden können (Anhang 10.11). Eine Zuordnung zu der Gruppe Krankenhauseinweisung innerhalb des nächsten Jahres erfolgt ab einer Vorhersagewahrscheinlichkeit von 0,5 oder größer. Bei den ermittelten Ausreißern erfolgt eine falsche Klassifizierung durch das Modell. Aufgrund der wenigen betroffenen Fälle ist davon auszugehen, dass es bei den Berechnungen der Koeffizienten nicht zu umfangreichen Verzerrungen und damit zu einer schlechten Modellanpassung gekommen ist.

### **7.3 Ausprägung wichtiger Variablen**

Da die Güte des finalen Modells im Validierungsdatensatz ausreichend ist, wird die Ausprägung der Variablen am aktuelleren Validierungsdatensatz ( $n=45.739$ ) betrachtet. Der Validierungsdatensatz enthält GKV-Routinedaten aus den Jahren 2007-2010. Den Abbildungen liegen die Vorhersagewahrscheinlichkeiten zugrunde, die bei der Anwendung des entwickelten Modells am Validierungsdatensatz für jeden Patienten berechnet wurden (PRE-Werte). Für ausgewählte Prädiktoren sind im Folgenden die Wahrscheinlichkeiten für eine (potenziell arzneimittelbedingte) Hospitalisie-

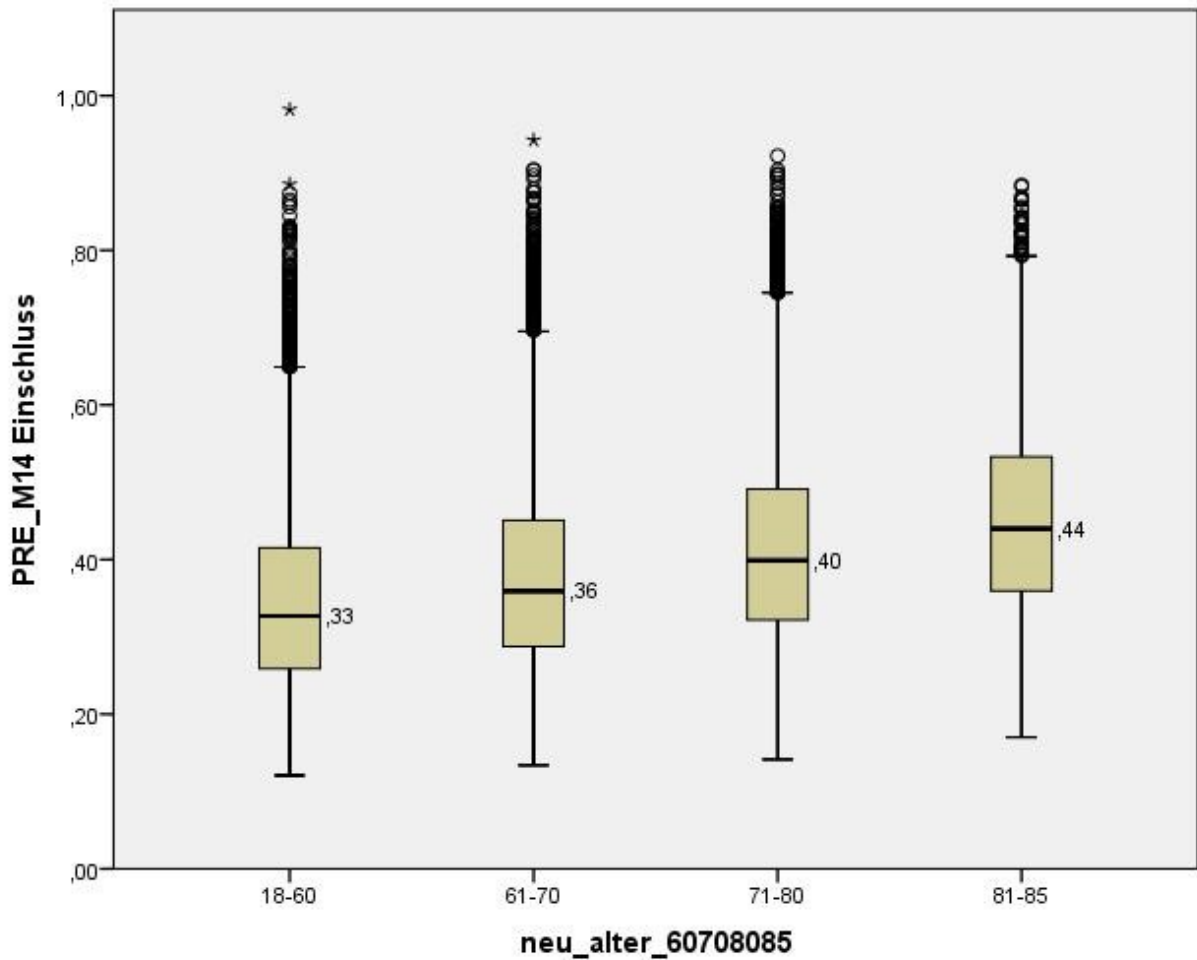
rung nach Fallgruppen im Boxplot-Diagramm dargestellt (Abbildung 33). Für alle dargestellten Variablen wurden ein Vergleich der Mittelwerte und eine Prüfung auf signifikante Unterschiede durchgeführt (siehe beispielhaft Anlage 10.13).



ist_Mann	weiblich	männlich
Anzahl	29.155	16.584
%	63,7	36,3

**Abbildung 33: Vorhersagewahrscheinlichkeit für 2010 nach Geschlecht**

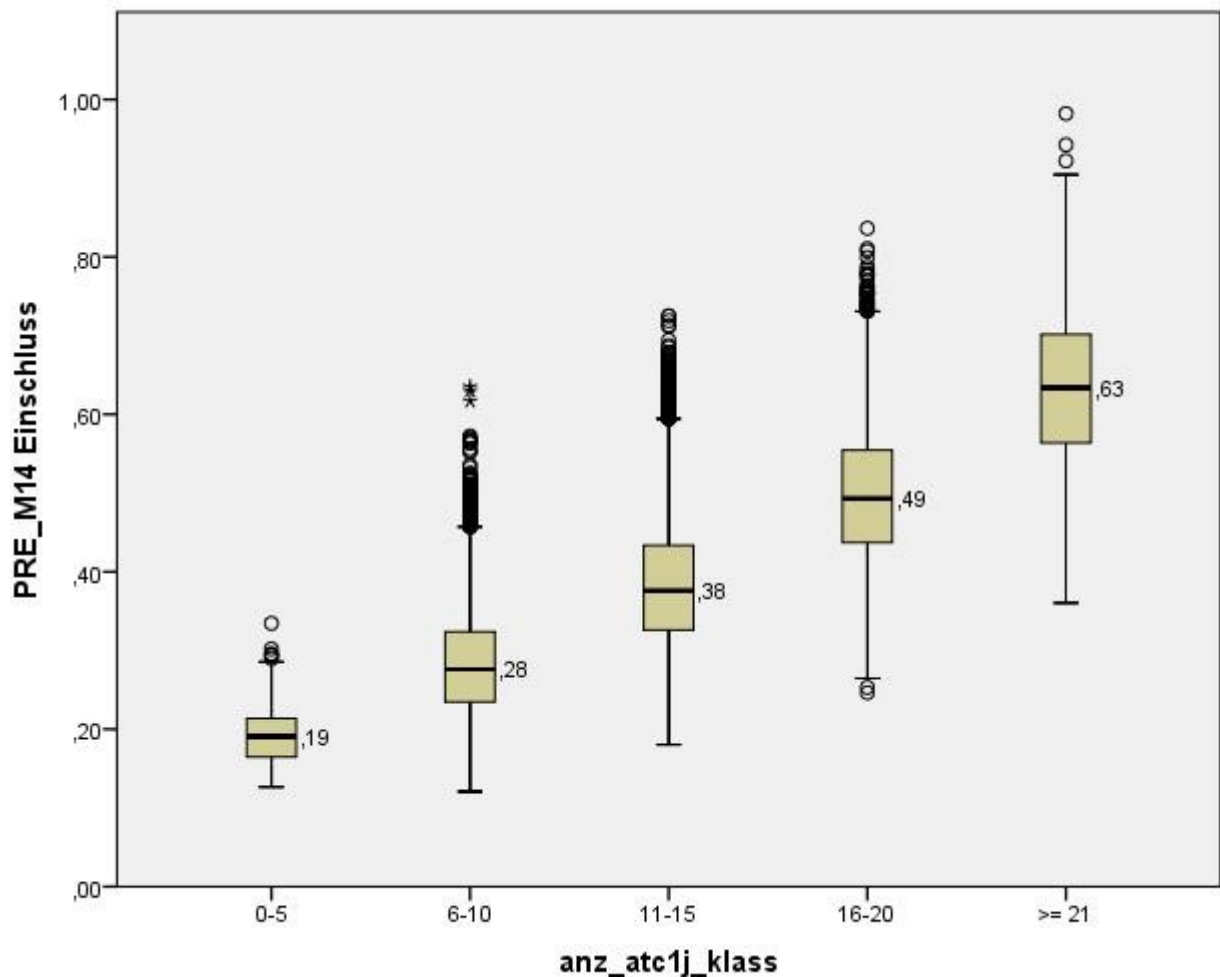
Im Validierungsdatensatz sind deutlich mehr weibliche als männliche Patienten vertreten. Im Boxplot (Abbildung 34) zeigen die männlichen Patienten mit Polypharmazie im Schnitt eine höhere Vorhersagewahrscheinlichkeit (PRE-Wert) für eine potenzielle ABH. Die Überprüfung mittels Varianzanalyse bestätigt, dass der Unterschied signifikant ist.



Alter_60708085	18-60	61-70	71-80	81-85
Anzahl	8.529	14.488	17.190	5.532
%	18,6	31,7	37,6	12,1

**Abbildung 34: Vorhersagewahrscheinlichkeit für 2010 nach Alter zum 31.12.2009**

Die Altersgruppe der 61-80 jährigen ist in dem Analysedatensatz am stärksten vertreten. Im Boxplot (Abbildung 35) zeigen sich in den höheren Altersgruppen höhere durchschnittliche PRE-Werte. Die Unterschiede sind bei genauere Prüfung mittels Varianzanalyse und Tamhane-T2-Mehrvergleichstest signifikant ( $p=0,000$ ).

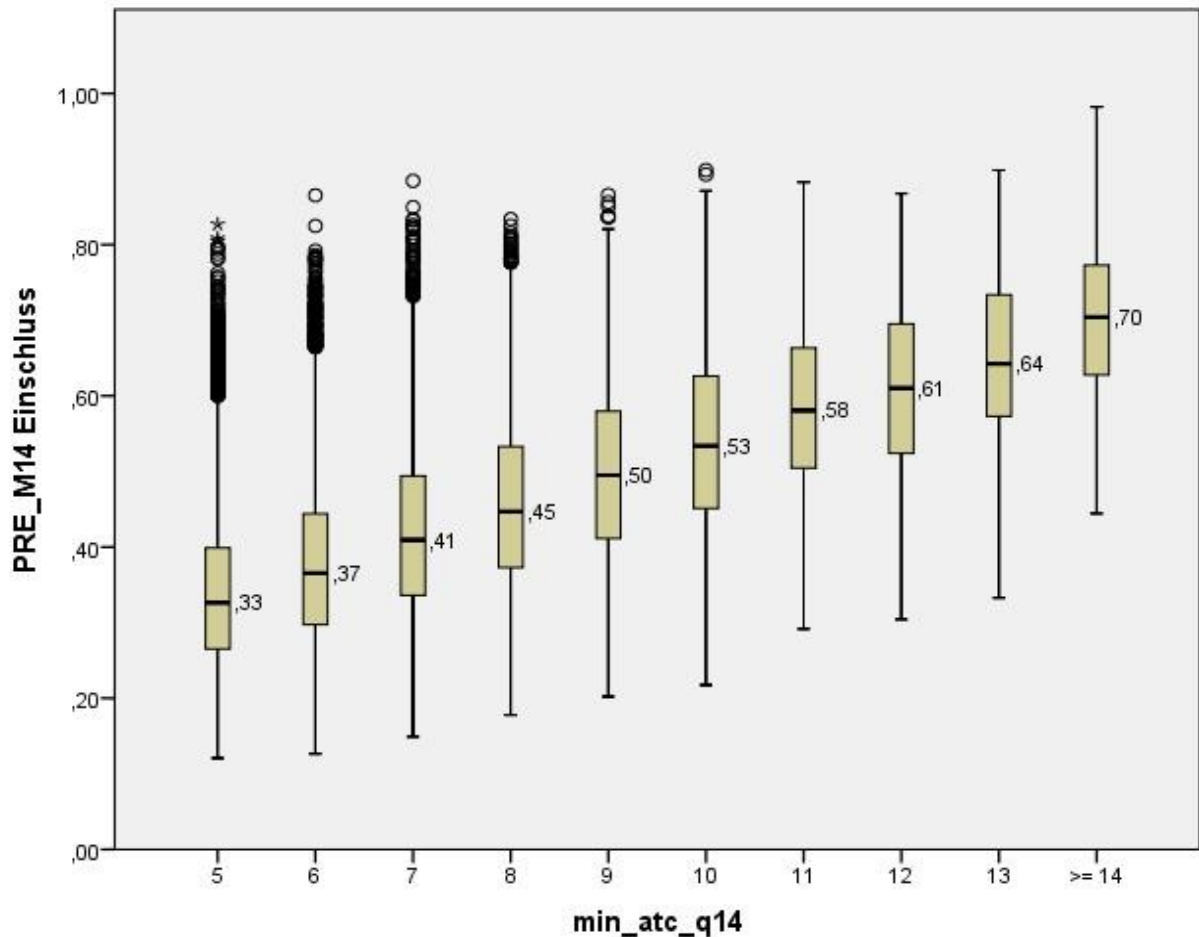


Anz_atc1j_klassiert	0-5	6-10	11-15	16-20	>=21
Anzahl	156	13.080	20.029	8.923	3.551
%	0,3	28,6	43,8	19,4	7,8

Abbildung 35: Vorhersagewahrscheinlichkeit für 2010 nach Anzahl ATC (2009)

Bei Betrachtung der Anzahl an Arzneistoffen pro Jahr, für die Darstellung gruppiert in 5 Klassen, zeigt das Boxplot-Diagramm (Abbildung 36) mit zunehmender Anzahl an ATC einen Anstieg der PRE-Werte. Die Unterschiede sind signifikant ( $p=0,000$ ). Die Variable zeigt im Modell einen relevanten prädiktiven Effekt (Kap. 8.1.).

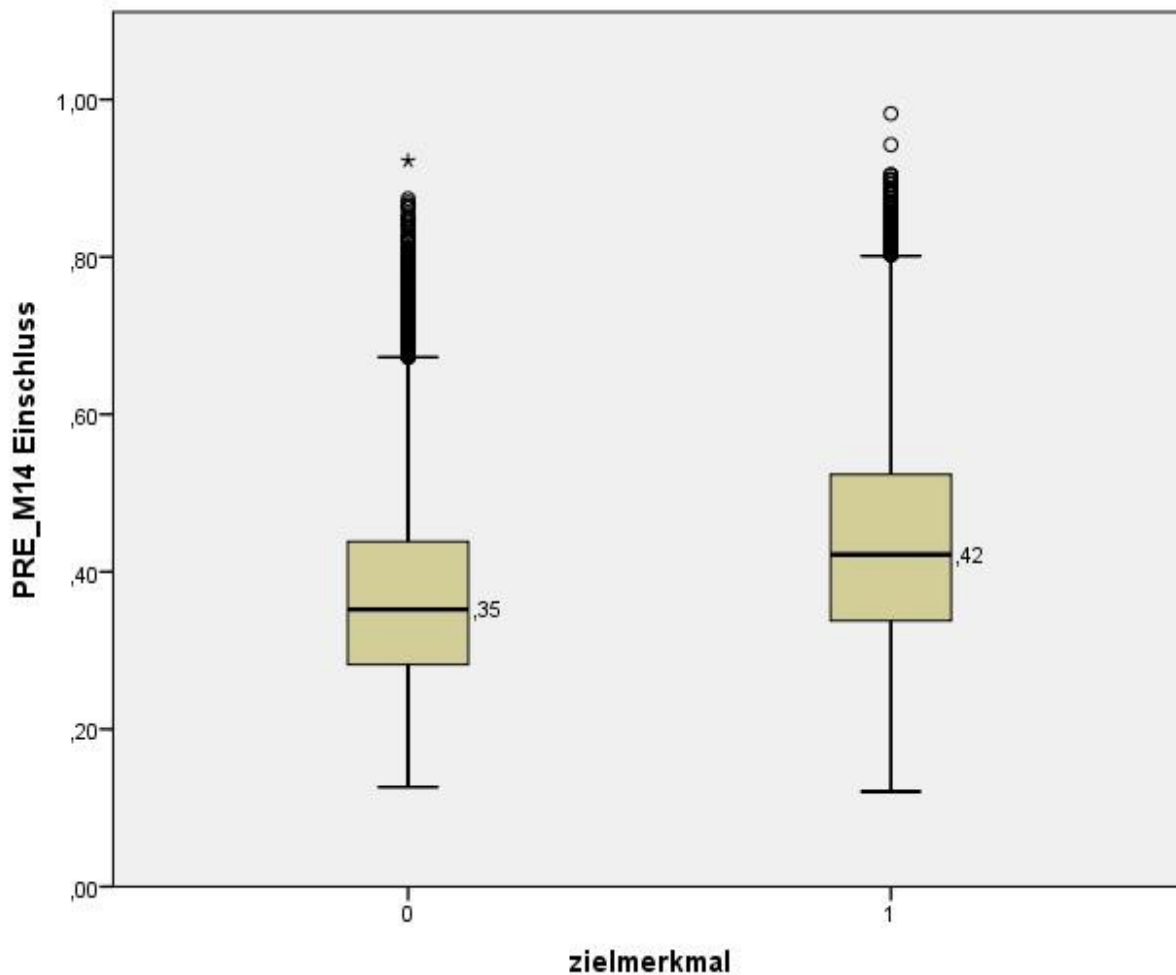




min_ATC_Q14	5	6	7	8	9	10	11	12	13	≥14
Anzahl	17.884	11.331	7.041	4.070	2.411	1.411	709	393	225	264
%	39,1	24,8	15,4	8,9	5,3	3,1	1,6	0,9	0,5	0,6
% (Gruppe)				88,2				10,8		1,1

**Abbildung 36: Vorhersagewahrscheinlichkeit für 2010 nach Polypharmazie-Gruppe (2009)**

Die Variable min\_ATC\_Q14 hat als Prädiktor keinen Eingang in das Modell gefunden, dennoch wird zum Vergleich mit anderen Variablen ein Diagramm dargestellt. Das Boxplot-Diagramm (Abbildung 37) zeigt mit zunehmendem Grad der Polypharmazie, im Sinne einer zunehmenden Anzahl an Arzneistoffen pro Quartal gemäß der entwickelten Definition, eine steigende Wahrscheinlichkeit für eine Krankenhaus-einweisung. Der Vergleich der Mittelwerte mit Prüfung auf signifikante Unterschiede der Fallgruppen bestätigt, dass die Unterschiede zwischen den Mittelwerten auf einem Niveau von 0.05 signifikant sind (Anhang 10.13).



		Zielmerkmal=0	Zielmerkmal=1	Gesamt
Gesamt	Anzahl	27.640	18.099	45.739
	%	60,4	39,6	100%
PRE < 0,5	Anzahl	23.887	12.733	36.620
	%	65,2	34,8	100%
PRE ≥ 0,5	Anzahl	3.753	5.366	9.119
	%	41,2	58,8	100%

\* Zielmerkmal=1: Krankenhauseinweisung innerhalb von 12 Monaten

Abbildung 37: Vergleich von Vorhersage und tatsächlich eingetretenem Zielmerkmal in 2010

Bei Betrachtung der ermittelten Vorhersagewahrscheinlichkeiten in Abhängigkeit des tatsächlich eingetretenen, vorherzusagenden Zielmerkmals zeigt sich im Boxplot-Diagramm ein höherer durchschnittlicher PRE-Wert in der Gruppe der Patienten mit Polypharmazie und einem Krankenhausaufenthalt im Folgejahr (Zielmerkmal=1). Der Unterschied ist mit  $p=0,000$  höchst signifikant. Während mittels einer Kreuztabelle über alle Fälle hinweg betrachtet 39,6% der Patienten tatsächlich ins Krankenhaus eingewiesen werden, erhöht sich dieser Prozentsatz auf 58,8% wenn ausschließlich die durch das Modell ermittelten Patienten mit Polypharmazie mit einem PRE-Wert von  $\geq 0,5$  betrachtet werden. Dieser positive prädiktive Wert (PPV) bedeutet einerseits eine verbesserte Erkennung der Zielgruppe, andererseits verbleibt noch ein großer Anteil an Patienten, der durch das Modell falsch vorhergesagt wird (Kapitel 7.4).

#### 7.4 Klassifikationsmatrix (Zielerreichung)

Für den angestrebten praktischen Einsatz des entwickelten Modells zur Erkennung von Patienten mit einem Bedarf an klinisch-pharmazeutischer Betreuung gilt es zu prüfen, zu welchem Grad das Modell Patienten mit Polypharmazie erkennen kann, die tatsächlich innerhalb eines Jahres eine Hospitalisation erleiden. Wie zu Beginn der Data-Mining Analyse herausgearbeitet, wird dabei angenommen, dass die durch das Modell vorhergesagte Hospitalisation potenziell arzneimittelbedingt ist und einen Bedarf an klinisch-pharmazeutischer Betreuung ausdrückt. Das Modell liefert für jeden Patientenfall im Validierungsdatensatz eine Vorhersagewahrscheinlichkeit (PRE-Wert) (Abbildung 38).

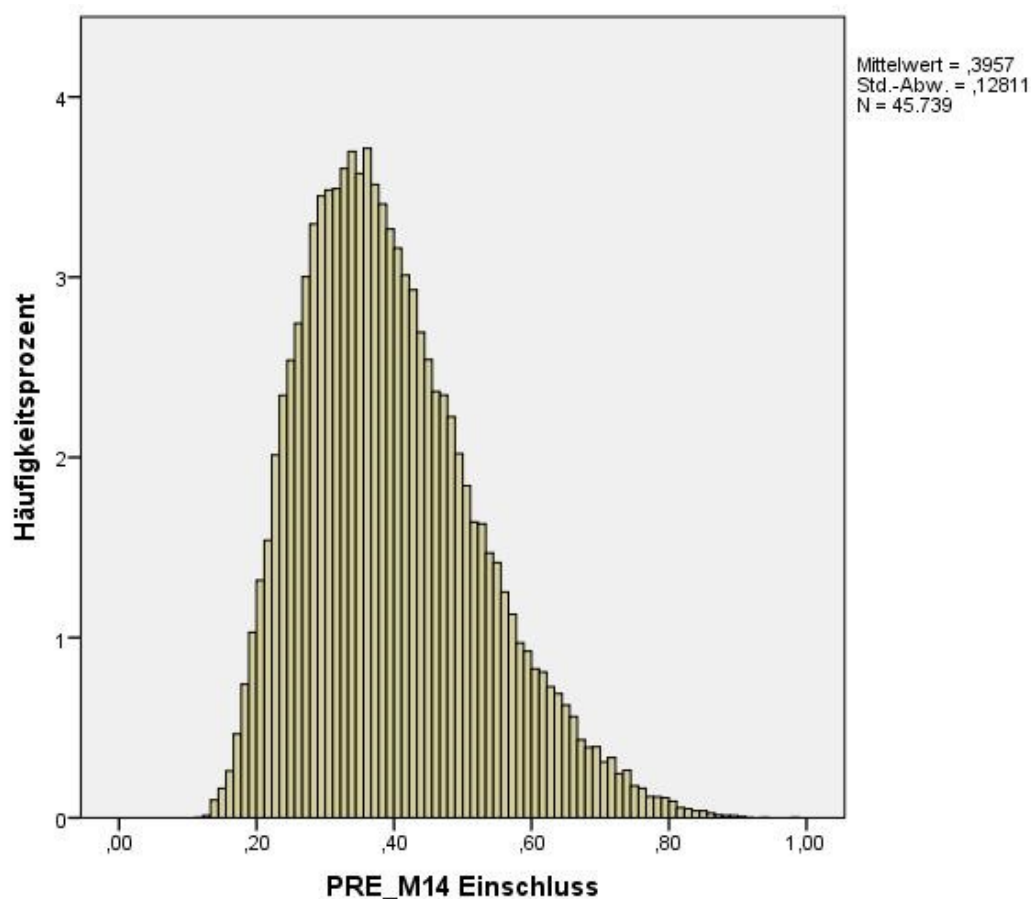
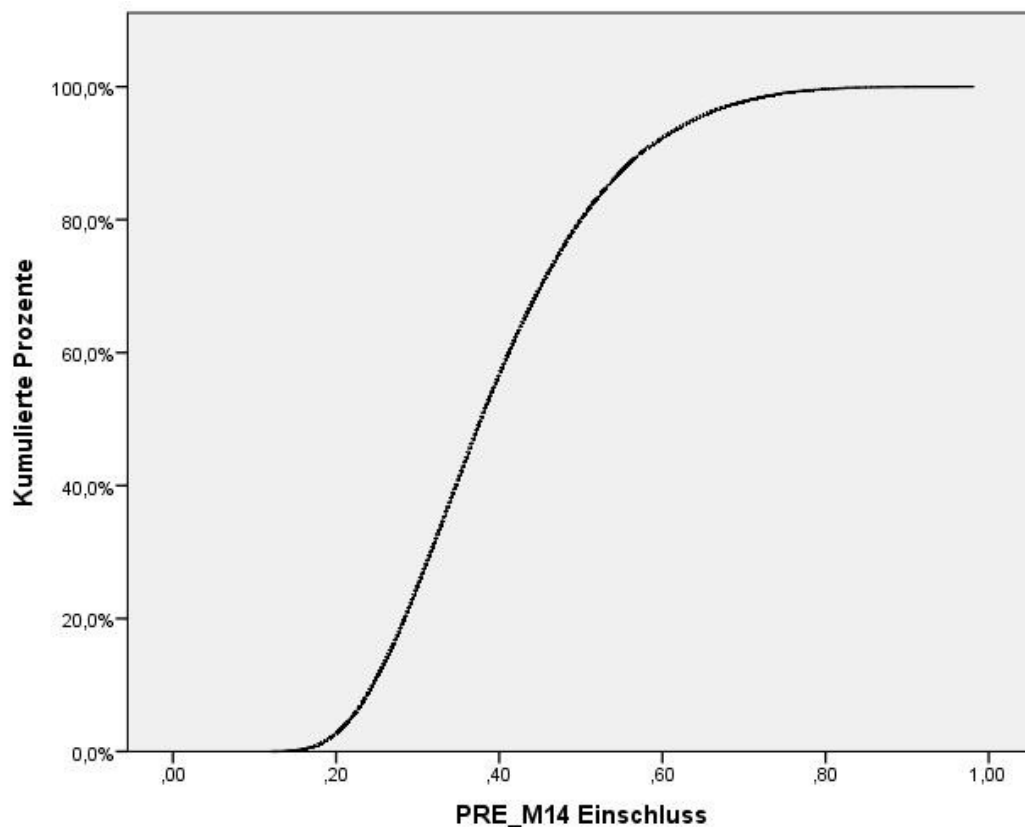


Abbildung 38: Häufigkeitsprozente der Vorhersagewahrscheinlichkeiten für 2010 (Datenbasis Validierungsdatensatz: 2007-2010)

Werden die Häufigkeitsprozentage kumuliert dargestellt, ergibt sich nachfolgendes Bild (Abbildung 39).



