

# Statistische Evaluation der Medikamentendaten – Cut-off Bestimmung für COPD

Peter Filzmoser\*, Can Mert

Department of Statistics and Probability Theory  
Vienna University of Technology, Austria

23. August 2012

---

\*Correspondence to P. Filzmoser, Vienna University of Technology, Department of Statistics and Probability Theory, Wiedner Hauptstraße 8-10, A-1040 Vienna, Austria; Email: P.Filzmoser@tuwien.ac.at, Tel.: +43 1 58801 10733, Fax: +43 1 58801 10799

# Inhaltsverzeichnis

<b>1</b>	<b>Problemstellung</b>	<b>3</b>
<b>2</b>	<b>Erstellung der Population</b>	<b>3</b>
<b>3</b>	<b>Kombinationen von ATC codes</b>	<b>5</b>
<b>4</b>	<b>Berücksichtigung der ICD codes</b>	<b>10</b>
<b>5</b>	<b>Zusammenfassung von Medikamenten-Gruppen</b>	<b>14</b>
<b>6</b>	<b>Isolation von COPD</b>	<b>17</b>
<b>7</b>	<b>Zusammenfassung der Ergebnisse</b>	<b>21</b>

# 1 Problemstellung

COPD (Chronic Obstructive Pulmonary Disease) und Asthma sind bezüglich der Medikation relativ ähnlich, was eine Identifikation von COPD schwierig macht. Das Ziel dieser Studie sollte es sein, anhand der Medikation und der Anzahl der Verordnungen Unterschiede herauszufiltern um COPD zu isolieren. Das Betrachten von Medikamenten-Kombinationen bzw. anderen Informationen wie Alter der Personen könnte zu diesem Ziel beitragen.

Die vorliegende Datenbasis geht über den Zeitraum 2006 und 2007. Die Medikation bezieht sich auf diese Daten, und wird personenbezogen ausgewertet. Folgende ATC codes (bzw. Gruppen von ATC codes) werden betrachtet:

Kurzwirksame Anticholinergica: R03BB01

Langwirksame Anticholinergica: R03BB04

Inhalative Corticoide: R03BA01, R03BA02, R03BA05

Kurzwirksame  $\beta$ -2-Mimetika: R03AC02, R03AC03, R03AC04, R03AK03, R03CC03

Langwirksame  $\beta$ -2-Mimetika: R03AC12, R03AC13, R03CC12, R03CC13

Kombination langwirksame  $\beta$ -2-Mimetika + inhalative Corticoide: R03AK06, R03AK07

Theophylline: R03DA04

Leukotrienrezeptor-Antagonisten: R03DC03

Typischerweise wird

**Langwirksame Anticholinergica: R03BB04**

für **COPD** verordnet, und

**Leukotrienrezeptor-Antagonisten: R03DC03**

für **Asthma**. Ein Anhaltspunkt für die Unterscheidung zwischen COPD und Asthma, bzw. für die Isolierung von COPD, könnte die Kombination solcher ATC-Gruppen sein.

## 2 Erstellung der Population

In der gesamten Datenbasis werden jene Personen herausgefiltert, bei denen mindestens einmal einer der oben gelisteten ATC codes auftritt. Dieser Datensatz definiert die *Population*, die Personen enthält, die potentiell and COPD bzw. Asthma erkrankt sind, die aber auch Akutpatienten mit Bronchitis oder grippalem Infekt sein können. Insgesamt sind etwa 800.000 verschiedene Personen in dieser Population. Als Information speichern wir die Personen-ID, die Versicherungskategorie, das Geburtsjahr, das Geschlecht, den Wohnbezirk, sowie die gesamten Häufigkeiten, mit denen jeder einzelne dieser ATC codes vorkommt. Abbildung 1 zeigt die Altersverteilung der Population. Man wird später an der Grenze 40 Jahre interessiert sein.

Im nächsten Schritt interessieren wir uns für die gesamte Anzahl an Verordnungen (dieser selektierten ATC codes). Es könnte sein, dass ein bestimmter Personenkreis (wie chronische erkrankte Personen) eine hohe Anzahl solcher Medikamente erhalten. Abbildung 2 stellt die Anzahl der Personen dar, die eine bestimmte Anzahl an Verordnungen erhalten haben. Klarerweise haben sehr viele Personen nur eine oder wenige Verordnung erhalten, wodurch die Verteilung extrem rechtsschief wird. In der rechten Grafik wird auf einen interessanten Bereich gezoomt, der aber keine klaren Strukturen (Brüche, Bimodalität) erkennen lässt.

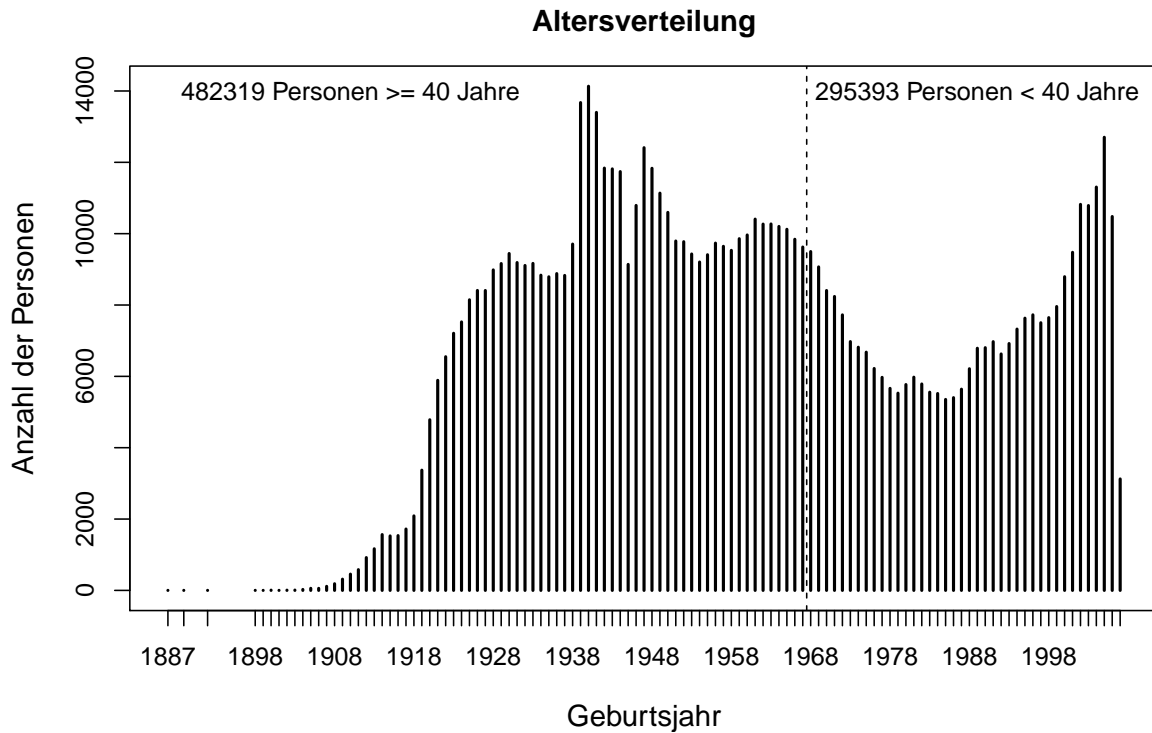


Abbildung 1: Altersverteilung der Population

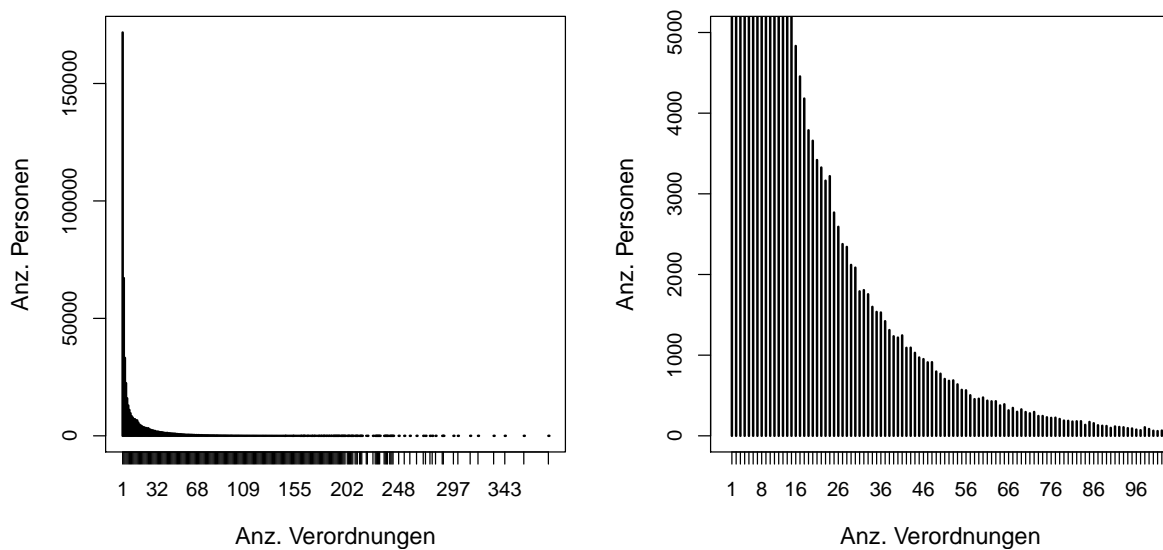


Abbildung 2: Anzahl der Personen mit einer bestimmten Anzahl von Verordnungen: links alle Daten, rechts gezoomt.

Eine bessere Visualisierung kann erreicht werden, indem die Anzahl der Personen in einer logarithmierten Skala aufgetragen wird. Abbildung 3 zeigt diese Darstellung. In einem weiten Bereich ist nun ein linearer Abfall sichtbar, der ab einer bestimmten Anzahl (136) von Verordnungen instabil wird. Es stellt sich somit die Frage, ob dieser Bruch bereits die akuten von den chronischen Personen unterscheidet. Diese Behauptung wäre allerdings unsicher, weil ein Bruch auch deshalb entstehen kann, weil die Häufigkeiten

(von Personen) mit dieser hohen Anzahl von Verordnungen mehr dem Zufall unterworfen sind. Die Häufigkeiten bewegen sich hier im Bereich 1-12, was natürlich zufällig schwanken kann.

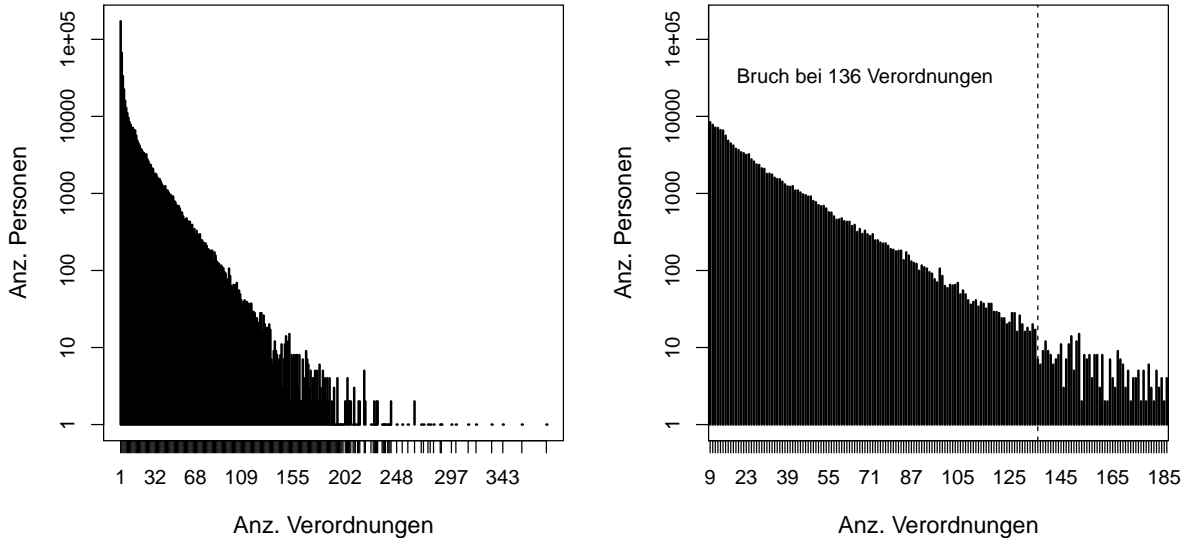


Abbildung 3: Logarithmierte Anzahl von Personen mit einer bestimmten Anzahl von Verordnungen: links alle Daten, rechts gezoomt.