Perzentile	PRE-Wert	Anzahl Patienten (PRE $\geq$ Schwelle)	%-Anteil von Gesamt
80%	0,50	9.119	20%
85%	0,53	6.675	15%
90%	0,57	4.491	10%
95%	0,63	2.367	5%
97%	0,69	1.161	3%
99%	0,74	474	1%

**Abbildung 39: Kumulierte Häufigkeitsprozentage der Vorhersagewahrscheinlichkeiten für 2010 (Datenbasis Validierungsdatensatz: 2007-2010)**

Aus der Abbildung der kumulierten Häufigkeitsprozente lässt sich abschätzen, dass ein PRE-Wert von 0,5 oder mehr bei etwa 20% der Patienten im Validierungsdatensatz auftritt. In SPSS 19.0 ist für log. Regressionsmodelle ein PRE-Wert von 0,5 als Schwellenwert voreingestellt. Von den 45.739 Patienten im Validierungsdatensatz wird durch das Modell für 9.119 Patienten mit Polypharmazie ein PRE-Wert von  $\geq 0,5$  ermittelt und für diese ein Krankenhausaufenthalt im Folgejahr vorhergesagt.

Eine Aussage über die Vorhersagegenauigkeit des Modells lässt sich mit Hilfe des PPV treffen. In einer Klassifikationsmatrix lassen sich für den Schwellenwert von PRE  $\geq 0,5$  die genauen Zahlenverhältnisse der Patienten zur Berechnung der PPV gegenüberstellen (Tabelle 22). Neben der Ermittlung der PPV und der Größe der Interventionsgruppe lassen sich zusätzlich weitere Gütemaße, wie Sensitivität und Spezifität, bestimmen.

Schwellenwert für Patientenauswahl	Anzahl Patienten Zielmerkmal=0	Anzahl Patienten- Zielmerkmal=1	Gesamt	Bewertung
PRE <0,5 (nein)	23.887	12.733	36.620	PPV: 58,8%
PRE $\geq 0,5$ (Ja)	3.753	5.366	9.119	
Gesamt	27.640	18.099	45.739	
Bewertung	Spezifität: 86,4%    Sensitivität: 29,6%			

Spezifität=Richtig-Negative; Sensitivität=Richtig-Positive; PPV=Richtig-Positive/ (Richtig-Positive +Falsch-Negative)

**Tabelle 22: Klassifikationsmatrix für den Schwellenwert PRE $\geq 0,50$  und Modellanwendung am Validierungsdatensatz (2007-2010)**

Von den 9.119 Patienten mit Polypharmazie und einer vorhergesagten Hospitalisation haben 58,8% (5.366 von 9.119) tatsächlich einen Krankenhausaufenthalt im Folgejahr. Die PPV für das Modell bei dem Schwellenwert von PRE  $\geq 0,5$  liegt demnach bei 58,8%, die Spezifität bei 86,4% (23.887 von 27.640) und die Sensitivität bei 29,6% (5.366 von 18.099).

Die PPV ändert sich mit Verschiebung des Schwellenwerts. Dadurch lassen sich PPV und damit die Vorhersagegenauigkeit optimieren, wobei zugleich die erkannte Patientengruppe (Interventionsgruppe) nicht zu klein werden soll.

Schwellenwert für Patientenauswahl	Anzahl Patienten Zielmerkmal=0	Anzahl Patienten Zielmerkmal=1	Gesamt	Bewertung
PRE <0,69 (Nein)	27.310	17.268	44.578	PPV: 71,6%
PRE ≥0,69 (Ja)	330	831	1.161	
Gesamt	27.640	18.099	45.739	
Bewertung	Spezifität: 98,8%	Sensitivität: 4,6%		

Spezifität=Richtig-Negative; Sensitivität=Richtig-Positive; PPV=Richtig-Positive/(Richtig-Positive+ Falsch-Negative)

**Tabelle 23: Klassifikationsmatrix für den Schwellenwert PRE<sub>≥0,69</sub> und Modellanwendung am Validierungsdatensatz (2007-2010)**

Wird ein höherer Schwellenwert von PRE <sub>≥0,69</sub> gewählt (Tabelle 23), werden 1.161 Patienten mit Polypharmazie und einem vorhergesagten Krankenhausaufenthalt identifiziert. Von diesen sind tatsächlich 71,6% ins Krankenhaus eingewiesen worden. Es ergibt sich im Vergleich eine höhere PPV bei einer ausreichend großen, verbleibenden Interventionsgruppe. Die Sensitivität der Vorhersage sinkt, vor dem Hintergrund der Zielsetzung ist jedoch eine möglichst hohe PPV entscheidender.

Zum Vergleich kann ohne Anwendung eines Vorhersagemodells allein aufgrund der Anzahl an verschiedenen Arzneistoffen pro Patient eine Patientenauswahl erfolgen. Nachfolgend ist die entsprechende Klassifikationsmatrix dargestellt für das Aufgreifen von Patienten ab 13 oder mehr Arzneistoffen (Tabelle 24).

Schwellenwert für Patientenauswahl	Anzahl Patienten Zielmerkmal=0	Anzahl Patienten Zielmerkmal=1	Gesamt	Bewertung
ATC/Quartal <13	27.444	17.806	45.420	PPV: 59,9%
ATC/Quartal ≥13	196	293	489	
Gesamt	27.640	18.099	45.739	
Bewertung	Spezifität: 99,3%	Sensitivität: 1,6%		

Spezifität=Richtig-Negative; Sensitivität=Richtig-Positive; PPV=Richtig-Positive/(Richtig-Positive+ Falsch-Negative)

**Tabelle 24: Klassifikationsmatrix für den Schwellenwert <sub>≥13</sub> ATC/Quartal in 2007 und Ermittlung des Zielmerkmals in 2010 (Datenbasis Validierungsdatensatz: 2007-2010)**

Für eine solche Auswahl der Patienten ergibt sich mit einem PPV von 59,9% ein ähnlich niedriger Wert wie bei Anwendung des Modells mit einem Schwellenwert von  $PRE \geq 0,50$  mit 58,8%. Gleichzeitig wird jedoch eine sehr viel kleinere Interventionsgruppe erhalten, was nachteilig ist. Die Sensitivität sinkt auf den im Vergleich kleinsten Wert.

Angesichts der jeweiligen Klassifikationsmatrix und der jeweiligen Güteparameter für die Vorhersagen lassen sich Vorteile des Modells erkennen. Im Gegensatz zu einer Patientenauswahl anhand der Anzahl an Arzneistoffen allein, zeigen sich für die Modellvorhersage bessere Güteparameter. Voraussetzung dafür ist, dass ein Schwellenwert von über 0,5 gewählt wird. Die verbesserte Erkennung von Patienten mit Polypharmazie und einem potenziellen Krankenhausaufenthalt beziehungsweise einem Bedarf an klinisch-pharmazeutischer Betreuung erscheint damit durch das entwickelte Modell gelungen.

## 8 Schlussfolgerung (Deployment)



### 8.1 Diskussion

Arzneimittel sind wirksam, effizient, aber auch potenziell gesundheitsgefährdend und können zu Krankenhauseinweisungen führen. Das Bewusstsein für Arzneimittelrisiken ist insbesondere bei den besonders gefährdeten älteren Patienten mit Polypharmazie gering und eine bessere Aufklärung zu Arzneimittelrisiken dringend notwendig. Diese ist in erster Linie die Aufgabe der behandelnden Ärzte. Klinische Pharmazeuten können diese ergänzen und dadurch sowohl zur Verbesserung der Arzneimitteladhärenz von Patienten als auch zur Reduktion von arzneimittelbedingte Krankenhauseinweisungen beitragen. Die Datengrundlage aus deutschen Arztpraxen ist „dünn“ (SVR, 2009). Es wurde deswegen in der vorliegenden Arbeit der Versuch unternommen, verfügbare GKV-Routinedaten zu nutzen, um Patienten mit Fokus auf einen klinisch-pharmazeutischen Betreuungsbedarf zu erkennen und poten-



zielle ABH zu verhindern. Eine bessere und patientenorientiertere Versorgung gerade älterer Patienten mit Polypharmazie und höherem Risiko für UAE wird immer wieder gefordert (Glaeske, 2011).

### **Vorgehensweise nach dem CRISP kann Massedaten erfolgreich eingrenzen**

Zur Strukturierung der Data-Mining Analyse wurden die Phasen des CRISP-DM 1.0 zugrunde gelegt. Dieses erleichterte die systematische Bearbeitung und Orientierung während der komplexen Vor- und Hauptanalysen. Der in anderen Anwendungsbereichen etablierte CRISP-DM 1.0 hat sich damit bei der behandelten Fragestellung als sehr hilfreich erwiesen.

Bei den fachlichen Vorüberlegungen wurden Fakten aus der Literatur zusammengeführt und bezüglich der Nützlichkeit für das Analyseziel hin diskutiert (Business Understanding). Im Zusammenspiel mit den **Voranalysen** (Data Understanding) erfolgte die schrittweise Eingrenzung der Zielgruppe und des Zielmerkmals, was den Umfang der Daten für den Entwicklungsdatensatz bei der Datenzusammenstellung deutlich reduziert hat. Gleichzeitig konnte frühzeitig abgeschätzt werden, ob ausreichende Patientenzahlen für die zur Wahl stehenden Variablen erreicht werden, so dass ein Modell mit ausreichender Güte entstehen konnte. Nachteilig waren die eingeschränkten Vergleichsmöglichkeiten zwischen den Analysen zu Beginn und am Ende der Untersuchung, da durch die anfangs noch unklaren Ein- und Ausschlusskriterien unterschiedliche Daten zugrunde gelegt wurden. Inhaltlich konnten in den Voranalysen die gemäß Literatur erwarteten Charakteristika der Zielgruppe (Kap. 2) in den verwendeten Datensätzen erfolgreich reproduziert werden.

Die Anzahl der für die **Modellierung** eingesetzten Variablen ist mit etwa *78 Merkmalen* hoch. Insgesamt sind in jeder Phase der Analyse die Auswertungen dadurch komplex und anfällig für Fehler. Die Nutzung verschiedener Datenquellen aus dem Data Warehouse des Datenlieferanten machen die Datenzusammenstellung (Data Preparation) zudem aufwendig. Die nach einer Vorauswahl für die Modellierung verbleibenden zirka *50 Variablen* erlaubten viele Variationsmöglichkeiten, führten jedoch auch zu Störeffekten und erschwerten die Interpretation der Variableneffekte und des Gesamtmodells. Die Anzahl der Variablen im finalen Modell bleibt mit *24 Prädiktoren* hoch, konnte jedoch im Vergleich zur Ausgangssituation deutlich reduziert werden (Anhang 10.7).

Die Maßnahmen zur **Optimierung der Variablen** zeigten positive Effekte auf die Modellgüte und waren somit erfolgreich. Zudem konnten Variablen mit einer besseren fachlichen Aussagekraft in das Modell eingeführt werden. Die durchschnittliche

Anzahl an DDD pro ATC ist beispielsweise ein genaueres Maß für eine angemessene Dosis und erlaubt eine bessere Abschätzung eines Arzneimittelrisikos als die Summe an DDD pro Jahr insgesamt. Zu testen bleibt, ob sich weitere Verbesserungen des Modells durch andere Variationen der DDD-Variablen, wie beispielsweise die auf Tageshöchstdosen berechnete DDD pro ATC (Glaeske et al., 2009), erzielen lassen würden. Bei der Vielzahl an binären Merkmalen, die in die Modellierung eingegangen sind, wäre zu erwarten, dass metrische Variablen mit einer differenzierten Ausprägung an Werten positive Effekte auf die Modellgüte haben. Bei der Optimierung der Variablen setzen sich jedoch auch klassierte Variablen durch, beispielsweise bei der Altersvariablen (`alter_60708085`). Die Einzelbetrachtungen von Merkmalen zur Optimierung hinsichtlich des Zielmerkmals und die dafür separat berechneten Korrelationskoeffizienten erscheinen daher sinnvoll und nützlich als Entscheidungshilfe bei der Auswahl von Variablen.

### **Merkmale im Modell sind fachlich interpretierbar und haben statistische Aussagekraft**

Die durchgeführte Data-Mining Analyse lässt, so wie angestrebt, Einflussfaktoren für die Hospitalisierung von Patienten mit Polypharmazie erkennen. Der Anspruch, einen kausalen Zusammenhang nachzuweisen, besteht hier nicht. Die fachliche Interpretierbarkeit untermauert jedoch die statistische Aussagekraft. Erwartungsgemäß spielen von den **Patienten-Merkmalen** das Alter und, damit zusammenhängend, die Gruppe der Rentner eine Rolle im Modell. Dieses entspricht den Erwartungen aus der Literatur (Knopf & Melchert, 2003; Keller, 2006; Mertens, 2009) und den Ergebnissen aus den Voranalysen. Unerwartet war, dass der Anteil an weiblichen Patienten in der untersuchten Population mit und ohne Krankenhausaufenthalt zwar höher war, für Männer jedoch eine positive Korrelation und höhere durchschnittliche Vorhersagewahrscheinlichkeiten in Bezug auf Krankenseinweisungen ermittelt wurde.

Positiv ist, dass zwei Drittel der ausgewählten und als **ATC-Merkmale** definierten Arzneistoffe, in das finale Modell aufgenommen wurden. Damit wird die fachliche Auswahl der relevanten ATC, unterstützt durch die Voranalysen, als bestätigt betrachtet. So sind *Phenprocoumon*, *Clopidogrel* und *Diuretika* als relevante Risikoarzneistoffe im Zusammenhang mit ABH in das Modell eingegangen. Ebenso erscheinen *Spirolacton*, *Amiodaron* und *Opioide* im Modell, was angesichts ihres Risikopotenzials plausibel erscheint, die allerdings in der betrachteten Literatur weniger prominent sind. *Insulin* dagegen fehlt entgegen den Erwartungen aus der Literatur, stellte sich jedoch auch bereits in den Voranalysen als unauffällig dar und ist daher vielleicht

eine Charakteristik des verwendeten Patientenpools. Der umgekehrte Fall gilt für *Metamizol*, das in den Voranalysen und auch im Modell überraschenderweise erscheint und eine neuere, zunehmende, aber bisher nur wenig beschriebene Problematik bestätigen könnte (AkdÄ, 2011). Entwicklungen wie die PRISCUS-Liste (Holt et al., 2010), die die Anwendung von Metamizol empfiehlt, und günstige Generika können hier zukünftig noch zu einer Verschärfung beitragen. Des Weiteren erscheinen *PPI* im Modell. Vermutlich jedoch nicht wegen ihres Risikopotenzials, sondern wegen ihres sehr häufigen Vorkommens im Datensatz, was andere Effekte zu überlagern scheint. Zwei ATC-Merkmale, Metformin und DDD pro ATC, zeigen negative Regressionskoeffizienten. Für beide Merkmale haben sich die negativen Korrelationen ebenfalls bereits in den Voranalysen abgezeichnet und werden daher als plausibel bewertet. Als fachliche Interpretation wäre denkbar, dass das zum einen betroffene *Metformin* mit vergleichsweise geringerer Morbidität einhergeht oder sicherer ist als andere Therapiealternativen. So werden Insulin und andere orale Antidiabetika wegen ausgelöster Hypoglykämien häufig mit ABH in Zusammenhang gebracht (Anhang 10.2). Zum anderen ist die Variable DDD pro ATC betroffen. Die überwiegend niedrigen *DDD pro ATC* (DDDproATC\_12345) in Verbindung mit Krankenhauseinweisungen könnten ein Ausdruck von Unterversorgung mit Therapieversagen sein, beispielsweise ausgelöst durch schlechte Compliance oder Unterdosierung. Eine gleichzeitige Polypharmazie widerspricht diesem Effekt nicht. *Molsidomin* erscheint unerwartet im Modell und ist schlecht erklärbar. Insgesamt bleibt zu bedenken, dass Diagnose-Daten von der Modellierung ausgeschlossen wurden, um ATC-Merkmale im Modell zu erhalten und nicht zu verdrängen. Durch die Vielzahl der eingegangenen ATC-Merkmale erhält das Modell zur Vorhersage von Krankenhauseinweisungen dadurch einen starken Bezug zu Arzneistoffen. Die für die zentrale Fragestellung wichtigen Korrelationen zwischen einzelnen *ATC und Krankenhauseinweisungen* erscheinen mit niedrigen Korrelationskoeffizienten insgesamt jedoch schwach ausgeprägt. Das Optimieren der ATC-Variablen, beispielsweise durch Kombinieren mit Diagnosen, lässt daher keine umfangreichen Modellverbesserungen erwarten. Das zeigt sich entsprechend im weiteren Verlauf der Modellierung. Ursache für die schwachen Korrelationen könnte sein, dass das Modell bereits innerhalb einer Risikogruppe von Patienten mit Polypharmazie angewendet wird.

Die fachlich als relevant betrachteten **ATC-Kombinationen** wurden überwiegend ausgeschlossen und konnten das Modell nur geringfügig verbessern. Lediglich 3 von 11 kombinierten ATC-Merkmalen finden sich im finalen Modell. Die Kombination von *Allopurinol und Ramipril* zeigt einen niedrigen Korrelationskoeffizienten nach Kendalls (0,039), weswegen der Einschluss ins Modell nicht erwartet wurde. Erwartungsgemäß riskantere Kombinationen, wie *Spironolacton und Ramipril*, werden dagegen trotz höherer Korrelationskoeffizienten (0,065) durch den Modellierungsalgo-

rhythmus ausgeschlossen. Für diese Kombinationen sind jedoch potenziell tödliche Hyperkaliämien beschrieben. Die Variable *ATC\_Risiko*, die für mindestens einen Risikoarzneistoff steht, findet allerdings mit dem höchsten Korrelationskoeffizienten von 0,070 Eingang in das finale Modell. Damit lässt sich die erwartete fachliche Relevanz der ATC-Kombinationen im Modell insgesamt nicht zeigen. Die Untersuchung von Störeffekten ist wegen der Vielzahl möglicher direkter und indirekter Störungen zwischen den ATC-Kombinationen und den übrigen Merkmalen erschwert.

Die fachliche **Interpretierbarkeit des Gesamtmodells** wird durch die Autorin als gut bewertet. Die eingegangenen Variablen entsprechen weitestgehend den Erwartungen aufgrund der Hinweise aus der Literatur und den Ergebnissen der Voranalysen. Fachliche Erklärungen für 22 von 24 Variablen erscheinen vorhanden. Weiterführende Analysen wären nötig, um die korrekte Interpretation aller Effekte zu bestätigen.

Insgesamt ist die **Modellgüte** mit 64,7% befriedigend und scheint mit den vorhandenen Variablen, die die GKV-Routinedaten bieten, nicht weiter gesteigert werden zu können. Dafür spricht, dass die Variationen der Variablen-Sets bei der Modellierung die Güte nur in sehr kleinem Ausmaß verändern. Störeffekte zwischen Variablen konnten, wo diese vermutet und überprüft wurden, aufgelöst werden. Bei guter fachlicher Interpretierbarkeit der Variablen im Modell haben alle Variablen zugleich einen statistisch signifikanten Einfluss auf das Zielmerkmal. Damit bieten diese Ansatzpunkte für eine klinisch-pharmazeutische Beratung zur Vermeidung von ABH. Das entstandene Modell bietet für die gewünschte Fragestellung eine praktikable Lösung und ermöglicht die Identifikation einer ausreichend großen Interventionsgruppe mit einer hohen statistischen Wahrscheinlichkeit für eine Hospitalisation, die potenziell arzneimittelbedingt sein kann. Nichtsdestotrotz ist das Zielmerkmal, Krankenhaus-einweisung innerhalb von einem Jahr ab Auswertungstichtag, grob und eine genauere Eingrenzung auf ABH kaum möglich.

Das statistische Modell verbessert trotz vorhandener Limitationen die Auswahl von Patienten mit Polypharmazie und einem klinisch-pharmazeutischen Betreuungsbedarf im Vergleich zu einer Auswahl nach Polypharmazie allein. Dieses steht im Einklang mit Ergebnissen aus Primärdatenanalysen, die eine bestimmte Anzahl an Arzneistoffen allein als Grenzwert und Indikator für arzneimittelbedingte Probleme für wenig nützlich erachten (Viktil et al., 2006). Das entwickelte Modell bietet einen darstellbaren Zusatznutzen.

### **Geläufige Polypharmazie-Definition bestätigt, aber nur begrenzt nützlich**

*Polypharmazie, im Sinne von  $\geq 5$  ATC pro Quartal über ein Jahr, steht im Zusammenhang mit (arzneimittelbedingten) Krankenhauseinweisungen.*

Ein besonderes Problem in der Gesundheitsversorgung älterer Patienten stellt Polypharmazie dar, für die Art, Menge und Zeitraum der verschiedenen, gleichzeitig eingenommenen Arzneistoffe unklar definiert ist (Glaeske, 2011). Die Data-Mining Analyse erfordert eine konkrete *Polypharmazie-Definition*, die nach den Voranalysen mit „mindestens 5 verschiedenen Arzneistoffen pro Quartal über 4 Quartale“ formuliert wurde. Diese Definition beschreibt eine *kontinuierliche Polypharmazie*. Auch wenn Polypharmazie-Definitionen in der Literatur im Detail variieren und nicht direkt vergleichbar sind, lässt sich feststellen, dass überwiegend der Schwellenwert bei 5 Arzneistoffen gewählt wird. Dieser kann nach den Voranalysen als sinnvoll bestätigt werden. Die Analysen zeigen, dass sich der Anteil an Patienten mit Krankenhauseinweisungen ab 5 oder mehr Arzneistoffen etwa verdoppelt und Krankenhauseinweisungen insgesamt signifikant häufiger sind als bei Patienten mit  $< 5$  Arzneistoffen (Kapitel 4.1 und 4.5). Polypharmazie, gemäß der verwendeten Definition, ist somit für sich allein bereits für die Erkennung der Zielgruppe ein nützliches, charakteristisches Patientenmerkmal. Dementsprechend wird es als Einschlusskriterium gewählt. Allerdings ist der Nutzen dieser Vorgehensweise begrenzt, da die betroffene Patientengruppe groß und eine weitere Eingrenzung der Patienten mit Polypharmazie wünschenswert ist. Die Ergebnisse der Analysen lassen allerdings auch andere mögliche Schwellenwerte erkennen. Bei 13 oder mehr Arzneistoffen beispielsweise steigt der Anteil an Patienten mit Krankenhauseinweisungen noch einmal deutlich (Abbildung 8) und internistische Krankenhauseinweisungen, die eher mit arzneimittelbedingten Hospitalisationen assoziiert werden, überwiegen (Abbildung 9). Die mit diesem Schwellenwert identifizierbare kleinere Interventionsgruppe ist jedoch nicht geeignet für eine anschließende Data-Mining Analyse, ist möglicherweise aber unter anderen Rahmenbedingungen interessant.

Mit dem signifikanten Zusammenhang zwischen der definierte *kontinuierlichen Polypharmazie* mit  $\geq 5$  Arzneistoffen und Krankenhauseinweisungen im Folgejahr ist die Verwendung von Krankenhauseinweisungen als Zielmerkmal für das Vorhersagemodell möglich. Krankenhauseinweisungen werden als Proxy für den Bedarf einer klinisch-pharmazeutischen Betreuung gewählt, da sich eine suboptimale oder risiko-

reiche Arzneimitteltherapie in gesteigerten Hospitalisationen ausdrücken sollte. Die Wahl des Zielmerkmals erfolgt deshalb, da besser geeignete Indikatoren für ABH, wie internistische Krankenhauseinweisungen und Notfalleinweisungen (Schneeweiss et al., 2002), sich aufgrund der geringen Häufigkeiten im Datensatz nicht für eine Data-Mining Analyse eignen. Die Anwendung komplexer Algorithmen zur Erkennung von arzneimittelbedingten Problemen (Hanlon, 1992) wird wegen zu vieler fehlender Informationen in der Datenbasis verworfen. Ein Zusammenhang zwischen Polypharmazie, im Sinne der entwickelten Definition, und Krankenhauseinweisungen generell kann zwar mit der vorliegenden Untersuchung bestätigt, jedoch mangels spezifischerer Daten nicht explizit für ABH demonstriert werden.

Zu der Modellbildung schließlich leistet die aufbauend auf die entwickelte Polypharmazie-Definition gebildete Variable `min_ATC_Q14` keinen Beitrag. Das **Polypharmazie-Merkmal** `min_ATC_Q14` stellt sich nicht als ein Prädiktor mit positivem Einfluss auf die Modellgüte dar. Für die Variable lässt sich anhand der ermittelten Korrelationskoeffizienten zwar ein signifikanter Zusammenhang mit Krankenhauseinweisungen innerhalb des Folgejahres erkennen, insgesamt ist diese Korrelation aber sehr gering ausgeprägt (Kapitel 6.4.). Bei der rechnergestützten Modellierung wird diese Variable wegen geringer Effekte verworfen. Auch die Transformation in die klassierte Variable `neu_ATCGr_5912` führte nicht zu einer Verbesserung. Ein Grund für den geringen Nutzen dieser Polypharmazie-Merkmale als Prädiktoren ist möglicherweise eine zu starke inhaltliche Überschneidung mit den Einschlusskriterien beziehungsweise zu der gleichzeitig im Modell verwendeten Variablen „Summe der ATC pro Jahr“ (`Anz_ATC_1J`).

### **Summe der Arzneistoffe pro Jahr ist ein guter Prädiktor**

*Mit zunehmender Anzahl an Arzneistoffen pro Jahr nimmt die Wahrscheinlichkeit für Krankenhauseinweisungen zu und eine Vorhersage ist möglich.*

Bei der Untersuchung galt es, die Kernhypothesen zu bestätigen, dass bei zunehmender Polypharmazie auch die Wahrscheinlichkeit von (arzneimittelbedingten) Krankenhauseinweisungen zunimmt und sich diese vorhersagen lassen (Kapitel 6.1). Wie zuvor ausgeführt, waren die dazu gebildeten Polypharmazie-Variablen wegen geringen Nutzens nicht als Prädiktoren im das Vorhersagemodell aufgenommen worden. Die pauschal über ein Jahr aufsummierte Anzahl verschiedener ATC (`Anz_ATC_1J`) erweist sich dagegen, dargestellt durch den ermittelten Korrelations-

koeffizient nach Pearson, als ein besserer Prädiktor (Tabelle 15). Diese Variable wird als einziges verbleibendes Polypharmazie-Merkmal durch den Modellierungsalgorithmus eingeschlossen (Tabelle 21). Die Anzahl verschiedener ATC wird mit der größten Chi-Quadrat-Veränderung als wichtigster Prädiktor in der logistischen Regressionsanalyse ausgewiesen (Anhang 10.10). Damit kann lediglich gezeigt werden, dass der Grad der Polypharmazie nur im weiteren Sinne, nämlich pauschal ausgedrückt durch eine hohen Anzahl an Arzneistoffen pro Jahr, für die Vorhersagewahrscheinlichkeit einer Krankenhauseinweisung eine Rolle spielt. Es ist anzunehmen, dass hier wahrscheinlich die gleichzeitig zunehmende Morbidität hineinspielt.

Die Kernhypothese, dass mit jedem weiteren Arzneistoff auch die Wahrscheinlichkeit für eine Krankenhauseinweisung steigt, wird dennoch als bestätigt angesehen. Dieses steht im Einklang mit Hinweisen aus der Literatur, nach der jeder zusätzliche Arzneistoff das Risiko für eine UAE um 14% steigert (Varallo et al., 2014). Damit werden auch ABH wahrscheinlicher. Unter Umständen ist jedoch ein entscheidenderer Aspekt als Polypharmazie im Sinne der entwickelten Definition (*min\_ATC\_Q14*) eine hohe Arzneimittelvielfalt (*Anz\_ATC\_1J*) oder der Anteil an möglicherweise unangebrachten Arzneistoffen (PIM) in der Medikation (Tamura et al., 2012).

### **Data-Mining Analyse identifiziert Risikopatienten, weniger Risikoarzneistoffe**

*Patienten mit Polypharmazie die innerhalb von einem Jahr (arzneimittelbedingt) in ein Krankenhaus eingewiesen werden, lassen sich aus GKV-Routinedaten vorhersagen.*

Die GKV-Routinedaten weisen als Sekundärdaten Limitationen auf und die Wahl der logistischen Regressionsanalyse als Methode liefert lediglich den Nachweis für statistische, jedoch nicht kausale Zusammenhänge. Dennoch kann gezeigt werden, dass mit der Untersuchungsmethode fachlich nutzbare Ergebnisse erbracht werden können. Eine wichtige Voraussetzung dafür ist die korrekte fachliche Auswahl von prädiktiven Variablen aus den GKV-Leistungsdaten. Neben der richtigen fachlichen Datenselektion sind zudem spezielle Kenntnisse für die korrekte Interpretation der GKV-Leistungsdaten gefordert (Hoffmann, 2008) und werden genutzt für eine bevorzugte Auswahl von Daten mit besserer Datenqualität. Während der schrittweisen Analysen wurde zusätzlich wiederholt die Plausibilität der Einzelergebnisse untereinander überprüft, was im Sinne einer empfohlenen internen Datenvalidierung ist (Hoffmann et al., 2008). So konnte gezeigt werden, dass beispielsweise die ermittelten häufig angewendeten Arzneistoffe und dokumentierten Diagnosen bei Patienten

mit Polypharmazie fachliche Ähnlichkeiten aufweisen und unter Umständen identische Aspekte aus verschiedenen Perspektiven zeigen. Infolge der Bemühungen, die Limitationen von GKV-Daten auszugleichen, kann ein interpretierbares finales Modell präsentiert werden. Das finale Modell enthält Prädiktoren, die zum größten Teil den Erwartungen entsprechen. So werden der Grad der Polypharmazie und das Patientenalter als wichtige Merkmale ausgewiesen, und unter den weiteren Prädiktoren finden sich viele bekannte Risikoarzneistoffe, wie Phenprocoumon, Spironolacton und Diuretika (Keller, 2006). Für weniger gut erklärbare Prädiktoren im Zusammenhang mit Krankenhauseinweisungen, wie PPI, können andere, fachlich aber durchaus plausible Erklärungen gefunden werden. Das Modell wird daher als gut erklärbar bewertet.

Neben der inhaltlichen kann die statistische Güte der Vorhersage gezeigt werden. Bei der Untersuchung erfolgt neben der Entwicklung auch die Validierung des entwickelten Vorhersagemodells. Anhand der überprüften Gütemaße, insbesondere des AUC-Wertes und des PPV lässt sich darstellen, dass das entwickelte Modell insgesamt eine ausreichende und zuverlässige Vorhersagequalität aufweist und, bei geeigneter Wahl eines Schwellenwerts, die gemachten Vorhersagen zu einem hohen Prozentsatz zutreffen werden.

Damit ist gezeigt, dass GKV-Routinedaten trotz ihrer Limitationen eine ausreichende Datenbasis für Data-Mining bieten, ausreichend zuverlässige Vorhersagen zur gesuchten Zielgruppe möglich sind und fachlich verwendbare Ergebnisse erzielt werden können. Die Überprüfung des Modells an einem aktuelleren Validierungsdatensatz konnte außerdem zeigen, dass die Modellgüte des entwickelten Modells über zwei Jahre stabil geblieben ist. Gleichzeitig wird deutlich, dass das Modell bei der routinemäßigen Anwendung an GKV-Daten in bestimmten Zeitabständen erneut zu überprüfen ist, um von einer dauerhaft ausreichenden Modellgüte ausgehen zu können.



## Data-Mining Analyse liefert Hinweise für Handlungsempfehlungen

*Krankenhauseinweisungen bei Patienten mit Polypharmazie stehen im Zusammenhang mit bestimmten Patientenmerkmalen und Arzneistoffen. Daraus lassen sich ein Bedarf und konkrete Handlungsempfehlungen für die klinisch-pharmazeutische Betreuung ableiten.*

Die Ergebnisse der Untersuchung sollen im Sinne von Public Health unter anderem beispielsweise Gesundheit fördern, die Erkrankungswahrscheinlichkeit reduzieren und Ressourcen im Gesundheitswesen sinnvoll nutzen helfen. Diesen Aspekten wird mit dem entwickelten Vorhersagemodell Rechnung getragen. Die gezielte Erkennung von Patienten mit Polypharmazie und einem absehbaren, potenziell arzneimittelbedingten Krankenhausaufenthalt macht das effiziente und bedarfsgerechte Angebot einer klinisch-pharmazeutische Betreuung möglich. Dieses ist von praktischen Nutzen, da bekannt ist und bestätigt werden konnte, dass die Anzahl an Arzneistoffen allein keinen ausreichenden Indikator für UAW und damit einen Betreuungsbedarf darstellen (Viktil et al., 2006). Zudem wird deutlich, dass die Gruppe der Polypharmaziepatienten insgesamt mit zirka 25% bereits groß ist, tendenziell noch größer wird und zu etwa 80-90% zwischen 5-8 Arzneistoffe gebraucht werden. Aus dieser Gruppe heraus gelingt dem Modell eine weitere Differenzierung der Patienten. Damit können personelle und monetäre Ressourcen, wo GKV-Routinedaten verfügbar sind, gezielt für eine intensivere Patientenbetreuung zu Arzneimittelrisiken und für die Vermeidung von potenziell arzneimittelbedingten Hospitalisationen eingesetzt werden.

Für die Beratungsinhalte geben die ermittelten Prädiktoren im Vorhersagemodell Anhaltspunkte. Der gezeigte Zusammenhang zwischen der zunehmenden Anzahl an Arzneistoffen pro Jahr und Krankenhausaufenthalten macht deutlich, dass der Fokus der Beratung auf der Reduktion der Anzahl an Arzneistoffen liegen sollte. Mit „*A practical guide to stopping medicines in older people*“ (BPAC, 2010) gehen Bestrebungen bereits in diese Richtung. Dabei ist jeder Arzneistoff, auf den bei gleichbleibend hoher Therapiequalität verzichtet werden kann, ein Erfolg im Sinne der erhöhten Arzneimittelsicherheit. Dieses konnte hier statistisch durch die hohe Relevanz des entsprechenden Prädiktors (*Anz\_ATC\_1J*) bestätigt werden.

Die prädiktiven Patientenmerkmale wie Alter und männliches Geschlecht helfen ohne die Möglichkeit das Modell anzuwenden zwar weiter, aber nur begrenzt. Denkbar

wäre allerdings eine intensivere Betreuung von männlichen Patienten, bei denen leicht höhere durchschnittliche Vorhersagewahrscheinlichkeiten gezeigt werden konnten. Dieses gilt insbesondere vor dem Hintergrund, dass für Männer beispielsweise auch eine seltenere Kontaktaufnahme zu Ärzten beobachtet werden kann. Inhaltlich bieten die in der Literatur beschriebenen und durch die Prädiktoren im Modell bestätigten Risikoarzneistoffe einen Ansatzpunkt für die Beratung. Bei der Beratung sollte auf die arzneimittelspezifischen Sicherheitsaspekte von beispielsweise *Blutgerinnungshemmern* und *Diuretika* eingegangen werden. Unter Umständen sollte auch die häufigere Verwendung von *Metformin* und die Durchführung von *Impfungen* angesichts der negativen Korrelation mit einem absehbaren Krankenhausaufenthalt stärker gefördert werden. Dieses entspräche nicht nur gängigen Arzneimittelvereinbarungen und hätte einen positiven, wirtschaftlichen Effekt, sondern wäre auch angesichts der zahlreichen, neuen oralen Antidiabetika mit umstrittenen Sicherheitsprofil ein Beitrag zur Erhöhung der Arzneimittelsicherheit.

### **GKV-Routinedaten unzureichend, um Potenzial von Data-Mining Analysen voll auszuschöpfen**

*Für die Erkennung eines Bedarfs an klinisch-pharmazeutischer Betreuung aus GKV-Routinedaten ist eine bessere Codierung von arzneimittelbedingten Problemen beziehungsweise Hospitalisationen notwendig.*

Für das mit Hilfe der logistischen Regressionsanalyse entwickelte Modell ist eine höhere Modellgüte bei weniger Prädiktoren anzustreben. Dazu sind optimaler Weise Merkmale mit starkem kausalem Zusammenhang zu arzneimittelbedingten Hospitalisationen (ABH) zu verwenden. Aus der Untersuchung ergibt sich, dass die Korrelationskoeffizienten der einzelnen Prädiktoren im entwickelten Modell durchweg sehr niedrig sind. Die Bildung eines Modells mit besserer Vorhersagequalität und nur wenigen, aber einflussreichen Merkmalen ist daher, zumindest aus den verwendeten GKV-Routinedaten, schwierig und unwahrscheinlich.

Zu erwarten war, wie auch im erstellten Modell erkennbar, ein komplexer Zusammenhang zwischen Arzneistoffen und ABH mit vielen unabhängigen Variablen. Statt einer systematischen Herangehensweise mit Gruppierung von ATC, wie hier durchgeführt, wäre die gezielte Auswahl und Testung einzelner ATC oder in zuvor ermittelten ATC-Clustern unter Umständen zielführender. Ebenso wäre die Anwendung weiterer Methoden des Data-Mining zu prüfen, wie die Entwicklung von Entscheidungsbäumen.

Eine andere Möglichkeit, um GKV-Routinedaten besser für die Beurteilung der Arzneimittelversorgung nutzen zu können, wäre, eine umfassendere Codierung von ABH zu fordern. Die Erfassung sollte häufiger, vollständiger und mit einer differenzierteren Systematik erfolgen. Gegebenenfalls sollte für UAEs dazu sogar ein separater Code entwickelt und getrennt von den ICDs dokumentiert werden. Untersuchungen mit internistischen Krankenhauseinweisungen oder Notfalleinweisungen sind lediglich eine Notlösung und bedürfen wegen der seltenen Codierung in GKV-Routinedaten geeigneter Untersuchungsmethoden oder ebenfalls einer besseren Qualität der Datenbasis.

Die Entwicklung eines Modells nach der gleichen Vorgehensweise, jedoch auf Grundlage von Primärdaten mit aussagekräftigeren klinischen Daten, die Zusammenhänge von Arzneistoffen mit UAW und ABH besser abbilden, wäre vielversprechender für die Erzielung einer besseren Modellgüte. Für die direkte Ermittlung eines Bedarfs an klinisch-pharmazeutischer Beratung wären neben objektiven Laborparametern, wie der Nierenfunktion, auch Informationen zum Gesundheitsverhalten der Patienten notwendig. Ein umfangreichere Datenerfassung in Praxen und ein besserer Zugang zu den fehlenden Informationen in den Routinedaten wären für die Versorgungsforschung notwendig.

### *Limitationen*

Trotz der nach statistischen Gesichtspunkten erfolgreichen Entwicklung und Anwendung des Modells zur Erkennung von Patienten mit einem klinisch-pharmazeutischen Betreuungsbedarf, haben die angewandte Analyseverfahren und der Nutzen des Modells Grenzen.

- Für die Data-Mining Analyse mussten die Begrifflichkeiten „klinisch-pharmazeutischer Betreuungsbedarf“ und „Polypharmazie“ exakter definiert werden. Beide Begriffe sind komplex und wegen fehlender benötigter Details in den GKV-Routinedaten nur grob vereinfacht und indirekt zu ermitteln. Die durch das Modell erkannte Patientenklientel kann daher auch nur mit einer gewissen Unschärfe der tatsächlichen Zielgruppe entsprechen.
- GKV-Routinedaten enthalten darüber hinaus nur wenige Informationen über UAW, die zudem nicht mit direkt zusammenhängenden Begleitumständen, wie verursachender Wirkstoff, Co-Medikation oder Gesundheitszustand des betroffenen Patienten verknüpft sind. Zudem wird bei UAW von einem „Underreporting“ ausgegangen. Anders als bei Primärdaten ist daher ein Arzneimittelbezug von Krankheitsereignissen und Leistungsansprüchen nur schwer herzustellen. Damit ist auch ein klinisch-pharmazeutischer Betreuungsbedarf trotz bester Analysemethoden nur schwer retrospektiv aus GKV-Daten zu erkennen.
- GKV-Arzneimitteldaten enthalten keine Informationen zur Selbstmedikation. Der Grad der Polypharmazie und damit die ermittelte Wahrscheinlichkeit für einen Krankenhausaufenthalt innerhalb eines Jahres wird bei der Bestimmung aus Routinedaten daher tendenziell eher noch unterschätzt.
- Das Modell wurde entwickelt für Patienten im Alter von 18-85 Jahren mit Polypharmazie in den letzten 12 Monaten. Die Verwendung für eine andere Patientenklientel, beispielsweise für Kinder oder Patienten mit einzelnen verordneten Risikoarzneistoffen, ist nicht möglich. Dafür wären eigene Modelle zu entwickeln.
- Zwar lassen sich aus den GKV-Routinedaten durch das Modell Patienten mit hohem Arzneimittelverbrauch ermitteln, für die eine Krankenseinweisung erwartet werden kann, unklar bleibt aber, ob ein kausaler Zusammenhang zwischen den verordneten Arzneistoffen und der zeitlich folgenden Krankenseinweisung besteht.
- Die Patienten in der Studienpopulation weisen eine Polypharmazie über ein Jahr auf, für die das entwickelte Modell eine Abschätzung über bevorstehende

Krankenhauseinweisungen im Folgejahr trifft. Das Modell ist nicht darauf ausgerichtet, frühzeitiger auftretende UAE, abzuschätzen.

- GKV-Routinedaten als Datenbasis erlauben keine sichere Abgrenzung von Krankenhauseinweisungen wegen höherer Morbidität von denen anderer Ursachen, wie Arzneimitteln. Dieses ist jedoch nicht nur ein Problem der Datenbasis, sondern auch eins des zunehmend komplexeren Krankheitsgeschehens (Multimorbidität) bei durchschnittlich immer älter werdenden Patienten.
- Bei einer Data-Mining Analyse werden statistische Zusammenhänge, keine kausalen ermittelt. Für die Analyse zugrunde zu legende Annahmen und die Interpretation der Analyseergebnisse bleiben damit Aufgabe des Forschers und können Fehler behaftet sein. So wurde vorausgesetzt, dass Polypharmazie an sich ein Risiko für ABH darstellt, was nicht zwangsläufig der Fall sein muss.
- Das entwickelte Modell ist für den verwendeten Datenpool entwickelt worden und nicht ohne weiteres auf beliebige andere Datensätze anwendbar. Selbst bei Anwendung auf GKV-Routinedaten des gleichen Datenlieferanten hat das Vorhersagemodell eine begrenzte „Haltbarkeit“. Die Ausprägung der im Modell verwendeten Merkmale verändert sich meist mit der Zeit im Datensatz und verschlechtert in der Regel die Modellgüte kontinuierlich. Nach etwa 1-3 Jahren wäre eine Neuentwicklung des Modells oder zumindest die Überprüfung der Modellgüte erforderlich.

Anfangs befürchtete Einschränkungen aufgrund zu geringer Patientenzahlen im Datensatz oder Subgruppen hat es nicht gegeben, überlagernde Effekte von Merkmalen haben die Modellentwicklung erschwert, aber nicht verhindert. Aufgrund der beschriebenen Limitationen bleibt festzulegen, wie das entwickelte Modell in der Praxis optimaler Weise eingesetzt werden sollte und worin noch weiterer Forschungsbedarf besteht.

## 8.2 *Fazit und Ausblick*

Die durchgeführte Untersuchung stellt einen Beitrag zu den ansonsten eher wenigen deutschen Studien zur Polypharmazie dar, reiht sich aber ein in die in diesem Themenfeld überwiegenden Studien auf Basis von Routinedaten. Die Ermittlung der Polypharmazie erfolgt, wie auch bei vielen anderen Studien, zwar kumulativ über jeweils 3 Monate, jedoch über einen zusammenhängenden Zeitraum von einem Jahr. Damit stehen Patienten mit einer *kontinuierlichen Polypharmazie* im Mittelpunkt der Untersuchung. Der verwendete Datenpool ist im Vergleich zu anderen Studien auf Basis von Routinedaten (Kapitel 1.1) mit rund 44.000 Patienten vergleichsweise groß.

Das gebildete Regressionsmodell kann Patienten mit Polypharmazie und einer erhöhten Wahrscheinlichkeit für eine Krankenhauseinweisung innerhalb eines Jahres erkennen. Dieses kann dazu beitragen, Patienten mit Polypharmazie zielgenauer und automatisiert zu identifizieren, die am ehesten von einer Beratung mit klinisch-pharmazeutischem Fokus zur Vermeidung von potenziell arzneimittelbedingten Hospitalisationen profitieren könnten. Die Vorgehensweise eignet sich dazu, Arzneistoffe als Prädiktoren zu ermitteln, untereinander zu gewichten und in Handlungsempfehlungen für beispielsweise Ärzte oder Apotheker einzubeziehen. Der Nachweis eines Kausalzusammenhangs zwischen Polypharmazie und einer ABH ist mit einer logistischen Regressionsanalyse nicht möglich und aufgrund der begrenzten Datenqualität der GKV-Routinedaten erschwert. Unklar bleibt also, ob die identifizierten Patienten lediglich morbider oder tatsächlich höheren Arzneimittelrisiken ausgesetzt sind und von einer klinisch-pharmazeutischen Betreuung profitieren würden. Die Anwendung des Modells sollte daher in der Praxis nur durch kundige Hände erfolgen, um die identifizierte Risikokonstellation bei den ermittelten Patienten fachlich differenzierter einschätzen zu können. Um Fehlinterpretationen vorzubeugen, sollte zudem ein erfahrener Biometriker Modellentwicklung und -pflege unterstützen.

### *Ausblick*

Eine umfangreichere und arzneimittelbezogene Codierung von ABH, sowie die Ergänzung von klinischen Daten, wie beispielsweise Nierenfunktion und Gesundheitsverhalten der Patienten, in GKV-Routinedaten wären notwendig, um ein prädiktives Modell mit besserer Güte entwickeln und einen Bedarf an klinisch-pharmazeutischer Betreuung direkter ermitteln zu können. UAE müssten häufiger erfasst werden und zusätzliche Daten, wie auslösende Arzneimittel und resultierende Leistungsanspruchnahme, erhoben werden. Die Zunahme der Selbstmedikation ist gerade bei

Älteren ein Sicherheitsrisiko (Albert et al., 2014) und die Einbeziehung von OTC in die systematische Datenerfassung zu fordern. Apotheker, als häufig die ersten Ansprechpartner für OTC, wären hierfür prädestiniert. Eine Verknüpfung von Primär- und Sekundärdaten zur Verbesserung der Aussagekraft von Massendaten ist dringend zu begrüßen, wie dieses beispielsweise in Großbritannien über den „*Clinical Practice Research Datalink*“ erfolgt. Welche zusätzlichen Daten für die bessere Beurteilung von UAE, ABH und Risikoarzneistoffen erfasst werden müssten, um aussagekräftigere Vorhersagemodelle erstellen zu können, bleibt zu untersuchen. Zusätzlich sollte auch die Auswertung und Interpretation von Gesundheitsdaten in die Ausbildung von Arzneimittelexperten, wie beispielsweise Apothekern, stärker integriert werden.