Die Schlussfolgerung ist, dass man in einem Plot der Anzahl von Personen gegen die Anzahl von Verordnungen *keinen Hinweis* erhält zur Unterscheidung von chronischer und akuter Krankheit. So ein Hinweis kann vielleicht dann gefunden werden, wenn nicht alle ATC codes gemeinsam, sondern nur selektierte codes genommen werden. Allerdings ist unklar, welche codes hier betrachtet werden sollen. Dieser Frage wird im nächsten Abschnitt nachgegangen.

### 3 Kombinationen von ATC codes

Wir schränken die Untersuchungen nun auf die 8 Gruppen von ATC codes ein, die in Abschnitt 1 definiert wurden. Von diesen Gruppen werden nun speziell Kombinationen betrachtet, die im Zusammenhang mit

**Langwirksame Anticholinergica: R03BB04**

(COPD) stehen, bzw. solche, bei denen immer

**Leukotrienrezeptor-Antagonisten: R03DC03**

(Asthma) vorkommen. Wir werden uns außerdem auf die Personengruppe mit Mindestalter 40 Jahre einschränken. Die Schwierigkeit ist, dass es nicht sinnvoll ist, nur nach Vorkommen TRUE/FALSE zu suchen, sondern auch die Häufigkeiten der Verordnungen in den Kombinationen zu betrachten. Ein erster Ansatz dazu findet sich in Abbildung 4 für COPD. Wir betrachten alle jene Personen, die eine feste Anzahl  $x > 0$  von Verordnungen für R03BB04 erhalten haben ( $x$  wird dann bis zum maximal auftretenden Wert variiert). Man erhält somit eine Matrix, bei der die Zeilen die Personen bilden, und die Spalten die Anzahlen der Verordnungen für die jeweilige ATC Gruppe (bei R03BB04 steht immer ein Wert von  $x$ ). Diese Matrix enthält sehr viele Nullen, weil bestimmte Kombinationen eben

nicht vorkommen. Für größere Werte von  $x$  wird die Matrix außerdem sehr klein, weil nicht mehr viele Personen eine so hohe Anzahl von Verordnungen für R03BB04 haben. Die Frage ist nun, wie man diese Information darstellt.

Abbildung 4 zeigt eine mögliche Darstellungsform. Für jede Matrix (entsprechend einem Wert  $x$ ) wurde der Spaltenmedian (linker Plot) und das Spaltenmittel (rechter Plot) berechnet. Die erhaltenen Werte wurden als Linien in die Grafiken eingezeichnet. Eine Linie entspricht dabei einem festgehaltenen Wert von  $x$ . Abbildung 4 bezieht sich auf den COPD-typischen code R03BB04, der ganz links in den Grafiken steht. Man bemerke hier die regelmäßigen Werte, die jeweils einem festen  $x$  entsprechen. Für eine bessere Unterscheidung der Linien wurde eine Farbskala verwendet, die von der Höhe von R03BB04 Verordnungen abhängt. Das Bild für Median und Mittel ist nicht sehr unterschiedlich: Es treten etwa die gleichen Kombinationen häufig auf. Für sehr hohe Anzahlen von R03BB04 Verordnungen wird das Bild eher chaotisch, was aber auch an der geringen Anzahl von Personen, die dahinter stehen, liegen kann. Abbildung 5 zeigt schließlich noch die Plots,