Trotz Vereinfachung der Zielvorgaben und der Eingrenzung der Daten für die Modellentwicklung durch die Voranalysen bleibt die durchgeführte Analyse komplex. Die hohe Komplexität entspricht jedoch der Versorgungsrealität. Ob die isolierte Untersuchung einzelner Arzneistoffe in einer enger definierten und homogeneren Patientengruppe bei gleicher Methode bessere Vorhersageergebnisse erzielen kann, wäre ein Feld für weitere Forschungen. Wegen des höheren Aufwandes und meist geringeren Patientenzahlen ergeben sich bei der verwendeten Methode jedoch vermutlich ebenfalls rasch Grenzen. Eine Verringerung der Komplexität in der Versorgungspraxis, beispielsweise durch Einschränkungen der Präparatevielfalt, wäre wünschenswert. Dieses würde nicht nur Datenanalysen erleichtern, sondern an sich bereits einen Beitrag zur Verbesserung der AMTS darstellen.

Polypharmazie allein ist nur begrenzt als prädiktives Merkmal für Krankenhauseinweisungen geeignet. Welche alternativen Merkmale in Kombination von Nutzen sind, gilt es weiter zu untersuchen. Denkbar wären die mit Hospitalisationen assoziierte Anzahl an Verordnungsänderungen (Sino et al., 2013), Adhärenz-Probleme in Form von Therapieunterbrechungen oder ein zu ermittelnder *Chronic Disease Score* (Clark et al., 1995, Putnam et al., 2002).

In der Untersuchung ergebende sich Wirkstoffe mit prädiktivem Wert für potenzielle ABH. Um das Risiko für (arzneimittelbedingte) Krankenhauseinweisungen zu senken, könnten die ermittelten Wirkstoffe quasi als „potenziell unangebrachte Medikation“ (PIM) bei Polypharmazie betrachtet werden. Ähnlich der PRISCUS-Liste könnte der Anteil dieser Wirkstoffe in der Verordnung als Qualitätsindikator genutzt werden. Limitationen bei der Aussagekraft sind dabei allerdings zu berücksichtigen.

Die Bestätigung zugrunde gelegter und neu vermuteter kausaler Zusammenhänge zwischen Arzneistoffen und UAE beziehungsweise ABH sind mit anderen Methoden als dem Data-Mining zu erbringen, beispielsweise mit Spontanmeldesystemen oder, wenn retrospektiv aus Massendaten, mit Faktoren- oder Clusteranalysen.

## 9 Literatur

ACCP - American College of Clinical Pharmacy (2008): The Definition of Clinical Pharmacy. *Pharmacotherapy*, 28(6): 816-17.

AGENS & Arbeitsgruppe Epidemiologische Methoden - Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten der DGSMP, Arbeitsgruppe Epidemiologische Methoden DGEpi, GMDS, DGSMP (2008): GPS - Gute Praxis Sekundärdatenanalyse: Revision nach grundlegender Überarbeitung. Version 2. Verfügbar unter: [http://www.gesundheitsforschung-bmbf.de/\\_media/GPS.pdf](http://www.gesundheitsforschung-bmbf.de/_media/GPS.pdf) (letzter Zugriff: 25.02.2013)

AkdÄ – Arzneimittelkommission der deutschen Ärzteschaft (2011): Aus der UAW-Datenbank. Agranulozytose nach Metamizol - sehr selten, aber häufiger als gedacht. *Dtsch Arztebl*, 108(33): 1758-59.

Albert SM, Bix L, Bridgeman MM, Carstensen LL, Dyer-Chamberlain M, Neafsey PJ, Wolf MS (2014): Promoting Safe and Effective Use of OTC Medications: CHPA-GSA National Summit. *Gerontologist*. 2014 May 20. pii: gnu034. [Epub ahead of print]

Anderson F, Garbe E (2008): Pharmakoepidemiologische Forschung mit Routinedaten des Gesundheitswesens. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 51(10): 1135-44.

Ashby C, Simms S (1998): Data Mining. Research Brief. In Verfügbar unter: <http://theweb.badm.sc.edu/798gstud/simsj/Datamininggrb.html> (letzter Zugriff 18.12.2013)

AVWG - Arzneimittelversorgungs-Wirtschaftlichkeitsgesetz (2006): Gesetz zur Verbesserung der Wirtschaftlichkeit in der Arzneimittelversorgung vom 26.04.2006. *BGBl. I*, 21: 984-987. Geltung ab 01.05.2006, abweichend siehe Artikel 3.

Backhaus K, Erichson B, Plinke W, Weiber R (2008): *Multivariate Analysemethoden – Eine Anwendungsorientierte Einführung*. 12. Auflage. Springer-Verlag: Berlin, Heidelberg.

Beers MH (1997): Explicit criteria for determining potentially inappropriate medication use by the elderly. *Arch Intern Med*, 157: 1531-6.

Beney J (2010): Clinical Pharmacy - a Definition. ESCP - European Society of Clinical Pharmacy vom 12.11.2010. Verfügbar unter: [http://www.escpweb.org/cms/Clinical\\_pharmacy](http://www.escpweb.org/cms/Clinical_pharmacy) (letzter Zugriff 31.5.2014)

Berry MJA, Linoff G (1997): *Data Mining Techniques for Marketing, Sales and Customer Support*. Wiley: New York.

Beyer M, Otterbach I, Erler A, Muth C, Gensichen J, Gerlach FM (2007): Multimorbidität in der Allgemeinpraxis - Teil I: Pragmatische Definition, Epidemiologie und Versorgungsprämissen. *Z Allg Med*, 83: 310-315.



BMFSFJ – Bundesministerium für Familie Senioren Frauen Jugend (2010): Sechster Bericht zur Lage der Älteren Generation in der Bundesrepublik Deutschland – Altersbilder in der Gesellschaft und Stellungnahme der Bundesregierung. Verfügbar unter: <http://dip21.bundestag.de/dip21/btd/17/038/1703815.pdf> (letzter Zugriff 25.05.2014)

BMG - Bundesministerium für Gesundheit (2007): Aktionsplan 2008/2009 zur Verbesserung Arzneimitteltherapiesicherheit (AMTS) in Deutschland. Verfügbar unter: <http://www.akdae.de/AMTS/Aktionsplan/Aktionsplan-AMTS-2008-2009.pdf> (letzter Zugriff 25.02.2013)

BMG - Bundesministerium für Gesundheit (2010): Aktionsplan 2010-12 zur Verbesserung der Arzneimitteltherapiesicherheit (AMTS) in Deutschland. Verfügbar unter: [http://www.bmg.bund.de/fileadmin/dateien/Publikationen/Gesundheit/Broschueren/Broschuere\\_Aktionsplan\\_2010\\_-\\_2012\\_zur\\_Verbesserung\\_der\\_Arzneimitteltherapiesicherheit\\_\\_in\\_Deutschland.pdf](http://www.bmg.bund.de/fileadmin/dateien/Publikationen/Gesundheit/Broschueren/Broschuere_Aktionsplan_2010_-_2012_zur_Verbesserung_der_Arzneimitteltherapiesicherheit__in_Deutschland.pdf) (letzter Zugriff 25.02.2013).

BMG - Bundesministerium für Gesundheit (2013): Aktionsplan 2013-2015 des Bundesministeriums für Gesundheit zur Verbesserung der Arzneimitteltherapiesicherheit in Deutschland. Verfügbar unter: <http://www.akdae.de/AMTS/Aktionsplan/Aktionsplan-AMTS-2008-2009.pdf> (letzter Zugriff 30.05.2014)

Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW (2005): Clinical Practice Guidelines and Quality of Care for Older Patients with Multiple Comorbid Diseases: Implications for Pay for Performance. *J Am Med Assoc*, 294(6): 716-724.

BPAC (2010): A practical guide to stopping medicines in older people. *Best Practice Journal*, 27: 10-23.

Brekke M, Hunskaar S, Straand J (2006): Self-reported drug utilization, health and lifestyle factors among 70-74 year old community dwelling individuals in Western Norway. The Hordaland Health Study (HUSK). *BMC Public Health* 6:121-128.

Brenghoj L, Thirstrup S, Kristensen MB, Sonne J (2005): Reliability of a modified medication appropriateness index in primary care. *Eur J Clin Pharmacol*, 61: 769-73.

Bühl A (2009): SPSS 18 - Einführung in die moderne Daten-analyse unter Windows. 12., aktualisierte Auflage. *Persons Studium*: München.

Chapmann P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (CRISP-DM Consortium) (2000): CRISP-DM 1.0 - Step-by-step data mining guide. USA: SPSS 19.0 Inc.: 1-73. Verfügbar unter: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> (letzter Zugriff am 29.05.2014)

Chou IC, Voit EO (2009): Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci*, 219(2): 57-83.

Clark DO1, Von Korff M, Saunders K, Baluch WM, Simon GE (1995): A chronic disease score with empirically derived weights. *Med Care*, 33(8):783-95.

Cousins D (2009): Current status of the monitoring of medication practice. *Am J Health Syst Pharm*, 66(3): 49-56.

Defort R (2011): Kostenreduktion und Qualitätssteigerung in der Pharmakotherapie. Präsentation zum Vortrag zur Veranstaltung "Ist weniger mehr? - Arzneimittelsicherheit bei älteren Menschen" von der Landesvereinigung für Gesundheit und Akademie für Sozialmedizin Niedersachsen e.V. am 19.02.2011 in Hannover. Verfügbar unter: [http://www.gesundheit-nds.de/CMS/images/stories/PDFs/11\\_02\\_19%20Arzneimittelsicherheit/2011-02-19%20Vortrag%20Defort.pdf](http://www.gesundheit-nds.de/CMS/images/stories/PDFs/11_02_19%20Arzneimittelsicherheit/2011-02-19%20Vortrag%20Defort.pdf) (letzter Zugriff 27.05.2014)

Delen D, Walker G, Kadam A (2005): Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med*, 34(2): 113-27.

DIMDI - Deutsches Institut für Medizinische Dokumentation und Information (Hrsg.) (2010): Amtliche Deutsche Fassung für 2010. Anatomisch-therapeutisch-chemische Klassifikation mit Tagesdosen - Amtliche Fassung des ATC-Index mit DDD-Angaben für Deutschland im Jahre 2010. Verfügbar unter: <http://www.dimdi.de/dynamic/de/klassi/downloadcenter/atcddd/vorgaenger/atc-ddd-amtlich-2010.pdf> (letzter Zugriff 31.05.2014)

DIMDI - Deutsches Institut für Medizinische Dokumentation und Information (Hrsg.) (2010): ICD-10-GM Version 2010 Systematisches Verzeichnis - Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, German Modification. Systematisches Verzeichnis. Stand 02. November 2009. *Bundesanzeiger*, 61(172): 3885.

Department of Health (2000): *An Organisation with a Memory*. Im Eigenverlag: London.

Ebbesen J, Buajordet I, Erikssen J, Brørs O, Hilberg T, Svaar H, Sandvik L (2002): Drug-related deaths in a department of internal medicine. *Arch Intern Med*, 161(19): 2317-23.

Edwards IR, Aronson JK (2000): Adverse drug reactions: definitions, diagnosis, and management. *Lancet*, 356:1255-1259.

Fachkommission Klinische Pharmazie der Bundesapothekerkammer (2005): Stoffkatalog zur Weiterbildung im Gebiet Klinische Pharmazie. Verfügbar unter: [http://www.abda.de/fileadmin/assets/Weiterbildung/Spezialisierungsrichtungen/Spezialisierungsgebiete/Klinische\\_Pharmazie/stoffkatalog\\_klinische\\_pharmazie\\_2005.pdf](http://www.abda.de/fileadmin/assets/Weiterbildung/Spezialisierungsrichtungen/Spezialisierungsgebiete/Klinische_Pharmazie/stoffkatalog_klinische_pharmazie_2005.pdf) (letzter Zugriff: 29.05.2014)

Ferber R (2003): *Information Retrieval - Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. dpunkt Verlag: Heidelberg.

Fick DM, Cooper JW, Wade WE, Waller JL, Maclean JR und Beers, M.H. (2003): Updating the Beers criteria for potentially inappropriate medication use in older adults. Results of a US consensus panel of experts. *Arch Intern Med*, 163: 2716-24.

FIP – Federation International Pharmaceutical, WSMI – World Self-Medication Industry (1999): Gemeinsame Erklärung seitens der Internationalen Pharmazeutischen Föderation und der World Self-Medication Industry: Verantwortungsvolle Selbstmedikation vom 10.06.1999. Verfügbar unter: [http://www.fip.org/www/uploads/database\\_file.php?id=243&table\\_id=#page=1&zoom=auto,-73,798](http://www.fip.org/www/uploads/database_file.php?id=243&table_id=#page=1&zoom=auto,-73,798) (letzter Zugriff 31.05.2014)

Flick U (Hrsg.)(2002): Innovation durch New Public Health. Hogrefe Verlag GmbH & Co. KG: Göttingen.

Flores LM, Mengue SS (2005): Drug use by the elderly in Southern Brazil. *Rev Saúde Pública*, 39 (6):1-6.

Foster C, Hillsdon M, Thorogood, M (2005): Interventions for promoting physical activity. *Cochrane Database Syst Rev*, Issue 1, Art. No.: CD003180.

Frölich JC (2000): Besonderheiten der Pharmakotherapie in der Geriatrie. In: Frölich JC, Kirch W (Hrsg.). *Praktische Arzneitherapie*, 2. Auflage. Springer: Berlin, Heidelberg, S. 80-94.

Fromm S (2005): Binäre logistische Regression - Eine Einführung für Sozialwissenschaftler mit SPSS 19.0 für Windows. *Bamberger Beiträge zur empirischen Sozialforschung*: 1-35.

Gallagher P, O' Mahony, D (2008): STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions): application to acutely ill elderly patients and comparison with Beers' criteria. *Age & Ageing*, 37(6): 673-679.

Garbade S (2009): Einführung in die logistische Regression. InterComponentWare AG. Verfügbar unter: [http://www.metabnet.de/fileVault/project\\_51/instrumente/Logistische\\_Regression.pdf](http://www.metabnet.de/fileVault/project_51/instrumente/Logistische_Regression.pdf). (letzter Zugriff: 06.09.2013).

Garfinkel D, Zur-Gil S, Ben-Israel J (2007): The war against polypharmacy: a new cost-effective geriatric-palliative approach for improving drug therapy in disabled elderly people. *Isr Med Assoc J*, 9(6): 430-4.

Gillespie U, Alassaad A, Henrohn D, Garmo H, Hammarlund-Udenaes M, Toss H, Kettis-Lindblad A, Melhus H, Mörlin C (2009): A comprehensive pharmacist intervention to reduce morbidity in patients 80 years or older: a randomized controlled trial. *Arch Intern Med*, 169: 894-900.

GKV-WSG (2007): Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz – GKV-WSG) vom 26.03.2007. *BGBl. I*, 11: 378-473. Geltung ab 01.04.2007, abweichend siehe Artikel 46.

Glaeske G, Janhsen K (2006): GEK-Arzneimittel-Report 2006. Auswertungsergebnisse der GEK-Arzneimitteldaten aus den Jahren 2004 bis 2005. GEK, 44, Asgard: St. Augustin.

Glaeske G., Schick Tanz, C., Janhsen, K (2008): GEK-Arzneimittelreport 2008. Asgard: St. Augustin.

Glaeske G, Hoffmann F, Windt R (2009): Arzneimittelbezogene Module für die Effektivitätsprüfung von Hausarztverträgen gem. § 73b SGB-V. Projektbericht. Zentrum für Sozialpolitik: Bremen.

Glaeske G, Rebscher H, Willich S (2010): Versorgungsforschung: Auf gesetzlicher Grundlage systematisch ausbauen. Dtsch Arztebl, 107(26): A1295–7.

Glaeske G (2011): Patientenorientierung in der medizinischen Versorgung - Vorschläge zur notwendigen Weiterentwicklung und Umgestaltung unseres Gesundheitswesens. Expertise im Auftrag der Abteilung Wirtschafts- und Sozialpolitik der Friedrich-Ebert-Stiftung: Bonn. Verfügbar unter: <http://library.fes.de/pdf-files/wiso/07806.pdf> (letzter Zugriff 27.05.2013)

GMG (2003): GKV-Modernisierungsgesetz (Gesetz zur Modernisierung der gesetzlichen Krankenversicherung) vom 19.11.2003. BGBl. I, 55: 2190-2258. Geltung ab 01.01.2004, abweichend siehe Artikel 37.

Graziano (2010): Development and Validation of a Score to Assess Risk of Adverse Drug Reactions Among In-Hospital Patients 65 Years or Older. Arch Intern Med, 170 (13): 1142-48.

Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, Cadoret C, Fish LS, Garber L, Kelleher M, Bates DW (2003): Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA, 289(9): 1107-16.

Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C (2012): Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. Clin Pharmacol Ther, 91(6): 1010-21.

Hanlon JT, Schmader KE, Samsa GP, Weinberger M, Uttech KM, Lewis IK, Cohen HJ, Feussner JR (1992): A Methode for assessing drug therapy appropriateness. J Clin Epidemiol, 45(10): 1045-51.

Hasford J, Schubert I, Garbe E, Dietlein G, Glaeske G (2004): Memorandum zu Nutzen und Notwendigkeit Pharmakoepidemiologischer Datenbanken in Deutschland. Asgard: St. Augustin.

Haynes RB, McDonald H, Garg AX, Montague P (2001): Interventions for helping patients to follow prescriptions for medications. Cochrane Database Syst Rev, Issue 2, Art. No.: CD000011.

Hepler CD, Strand LM (1990): Opportunities and responsibilities in pharmaceutical care. Am J Hosp Pharm, 47: 533-543.

Hoffmann F (2008): (Arzneimittel)Routinedaten als Datenbasis für die Versorgungsforschung und Pharmakoepidemiologie. Dissertation. Bremen.

Hoffmann F, Pfannkuche M, Glaeske G (2008): Validität forschungsrelevanter Informationen in Arzneimittelroutinedaten von 2000 bis 2006. Dtsch Med Wochenschr, 133(18): 943-49.

Hoffmann F, Pfannkuche M, Glaeske G (2007): Wie häufig sind Verordnungs- und Abgabedatum von Arzneimitteln in Routinedaten korrekt erfasst?. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitschutz, 51(10): 1418-23.

Holt S, Schmiedl S, Thürmann PA (2010): Potenziell inadäquate Medikation für ältere Menschen: Die PRISCUS-Liste. Dtsch Arztebl Int, 107(31-32): 543-51.

IBM Corp. (2010): IBM SPSS Statistics for Windows, Version 19.0. IBM Corp.: Armonk, NY.

Iniesta-Navalon C, Urbieto-Sanz E, Gascón-Cánovas JJ (2011): Analysis of the drug interactions associated to domiciliary drug therapy in elderly hospitalized patients. Rev Clin Esp, 211(7): 344-51.

IOM - Institute of Medicine: Kohn LT, Corrigan JM, Donaldson MS (1999): To err is human: Building a safer health system. Washington, D.C.: National Academy Press, 1999.

Jaehde U, Ammon HPT (1999): Klinische Pharmazie – eine neue Fachdisziplin an Pharmazeutischen Instituten aus: DGPT-Forum 25: 27-28.

Jaehde U (2009): Klinisch-pharmazeutische Ausbildung an der Universität Bonn. Krankenhauspharmazie, 30: 205-8.

Jaehde U, Thürmann PA (2012): Arzneimitteltherapiesicherheit in Alten- und Pflegeheimen. Z. Evid.Fortbild. Qual. Gesundh. wesen (ZEFQ), 106, 712-716.

Jorgensen F (2010): ESCP News - Clinical pharmacy - Pharmaceutical care - is there a difference?. ESCP News, 147: 1-2.

Keller E (2006): Unerwünschte Arzneimittelwirkungen: Ansätze zur Inzidenzschätzung und Signalgenerierung mit fallbasierten Pharmakovigilanzdaten. Dissertation. München.

Knopf H, Melchert HU (2003): Bundes-Gesundheitssurvey: Arzneimittelgebrauch Konsumverhalten in Deutschland. Beiträge zur Gesundheitsberichterstattung des Bundes. RKI: Berlin.

Koordinierungsgruppe AMTS - Koordinierungsgruppe zur Umsetzung und Fortschreibung des Aktionsplans zur Verbesserung der Arzneimitteltherapiesicherheit in Deutschland (2011): Memorandum zur Entwicklung der Forschung auf dem Gebiet der Arzneimitteltherapiesicherheit (AMTS). Verfügbar unter: <http://www.akdae.de/AMTS/Massnahmen/docs/Forschungsmemorandum.pdf> (letzter Zugriff: 30.05.2014)

Krska J, Cromarty JA, Arris F, Jamieson D, Hansford D, Duffus PRS, Downie G, Seymour DG (2001): pharmacist-led medication review in patients over 65: a randomized controlled trial in primary care. *Age Aging*, 30(3): 205-211.

Lazarou J, Pomeranz BH, Corey PN (1998): Incidence of adverse drug reactions in hospitalized patients: a metaanalysis of prospective studies *JAMA*, 279: 1200–5.

Linjakumpu T, Hartikainen S, Klaukka T, Veijola J, Kivelä SL, Isoaho R (2002): Use of medications and polypharmacy are increasing among the elderly. *J Clin Epidemiol*, 55(8): 809-817.

LGH - Leitliniengruppe Hessen Hausärztliche Pharmakotherapie, DEGAM - Deutsche Gesellschaft für Allgemein- und Familienmedizin, PMV forschungsgruppe, ÄZQ - Ärztliches Zentrum für Qualität in der Medizin (2013): Hausärztliche Leitlinie Multimedikation: Empfehlungen zum Umgang mit Multimedikation bei Erwachsenen und geriatrischen Patienten (äzq Schriftenreihe; 41). ÄZQ: Berlin. Verfügbar unter: <http://www.aezq.de/mdb/edocs/pdf/schriftenreihe/schriftenreihe41.pdf> (letzter Zugriff: 29.5.2014)

Mangoni AA, Jackson SHD (2003): Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. *Br J Clin Pharmacol*, 57: 6-14.

Mertens G (2009): Polypharmazie im Alter. Magisterarbeit. TU Berlin.

Mühlberg W (2002): Arzneimittelnebenwirkungen im Alter. Autoreferat vom 41. Bayerischen Internistenkongress des VBI/BDI vom 8-10.11.2002 in München. Verfügbar unter: <http://www.bayerischerinternistenkongress.de/autoreferate-2002/autoreferat-muehlberg.html> (letzter Zugriff: 25.5.2014)

Mühlberg W, Platt D, Mutschler E (1999): Neben- und Wechselwirkungen von Pharmaka im Alter. In: Platt D, Mutschler E (Hrsg.). *Pharmakotherapie im Alter*. Wissenschaftliche Verlagsgesellschaft: Stuttgart, S. 21-32.

Mukhtar AM (2010): Methodische Aspekte der Datenanalyse zur Polypharmazie. Präsentation zum Vortrag bei der Fachtagung „Polypharmazie und PRISCUS-Liste“ des ZI am 18.11.2010 in Berlin. Verfügbar unter: [http://www.zi-berlin.de/cms/fileadmin/images/content/PDFs\\_alle/Mukhtar.pdf](http://www.zi-berlin.de/cms/fileadmin/images/content/PDFs_alle/Mukhtar.pdf) (letzter Zugriff 14.5.2014)

Mühlberger N, Schneeweiss S and Hasford J (1997): Adverse drug reaction monitoring – cost and benefit considerations. Part I: frequency of adverse drug reactions causing hospital admissions. *Pharmacoepidemiology and Drug Safety*, 6(S3): S71-S77.

Naranjo B, Sellers A (1981): A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*, 30(2): 239-45.

NCC MERP - National Coordinating Council for Medication Error Reporting and Prevention (2013): What is a Medication Error?. Verfügbar unter: <http://www.nccmerp.org/aboutMedErrors.html> (letzter Zugriff: 23.03.2013)

NIHS - National Institute for Health Research, MHRA - Medicines and Healthcare products Regulatory Agency: Clinical Practice Research Datalink. Verfügbar unter: <http://www.cprd.com/intro.asp> (letzter Zugriff: 31.05.2014)

Oztekin A, Delen D, Kong ZJ (2009): Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology. *Int J Med Inform*, 78(12): e84-96.

Pepels W (2008): Marktforschung. Verfahren, Datenauswertung, Ergebnisdarstellung. 2. Überarbeitete Auflage. 2. Überarbeitete Ausgabe. Symposium Publishing: Krefeld.

Pfaff H, Glaeske G, Neugebauer EAM, Schrappe M (2009): Memorandum III: Methoden für die Versorgungsforschung (Teil 1). *Gesundheitswesen*, 71: 505-10.

Picksak G (2009): Arzneimittelsicherheit – eine wissenschaftliche Analyse auf Grundlage von Verordnungsdaten. Dissertation. Hannover.

Pintor-Mármol A, Baena MI, Fajardo PC, Sabater-Hernández D, Sáez-Benito L, García-Cárdenas MV, Fikri-Benbrahim N, Azpilicueta I, Faus MJ (2012): Terms used in patient safety related to medication: a literature review. *Pharmacoepidemiol Drug Saf*, 21(8): 799-809.

Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, Breckenridge AM (2004): Adverse drug reactions as cause of admission to hospital: prospective analysis of 18820 patients. *BMJ*, 329: 15-19.

Pitkala KH1, Strandberg TE, Tilvis RS (2001): Is it possible to reduce polypharmacy in the elderly? A randomised, controlled trial. *Drugs Aging*, 18(2): 143-9.

Putnam KG1, Buist DS, Fishman P, Andrade SE, Boles M, Chase GA, Goodman MJ, Gurwitz JH, Platt R, Raebel MA, Arnold Chan K (2002): Chronic disease score as a predictor of hospitalization. *Epidemiology*, 13(3): 340-6.

Rothschild JM, Churchill W, Erickson A, Munz K, Schuur JD, Salzberg CA, Lewinski D, Shane R, Aazami R, Patka J, Jagers R, Steffenhagen A, Rough S, Bates DW (2010): Medication errors recovered by emergency department pharmacists. *Ann Emerg Med*, 55(6): 513-21.

Roughead EE, Gilbert AL, Primrose JG, Sansom LN (1998): Drug-related hospital admissions: a review of Australian studies published 1988–1996. *Med J Aust*, 165: 405–8.

Reinboth C (2006): Multivariate Analyseverfahren in der Marktforschung. Lu-Lu Verlag: Morisville.

Rivo E de la Fuente J, Rivo Á, García-Fontán E, Cañizares MÁ, Gil P (2012): Cross industry standard process for data mining is applicable to lung cancer surgery domain, improving decision making as well as knowledge and quality management. *Clin Transl Oncol*, 14(1): 73-9.

Rochon PA, Gurwitz JH (1997): Optimising drug treatment for elderly people: the prescribing cascade. *BMJ*, 315: 1096-1099.

Sackett D, Haynes RB, Gibson ES, Tylor DW, Roberts RS, Johnson AL (1978): Patient compliance with antihypertensive regimens. *Patient Counselling & Health Education*, 11: 18-21.

Schmiedl S, Hanke F, Thürmann P (2010): Polypharmazie und Nebenwirkungen bei Älteren – eine Auswahl aktueller Studien in deutschen Altenheimen. *Phys Rehab Kur Med*, 20 - A33.

Schneeweiss S, Hasford J, Gottler M, Hoffmann A, Riethling AK, Avorn J (2002): Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. *Eur J Clin Pharmacol*, 58: 285-91.

Schnurrer JU, Frölich JC (2003): Zur Häufigkeit und Vermeidbarkeit von tödlichen unerwünschten Arzneimittelwirkungen. *Internist*, 44: 889-895.

Schulz M (2012): Selbstmedikation: Beratung bei der Selbstbehandlung. *Pharm Unserer Zeit*, 41: 294-301.

SGB-V – Sozialgesetzbuch Fünftes Buch Gesetzliche Krankenversicherung (2012): Stand: Zuletzt geändert durch Art. 4 G v. 21.7.2012 I 1613. Verfügbar unter: <http://www.sozialgesetzbuch-sgb.de/sgbv/12.html> (letzter Aufruf: 25.02.2013)

Statistisches Bundesamt (2009): 12. koordinierte Bevölkerungsvorausberechnung. Verfügbar unter: [www.destatis.de/Bevölkerungspyramide](http://www.destatis.de/Bevölkerungspyramide) (letzter Zugriff: 10.02.2013)

Stausberg J, Hasford J (2010): Erkennung von unerwünschten Arzneimittelereignissen: Nutzung von ICD-10-codierten Diagnosen in Routinedaten von Krankenhäusern *Dtsch Arztebl*, 107(3): 23-29.

SVR - Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. Bedarfsgerechtigkeit und Wirtschaftlichkeit (2002): Gutachten 2000/2001 - Band III: Über-, Unter- und Fehlversorgung. Baden-Baden: Nomos.

SVR - Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2003): Finanzierung, Nutzerorientierung und Qualität. Gutachten 2003 – Band I & Band II, Kurzfassung. Verfügbar unter: [http://www.svr-gesundheit.de/fileadmin/user\\_upload/Gutachten/2003/kurzf-de03.pdf](http://www.svr-gesundheit.de/fileadmin/user_upload/Gutachten/2003/kurzf-de03.pdf) (letzter Zugriff: 30.05.2014)

SVR - Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2009): Koordination und Integration - Gesundheitsversorgung in einer Gesellschaft des längeren Lebens. Bundestags-Drucksache 16/13770. Verfügbar unter: <http://dip21.bundestag.de/dip21/btd/16/137/1613770.pdf> (letzter Zugriff: 26.02.2013)

Tamura BK1, Bell CL, Inaba M, Masaki KH (2012): Outcomes of polypharmacy in nursing home residents. *Clin Geriatr Med*, 28(2): 217-36.



Thomas HF, Sweetnam PM, Janchawee B, Luscombe DK (1999): Polypharmacy among older men in South Wales. *European Journal of Clinical Pharmacology*, 55 (5): 411-415.

Thürmann, P et al. (2007): Arzneimittelrisiken bei hochbetagten Patienten: Ergebnisse deutscher Studien. In: BÄK (Hrsg.), Fortschritt und Fortbildung in der Medizin, Bd. 31, Dt. Ärzte Verlag: Köln, S. 216-224.

Thürmann PA, Holt S, Nink K, Zawinell A (2012): Arzneimittelversorgung älterer Patienten, In: Günster C, Klose J, Schmacke N (Hrsg.). Versorgungs-Report 2012. Schattauer-Verlag: Stuttgart, 111-130.

Timischl (2000): Biostatistik: Eine Einführung für Biologen und Mediziner. 2. neu bearbeitete Auflage. Springer: Wien, New York.

Turnheim K (2004): Drug therapy in the elderly. *Exp Gerontol*, 39: 1731-1738.

Varallo FR, Capucho HC, Planeta CS, Mastroianni PC (2014): Possible adverse drug events leading to hospital admission in a Brazilian teaching hospital. *Clinics*, 69(3): 163-167.

Varela G Jiménez MF, Novoa N. (2003): A model to predict death after lung cancer resection: applicability to individual cases. *Arch Bronconeumol*, 39(6): 249-52.

Veehof LJG, Stewart RE, Haaijer-Ruskamp FM, Meyboom-de Jong B (2000): The development of polypharmacy. A longitudinal study. *Family Practice*, 17(3):261-267.

Viktil KK, Blix HS, Moger TA, Reikvam A (2006): Polypharmacy as commonly defined is an indicator of limited value in the assessment of drug-related problems. *Br J Clin Pharmacol*, 63(2): 187-195.

Wehling M, Peiter A (2003): Arzneimitteltherapie im Alter aus der Sicht des klinischen Pharmakologen. *Internist*, 44: 1003-1009.

WHO - World Health Organization Regional, Office for South-East Asia (2006): The Role of Education in the Rational Use of Medications. SEARO Technical Publication Series No. 45. Im Eigenverlag: New Delhi.

WHO - World Health Organization (2009): WHO Patient Safety Research. WHO/IER/PSP/2009.10. Verfügbar unter: [http://whqlibdoc.who.int/hq/2009/WHO\\_IER\\_PSP\\_2009.10\\_eng.pdf](http://whqlibdoc.who.int/hq/2009/WHO_IER_PSP_2009.10_eng.pdf) (letzter Zugriff: 30.05.2014)

WIdO – Wissenschaftliches Institut der AOK (2010): Anatomisch-therapeutisch-chemische Klassifikation mit Tagesdosen - Amtliche Fassung des ATC-Index mit DDD-Angaben für Deutschland im Jahre 2010. DIMDI: Köln.

WIdO – Wissenschaftliches Institut der AOK (2012): Wenn fünf Pillen oder mehr normal sind – Älterer Patienten unterschätzen die Risiken bei der Einnahme mehrerer Arzneimittel. Pressemitteilung vom 21.06.2012. WIdO: Berlin.

Wiedmann KP, Buckler F (Hrsg.) (2003): Neuronale Netze im Marketing Management - Praxisorientierte Einführung in modernes Data-Mining. 2. Auflage. Gabler Verlag: Wiesbaden.

Winslow CEA (1920): The unfilled field of public health. *Modern Medicine*, 2: 183-191.

Yarnall, KS, Pollak, KI, Ostbye, T, Krause, KM und Michener, JL (2003): Primary care: is there enough time for prevention?. *Am J Public Health*, 93: 635-41.

## 10 ANLAGEN

### 10.1 Analyseplan

Betreff	Beschreibung
Fragestellung(en):	Lassen sich Patienten mit einem Bedarf an klinisch-pharmazeutischer Betreuung aus GKV-Routinedaten erkennen? Wie kann dieser Bedarf erkannt werden? Welche Einflussfaktoren spielen eine Rolle?,
Intention:	Entwicklung und Validierung eines Vorhersagemodells für den Bedarf an klinisch-pharmazeutischer Betreuung bei Polypharmazie aus GKV-Routinedaten mittels Data-Mining
Motivation:	Polypharmazie nimmt zu und verursacht vermutlich zunehmend arzneimittelbedingte Hospitalisationen (ABH). Ein steigender Bedarf an klinisch-pharmazeutischer Betreuung bei den Betroffenen wird angenommen und die Möglichkeit gesehen, GKV-Routinedaten zur Vorhersage eines Betreuungsbedarfs zu nutzen.
Untersuchungsdesign:	Retrospektive Datenbankanalyse mit Sekundärdaten einer GKV
Zielvariable	Potenzielle ABH innerhalb von einem Jahr ab Auswertungstichtag
Datengrundlage	Pseudonymisierte Routinedaten, einer bundesweiten, mittelgroßen Gesetzlichen Krankenversicherung von 2005-2008 (Entwicklungsdatensatz) und 2007-2010 (Validierungsdatensatz)
Ein- und Ausschlusskriterien	Patienten zwischen 18-85 Jahren mit Polypharmazie im Sinne von 5 oder mehr verschiedene Arzneistoffe pro Quartal über 1 Jahr ohne schwerwiegende Erkrankungen
Qualitätssicherung	<u>inhaltlich</u> : Analyse mit eindeutiger Zielgruppe und definiertem Zielmerkmal, sowie vorab formulierten Hypothesen und Analyseplan <u>methodisch</u> : Systematische Data-Mining Analyse nach etabliertem Standard (CRISP) und mittels logistischer Regression als geeignetes Strukturen prüfendes Verfahren des Data-Minings <u>praktisch/ ethisch</u> : Analyse auf Grundlage §303e SGB-V (Nutzungsrecht von Leistungsdaten) unter Beachtung des Datenschutzes und mit vollständiger Veröffentlichung der Ergebnisse.
Ergebnissicherung:	<u>interne Validität (statistische Güte)</u> : Prädiktoren: Signifikanzprüfung von Regressions- und Effektkoeffizienten, Chi-Quadrat-Veränderungen, Modell: Omnibus-Test: Pseudo-R-Quadrat, AUC, Ausreißer-Bestimmung, Spezifität, Sensitivität, Trefferquote, PPV <u>externe Validität (Verallgemeinerbarkeit)</u> : Prüfung von Modellstabilität und -güte am Validierungsdatensatz.

**10.2 UAW-Triggersymptome**

ICD-10	Bezeichnung	Quelle
A04.7	Enterokolitis durch Clostridium difficile*	Stausberg & Hasford, 2010
T88.7	nicht näher bezeichnete unerwünschte Nebenwirkung eines Arzneimittels oder einer Droge	
I95.2	Hypotonie durch Arzneimittel	
D61.10	aplastische Anämie infolge zytostatischer Therapie	
D69.0	Purpura anaphylactoides	
D69.58	sonstige sekundäre Thrombozytopenie, nicht als transfusionsrefraktär bezeichnet	
D69.59	sekundäre Thrombozytopenie, nicht näher bezeichnet	
F13.7	psychische und Verhaltensstörungen durch Sedativa oder Hypnotika: Restzustand und verzögert auftretende psychotische Störung	
T50.9	Vergiftung: Sonstige und nicht näher bezeichnete Arzneimittel, Drogen und biologisch aktive Substanzen	
T78.3	angioneurotisches Ödem	
L27.0	generalisierte Hauteruption durch Drogen oder Arzneimittel	
K52.1	toxische Gastroenteritis und Kolitis	
D70.10	Arzneimittel-induzierte Agranulozytose und Neutropenie	
T78.4	Allergie, nicht näher bezeichnet	
Y57.9	Komplikationen durch Arzneimittel oder Drogen	
T80.1	Gefäßkomplikationen nach Infusion, Transfusion oder Injektion zu therapeutischen Zwecken	

ICD-10	Bezeichnung	Quelle
I	Kardiovaskuläre Reaktionen	Keller, 2006
R0	Bradykardie (siehe „Sonstigen Reaktionen“)*	
R5	Synkope (siehe „Sonstigen Reaktionen“)*	
I4/R0	Arrhythmien (siehe „Sonstigen Reaktionen“)*	
I1	Schwere Hypotonie	
I1	Hypertone Krise	
K5	Akute gastrointestinale Reaktionen	

K2	Gastrointestinale Blutungen
K2	Gastrointestinale Ulcera
K2	Gastroenteritis
A0(K5)	Pseudomembranöse Kolitis
D 5,6	Blutbildungsstörungen, Blutgerinnungsstörungen, Gefäßreaktionen
D6/D5	Anämie
D6	Thrombozytopenie
D7	Leukozytopenie
D70	Agranulozytose
D61	Panzytopenie
D6	Hämorrhagie
D6	Purpura
I7/M/L	Vaskulitis
I6-18	Thrombose/Thrombophlebitis (siehe „Kardiovaskuläre Reaktionen“)*
I2	Lungenembolie
T	Allergische Reaktionen
J9	Bronchospasmus/ Laryngospasmus (siehe „Lungenerkrankungen“)*
R0	Dyspnoe (siehe „Sonstigen Reaktionen“)*
J8/I5	Lungenödem (siehe „Lungen-“, „Kardiovaskuläre Reaktionen“)*
J4	Asthma allergischer Genese (siehe „Lungenerkrankungen“)*
T7	Anaphylaxie
L/R	Schwere Hautreaktionen (siehe „Sonstigen Reaktionen“)*
K7	Leberfunktionsstörungen/ Leberparenchymschäden
B1	Ikterus
K7	Intrahepatische Cholestase
K7	Nichtinfektiöse Hepatitis
K7	Toxische Hepathopathie
N	Nierenfunktionsstörungen/ Nephropathien
R3	Oligurie/Anurie
E7	Hypoproteinämie (gemäß ICD: „Stoffwechselstörungen“)*

N1	Akutes/chronisches Nierenversagen
N1/N0	Nephropatie
(F)	ZNS-Symptomatik und psychische Reaktionen
G20	Parkinson-Syndrom
R4	Verwirrtheit, Desorientierung (siehe „Sonstigen Reaktionen“)*
R4	Halluzinationen (siehe „Sonstigen Reaktionen“)*
F3	Akute Psychosen
F3	Depression
F0/F1	Delirium
F1/F5	Medikamentenabhängigkeit
E	Stoffwechselstörungen
E1	Hypoglykämie
E1	Hyperglykämie
E8	Hyperkaliämie
E8	Hypokaliämie
E2	Cushing-Syndrom
	Sonstige Reaktionen
H	Akuter Hörverlust
H	Akute Sehstörung, Glaukom
T6/L5	Photosensibilität
K8	Pankreatitis
-	Gynäkomastie
F5	Potenz- und Libidostörungen

\*=Anmerkung: Einordnung der ICD inhaltlich zu einer alternativen Gruppe oder Gruppen denkbar

### 10.3 UAW-Triggersymptome nach ATC

UAW-Triggersymptome nach ATC		Quelle
Analgesics and antirheumatics (n=267)		
NSAIDs (n=168)	95 gastroduodenal bleeding (57%), 41 gastroduodenal lesions (24%), 5 diarrhea, (3.0%), 5 urticaria/anaphylaxis (3.0%), 3 agranulocytosis/ pancytopenia (1.8%)	Schneeweiss et al., 2002
Salicylates (n=99)	57 gastroduodenal bleeding (58%), 26 gastroduodenal lesions (26%), 4 urticaria/anaphylaxis (4%), 3 hypoglycemia (3%)	
Cardiovascular drugs (n=450)		
Digoxin/digitoxin (n=96)	31 bradycardia (32%), 20 syncope (21%), 6 AV block (6%), 6 Stoke-Adams syndrome (6%), 3 torsades de pointes (3%), 9 other arrhythmia (9%), 5 vomiting (5%),	
Calcium antagonists (n=91)	12 syncope (13%), 9 bradycardia (9.9%), 4 Stoke-Adams syndrome (4.4%), 8 other arrhythmia (8.8%), 29 gastroduodenal bleeding	
ACE inhibitors (n=73)	15 hypoglycemia (21%), 12 syncope (16%), 3 bradycardia (4.1%), 6 other arrhythmia (8.2%), 10 hypotension (7.3%), 3 hyperkalemia (4.1%)	
Beta blockers (n=73)	15 syncope (21%), 17 bradycardia (23%), 8 Stoke-Adams syndrome (11%), 4 AV block (5.5%), 8 other arrhythmia (11%), 13 hypoglycemia (18%), 5 hypotension (6.8%)	
Diuretics (n=50)	13 syncope (26%), 12 electrolyte disturbances (24%), 5 hypotension (10%), 5 bradycardia/ AV block (10%), 3 dehydration or dizziness (6%)	
Nitrates (n=43)	8 hypotension (19%), 10 syncope (23%), 7 bradycardia (16%), 4 Stoke-Adams syndrome (9.3%), 4 other arrhythmia (9.3%), 7 gastroduodenal lesions (16%)	
Other antihypertensives (n=14)	4 syncope (17%), 3 hypotension (21%), 4 arrhythmia (17%)	

Angiotensin-2 blockers (n=10)	3 bradycardia/ AV block (30%), 3 syncope/ hypotension (30%),
Antidiabetics (n=151)	
Insulin (n=98)	68 hypoglycemia (69%), 29 hypoglycemic coma (30%)
Oral antidiabetics (n=53)	30 hypoglycemia (57%), 21 hypoglycemic coma (40%)
Antithrombotics (n=293)	
Antithrombotics (n=293)	156 gastrointestinal hemorrhage (65%), 53 gastroduodenal lesions (18%), 8 intracerebral/ intracranial bleedings (2.7%), 37 other bleed- ing (13%), 5 hepatitis (2%), 3 anemia (1%), 3 thrombocytopenia (1%), 3 hypoglycemia (1%)
Hormones (n=57)	
Systemic corticosteroids (n=57)	12 gastrointestinal hemorrhage (21%), 10 gas- troduodenal lesions (18%), 11 osteoporosis (18%), 5 deep vein thrombosis (8.8%), 4 thrombophlebitis (7.0%), 3 diabetes (5.3%), gastroduodenal lesions include gastritis and ulcer



### 10.4 Analyse – Ausschlussdiagnosen

Die nachfolgende Tabelle zeigt die Verteilung der Ausschlussdiagnosen auf die verschiedenen ICD-Gruppen auf 3-stelliger Ebene betrachtet. Diagnosen aus Krankenhausdaten sind dann Ausschlussdiagnosen, wenn sie gelistet sind und es sich um eine Entlassungsdiagnose und zusätzlich um eine Haupt- oder Nebendiagnosen handelt.

ICD	ICD-Bezeichnung	Anzahl	Haupt- diagnose	Neben- diagnose
A00-B99	Schwere infektiöse Erkrankungen, HIV	19	Ja	Ja
C00-D48	Schwere Krebserkrankungen	225	Ja	Ja
D50-D90	Schwere Krebserkrankungen, Hämophilie	226	Ja	Ja
E00-E90	Seltene Stoffwechselerkrankungen	23	Ja	Ja
F00-F99	Psychische Erkrankungen, Demenz	554	Ja	Ja
G00-G99	Erkrankungen des Zentralnervensystems	530	Ja	Ja
H00-H59	Stark eingeschränkte Sehfähigkeit	5	Ja	Ja
H60-H95	Stark eingeschränkte Hörfähigkeit	4	Ja	Ja
I00-I99	Suchtproblematik, Kreislauferkrankung	110	Ja	Ja
J00-J99	Suchtproblematik, Atemwegserkrankung	109	Ja	Ja
K00-K93	Suchtproblematik, Magenerkrankung	25	Ja	Ja
M00-M99	Haut, Muskel-Skeletterkrankungen	26	Ja	Ja
N00-N99	Schwere Niereninsuffizienz: Dialysepflicht	820	Ja	Ja
P00-P96	Diagnose bezogen auf Geburt oder Neugeborene	1	Ja	Ja
Q00-Q99	Schwere Fehlbildungen	10	Ja	Ja
R00-R99	Senilität, Sprechstörungen, Todesfall	218	Ja	Ja
S00-T98	Schwere Verletzungen, Vergiftung	70	Ja	Ja
U00-U99	Sonstige schwere Erkrankungen	438	Ja	Ja
V01-Y84	Selbstmord oder Gewalttätigkeit	17	Ja	Ja
Z00-Z99	Sonstige hohe Inanspruchnahme med. Leistungen	490	Ja	Ja
Gesamt		3.920		

**10.5 Ergebnistabellen zu den Voranalysen***Voranalyse I – III*

<b>Anzahl an Patienten in 2005-2007</b>	<b>&lt;5 ATC</b>	<b>&gt;=5 ATC</b>	<b>5-8 ATC*</b>	<b>9-12 ATC</b>	<b>&gt;=13 ATC</b>
Anzahl (gesamt)	314.434	101.706	95.913	3.992	1.801
davon ohne Krhs. in 2008	247.256	62.642	59.720	2.291	631
davon mit Krhs. in 2008	67.178	39.064	36.193	1.701	1.170
davon sonstiger Krhs. in 2008	45.609	21.316	19.945	873	498
davon internistischer Krhs. in 2008	21.452	17.611	16.124	821	666
davon Notfall in 2008	117	137	124	7	6
<b>% Anteile</b>					
% Anteil (gesamt)	76%	24%	22,7%	1,3%	0,4%
% Anteil (>=5ATC)	-	100%	94%	4%	2%
<b>% Anteil in der ATC-Gruppe</b>					
% gesamt	100%	100%	100%	100%	100%
% ohne Krhs.	79%	62%	62%	57%	35%
% mit Krhs.	21%	38%	38%	43%	65%
% sonstiger Krhs.	15%	21%	21%	22%	28%
% internistischer Krhs.	7%	17%	17%	21%	37%
% Notfall	0,04%	0,13%	0,13%	0,18%	0,33%
<b>% Anteil in der Krhs. Gruppe</b>					
% mit Krhs.: alle Arten	100%	100%	100%	100%	100%
% mit Krhs.: sonstiger Krhs.	68%	55%	55%	51%	43%
% mit Krhs.: internistischer Krhs.	32%	45%	45%	48%	57%
% mit Krhs.: Notfall	0,17%	0,35%	0,34%	0,41%	0,51%
<b>Ø Anzahl Krhs. in 2008</b>					
Ø Anz. Krhs. (alle Pat.)	0,36	0,76	-	0,87	1,71
Ø Anz. Krhs. (Patienten mit Krhs.)	1,70	1,98	-	2,05	2,63
Ø Anz. internist. Krhs. (alle Pat.)*	0,10	0,28	-	0,33	0,68
Ø Anz. internist. Krhs. (Pat. mit Krhs.)	1,46	1,59	-	1,58	1,83

Ø Anz. Notfälle (Pat. mit Krhs.)	1,05	1,11	-	1,00	1,00
----------------------------------	------	------	---	------	------

ATC=Anatomisch-Therapeutisch-Chemischer Code; Intern.=internistisch; Krhs.=Krankenhausaufenthalt; Pat.=Patienten; grau=Differenzierte Angaben zu den Krankenhausaufenthalten; \*=aus Einzelergebnissen berechnetes Ergebnis;