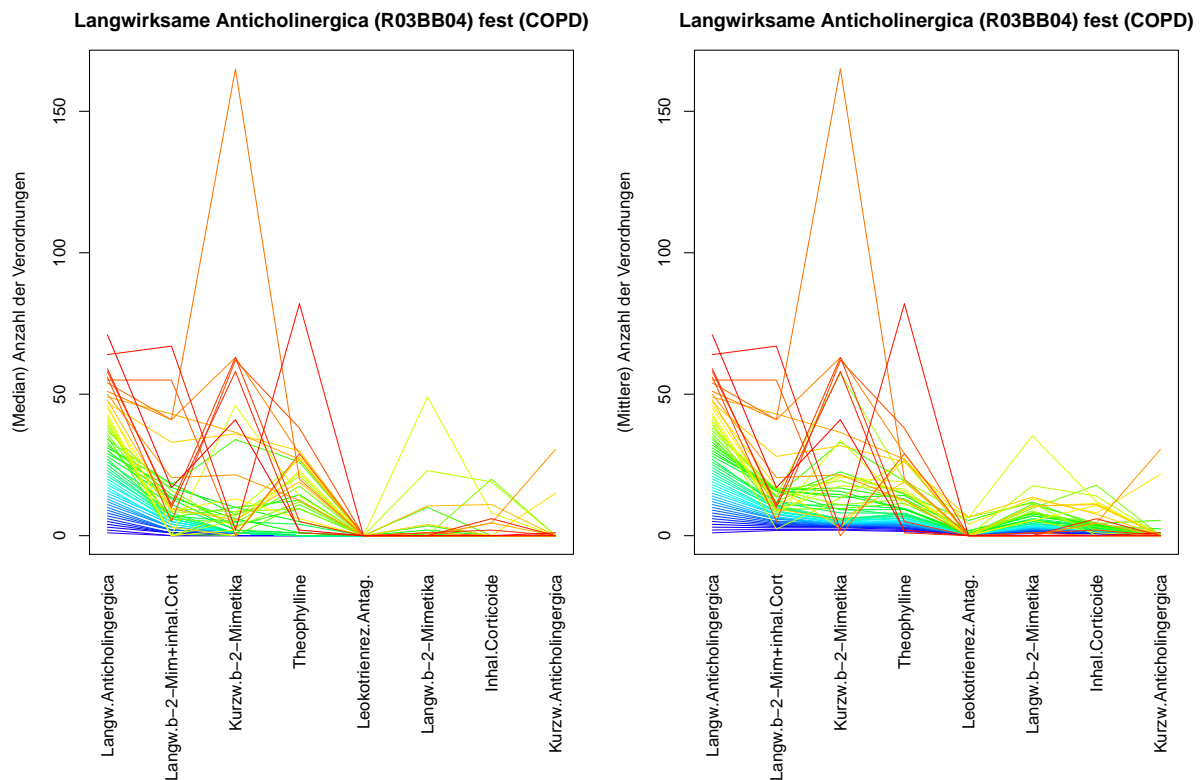


Abbildung 4: Kombinationen bezogen auf den ATC code R03BB04 (COPD). Links die Mediane zu anderen ATC codes, rechts die Mittelwerte.

die sich für die Kombinationen mit R03DC03 (Asthma) ergeben. Auch hier ist für hohe Anzahlen von Verordnungen ein chaotisches Bild, aber für geringere Anzahlen zeigt sich eine relativ klare Struktur.

Die Abbildungen sind offenbar noch nicht sehr zufriedenstellend, weil man die aktuellen Anzahlen von Kombinationen nur schwer abschätzen kann. Daher wird nun dazu übergegangen, die einzelnen Anzahlen von Kombinationen wirklich zu ermitteln. Wir unterscheiden dabei zwischen hoher Anzahl von Verordnungen (mindestens 10) und mittlerer Anzahl (5 bis 9). Jetzt kann man wiederum zählen, ob eine hohe Anzahl von Verordnungen bei R03BB04 (COPD) in Kombination mit einer hohen bzw. mittleren Anzahl der

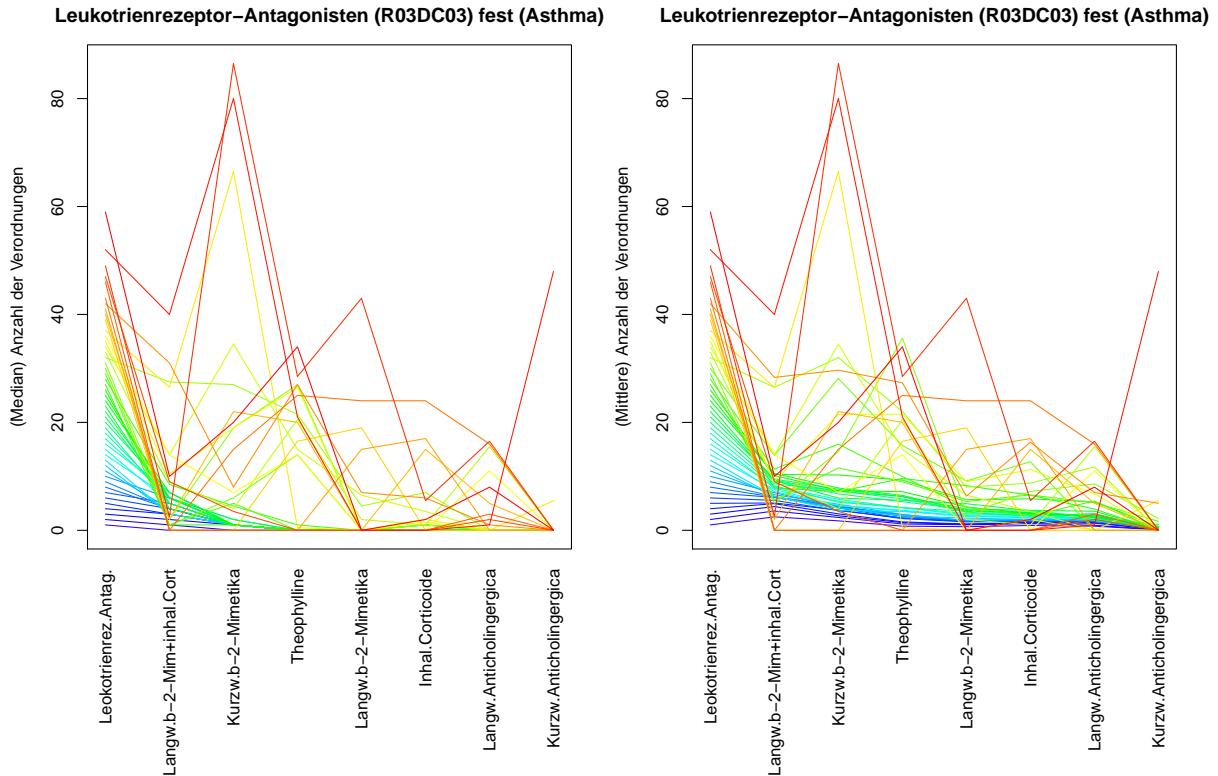


Abbildung 5: Kombinationen bezogen auf den ATC code R03DC03 (Asthma). Links die Mediane zu anderen ATC codes, rechts die Mittelwerte.