### Voranalyse IV-V

Anzahl an Patienten	<5	>=5	>=5	>=5
	ATC/4Q	ATC/4Q	ATC/8Q	ATC/12Q
Auswertungsjahr	2005-7	2005-7	2006-7	2005-7
Anzahl (gesamt)	314.434	101.706	37.839	26.728
davon ohne Krhs.	247.256	62.642	20.706	14.202
davon mit Krhs.	67.178	39.064	17.133	12.526
davon sonstiger Krhs.	45.609	21.316	8.448	6.083
davon internistischer Krhs.	21.452	17.611	8.608	6.380
davon Notfall	117	137	77	63
% Anteil in der ATC-Gruppe				
% gesamt	100%	100%	100%	100%
% ohne Krhs.	79%	62%	55%	53%
% mit Krhs.	21%	38%	45%	47%
% sonstiger Krhs.	15%	21%	22%	23%
% internistischer Krhs.	7%	17%	23%	24%
% Notfall	0,04%	0,13%	0,20%	0,24%
% Anteil in der Krhs. Gruppe				
% mit Krhs.: alle Typen	100%	100%	100%	100%
% mit Krhs.: sonstiger Krhs.	68%	55%	49%	49%
% mit Krhs.: internistischer Krhs.	32%	45%	50%	51%
% mit Krhs.: Notfall	0,17%	0,35%	0,45%	0,50%

Ø Anzahl Krhs. in 2008				
Ø Anz. Krhs. (alle Pat.)	0,36	0,76	0,95	0,99
Ø Anz. Krhs., intern.(alle Pat.)	0,47	0,72	0,83	0,84
Ø Anz. Krhs. (Pat. mit Krhs.)	1,70	1,98	2,10	2,11
Ø Anz. Krhs., intern. (Pat. mit Krhs.)	1,46	1,59	1,65	1,66
Ø Anz. Notfälle (Pat. mit Krhs.)	1,05	1,11	1,08	1,08

Anz.=Anzahl; Krhs.=Krankenhauseinweisung; Pat.=Patienten; Q=Quartal

### Anzahl der Krankenhausfälle in Gruppe A

SPSS 19.0 Output:

#### Häufigkeiten

[DatenSet1] E:\\_Promotion\_16Jan2012\3\_Ergebnisse\04 Datensatz Signifikanzprüfung\Datensatz Signifikanzprüfung.sav

##### Statistiken

ANZ\_KRKH\_FAELLE

N	Gültig	52160
	Fehlend	0
Mittelwert		,52
Median		,00
Standardabweichung		,718

##### ANZ\_KRKH\_FAELLE

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 0	30866	59,2	59,2	59,2
1	16411	31,5	31,5	90,6
2	4060	7,8	7,8	98,4
3	711	1,4	1,4	99,8
4	93	,2	,2	100,0
5	17	,0	,0	100,0
6	2	,0	,0	100,0
Gesamt	52160	100,0	100,0	

```
FREQUENCIES VARIABLES=GR
  /STATISTICS=STDDEV MEAN MEDIAN
  /ORDER=ANALYSIS.
```

**Merkmal Krankenhauseinweisungen in Gruppe A****Häufigkeiten**

[DatenSet1] E:\\_Promotion\_16Jan2012\3\_Ergebnisse\04 Datensatz Signifikanzprüfung\Datensatz Signifikanzprüfung.sav

**Statistiken**

KN\_EINWEISUNG

N	Gültig	52160
	Fehlend	0

**KN\_EINWEISUNG**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig J	21294	40,8	40,8	40,8
N	30866	59,2	59,2	100,0
Gesamt	52160	100,0	100,0	

**10.6 Felddescription für die Variablen**

Variable	Felddescription*	Ausprägung	Zeitraum**
Patienten-Merkmale			
VERSCHL_PID	Verschlüsselte Patienten-Identifikationsnummer (mindestens 5 ATC pro Quartal über 4 Quartale)	8-stellige Nummer	(1) 2007 (2) 2009
KN_MG	Kennzeichen für Mitglied (durchgängig versichert)	J für Ja; N für Nein	(1) 2005-2008 (2) 2007-2010
ALTER_ID	Alter des Versicherten (18-85 Jahre)	von 0 bis über 100	(1) 31.12.2007 (2) 31.12.2009
KN_GESCHLECHT	Geschlecht des Versicherten	M für Männlich; W für weiblich	(1) 31.12.2007 (2) 31.12.2009
Anz_FAVE	Anzahl der Familienversicherten	0 bis ...	(1) 31.12.2007 (2) 31.12.2009
BUNDESLAND_ID	Bundesland in dem Versicherter wohnt	0 unbekannt 1 Schleswig-H. 2 Hamburg 3 Niedersachsen 4 Bremen 5 Nordrhein-Westf. 6 Hessen 7 Rheinland-Pf. 8 Baden-W. 9 Bayern 10 Saarland 11 Berlin 12 Brandenburg 13 Mecklenburg-Vorp. 14 Sachsen 15 Sachsen-Anh. 16 Thüringen	(1) 31.12.2007 (2) 31.12.2009

BG_KUERZEL	Bestandsgruppe des Versicherten	---- : unbekannt ALU: Arbeitsloser ARN: Arbeitnehmer REN: Rentner REST: Sonstige SEL: Selbstständige SOZ: Sozialhilfe STU: Student	(1) 31.12.2007 (2) 31.12.2009
Polypharmazie-Merkmale			
Anz_ATC_1J	Anzahl verschiedener Arzneistoffe pro Jahr (ohne Hilfsmittel)	Anzahl: 0 - n	(1) 2007 (2) 2009
Anz_ATC_3J	Anzahl verschiedener Arzneistoffe in 3 Jahren (ohne Hilfsmittel)	Anzahl: 0 - n	(1) 2005-2007 (2) 2007-2009
min_ATC_Q	Mindestanzahl an ATC pro Quartal über Quartal 1-4 im Auswertungsjahr	Anzahl: 0 - n	(1) 2007 (2) 2009
Rezept-Merkmale			
Anz_DDD_1J	Anzahl der DDD aller Arzneistoffe pro Jahr	Anzahl: 0 - n	(1) 2007 (2) 2009
Anz_VO_1J	Anzahl Arzneimittelverordnungen pro Jahr	Anzahl: 0 - n	(1) 2007 (2) 2009
Anz_REZEPTE_1J	Anzahl der Arzneimittelrezepte pro Jahr	Anzahl: 0 - n	(1) 2007 (2) 2009
ARZN_KOSTEN_1J	Arzneimittelkosten pro Jahr (ohne Hilfsmittel)	Summe in Euro: 0 - n	(1) 2007 (2) 2009
Anz_DDD_3J	Summe der DDD aller Arzneistoffe in 3 Jahren	Summe: 0 - n	(1) 2005-2007 (2) 2007-2009
Anz_VO_3J	Summe der Arzneimittelverordnungen in 3 Jahren	Anzahl: 0 - n	(1) 2005-2007 (2) 2007-2009
Anz_REZEPTE_3J	Anzahl der Arzneimittelrezepte in 3 Jahren	Anzahl: 0 - n	(1) 2005-2007 (2) 2007-2009
ARZN_KOSTEN_3J	Arzneimittelkosten in 3 Jahren (ohne Hilfsmittel)	Summe in Euro: 0 - n	(1) 2005-2007 (2) 2007-2009

Zielmerkmal			
ZIELMERKMAL	Krankenhausaufenthalt innerhalb von einem Jahr ab dem Auswertungstichtag	0: kein Krankenhaus 1: $\geq 1$ Krankenhaus- einweisung	(1) 2008 (2) 2010
ATC-Merkmale (26)			
ATC_GRP_A02BC01	Omeprazol	0: keine ATC,	(1) 2005-2007
ATC_GRP_A03FA01	Metoclopramid	1: mindestens eine ATC vorhanden	(2) 2007-2009
ATC_GRP_A10BA02	Metformin		
ATC_GRP_B01AA04	Phenprocoumon		
ATC_GRP_B01AC04	Clopidogrel		
ATC_GRP_B01AC05	Ticlopidin		
ATC_GRP_C01AA04	Digitoxin		
ATC_GRP_C01BD01	Amiodaron		
ATC_GRP_C01DX12	Molsidomin		
ATC_GRP_C03CA	Furo-, Torasemid		
ATC_GRP_C03DA01	Spironolacton		
ATC_GRP_C09AA05	Ramipril		
ATC_GRP_J01EE01	Sulfamethoxazol+ Trimethoprim		
ATC_GRP_J01_SPEZ	Makrolid (Clindamycin) oder Chinolon (Levofloxacin, Moxifloxacin)		
ATC_GRP_M04AA01	Allopurinol		
ATC_GRP_NSAR_SPEZ	Diclofenac, Ibuprofen oder ASS		
ATC_GRP_N02A_SPEZ	Tramadol, Tilidin- Kombinationen oder Fentanyl		
ATC_GRP_N02BB02	Metamizol		
ATC_GRP_N02BG07	Flupirtin		



ATC_GRP_N_SPEZ	Gabapentin, Pregabalino- der L-Dopa+Benserazid
ATC_GRP_N05_SPEZ	Zopiclon, Zolpidem, O- xazepam, Lorazepam, Bromazepam oder Pro- methazin
ATC_GRP_N06_SPEZ	Amitriptylin, Citalopram, Mirtazapin, Doxepin oder Pipofezin
ATC_GRP_R03DA04	Theophyllin
ATC_GRP_R03_SPEZ	Tiotropium, Salmeterol- Kombi, Formoterol (- Kombi), Fenoterol, For- moterol + Budesonid
ATC_GRP_R05CB01	Acetylcystein
ATC_GRP_S01_SPEZ	Prednisolon oder Dexa- methason + Antiinfektiva (Ophthalmika)
ATC_GRP_Risiko	Risikoarzneistoffe: Phen- procoumon, Clopidogrel, Ticlopidin, Digitoxin, Spi- ronolacton, NSAR (Dicl- ofenac, Ibuprofen, ASS), Allopurinol oder Theophyl- lin
ICD-Merkmale	
ICD Grp A	<p>Infektiöse Erkrankungen: 0: keine ICD, (1) 2005-2007</p> <p>Gastroenteritis und Kolitis 1: mindestens eine (2) 2007-2009</p> <p>A09 (UAW-Trigger) ICD vorhanden</p>
ICD Grp B	Kandidosen (B37) oder Streptokokken oder Staphylokokken-Infektion (B95)
ICD Grp D	Anämien D50, D62, D64, D68

ICD Grp E	Sonstige Hypothyreosen E03
ICD Grp E7	Purin-Pyrimidin- stoffwechselstörung E79
ICD Grp E8	Volumenmangel E86 oder Elektrolytstörungen E87
ICD Grp F	Depression F32 oder F45 oder F03,F06,F33,F43
ICD Grp G	Schlafstörungen G47oder Polyneuropathien G62-63 oder G20, G25, G40, G45, G55, G56, G81, G93
ICD Grp H	Augenerkrankungen bzw. Katarakt H04, H25, H26, H35, H52, H53 oder Hör- schwäche H61, H90, H91
ICD Grp I	Hypertensive Herzkrank- heit I11 oder Angina pec- toris I20 oder Herzinsuffi- zienz oder weitere Herz- erkrankungen I27, I34, I35, I51, I65
ICD Grp I4	Vorhofflattern I48 oder kardiale Arrhythmien I49 oder AV-Block I44
ICD Grp I6-9	Atherosklerose I70 oder sonstige periphere Ge- fäßerkrankung I73 oder weitere I67, I69, I83, I87, I89, I95
ICD Grp J	COPD J44 oder J96, Pneumonie J18, Bronchi- tis J20
ICD Grp K	Gastritis K29 oder funkti- onelle Darmstörung K59 oder weitere K52, K57, K92

ICD Grp L	Dermatitis L23 oder Dekubitus L89 oder weitere L29, L30, L97, L98
ICD Grp M	Osteoporose M81 oder Spondylose M47 oder weitere M06, 13, 15-17, 19, 25, 42, 48, 51, 53, 54, 62, 75, 79, 99
ICD Grp N	Chronische Nierenerkrankung N18 oder weitere N08, N19, Zystitis N30, N39 oder Prostatahyperplasie N40
ICD Grp R	Störung der Atmung R06 oder Schmerz R55 oder weitere R00, R07, R10, Übelkeit R11, R15, R26, R32, R39, Schwindel R42, R53, R55, Ödem R60
ICD Grp T	Verletzungen T14 oder Komplikationen nach Eingriffen T81, T82, T88
ICD Grp Z	kardiale, vaskuläre, funktionelle Implantate Z92, Z95, Z96 oder medizinische Behandlung in Eigenanamnese Z92 oder weitere Z03, Z51, Z74, Z85, Z90, unterstützende Apparate Z99

\*=Angaben in % stellen die Differenz der relativen Häufigkeiten in der Gruppe der Polypharmaziepatienten mit und ohne Krankenhausaufenthalt dar; (1)=Auswertungszeitraum für Traindatensatz; (2)=Auswertungszeitraum für Validierungsdatensatz

### 10.7 Übersicht der Variablen für Datensatz, Modellierung und finales Modell

Die folgende Tabelle enthält eine Übersicht über die Auswahl der Variablen in den unterschiedlichen Phasen der Modellentwicklung, das heißt bei der Datensatzzusammenstellung, bei der Modellierung und letztendlich im finalen Modell.

Variable	Feldbeschreibung (Kurzform)	Auswahl für Daten- satz	Auswahl für Mode- llierung	Auswahl für finales Modell
Patienten-Merkmal (7)				
VERSCHLÜSSELTE_PID	Patientennummer	X	--	--
KN_MG	Mitglied	X	--	--
ALTER_ID	Alter	X	X	X*
KN_GESCHLECHT	Geschlecht	X	X	X*
AnzAHL_FAVE	Familienversicherte	X	X	--
BUNDESLAND_ID	Bundesland d. Wohnorts	X	X*	--
BG_STATUS	Berufsgruppe	X	X	X*
Polypharmazie-Merkmale				
Anz_ATC_1_J	ATC pro 1 J.	X	X	X
Anz_ATC_3_J	ATC pro 3 J.	X	--	--
min_ATC_Q14 *		--	X	
Rezept-Merkmale				
Anz_DDD_1_J	DDD pro 1 J.	X	X	--
DDDproATC_12345	Ø DDD pro ATC in 1 J.			X*
Anz_VO_1_J	Verordnungen pro 1 J.	X	X	--
Anz_REZEPTE_1_J	Rezepte pro 1 J.	X	X	X
ARZN_KOSTEN_1_J	Arzneikosten pro 1 J.	X	X	--
LnKOST1Jplus1	Log. Kosten pro 1 J. +1			X*
Anz_DDD_3_J	DDD pro 3 J.	X	--	--
Anz_VO_3_J	Verordnungen pro 3 J.	X	--	--
Anz_REZEPTE_3_J	Rezepte pro 3 J.	X	--	--
ARZN_KOSTEN_3_J	Kosten pro 3 J.	X	--	--

Krankenhaus-Merkmale				
Anz_KRKH_F_GES_1_J	Krhs.-Fälle pro 1 J.	X	--	--
Anz_KRKH_T_GES_1_J	Krhs.-Tage pro 1 J.	X	--	--
Anz_KRKH_F_INT_1_J	Krhs.-Fälle, internist. 1 J.	X	--	--
Anz_KRKH_T_INT_1_J	Krhs.-Tage, internist. 1 J.	X	--	--
KOSTEN_KRKH_GES_1_J	Krhs.-Kosten 1 J.	X	--	--
Anz_KRKH_F_GES_3_J	Krhs.-Fälle pro 3 J.	X	--	--
Anz_KRKH_T_GES_3_J	Krhs.-Tage pro 3 J.	X	--	--
Anz_KRKH_F_INT_3_J	Krhs.-Fälle, internist. 3 J.	X	--	--
Anz_KRKH_T_INT_3_J	Krhs.-Tage, internist. 3 J.	X	--	--
KOSTEN_KRKH_GES_3_J	Krhs.-Kosten pro 3 J.	X	--	--
Abhängige Variable (1)				
ZIELMERKMAL	Krankenhaus innerh. 1 J.	X	X	X
ATC-Gruppierung aus Rezeptdaten				
ATC__A02BC01	Omeprazol	X	X	X
ATC_A03FA01	Metoclopramid	X	X	--
ATC_A10BA02	Metformin	X	X	X
ATC_B01AA04	Phenprocoumon	X	X	X
ATC_B01AC04	Clopidogrel	X	X	X
ATC_B01AC05	Ticlopidin	X	X	X
ATC_C01AA04	Digitoxin	X	X	--
ATC_C01BD01	Amiodaron	X	X	X
ATC_C01DX12	Molsidomin	X	X	X
ATC_C03CA	Furo-, Torasemid	X	X	X
ATC_C03DA01	Spironolacton	X	X	X
ATC_C09AA05	Ramipril	X	X	--
ATC_J01EE01	Sulfam.+Trimethoprim	X	X	--
ATC_J01_SPEZ	i.v. Antibiotika	X	X	X
ATC_M04AA01	Allopurinol	X	X	--
ATC_NSAR_SPEZ	NSAR-Auswahl	X	X	--

ATC_N02A_SPEZ	Opioid-Auswahl	X	X	X
ATC_N02BB02	Metamizol	X	X	X
ATC_N02BG07	Flupirtin	X	X	--
ATC_N_SPEZ	Psychopharmaka	X	X	X
ATC_N05_SPEZ	Sedativa	X	X	--
ATC_N06_SPEZ	Antidepressiva	X	X	X
ATC_R03DA04	Theophyllin	X	X	--
ATC_R03_SPEZ	Inhalativa für COPD	X	X	--
ATC_R05CB01	ACC	X	X	--
ATC_S01_SPEZ	Ophthalmika	X	X	--
ATC-Kombinationen*				
B01AC04_05	Clopidogrel + ASS	X*	X*	--
C01_BD01_AA04	Digitoxin + Amiodaron	X*	X*	--
C03DA01_CA	Schleifendiuretika + Spi- ronolacton	X*	X*	--
C09AA05_03CA	Furosemid + Ramipril	X*	X*	X*
M04AA01_C03CA	Schleifendiuretika + Al- lopurinol	X*	X*	--
R03DA04_M04AA01	Theophyllin + Allopurinol	X*	X*	--
C09AA05_03DA01	Spironolacton + Ramipril	X*	X*	--
M04AA01_C09AA05	Allopurinol + Ramipril	X*	X*	X*
R03DA04_B01AC04	Theophyllin + Clopidogrel	X*	X*	--
C01BD01_N02Aspz	Amiodaron + Opiode	X*	X*	--
ATC_Risiko	Risikoarzneistoff-Gruppe	X*	X*	X*
ICD-Gruppierung aus Arztdate				
ICD Grp A		X	--	--
ICD Grp B		X	--	--
ICD Grp D		X	--	--
ICD Grp E		X	--	--
ICD Grp E7		X	--	--
ICD Grp E8		X	--	--

ICD Grp F	X	--	--
ICD Grp G	X	--	--
ICD Grp H	X	--	--
ICD Grp I	X	--	--
ICD Grp I4	X	--	--
ICD Grp I6-9	X	--	--
ICD Grp J	X	--	--
ICD Grp K	X	--	--
ICD Grp L	X	--	--
ICD Grp M	X	--	--
ICD Grp N	X	--	--
ICD Grp R	X	--	--
ICD Grp T	X	--	--
ICD Grp Z	X	--	--

\*=optimierte oder neu erstellte Variable

## 10.8 Gruppierte ATC-Kombinationen (Matrix)

Nr.	Variable	Arzneimittel/-gruppe	risikoreich*	häufig**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	A02BC	Protonenpumpenhemmer		X		X													X											
2	A03FA01	Metoclopramid			X																									
3	A10BA02	<b>Metformin</b>	X	X																										
4	B01AA04	<b>Phenprocoumon</b>	X						X	X																				
5	B01AC04	<b>Clopidogrel</b>	X					X	X																					
6	B01AC05	<b>Acetylsalicylsäure (TAH)</b>					X	X																						
7	C01AA04	<b>Digitoxin</b>	X								XX	X	X	X	X															
8	C01BD01	<b>Amiodaron</b>	X							XX		X	X	X	X															
9	C01DX12	Molsidomin								X	X		X	X	X															
10	C03CA	<b>Schleifendiuretika</b>	X	X						X	X	X		X	X					X										
11	C03DA01	<b>Spironolacton</b>	X							X	X	X	X	X	XX															
12	C09AA05	Ramipril		X						X	X	X	X	XX						X										
13	J01EE01	Sulfamethoxazol und Trimethoprim																X												
14	J01_Spez	Antibiotika (systemische)															X													
15	NSAR_spez	<b>"NSAR"</b>	X	X		X															X	X								
16	M04AA01	<b>Allopurinol</b>	X											X	X													X		
17	N02A_spez	OPIOIDE		X															X			X	X							
18	N02BB02	Metamizol-Natrium		X															X		X		X							
19	N02BG07	Flupirtin																			X	X								
20	N_Spez	NERVENSYSTEM																						X	X					
21	N05_Spez	PSYCHOLEPTIKA ( ANTIPSYCHO																						X		X				
22	N06_Spez	PSYCHOANALEPTIKA (ANTIDEP																						X	X					
23	R03_Spez	Mittel bei Lungenerkrankungen																									X	X		
24	R03DA04	<b>Theophyllin</b>	X																	X	X					X		X		
25	R05CB01	Acetylcystein (ACC)		X																						X	X			
26	S01_Spez	OPHTHALMIKA*																												

\* = Auffälliger Arzneimittelwirkstoff bzw. Wirkstoffgruppe bezüglich der Anzahl an gleichzeitigen Krankenhauseinweisungen in den GKV-Routinedaten

\*\* = Auffälliger Arzneimittelwirkstoff bzw. Wirkstoffgruppe bezüglich der Häufigkeit der Verordnung in den Rezeptdaten

X, X, X = Zusammenhang mit zunehmender Stärke vermutet gemäß Angaben in der Literatur



## 10.9 Dokumentation zur Prüfung der Variablen

### Abgleich des Entwicklungsdatensatzes mit den Voranalyse-Ergebnissen

\*Datenbasis:

F:\20121114\_Promotion\3\_Ergebnisse\01 Datensatz Train PPHZ\_2005  
2008\AZMS\_Train\_opw\_v05\_logreg schritt.w.sav

\*Geschlechtsverteilung

**ist\_weibl**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 0	15926	36,1	36,1	36,1
1	28182	63,9	63,9	100,0
Gesamt	44108	100,0	100,0	

\*(prozentuale) Anteile der Polypharmazie-Gruppen im Datensatz

**neu\_atc\_gr5913**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ATC 5-8	39321	89,1	89,1	89,1
ATC 9-12	4348	9,9	9,9	99,0
ATC >=13	439	1,0	1,0	100,0
Gesamt	44108	100,0	100,0	

\*Häufigkeit des Zielmerkmals bei Polypharmazie (>=5 ATC/Q über 4Q)

**zielmerkmal**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 0	27089	61,4	61,4	61,4
1	17019	38,6	38,6	100,0
Gesamt	44108	100,0	100,0	

\*Häufigkeit des Zielmerkmals in den Polypharmazie-Gruppen

**zielmerkmal \* neu\_atc\_gr5913 Kreuztabelle**

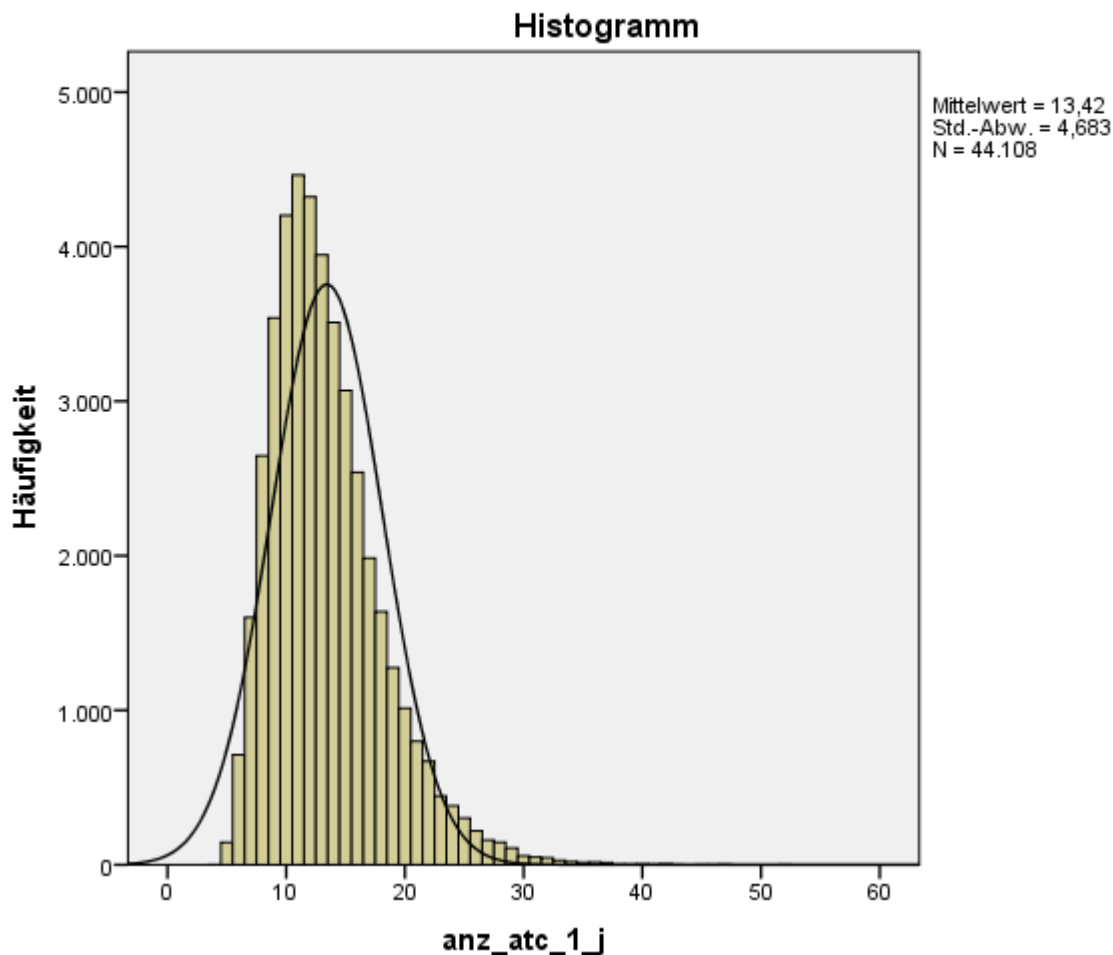
Anzahl

		neu_atc_gr5913			Gesamt
		ATC 5-8	ATC 9-12	ATC >=13	
zielmerkmal	0	24773	2137	179	27089
	1	14548	2211	260	17019
Gesamt		39321	4348	439	44108

**Beispiel: Prüfung auf Normalverteilung**

anz\_atc\_1\_j

N	Gültig	44108
	Fehlend	0

**Tests auf Normalverteilung**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
anz_atc_1_j	,115	44108	,000

a. Signifikanzkorrektur nach Lilliefors

**Erläuterung:**

Gemäß dem „Kolmogorov-Smirnov-Test“ ist die Variable Anz\_ATC\_1J, das heißt die Anzahl an verordneten Wirkstoffen pro Jahr, mit einer Wahrscheinlichkeit von  $p=0,000$  normalverteilt. Das heißt, die Variable ist nicht normalverteilt, auch wenn das Histogramm eine annähernde Normalverteilung zeigt.