anderen ATC Gruppen vorkommt. Abbildung 6 zeigt das Resultat für die Kombination  $\leq 10$  Verordnungen von R03BB04 (COPD) und  $\leq 10$  Verordnungen der anderen ATC Gruppen. Im linken Teil wurden die Personen unter 30 Jahre betrachtet, im rechten Teil alle mit mindestens 40 Jahren. Die aufgelisteten Kombinationen sind von oben nach unten der Häufigkeit nach sortiert. Die Liste wurde nach einer bestimmten Länge abgeschnitten. Es wurden Kombinationen von R03BB04 (COPD) mit bis zu 4 anderen ATC Gruppen betrachtet. Am linken Rand stehen die Häufigkeiten der Kombinationen. Diese Häufigkeiten entsprechen der Anzahl von Personen, bei denen diese Kombination auftritt. Am rechten Rand stehen die relativen Häufigkeiten, also die Häufigkeiten dividiert durch die Anzahl von Personen, die  $\leq 10$  Verordnungen von R03BB04 (COPD) haben. Man erkennt einen grossen Unterschied zwischen jungen (links) und älteren (rechts) Personen. Bei jungen Personen kommt R03BB04 (COPD) überhaupt nur sehr selten vor (einmal pro 2000). Die Häufigkeiten von Kombinationen fallen sehr rasch ab. Abbildung 7 zeigt ebenfalls die häufigsten Kombinationen mit R03BB04 (COPD), allerdings mit der Bedingung, dass R03BB04 (COPD)  $\geq 10$  Mal verordnet werden muss, aber die anderen ATC Gruppen nur 5-9 Mal. Im Spitzenfeld (der älteren Personen) findet man aber wieder analoge Kombinationen.

Die weiteren Analysen betreffen den ATC code R03DC03 (Asthma). Wir suchen wiederum nach häufigen Kombinationen von vielen Verordnungen ( $\geq 10$ ) von R03DC03 (Asthma) mit vielen ( $\geq 10$ ) bzw. mittleren (5-9) Verordnungen der anderen ATC Gruppen. Die Abbildungen 8 bzw. 9 zeigen die entsprechenden Ergebnisse. Man findet einen klaren Spitzenreiter, nämlich die Kombination aus R03DC03 (Asthma) und Langwirksame  $\beta$ -2-Mimetika, die allerdings sowohl bei den älteren als auch bei jüngeren Personen

**Unter 30 Jahre: Kombination mit mind. 10 Verordnungen**  
(mind. 10 bei Langw.A., das sind 17 Personen)

5		x				x			29.4%
2		x				x	x		11.8%
1		x	x		x				5.9%
1		x	x		x			x	5.9%
1		x		x		x			5.9%
1		x		x		x	x		5.9%
1		x			x			x	5.9%
1		x				x			5.9%
0	x	x							0.0%
0	x	x	x						0.0%
0	x	x		x					0.0%
0	x	x			x				0.0%
0	x	x				x			0.0%
0	x	x					x		0.0%
0	x	x						x	0.0%

Kurzw.-Anticholingergica  
Langw.-Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Langw.b-2-Mimetika  
Langw.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

**Ueber 40 Jahre: Kombination mit mind. 10 Verordnungen**  
(mind. 10 bei Langw.A., das sind 33802 Personen)

6491		x						x	19.2%
2765		x						x	8.2%
2047		x					x		6.1%
1976		x						x	5.8%
1346		x		x			x	x	4.0%
1135		x		x			x		3.4%
866		x	x		x				2.6%
739		x		x					2.2%
638		x				x		x	1.9%
504		x	x			x		x	1.5%
487		x	x						1.4%
434		x			x			x	1.3%
276		x	x	x	x			x	0.8%
218		x	x					x	0.6%
189		x		x	x			x	0.6%

Kurzw.-Anticholingergica  
Langw.-Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Langw.b-2-Mimetika  
Langw.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

Abbildung 6: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03BB04 (COPD) mit  $\leq 10$  Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit 5-9 Verordnungen**  
(mind. 10 bei Langw.A., das sind 17 Personen)

4		x		x					23.5%
2		x		x				x	11.8%
1		x				x			5.9%
0	x	x							0.0%
0	x	x	x						0.0%
0	x	x		x					0.0%
0	x	x			x				0.0%
0	x	x				x			0.0%
0	x	x					x		0.0%
0	x	x	x	x				x	0.0%
0	x	x	x	x					0.0%
0	x	x	x		x				0.0%
0	x	x	x			x			0.0%
0	x	x	x				x		0.0%
0	x	x	x					x	0.0%

Kurzw.-Anticholingergica  
Langw.-Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Langw.b-2-Mimetika  
Langw.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

**Ueber 40 Jahre: Kombination mit 5-9 Verordnungen**  
(mind. 10 bei Langw.A., das sind 33802 Personen)

2795		x						x	8.3%
2638		x		x					7.8%
1414		x						x	4.2%
1270		x				x			3.8%
890		x	x						2.6%
376		x		x			x		1.1%
293		x					x	x	0.9%
247		x		x				x	0.7%
236		x	x			x			0.7%
188		x				x	x		0.6%
150		x						x	0.4%
148		x		x	x				0.4%
144		x	x	x					0.4%
118	x	x							0.3%
94		x				x		x	0.3%

Kurzw.-Anticholingergica  
Langw.-Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Langw.b-2-Mimetika  
Langw.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

Abbildung 7: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03BB04 (COPD) mit 5-9 Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

dominant ist.



**Unter 30 Jahre: Kombination mit mind. 10 Verordnungen**  
(mind. 10 bei Leok.A., das sind 3497 Personen)

473					x		x	13.5%
196		x					x	5.6%
122			x				x	3.5%
65		x	x				x	1.9%
55			x		x		x	1.6%
41				x			x	1.2%
13		x		x			x	0.4%
6		x	x	x			x	0.2%
5			x	x			x	0.1%
5					x	x	x	0.1%
5						x	x	0.1%
4			x		x	x	x	0.1%
3	x						x	0.1%
3		x			x		x	0.1%
3			x			x	x	0.1%
3					x		x	0.1%

Kurzw.Anticholingergica  
Langw.Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Lang.w.b-2-Mimetika  
Lang.w.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