**Beispiel: Test auf Unterschied der Mittelwerte.**

Zum Test auf Unterschied der Mittelwerte wird der „*T-Test für unabhängige Stichproben*“ verwendet.

**Gruppenstatistiken**

zielmerkmal	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
anz_rezepte_1_j	17019	30,06	12,995	,100
0	27089	26,12	10,800	,066
anz_atc_1_j	17019	14,62	5,040	,039
0	27089	12,66	4,274	,026
anz_ddd_1_j	17019	3024,92	1435,887	11,007
0	27089	2830,15	1262,011	7,668
arzn_kosten_1_j	17019	2771,030968	4221,571867	32,3598998
0	27089	2234,280522	3385,745085	20,5711121

**Test bei unabhängigen Stichproben**

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere
anz_rezepte_1_j	451,452	,000	34,474	44106	,000	3,944	,114	3,720	4,168
Varianzen sind nicht gleich			33,063	31289,224				,000	3,944
anz_atc_1_j	458,606	,000	43,779	44106	,000	1,963	,045	1,875	2,051
Varianzen sind nicht gleich			42,177	31796,436				,000	1,963
anz_ddd_1_j	142,753	,000	14,951	44106	,000	194,765	13,027	169,232	220,297
Varianzen sind nicht gleich			14,519	32705,123				,000	194,765
arzn_kosten_1_j	174,272	,000	14,710	44106	,000	536,7504458	36,4890017	465,2313766	608,2695149
Varianzen sind nicht gleich			13,998	30429,510				,000	536,7504458

**Erläuterung:**

Gemäß dem Ergebnis des *Levene-Tests* (Signifikanz=0,000) sind die Varianzen aller geprüften Variablen verschieden. In der Spalte „Varianzen sind nicht gleich“ zeigt der „*T-Test*“ mit  $p < 0,001$  an, dass die Mittelwerte in der Patientengruppe mit und ohne Krankenhausaufenthalt nicht nur bei den untersuchten Patienten, sondern auch in der Grundgesamtheit verschieden sind.

**Beispiel Prüfung auf Unterschiede der Mittelwerte in Fallgruppen**

Zur Prüfung auf Unterschiede der Mittelwerte in Fallgruppen wird eine Varianzanalyse durchgeführt. Im Beispiel wird die Variable zur Berufsgruppe untersucht (BG\_KUERZEL): 1=Selbständig, 2=Arbeitnehmer, 3=Rentner, 4=Arbeitsloser,

5=Sozialhilfeempfänger, 6=<gestrichen>, 7=Sonstiges, 8=keine Angabe).

**ONEWAY deskriptive Statistiken**

zielmerkmal

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1,00	656	,31	,464	,018	,28	,35	0	1
2,00	3988	,29	,454	,007	,28	,30	0	1
3,00	35235	,40	,490	,003	,40	,41	0	1
4,00	1339	,33	,470	,013	,30	,36	0	1
5,00	169	,47	,500	,038	,39	,54	0	1
7,00	1055	,37	,484	,015	,34	,40	0	1
8,00	1666	,35	,476	,012	,32	,37	0	1
Gesamt	44108	,39	,487	,002	,38	,39	0	1

**ONEWAY ANOVA**

zielmerkmal

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	57,664	6	9,611	40,775	,000
Innerhalb der Gruppen	10394,583	44101	,236		
Gesamt	10452,247	44107			

**Post-Hoc-Tests**

**Mehrfachvergleiche**

zielmerkmal  
Tamhane

(I) neu_bg_ostud	(J) neu_bg_ostud	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall		
					Untergrenze	Obergrenze	
1,00	2,00	,024	,020	,994	-,04	,08	
	3,00	-,088*	,018	,000	-,14	-,03	
	4,00	-,016	,022	1,000	-,08	,05	
	5,00	-,153*	,043	,008	-,28	-,02	
	7,00	-,058	,023	,237	-,13	,01	
	8,00	-,031	,022	,966	-,10	,03	
2,00	1,00	-,024	,020	,994	-,08	,04	
	3,00	-,114*	,008	,000	-,14	-,09	
	4,00	-,040	,015	,125	-,08	,00	
...							
8,00	1,00	,031	,022	,966	-,03	,08	
	2,00	,055*	,014	,001	,01	,10	
	3,00	-,057*	,012	,000	-,09	-,02	
	4,00	,015	,017	1,000	-,04	,07	
	5,00	-,122	,040	,055	-,25	,00	
	7,00	-,027	,019	,965	-,08	,03	

\*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

**Beispiel: Prüfung auf Häufigkeitsverteilung und Signifikanz**

Es wird eine Prüfung auf Häufigkeitsverteilung und Signifikanz mittels einer Kreuztabelle und einem Chi-Quadrat-Test untersucht.

**Verarbeitete Fälle**

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
neu_weibl * zielmerkmal	44108	100,0%	0	,0%	44108	100,0%

**ist\_weibl \* zielmerkmal Kreuztabelle**

			zielmerkmal		Gesamt
			0	1	
ist_weibl	0	Anzahl	9628	6298	15926
		% innerhalb von zielmerkmal	35,5%	37,0%	36,1%
	1	Anzahl	17461	10721	28182
		% innerhalb von zielmerkmal	64,5%	63,0%	63,9%
Gesamt		Anzahl	27089	17019	44108
		% innerhalb von zielmerkmal	100,0%	100,0%	100,0%

**Chi-Quadrat-Tests**

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	9,705 <sup>a</sup>	1	,002		
Kontinuitätskorrektur <sup>b</sup>	9,642	1	,002		
Likelihood-Quotient	9,693	1	,002		
Exakter Test nach Fisher				,002	,001
Anzahl der gültigen Fälle	44108				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6145,02.

b. Wird nur für eine 2x2-Tabelle berechnet

Erläuterung:

Die Kreuztabelle zeigt eine Häufigkeit von 63% „weiblich“ im Zusammenhang mit dem Zielmerkmal. Das Ergebnis des Chi-Quadrat-Tests zeigt eine „Asymptotische Signifikanz“ dafür, dass es keinen Zusammenhang mit einer Irrtumswahrscheinlichkeit von 0,00% gibt. Damit ergibt sich, dass ein Zusammenhang zwischen dem

Merkmal „weiblich“ und dem Zielmerkmal für die Grundgesamtheit als sehr wahrscheinlich vorausgesetzt werden kann.

**Beispiel: Ermittlung der Korrelationskoeffizienten (metrische Variable)**

\*Kreuztabelle (Variable Alter\_id)

\*Datenbasis [DatenSet2] E:\20121028\_Promotion\3\_Ergebnisse\01 Datensatz Train PPHZ\_2005 2008\AZMS\_Train\_opw\_v05\_logreg schritt.w.sav

**Symmetrische Maße**

		Wert	Asymptotischer Standardfehler <sup>a</sup>	Näherungsweise T <sup>b</sup>	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,091			,000
	Cramer-V	,091			,000
Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	,070	,004	17,754	,000
	Korrelation nach Spearman	,084	,005	17,736	,000 <sup>c</sup>
Intervall- bzgl. Intervallmaß	Pearson-R	,081	,005	16,987	,000 <sup>c</sup>
Anzahl der gültigen Fälle		44108			

- a. Die Null-Hyphothese wird nicht angenommen.
- b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.
- c. Basierend auf normaler Näherung

**Richtungsmaße**

			Wert
Nominal- bzgl. Intervallmaß	Eta	alter_id abhängig	,081
		zielmerkmal abhängig	,091

**Beispiel: Ermittlung der Korrelationskoeffizienten (kategoriale Variable)**

\*Nicht-parametrische Korrelation (Variable Alter\_60708085)

\*Datenbasis

[DatenSet2] E:\20121028\_Promotion\3\_Ergebnisse\01 Datensatz Train PPHZ\_2005 2008\AZMS\_Train\_opw\_v05\_logreg schritt.w.sav

**Korrelationen**

			zielmerkmal	neu_alter607085
Kendall-Tau-b	zielmerkmal	Korrelationskoeffizient	1,000	,075**
		Sig. (2-seitig)	.	,000
		N	44108	44108
	neu_alter607085	Korrelationskoeffizient	,075**	1,000
		Sig. (2-seitig)	,000	.
		N	44108	44108
Spearman-Rho	zielmerkmal	Korrelationskoeffizient	1,000	,082**
		Sig. (2-seitig)	.	,000
		N	44108	44108
	neu_alter607085	Korrelationskoeffizient	,082**	1,000
		Sig. (2-seitig)	,000	.
		N	44108	44108

\*\* . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

**10.10 Dokumentation zur Modellbildung****SPSS 19.0-Syntax:**

```
LOGISTIC REGRESSION VARIABLES zielmerkmal
  /METHOD=FSSTEP(WALD) neu_lnkost1jplus1 neu_ist_rentner neu_alter60708085
anz_rezepte_1_j neu_ist_mann atc_grp_b01ac04 atc_grp_b01aa04
atc_grp_c03da01 atc_grp_n02a_spez atc_grp_c01bd01 atc_grp_a02bc
atc_grp_n_spez neu_Risiko_mind1 atc_grp_c01dx12
atc_grp_n02bb02 atc_grp_b01ac05 atc_gr_m04aa01_c09aa05 atc_gr_c09aa05_03ca
atc_grp_c01aa04 atc_grp_j01_spez atc_grp_c03ca atc_grp_n06_spez
atc_grp_a10ba02 anz_atc_1_j DDDproATC_12345
  /PRINT=SUMMARY
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

**Ergebnis (SPSS 19.0-Output):****Zusammenfassung der Fallverarbeitung**

Ungewichtete Fälle <sup>a</sup>	N	Prozent
Ausgewählte Fälle	44108	100,0
Einbezogen in Analyse		
Fehlende Fälle	0	,0
Gesamt	44108	100,0
Nicht ausgewählte Fälle	0	,0
Gesamt	44108	100,0

a. Wenn die Gewichtung wirksam ist, finden Sie die Gesamtzahl der Fälle in der Klassifizierungstabelle.

**Codierung abhängiger Variablen**

Ursprünglicher Wert	Interner Wert
0	0
1	1

Erläuterung:

Das Zielmerkmal „Krankenhauseinweisung (innerhalb der folgenden 12 Monate)“ ist codiert mit ja=1 und 0=nein. Das Modell schätzt stets die höhere Merkmalsausprägung, das heißt, ob eine Krankenhauseinweisung vorliegt. SPSS 19.0 übernimmt „intern“ dazu die ursprüngliche Codierung im Datensatz.

**Block 0: Anfangsblock****Klassifizierungstabelle<sup>a,b</sup>**

Beobachtet			Vorhergesagt		
			zielmerkmal		Prozentsatz der Richtigen
			0	1	
Schritt 0	zielmerkmal	0	27089	0	100,0
		1	17019	0	,0
Gesamtprozentsatz					61,4

a. Konstante in das Modell einbezogen.

b. Der Trennwert lautet ,500

Erläuterung:

Schritt 0 bezeichnet die Ausgangssituation der Modellbildung. Das Modell enthält nur das Intercept, eine Konstante, und noch keine Merkmale. Im Datensatz lassen sich 17.019 von 44.108 (38,6%) Patienten in der Gruppe „mit Krankenhauseinweisung“ beobachten. Vorhergesagt wird für alle überprüften Patienten mit Hilfe des Ausgangsmodells die Zugehörigkeit aller Patienten in die Gruppe „ohne Krankenhauseinweisung“. Demnach sind 27.089 von 44.108 Patienten richtig klassifiziert worden. Das Ausgangsmodell trifft damit in 61,4% (Cut-off) der überprüften Patienten eine richtige Vorhersage.

**Variablen in der Gleichung**

	Regressions-	Standardfehler	Wald	df	Sig.	Exp(B)
--	--------------	----------------	------	----	------	--------



		koeffizient B					
Schritt 0	Konstante	-,465	,010	2258,068	1	,000	,628

Erläuterung:

Für die einzige in Schritt 0 verwendete „Variable“ liegt der Effektkoeffizient  $\text{Exp}(B)$  bei 0,628, der sich aus den richtig im Verhältnis zu den falsch klassifizierten Patienten ergibt (Odds-Ratio bzw. relative Häufigkeit:  $17.019/27.089=62,8\%$ ). In logarithmierter Form ergibt sich daraus ein Regressionskoeffizient von - 0,465. Diese Koeffizienten sind Modell bezogen und verändern sich von Modell zu Modell bei Veränderung der Variablen-Sets.

### Block 1: Methode=Vorwärts Schrittweise (Wald)

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt		4,560	1	,033
Schritt 24	Block	3067,090	24	,000
	Modell	3067,090	24	,000

Modellzusammenfassung

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
24	55760,168 <sup>a</sup>	,067	,091

a. Schätzung beendet bei Iteration Nummer 4, weil die Parameterschätzer sich um weniger als ,001 änderten.

Klassifizierungstabelle<sup>a</sup>

Beobachtet		Vorhergesagt		
		zielmerkmal		Prozentsatz der Richtigen
		0	1	
Schritt 24	zielmerkmal 0	23692	3397	87,5
	1	12274	4745	27,9
Gesamtprozentsatz				64,5

a. Der Trennwert lautet ,500

Erläuterung:

Die Modellierung erfolgt in „Block 1“ unter Einbeziehung aller vorgegebenen Variablen, nach der Methode „Schrittweise nach Wald“. Angezeigt ist das Ergebnis des Omnibus-Tests für den letzten Schritt der Modellierung bei 24 eingeschlossenen Va-

riablen (Schritt 24) unter „Modell“. Aus der Modellzusammenfassung ist das *R-Quadrat nach Nagelkerke* (Gütemaß) ablesbar. Die Klassifizierungstabelle zeigt an, welcher Anteil der Patienten durch das Vorhersagemodell richtig klassifiziert wurden: 87,5% (23.692 von 27.089) in der Gruppe „ohne Krankenhauseinweisung“, 27,9% (12.274 von 17.019) in der Gruppe „mit Krankenhauseinweisung“. Die richtige Vorhersage in beiden Gruppen ist unterschiedlich gut, insgesamt werden 64,5% der Patienten richtig durch das Vorhersagemodell zugeordnet.

		Variablen in der Gleichung							
		Regres- sions- koeffizient B	Stan-- dard- fehler	Wald	df	Sig.	Exp(B)	95% Konfidenz- intervall für EXP(B)	
								Unte- rer Wert	Oberer Wert
Schritt 24 <sup>a</sup>	neu_Inkost1jplus1	,151	,019	60,850	1	,000	1,163	1,120	1,208
	neu_ist_rentner	,131	,029	20,156	1	,000	1,140	1,077	1,207
	neu_alter60708085	,128	,013	102,067	1	,000	1,137	1,109	1,165
	anz_rezepte_1_j	,007	,001	28,533	1	,000	1,007	1,004	1,010
	neu_ist_mann	,124	,022	30,771	1	,000	1,132	1,084	1,183
	atc_grp_b01ac04	,251	,032	62,909	1	,000	1,285	1,208	1,367
	atc_grp_b01aa04	,238	,029	66,938	1	,000	1,268	1,198	1,343
	atc_grp_c03da01	,193	,041	22,046	1	,000	1,213	1,119	1,315
	atc_grp_n02a_spez	,191	,023	66,043	1	,000	1,210	1,156	1,267
	atc_grp_c01bd01	,198	,066	9,063	1	,003	1,219	1,072	1,388
	atc_grp_a02bc	,170	,022	60,596	1	,000	1,186	1,136	1,238
	atc_grp_n_spez	,144	,030	23,292	1	,000	1,155	1,089	1,224
	neu_Risiko_mind1	,136	,037	13,195	1	,000	1,145	1,064	1,232
	atc_grp_c01dx12	,115	,041	7,985	1	,005	1,122	1,036	1,215
	atc_grp_n02bb02	,103	,023	20,837	1	,000	1,109	1,061	1,159
	atc_grp_b01ac05	,093	,025	13,853	1	,000	1,097	1,045	1,152
	atc_gr_m04aa01_c09aa05	,096	,045	4,574	1	,032	1,101	1,008	1,203
	atc_gr_c09aa05_03ca	,102	,040	6,438	1	,011	1,107	1,023	1,198
	atc_grp_j01_spez	,080	,023	12,167	1	,000	1,084	1,036	1,134
	atc_grp_c03ca	,063	,025	6,464	1	,011	1,065	1,015	1,118
	atc_grp_n06_spez	,066	,025	6,998	1	,008	1,068	1,017	1,122
	atc_grp_a10ba02	-,140	,024	34,060	1	,000	,869	,830	,911
	anz_atc_1_j	,036	,003	110,006	1	,000	1,037	1,030	1,044
	DDDproATC_12345	-,060	,009	41,292	1	,000	,942	,925	,959
Konstante	-3,091	,139	491,163	1	,000	,045			

a. In Schritt 24 eingegebene Variablen: atc\_gr\_m04aa01\_c09aa05.

Variablen nicht in der Gleichung			Wert	df	Sig.
Schritt 24	Variablen	atc_grp_c01aa04	2,752	1	,097
	Gesamtstatistik		2,752	1	,097

Zusammenfassung der Stufen<sup>a,b</sup>

Schritt	Verbesserung			Modell			Richtige Klassen in %	Variable
	Chi-Quadrat	df	Sig.	Chi-Quadrat	df	Sig.		
1	1837,387	1	,000	1837,387	1	,000	63,2%	IN: anz_atc_1_j
2	276,095	1	,000	2113,482	2	,000	63,5%	IN: neu_alter60708085
3	160,293	1	,000	2273,775	3	,000	63,6%	IN: neu_lnkost1jplus1
4	132,903	1	,000	2406,678	4	,000	63,9%	IN: atc_grp_n02a_spez
5	115,275	1	,000	2521,953	5	,000	64,0%	IN: atc_grp_b01aa04
6	107,818	1	,000	2629,771	6	,000	64,2%	IN: atc_grp_b01ac04
7	84,530	1	,000	2714,300	7	,000	64,1%	IN: atc_grp_a02bc
8	45,370	1	,000	2759,671	8	,000	64,0%	IN: atc_grp_a10ba02
9	41,495	1	,000	2801,166	9	,000	64,1%	IN: atc_grp_c03da01
10	34,696	1	,000	2835,862	10	,000	64,2%	IN: atc_grp_n_spez
11	29,180	1	,000	2865,042	11	,000	64,3%	IN: neu_ist_mann
12	28,985	1	,000	2894,027	12	,000	64,2%	IN: atc_grp_n02bb02
13	21,179	1	,000	2915,206	13	,000	64,3%	IN: atc_gr_c09aa05_03ca
14	23,528	1	,000	2938,734	14	,000	64,4%	IN: DDDproATC_12345
15	27,275	1	,000	2966,009	15	,000	64,4%	IN: anz_rezepte_1_j
16	20,887	1	,000	2986,896	16	,000	64,4%	IN: neu_ist_rentner
17	20,373	1	,000	3007,270	17	,000	64,4%	IN: atc_grp_b01ac05
18	15,054	1	,000	3022,323	18	,000	64,4%	IN: neu_Risiko_mind1
19	11,183	1	,001	3033,506	19	,000	64,4%	IN: atc_grp_j01_spez
20	9,296	1	,002	3042,802	20	,000	64,4%	IN: atc_grp_c01bd01
21	7,933	1	,005	3050,735	21	,000	64,4%	IN: atc_grp_c01dx12
22	6,280	1	,012	3057,014	22	,000	64,5%	IN: atc_grp_n06_spez
23	5,515	1	,019	3062,529	23	,000	64,4%	IN: atc_grp_c03ca
24	4,560	1	,033	3067,090	24	,000	64,5%	IN: atc_gr_m04aa01_c09aa05

a. Aus dem aktuellen Modell können keine weiteren Variablen entnommen, bzw. ihm hinzugefügt werden.

b. Ende Block: 1

#### Erläuterung:

Die eingegangenen und die nicht eingegangenen Variablen des Modells sind tabellarisch aufgeführt. Die angezeigte Reihenfolge der eingegangenen Variablen entspricht der Reihenfolge der Variablen in der Syntax. Angezeigt sind der Regressionskoeffizient B, der Effektkoeffizient  $\text{Exp}(B)$  und die dazu berechneten Signifikanzen. In der Zusammenfassung der Stufen sind die Chi-Quadrat Verbesserungen je Schritt und Variable, sowie die Verbesserung für das Modell insgesamt angegeben. Die angezeigte Reihenfolge entspricht dem Eingang der Variablen in das Modell. Zusätzlich

sind die Signifikanzen angegeben und die Verbesserung der richtigen Klassifizierung in Prozent.

Fallweise Liste<sup>b</sup>

Fall	Ausgewählter Status <sup>a</sup>	Beobachtet	Vorhergesagt	Vorhergesagte Gruppe	Temporäre Variable	
		zielmerkmal			Resid	ZResid
1405	S	1**	,129	0	,871	2,593
1669	S	1**	,121	0	,879	2,691
1677	S	1**	,135	0	,865	2,531
2029	S	1**	,128	0	,872	2,612
3814	S	1**	,134	0	,866	2,545
3831	S	1**	,122	0	,878	2,681
8743	S	0**	,899	1	-,899	-2,989
14805	S	0**	,881	1	-,881	-2,723
22154	S	0**	,871	1	-,871	-2,603
41830	S	0**	,920	1	-,920	-3,393
42827	S	1**	,127	0	,873	2,626
42868	S	1**	,121	0	,879	2,700
43564	S	1**	,127	0	,873	2,623

a. S=ausgewählte Fälle, U=nicht ausgewählte Fälle und \*\*=falsch klassifizierte Fälle.

b. Fälle mit studentisierten Residuen größer als 2,000 werden aufgelistet.

#### Erläuterung:

In der „Fallweisen Liste“ sind die Patientenfälle aufgeführt, die Ausreißer darstellen und Z-Residuen größer als den Betrag von zwei aufweisen. Die Z-Residuen stellen ein Maß für die Schwere des Klassifikationsfehlers dar. Für die Ausreißer sind die durch das Modell vorhergesagten Wahrscheinlichkeiten angezeigt, sowie die dadurch fälschlicherweise vorhergesagte Gruppe (0=ohne Zielmerkmal, 1=mit Zielmerkmal).

### 10.11 Validierungsdatensatz

#### Datengrundlage:

- Altersbeschränkung auf 18-85 Jahren
- Patienten sind durchgängig versichert in 2007-2010
- Patienten mit Polypharmazie:  $\geq 5$  ATC pro Quartal über 4 Quartale in 2009 (und keine Polypharmazie im 4. Quartal 2008)
- Patienten ohne Ausschlussdiagnosen

#### Datensatzbeschreibung (Validierungsdatensatz)

Merkmal in 2009 bzw. 2007-9	Mess- niveau	Pat. mit Krhs. in 2010 (%)	Pat. ohne Krhs. In 2010 (%)
Patienten-Merkmale			
n (Gesamt: 45.739)		18.099 (39,6%)	27.640 (60,4%)
Alter (18-85 Jahre)	metrisch	Ø 70,0	Ø 68,7
Geschlecht : weiblich *	kategorial	11.370 (62,8%)	17.785 (64,3%)
Berufsgruppe: Rentner *	kategorial	15.182 (83,9%)	21.978 (79,5%)
Polypharmazie-Merkmale			
Anz. ATC pro Jahr	metrisch	Ø 14,7	Ø 12,6
kleinste Anzahl ATC pro Quartal über Quartal 1-4 *	metrisch	Ø 6,7	Ø 6,2
Rezept-Merkmale			
DDD pro Jahr	metrisch	Ø 3.073	Ø 2.865
DDD pro ATC pro Jahr *	kategorial	Ø 222,2	Ø 242,9
Rezepte pro Jahr	metrisch	Ø 29,9	Ø 25,9
Arzneimittelkosten/ Jahr	metrisch	Ø 2.889	Ø 2.301
ATC-Merkmale (Auswahl)			
Schleifendiuretika *	kategorial	8.069 (44,6%)	9.778 (35,4%)
Phenprocoumon	kategorial	3.821 (21,1%)	4.210 (15,2%)

Risikoarzneistoffe *, ***	kategorial	16.659 (92,0%)	24.372 (88,2%)
ICD-Merkmale (Auswahl)			
Volumenmangel (E8)	kategorial	3.963 (21,9%)	3.644 (13,2%)
Schlafstörungen (G)	kategorial	10.159 (56,1%)	12.862 (46,5%)
Atherosklerose (I6-9)	kategorial	11.373 (62,8%)	15.096 (54,6%)
COPD (J)	kategorial	7.933 (43,8%)	10.090 (36,5%)
Gastritis (K)	kategorial	9.420 (52,0%)	11.799 (42,7%)
Nierenerkrankung (N)	kategorial	10.837 (59,9%)	13.596 (49,2%)
Übelkeit, u.a. (R)	kategorial	13.317 (73,6%)	17.742 (64,2%)
Eigenanamnese ... (Z)	kategorial	12.125 (67,0%)	14.593 (52,8%)

Krhs.= Krankenhausaufenthalt; Std.= Standardabweichung; \*= berechnete Variablen; \*\*= Risikoarzneistoffe= Allopurinol, Clopidogrel, Digitoxin, Phenprocoumon, Spironolacton, Theophyllin, Ticlopidin,

## 10.12 Dokumentation zur Modellvalidierung

### Bestimmung der Modellgüte am Validierungsdatensatz

\*Datenbasis

E:\20121114\_Promotion\3\_Ergebnisse\02 Datensatz Validierung PPHZ\Datensatz Validierung opw\_logRegr M14 Val.sav

\*Anwendung des Modells