**Ueber 40 Jahre: Kombination mit mind. 10 Verordnungen**  
(mind. 10 bei Leok.A., das sind 6486 Personen)

1278							x	19.7%
293							x	4.5%
267							x	4.1%
259				x			x	4.0%
207					x		x	3.2%
194		x			x		x	3.0%
173			x				x	2.7%
144		x					x	2.2%
141				x		x	x	2.2%
131	x					x	x	2.0%
109		x					x	1.7%
101				x			x	1.6%
100	x					x	x	1.5%
83		x	x			x	x	1.3%
66			x		x		x	1.0%

Kurzw.Anticholingergica  
Langw.Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Lang.w.b-2-Mimetika  
Lang.w.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

Abbildung 8: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit  $\leq 10$  Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit 5-9 Verordnungen**  
(mind. 10 bei Leok.A., das sind 3497 Personen)

376					x		x	10.8%
274			x				x	7.8%
245		x					x	7.0%
131		x	x				x	3.7%
63			x		x		x	1.8%
30		x			x		x	0.9%
17				x			x	0.5%
11		x	x		x		x	0.3%
8		x		x			x	0.2%
7						x	x	0.2%
6			x	x			x	0.2%
6				x	x		x	0.2%
2		x		x	x		x	0.1%
2			x	x	x		x	0.1%
1	x						x	0.0%

Kurzw.Anticholingergica  
Langw.Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Lang.w.b-2-Mimetika  
Lang.w.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

**Ueber 40 Jahre: Kombination mit 5-9 Verordnungen**  
(mind. 10 bei Leok.A., das sind 6486 Personen)

637							x	9.8%
433			x				x	6.7%
223			x				x	3.4%
204							x	3.1%
176					x		x	2.7%
122		x					x	1.9%
64				x		x	x	1.0%
63			x		x		x	1.0%
46			x	x			x	0.7%
33						x	x	0.5%
30	x						x	0.5%
29		x				x	x	0.4%
27			x			x	x	0.4%
26				x	x		x	0.4%
25				x			x	0.4%

Kurzw.Anticholingergica  
Langw.Anticholingergica  
Inhal.Corticoide  
Kurz.w.b-2-Mimetika  
Lang.w.b-2-Mimetika  
Lang.w.b-2-Mim+inhal.Cort  
Theophylline  
Leokotrienrez.Antag.

Abbildung 9: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit 5-9 Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

## 4 Berücksichtigung der ICD codes

Bei den Daten der Krankenanstalten wurden COPD und Asthma auch als Haupt- oder Nebendiagnose kodiert. Die ICD10-codes sind J44 (COPD) und J45 (Asthma). Als weitere Vorgangsweise kann man nun die Population erweitern um jene Personen, bei denen J44 oder J45 aufscheint. Diese Personen können Medikamente aus den betrachteten Gruppen erhalten haben, dies ist aber nicht zwingend. Jedenfalls kann diese Information der Diagnose verwendet werden, um Rückschlüsse auf die Medikation zu erhalten, und Zusammenhänge darzustellen.

Im Vergleich zur früheren Population kommen in der aktuellen Population 15174 Personen dazu, die mindestens einen Eintrag bei J44 oder J45 haben, aber keinen Eintrag bei einem der gewählten ATC codes. Die umgekehrte Variante, kein Eintrag bei J44 und J45, aber mindestens ein Eintrag bei den gewählten ATC codes, ergibt 713472 Treffer. In Tabelle 1 findet man die genaue Aufstellung, auch getrennt für Personen über 40 Jahren. Man erkennt aus Tabelle 1, dass der Großteil der Personen, die keinen der ATC codes

Tabelle 1: Anzahl der Verordnungen (Diagnosen) für die ATC Gruppen (ICD codes) bzw. Anzahl der betroffenen Personen.

Personengruppe	Filter	Anzahl der Personen
Population	kein ATC, mind. ein J44 oder J45	15174
Population >40 Jahre	kein ATC, mind. ein J44 oder J45	13310
Population	mind. ein ATC, weder J44 noch J45	713472
Population >40 Jahre	mind. ein ATC, weder J44 noch J45	425019

aber zumindest eine Diagnose J44 bzw. J45 haben, älter als 40 Jahre ist. Im umgekehrten Fall ist die Altersbedingung nicht relevant.

Die genauen Häufigkeiten für die einzelnen codes bzw. Gruppen sind aus Tabelle 2 ersichtlich.

Tabelle 2: Anzahl der Verordnungen (Diagnosen) für die ATC Gruppen (ICD codes) bzw. Anzahl der betroffenen Personen.

Gruppe	Anzahl Verordnungen/Diagnosen	Anzahl Personen
Kurzw.Anticholingergica	64091	14517
Langw.Anticholingergica	774279	96491
Inhal.Corticoide	687814	254844
Kurzw.b-2-Mimetika	1282708	401042
Langw.b-2-Mimetika	445101	71597
Langw.b-2-Mim+inhal.Cort	1287448	266786
Theophylline	751083	98710
Leukotrienrez.Antag.	355812	92364
J44	135430	65827
J45	22490	16455

Auf den folgenden beiden Seiten werden die Abbildungen 6–9 um zwei Spalten erweitert, nämlich um die Anzahlen der Diagnosen für J44 und J45 für die jeweiligen Kombinationen. Die erweiterten Tabellen sind dargestellt in Abbildung 10–13. Es ist dabei zu

bedenken, dass J44 in der Population wesentlich häufiger vorkommt als J45, siehe Tabelle 2.

Die **Resultate** von Abbildung 10–13 kann man folgendermaßen zusammenfassen:

- *Personen unter 30 Jahre:*

- Bei den typischen COPD Kombinationen gibt es kaum Personen bzw. Diagnosen, weder zu J44, noch zu J45.
- Bei den typischen Asthma Kombinationen tritt J45 viel öfter auf als J44 (obwohl J44 in der Population viel häufiger vorkommt). Somit kann man die häufigsten Asthma Kombinationen bei jungen Menschen ziemlich klar Asthma zuordnen.