```
LOGISTIC REGRESSION VARIABLES zielmerkmal
/METHOD=Enter neu_lnkost1Jplus1 neu_ist_rentner neu_alter_60708085
anz_rezepte_1_j neu_ist_mann atc_grp_b01ac04 atc_grp_b01aa04
atc_grp_c03da01 atc_grp_n02a_spez atc_grp_c01bd01 atc_grp_a02bc
atc_grp_n_spez neu_Risiko_mind1 atc_grp_c01dx12
atc_grp_n02bb02 atc_grp_b01ac05 atc_vgr_m04aa01_c09aa05
atc_vgr_c_09aa05_03ca atc_grp_j01_spez atc_grp_c03ca atc_grp_n06_spez
atc_grp_a10ba02 anz_atc_1_j DDDproATC_12345
/SAVE=PRED LRESID
/CLASSPLOT
/CASEWISE OUTLIER(2)
/PRINT=ITER(1) SUMMARY CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

SPSS 19.0-Output:

## Logistische Regression

**Zusammenfassung der Fallverarbeitung**

Ungewichtete Fälle <sup>a</sup>	N	Prozent
Ausgewählte Fälle	45739	100,0
Einbezogen in Analyse		
Fehlende Fälle	0	,0
Gesamt	45739	100,0
Nicht ausgewählte Fälle	0	,0
Gesamt	45739	100,0

a. Wenn die Gewichtung wirksam ist, finden Sie die Gesamtzahl der Fälle in der Klassifizierungstabelle.

**Modellzusammenfassung**

Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	58212,384 <sup>a</sup>	,067	,091

a. Schätzung beendet bei Iteration Nummer 4, weil die Parameterschätzer sich um weniger als ,001 änderten.

**Klassifizierungstabelle<sup>a</sup>**

Beobachtet		Vorhergesagt		
		zielmerkmal		Prozentsatz der Richtigen
		0	1	
Schritt 1	zielmerkmal	0	1	
		23887	3753	86,4
		12733	5366	29,6
	Gesamtprozentsatz			64,0

a. Der Trennwert lautet ,500



Variablen in der Gleichung

	Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95% Konfidenzintervall für EXP (B)	
							Unterer Wert	Oberer Wert
Schritt 1 <sup>a</sup>								
neu_Inkost1Jplus1	,136	,018	59,793	1	,000	1,146	1,107	1,186
neu_ist_rentner	,104	,029	12,733	1	,000	1,110	1,048	1,176
neu_alter_60708085	,093	,013	54,869	1	,000	1,097	1,071	1,125
anz_rezepte_1_j	,004	,001	10,149	1	,001	1,004	1,002	1,007
neu_ist_mann	,135	,022	38,000	1	,000	1,145	1,097	1,195
atc_grp_b01ac04	,278	,032	76,330	1	,000	1,321	1,241	1,406
atc_grp_b01aa04	,213	,028	59,331	1	,000	1,237	1,172	1,306
atc_grp_c03da01	,116	,039	8,908	1	,003	1,123	1,041	1,212
atc_grp_n02a_spez	,142	,023	38,282	1	,000	1,153	1,102	1,206
atc_grp_c01bd01	,337	,064	27,671	1	,000	1,401	1,236	1,589
atc_grp_a02bc	,150	,022	46,973	1	,000	1,162	1,113	1,213
atc_grp_n_spez	,146	,028	27,358	1	,000	1,158	1,096	1,223
neu_Risiko_mind1	,037	,036	1,028	1	,311	1,037	,966	1,113
atc_grp_c01dx12	,103	,045	5,298	1	,021	1,109	1,015	1,211
atc_grp_n02bb02	,089	,022	16,602	1	,000	1,093	1,047	1,141
atc_grp_b01ac05	,103	,025	17,258	1	,000	1,108	1,056	1,163
atc_vgr_m04aa01_c09aa05	,001	,025	,003	1	,957	1,001	,953	1,052
atc_vgr_c_09aa05_03ca	,076	,036	4,549	1	,033	1,079	1,006	1,158
atc_grp_j01_spez	,069	,022	9,385	1	,002	1,071	1,025	1,120
atc_grp_c03ca	,086	,033	6,848	1	,009	1,090	1,022	1,162
atc_grp_n06_spez	,038	,024	2,412	1	,120	1,038	,990	1,089
atc_grp_a10ba02	-,109	,023	22,991	1	,000	,897	,858	,938
anz_atc_1_j	,052	,003	221,574	1	,000	1,053	1,046	1,060
DDDproATC_12345	-,061	,009	43,524	1	,000	,941	,924	,958
Konstante	-3,025	,145	435,032	1	,000	,049		

a. In Schritt 1 eingegebene Variablen: neu\_Inkost1Jplus1, neu\_ist\_rentner, neu\_alter\_60708085, anz\_rezepte\_1\_j, neu\_ist\_mann, atc\_grp\_b01ac04, atc\_grp\_b01aa04, atc\_grp\_c03da01, atc\_grp\_n02a\_spez, atc\_grp\_c01bd01, atc\_grp\_a02bc, atc\_grp\_n\_spez, neu\_Risiko\_mind1, atc\_grp\_c01dx12, atc\_grp\_n02bb02, atc\_grp\_b01ac05, atc\_vgr\_m04aa01\_c09aa05, atc\_vgr\_c\_09aa05\_03ca, atc\_grp\_j01\_spez, atc\_grp\_c03ca, atc\_grp\_n06\_spez, atc\_grp\_a10ba02, anz\_atc\_1\_j, DDDproATC\_12345.

### Test auf Ausreißer

Fallweise Liste<sup>b</sup>

Fall	Ausgewählter Status <sup>a</sup>	Beobachtet	Vorhergesagt	Vorhergesagte Gruppe	Temporäre Variable	
		zielmerkmal			Resid	ZResid
11381	S	0**	,865	1	-,865	-2,534
14865	S	0**	,922	1	-,922	-3,439
16183	S	0**	,869	1	-,869	-2,581
22053	S	0**	,874	1	-,874	-2,632
24975	S	0**	,866	1	-,866	-2,540
28875	S	1**	,121	0	,879	2,700
34998	S	1**	,125	0	,875	2,649
44766	S	1**	,135	0	,865	2,536

a. S = Ausgewählte, U = Nicht ausgewählte Fälle und \*\* = Falsch klassifizierte Fälle.

b. Fälle mit studentisierten Residuen größer als 2,000 werden aufgelistet.

## Bestimmung des AUC-Werts am Validierungsdatensatz

\*Bestimmung von ROC-Kurve und AUC-Wert aus Vorhersagewahrscheinlichkeit (PRE\_5) gegen das Zielmerkmal

```
ROC PRE_5 BY zielmerkmal (1)
/PLOT=CURVE (REFERENCE)
/PRINT=SE
/CRITERIA=CUTOFF (INCLUDE) TESTPOS (LARGE) DISTRIBUTION (FREE) CI (95)
/MISSING=EXCLUDE.
```

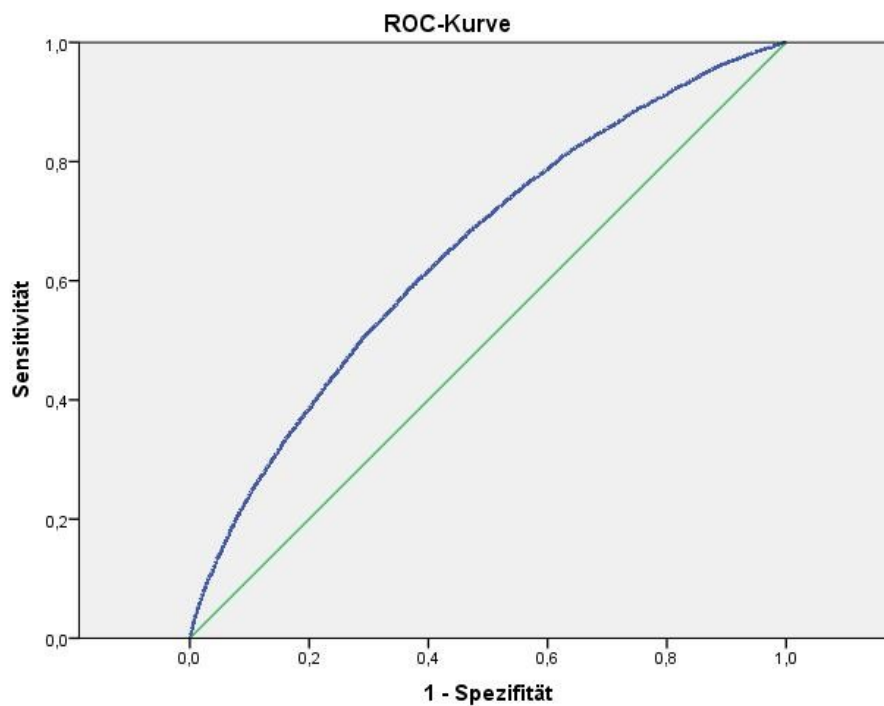
## ROC-Kurve

### Verarbeitete Fälle

zielmerkmal	Gültige N (listenweise)
Positiv <sup>a</sup>	18099
Negativ	27640

Größere Werte der Variable(n) für das Testergebnis deuten stärker auf einen positiven Ist-Zustand hin.

a. Der positive Ist-Zustand ist 1.



**Fläche unter der Kurve**

Variable(n) für Testergebnis:Modell30 einschluss

Fläche	Standardfehler $r^a$	Asymptotische Signifikanz <sup>b</sup>	Asymptotisches 95% Konfidenzintervall	
			Untergrenze	Obergrenze
,652	,003	,000	,647	,657

a. Unter der nichtparametrischen Annahme  
b. Nullhypothese: Wahrheitsfläche = 0.5

**10.13 Dokumentation zur Ergebnisprüfung (Variablen)**Prüfung von *min\_ATC\_Q14*:

Häufigkeitsverteilung:

**min\_atc\_q14**

	Häufigkeit	Prozent	Gültige Prozen- te	Kumulierte Pro- zente
5	17884	39,1	39,1	39,1
6	11331	24,8	24,8	63,9
7	7041	15,4	15,4	79,3
8	4070	8,9	8,9	88,2
9	2411	5,3	5,3	93,4
Gültig 10	1411	3,1	3,1	96,5
11	709	1,6	1,6	98,1
12	393	,9	,9	98,9
13	225	,5	,5	99,4
>=14	264	,6	,6	100,0
Gesamt	45739	100,0	100,0	

**min\_atcq14\_123**

	Häufigkeit	Prozent	Gültige Prozen- te	Kumulierte Pro- zente
1,00	40326	88,2	88,2	88,2
Gültig 2,00	4924	10,8	10,8	98,9
3,00	489	1,1	1,1	100,0
Gesamt	45739	100,0	100,0	

1=5-8 ATC/Quartal; 2=9-12 ATC/Quartal; 3=13 oder mehr ATC/Quartal

Vergleich der Mittelwerte und Prüfung auf signifikante Unterschiede:

Beispiel: Vorhersagewahrscheinlichkeit (PRE\_M14) je Fallgruppe von *min\_ATC\_Q14*

**ONEWAY deskriptive Statistiken**

PRE\_M14 Einschluss

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
5	17884	,3387	,10059	,00075	,3372	,3402	,12	,83
6	11331	,3778	,10797	,00101	,3758	,3798	,13	,87
7	7041	,4198	,11386	,00136	,4171	,4225	,15	,88
8	4070	,4571	,11731	,00184	,4535	,4607	,18	,83
9	2411	,4976	,11708	,00238	,4929	,5023	,20	,87
10	1411	,5398	,12025	,00320	,5335	,5461	,22	,90
11	709	,5830	,11537	,00433	,5745	,5915	,29	,88
12	393	,6092	,11192	,00565	,5981	,6203	,30	,87
13	225	,6466	,11011	,00734	,6322	,6611	,33	,90
>=14	264	,6998	,10707	,00659	,6868	,7128	,44	,98
Gesamt	45739	,3957	,12811	,00060	,3945	,3969	,12	,98

**Einfaktorielle ANOVA**

PRE\_M14 Einschluss

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	216,828	9	24,092	2063,927	,000
Innerhalb der Gruppen	533,791	45729	,012		
Gesamt	750,619	45738			

**Mehrfachvergleiche**

Abhängige Variable: PRE\_M14 Einschluss

Tamhane

(I) min_atc_q14	(J) min_atc_q14	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
5	6	-,03911*	,00126	,000	-,0432	-,0350
	7	-,08108*	,00155	,000	-,0861	-,0760
	8	-,11840*	,00199	,000	-,1249	-,1119
	9	-,15887*	,00250	,000	-,1670	-,1507
	10	-,20106*	,00329	,000	-,2118	-,1903
	11	-,24431*	,00440	,000	-,2587	-,2299
	12	-,27051*	,00570	,000	-,2892	-,2518
	13	-,30792*	,00738	,000	-,3322	-,2836
	>=14	-,36107*	,00663	,000	-,3829	-,3393
6	5	,03911*	,00126	,000	,0350	,0432
	7	-,04197*	,00169	,000	-,0475	-,0365
	8	-,07929*	,00210	,000	-,0861	-,0725

	9	-,11976*	,00259	,000	-,1282	-,1113
	10	-,16196*	,00336	,000	-,1729	-,1510
	11	-,20521*	,00445	,000	-,2197	-,1907
	12	-,23140*	,00574	,000	-,2502	-,2126
	13	-,26882*	,00741	,000	-,2932	-,2444
	>=14	-,32196*	,00667	,000	-,3439	-,3000
	5	,08108*	,00155	,000	,0760	,0861
	6	,04197*	,00169	,000	,0365	,0475
	8	-,03732*	,00229	,000	-,0448	-,0299
	9	-,07779*	,00274	,000	-,0867	-,0689
7	10	-,11998*	,00348	,000	-,1313	-,1087
	11	-,16323*	,00454	,000	-,1781	-,1484
	12	-,18943*	,00581	,000	-,2084	-,1704
	13	-,22684*	,00746	,000	-,2514	-,2023
	>=14	-,27999*	,00673	,000	-,3021	-,2579
	5	,11840*	,00199	,000	,1119	,1249
	6	,07929*	,00210	,000	,0725	,0861
	7	,03732*	,00229	,000	,0299	,0448
	9	-,04047*	,00301	,000	-,0503	-,0307
8	10	-,08267*	,00369	,000	-,0947	-,0706
	11	-,12592*	,00471	,000	-,1413	-,1106
	12	-,15211*	,00594	,000	-,1715	-,1327
	13	-,18953*	,00757	,000	-,2144	-,1646
	>=14	-,24267*	,00684	,000	-,2651	-,2202
	5	,15887*	,00250	,000	,1507	,1670
	6	,11976*	,00259	,000	,1113	,1282
	7	,07779*	,00274	,000	,0689	,0867
	8	,04047*	,00301	,000	,0307	,0503
9	10	-,04220*	,00399	,000	-,0552	-,0292
	11	-,08544*	,00495	,000	-,1016	-,0693
	12	-,11164*	,00613	,000	-,1317	-,0916
	13	-,14905*	,00772	,000	-,1744	-,1237
	>=14	-,20220*	,00701	,000	-,2252	-,1792
	5	,20106*	,00329	,000	,1903	,2118
	6	,16196*	,00336	,000	,1510	,1729
	7	,11998*	,00348	,000	,1087	,1313
	8	,08267*	,00369	,000	,0706	,0947
10	9	,04220*	,00399	,000	,0292	,0552
	11	-,04325*	,00539	,000	-,0608	-,0257
	12	-,06944*	,00649	,000	-,0907	-,0482
	13	-,10686*	,00801	,000	-,1332	-,0806
	>=14	-,16000*	,00733	,000	-,1840	-,1360
11	5	,24431*	,00440	,000	,2299	,2587
	6	,20521*	,00445	,000	,1907	,2197



5	17884	,0703	,25564	,00191	,0665	,0740	,00	1,00
6	11331	,1316	,33805	,00318	,1254	,1378	,00	1,00
7	7041	,2335	,42308	,00504	,2236	,2434	,00	1,00
8	4070	,3410	,47411	,00743	,3265	,3556	,00	1,00
9	2411	,4853	,49989	,01018	,4653	,5052	,00	1,00
10	1411	,5989	,49030	,01305	,5733	,6245	,00	1,00
11	709	,7673	,42286	,01588	,7361	,7985	,00	1,00
12	393	,8142	,38940	,01964	,7756	,8529	,00	1,00
13	225	,9156	,27867	,01858	,8789	,9522	,00	1,00
>=14	264	,9621	,19127	,01177	,9389	,9853	,00	1,00
Gesamt	45739	,1994	,39953	,00187	,1957	,2030	,00	1,00

### Einfaktorielle ANOVA

PRE\_2\_01

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	1508,452	9	167,606	1323,170	,000
Innerhalb der Gruppen	5792,490	45729	,127		
Gesamt	7300,942	45738			

### Ausprägung des Zielmerkmals in Abhängigkeit vom PRE-Wert

CROSSTABS

```

/TABLES=PRE_2_01 BY zielmerkmal
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ
/CELLS=COUNT ROW
/COUNT ROUND CELL
/BARCHART.

```

### PRE\_2 kategorial (1;0) \* zielmerkmal Kreuztabelle

			zielmerkmal		Gesamt
			0	1	
PRE_2 kategorial (1;0)	,00	Anzahl	23887	12733	36620
		% innerhalb von PRE_2 kategorial (1;0)	65,2%	34,8%	100,0%
	1,00	Anzahl	3753	5366	9119
		% innerhalb von PRE_2 kategorial (1;0)	41,2%	58,8%	100,0%
Gesamt		Anzahl	27640	18099	45739
		% innerhalb von PRE_2 kategorial (1;0)	60,4%	39,6%	100,0%

## Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	1769,457 <sup>a</sup>	1	,000		
Kontinuitätskorrektur <sup>b</sup>	1768,450	1	,000		
Likelihood-Quotient	1733,863	1	,000		
Exakter Test nach Fisher				,000	,000
Zusammenhang linear mit-linear	1769,418	1	,000		
Anzahl der gültigen Fälle	45739				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 3608,40.

b. Wird nur für eine 2x2-Tabelle berechnet

\*SPSS 19.0-Syntax für die Erstellung des Boxplots über „Diagrammerstellung“

GET

FILE='E:\20121114\_Promotion\3\_Ergebnisse\02 Datensatz Validierung PPHZ\Datensatz Validierung opw\_logRegr M30 -Val neu Kopie.sav'.

DATASET NAME DatenSet1 WINDOW=FRONT.

\* Diagrammerstellung.

GGRAPH

/GRAPHDATASET NAME="graphdataset" VARIABLES=zielmerkmal PRE\_5 MISSING=LISTWISE REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE.

BEGIN GPL

SOURCE: s=userSource(id("graphdataset"))

DATA: zielmerkmal=col(source(s), name("zielmerkmal"), unit.category())

DATA: PRE\_5=col(source(s), name("PRE\_5"))

DATA: id=col(source(s), name("\$CASENUM"), unit.category())

GUIDE: axis(dim(1), label("zielmerkmal"))

GUIDE: axis(dim(2), label("PRE\_M14 Einschluss"))

SCALE: linear(dim(2), include(0))

ELEMENT: schema(position(bin.quantile.letter(zielmerkmal\*PRE\_5)), label(id))

END GPL.



**Erklärung**

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unerlaubte fremde Hilfe angefertigt, keine anderen als die von mir angegebenen Quellen oder Hilfsmittel benutzt habe und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich erkläre außerdem, dass ich keine früheren Promotionen oder weiteren Promotionsversuche an dieser oder einer anderen Fakultät unternommen habe oder andere Promotionsvorhaben beantragt habe.

Hannover, den 15.10.2015

*Kerstin Boldt*

.....

Ort, Datum Unterschrift

## Abstract

### **Identifying patients receiving polypharmacy who are in need of pharmaceutical care – development and validation of a predictive model**

**objectives:** Polypharmacy is associated with adverse drug events, which can lead to hospitalization and even death. Providing pharmaceutical care can increase drug safety and reduce hospital admissions. As the number of drugs alone does not seem to be a sufficient predictor of risks the purpose of this study was to identify additional aspects of patients receiving polypharmacy who are at risk of hospitalization and in need of pharmaceutical care.

**methode:** A retrospective database analysis of two sets of routine health insurance data from 2005-2008 and 2007-2010 was performed. Patients aged 18-85 years on continuous polypharmacy ( $\geq 5$  drugs per quarter for 4 quarters) were included. Seriously ill patients were excluded. From the first data set (n=44.108) a predictive model was derived from 64 variables in a stepwise approach following the established data mining process CRISP-DM 1.0 using SPSS 19.0 and logistic regression. The final model containing 24 variables was validated using the second data set (n=45.739).

**results:** Of 45.739 patients on polypharmacy 39.6% (18.099 of 45.739) were admitted to hospital within one year. Most patients were on medium polypharmacy using 5-8 drugs for one year (88.1%; 40.326 von 45.739). Compared to using the number of drugs as a solely predictor ( $\geq 13$  drugs: n=489, PPV=59.9%) the model identified a larger group of patients with a higher probability of hospitalization and a presumed need of pharmaceutical care (n=1.161, PPV=71.6%). The quality of the predictive model was acceptabel (AUC=65.2%, 95% CI 64.7-65.7%) and stabel over a two year period. The strongest predictors for hospitalization among patients on polypharmacy appeared to be number of different drugs per year, age, drug costs and the use of metamizol, opioids, loop-diuretics, phenprocoumon und clopidogrel.

**conclusion:** The derived predictive model improves identification of patients on polypharmacy at risk of hospitalization and helps addressing phramaceutical care more exactly to patients in need. Data basis should be improved to increase quality of prediction by extended linkage of primary and secondary data records and intensified documentation of drug use and triggered events by drug experts.