- *Personen über 40 Jahre:*

- Bei den typischen COPD Kombinationen kommt J44 viel öfter (auch relativ gesehen) als J45 vor.
- Bei den typischen Asthma Kombinationen sind die Häufigkeiten von J44 und J45 vergleichbar. Bedenkt man aber, dass J44 in der Population fast 6 mal so häufig ist wie J45, sprechen diese Kombinationen klar für Asthma.



**Unter 30 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Leok.A., das sind 3497 Personen)**

473						x			x	2	48	13.5%
196		x							x	9	42	5.6%
122			x						x	3	29	3.5%
65		x	x						x	1	10	1.9%
55			x						x	1	18	1.6%
41				x					x	0	2	1.2%
13		x		x					x	0	0	0.4%
6		x	x	x					x	0	0	0.2%
5			x	x					x	0	2	0.1%
5						x	x		x	0	0	0.1%
5						x	x		x	0	6	0.1%
4				x			x	x	x	0	0	0.1%
3	x								x	0	1	0.1%
3			x						x	0	0	0.1%
3			x				x		x	0	3	0.1%

Kurzw.Anticholingergica  
 Langw.Anticholingergica  
 Inhal.Corticoide  
 Kurzw.b-2-Mimetika  
 Langw.b-2-Mimetika  
 Langw.b-2-Mim+inhal.Cort  
 Theophylline  
 Leukotrienrez.Antag.  
 J44  
 J45

**Ueber 40 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Leok.A., das sind 6486 Personen)**

1278									x		x	174	272	19.7%
293									x	x	x	103	79	4.5%
267									x	x	x	141	62	4.1%
259						x			x		x	34	49	4.0%
207									x			21	20	3.2%
194				x					x			31	33	3.0%
173						x						27	30	2.7%
144						x						41	14	2.2%
141						x			x	x		48	47	2.2%
131									x			74	23	2.0%
109						x			x			61	9	1.7%
101									x	x		50	48	1.6%
100						x			x	x	x	91	14	1.5%
83						x			x	x	x	109	30	1.3%
66						x			x		x	22	20	1.0%

Kurzw.Anticholingergica  
 Langw.Anticholingergica  
 Inhal.Corticoide  
 Kurzw.b-2-Mimetika  
 Langw.b-2-Mimetika  
 Langw.b-2-Mim+inhal.Cort  
 Theophylline  
 Leukotrienrez.Antag.  
 J44  
 J45

Abbildung 12: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit  $\leq 10$  Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit 5-9 Verordnungen  
(mind. 10 bei Leok.A., das sind 3497 Personen)**

376									x		x	2	27	10.8%
274									x	10	40	7.8%		
245									x	7	32	7.0%		
131									x	7	26	3.7%		
63									x	1	24	1.8%		
30									x	1	6	0.9%		
17									x	2	5	0.5%		
11									x	1	4	0.3%		
8									x	1	1	0.2%		
7									x	0	9	0.2%		
6									x	0	0	0.2%		
6									x	0	0	0.2%		
2									x	0	2	0.1%		
2									x	0	0	0.1%		
1	x								x	0	0	0.0%		

Kurzw.Anticholingergica  
 Langw.Anticholingergica  
 Inhal.Corticoide  
 Kurzw.b-2-Mimetika  
 Langw.b-2-Mimetika  
 Langw.b-2-Mim+inhal.Cort  
 Theophylline  
 Leukotrienrez.Antag.  
 J44  
 J45

**Ueber 40 Jahre: Kombination mit 5-9 Verordnungen  
(mind. 10 bei Leok.A., das sind 6486 Personen)**

637											x	169	112	9.8%
433									x		x	151	62	6.7%
223									x		x	59	50	3.4%
204											x	108	42	3.1%
176											x	52	20	2.7%
122									x		x	100	28	1.9%
64									x		x	19	19	1.0%
63									x		x	14	17	1.0%
46									x		x	14	5	0.7%
33											x	13	6	0.5%
30											x	23	10	0.5%
29									x		x	22	9	0.4%
27									x		x	3	14	0.4%
26									x		x	17	8	0.4%
25									x		x	30	6	0.4%

Kurzw.Anticholingergica  
 Langw.Anticholingergica  
 Inhal.Corticoide  
 Kurzw.b-2-Mimetika  
 Langw.b-2-Mimetika  
 Langw.b-2-Mim+inhal.Cort  
 Theophylline  
 Leukotrienrez.Antag.  
 J44  
 J45

Abbildung 13: Häufigste Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit 5-9 Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

## 5 Zusammenfassung von Medikamenten-Gruppen

Aus inhaltlicher Sicht macht es Sinn, die betrachteten Gruppen von Medikamenten weiter zusammenzufassen. Dies wird zu einer erheblichen Reduktion der Anzahl von Kombinationen, so wie sie im letzten Abschnitt betrachtet wurden, beitragen. Nach Rücksprache mit dem HVB werden folgende beiden Gruppen betrachtet, die mit den bisherigen Kenntnissen indikativ für COPD bzw. für Asthma sind:

### **COPD-Gruppe:**

Inhalative Corticoide: R03BA01, R03BA02, R03BA05

Langwirksame  $\beta$ -2-Mimetika: R03AC12, R03AC13, R03CC12, R03CC13

Kombination langwirksame  $\beta$ -2-Mimetika + inhalative Corticoide: R03AK06, R03AK07

### **Asthma-Gruppe:**

Theophylline: R03DA04

Wie im letzten Abschnitt werden diese Gruppen mit

**Langwirksame Anticholinergica: R03BB04** (typisch verordnet für COPD)  
bzw. mit

**Leukotrienrezeptor-Antagonisten: R03DC03** (typisch verordnet für Asthma) kombiniert. Die Resultate sind in Abbildungen 14–17 auf analoge Weise wie früher dargestellt. Man bemerke, dass nun *alle* 7 Kombinationen dargestellt sind. Die Ergebnisse können folgendermaßen zusammengefasst werden:

- *Personen unter 30 Jahre:*
  - Wie früher, gibt es für COPD kaum Einträge.
  - Bei den typischen Asthma Kombinationen tritt J45 viel öfter auf als J44 (obwohl J44 in der Population viel häufiger vorkommt). Allerdings trifft dies nur für die jeweils 1. Kombination mit der COPD-Gruppe zu, nicht aber bei der Kombination mit der Asthma-Gruppe. Die Ergebnisse sind somit für unseren Zweck nicht brauchbar.
- *Personen über 40 Jahre:*
  - Bei den typischen COPD Kombinationen kommt J44 viel öfter (auch relativ gesehen) als J45 vor. Die jeweils 1. Kombination ist auch die “richtige”, die mit der COPD-Gruppe kombiniert. Diese Kombination erscheint sehr sinnvoll.
  - Bei den typischen Asthma Kombinationen sind die Häufigkeiten von J44 und J45 vergleichbar. Bedenkt man aber, dass J44 in der Population fast 6 mal so häufig ist wie J45, sprechen diese Kombinationen klar für Asthma. Allerdings ist in Abbildung 16 (rechts) nur die 3. Kombination von oben jene mit der Asthma-Gruppe, und in Abbildung 17 (rechts) die 2. Kombination mit der Asthma-Gruppe. Diese Ergebnisse werden daher kaum hilfreich sein.

**Unter 30 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Langw.A., das sind 17 Personen)**

8	x		x		0	0	47.1%
3	x		x	x	0	0	17.6%
2	x	x	x		0	1	11.8%
0	x	x			0	0	0.0%
0	x	x		x	0	0	0.0%
0	x	x	x	x	0	0	0.0%
0	x			x	0	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

**Ueber 40 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Langw.A., das sind 33802 Personen)**

12640	x		x		10704	515	37.4%
6711	x		x	x	10573	389	19.9%
2128	x			x	2785	78	6.3%
304	x	x	x		201	48	0.9%
303	x	x	x	x	384	81	0.9%
112	x	x			69	8	0.3%
50	x	x		x	42	13	0.1%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

Abbildung 14: Alle Kombinationen von  $\leq 10$  Verordnungen von R03BB04 (COPD) mit  $\leq 10$  Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit 5–9 Verordnungen  
(mind. 10 bei Langw.A., das sind 17 Personen)**

2	x	x			0	0	11.8%
1	x		x		0	0	5.9%
0	x	x	x		0	0	0.0%
0	x	x		x	0	0	0.0%
0	x	x	x	x	0	0	0.0%
0	x		x	x	0	0	0.0%
0	x			x	0	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

**Ueber 40 Jahre: Kombination mit 5–9 Verordnungen  
(mind. 10 bei Langw.A., das sind 33802 Personen)**

4060	x		x		3844	147	12.0%
1906	x			x	3007	96	5.6%
399	x		x	x	685	59	1.2%
218	x	x			337	21	0.6%
33	x	x		x	27	4	0.1%
28	x	x	x		20	2	0.1%
2	x	x	x	x	1	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

Abbildung 15: Alle Kombinationen von  $\leq 10$  Verordnungen von R03BB04 (COPD) mit 5-9 Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Leok.A., das sind 3497 Personen)**

992		x	x		17	147	28.4%
17		x	x	x	0	6	0.5%
4		x		x	0	4	0.1%
2	x	x	x		0	1	0.1%
0	x	x			0	0	0.0%
0	x	x		x	0	0	0.0%
0	x	x	x	x	0	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

**Ueber 40 Jahre: Kombination mit mind. 10 Verordnungen  
(mind. 10 bei Leok.A., das sind 6486 Personen)**

2439		x	x		398	459	37.6%
779		x	x	x	291	232	12.0%
329		x		x	176	95	5.1%
304	x	x	x		201	48	4.7%
303	x	x	x	x	384	81	4.7%
112	x	x			69	8	1.7%
50	x	x		x	42	13	0.8%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

Abbildung 16: Alle Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit  $\leq 10$  Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.

**Unter 30 Jahre: Kombination mit 5-9 Verordnungen  
(mind. 10 bei Leok.A., das sind 3497 Personen)**

825		x	x		18	100	23.6%
6		x		x	0	8	0.2%
2		x	x	x	0	1	0.1%
1	x	x	x		0	0	0.0%
0	x	x			0	0	0.0%
0	x	x		x	0	0	0.0%
0	x	x	x	x	0	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

**Ueber 40 Jahre: Kombination mit 5-9 Verordnungen  
(mind. 10 bei Leok.A., das sind 6486 Personen)**

876		x	x		212	140	13.5%
262		x		x	168	53	4.0%
170	x	x			143	52	2.6%
48		x	x	x	19	7	0.7%
34	x	x	x		78	15	0.5%
15	x	x		x	13	0	0.2%
1	x	x	x	x	0	0	0.0%
	Langw.Anticholingergica	Leukotrienrez.Antag.	COPD-Gruppe	Asthma-Gruppe	J44	J45	

Abbildung 17: Alle Kombinationen von  $\leq 10$  Verordnungen von R03DC03 (Asthma) mit 5-9 Verordnungen der anderen ATC Gruppen. Links Personen  $< 30$  Jahre, rechts Personen  $\geq 40$  Jahre.



## 6 Isolation von COPD

Im letzten Abschnitt konnte in Abbildung 14 (rechts) bzw. 15 (rechts) die jeweils 1. Kombination als mögliche Kombination zur Isolation für COPD ausfindig gemacht werden. Wir konzentrieren uns daher in Folge nur mehr auf diese Kombination, die von der **Population mit Alter über 40 Jahre** kommt. Diese Kombination wird deshalb sinnvoll sein, weil sie aus typischen Medikamenten besteht, die bei COPD verabreicht werden, und weil sehr klar die Diagnose J44 bei den betroffenen Personen auftritt. Diese Kombination für COPD enthält somit:

### COPD:

Langwirksame Anticholinergica: R03BB04

Inhalative Corticoide: R03BA01, R03BA02, R03BA05

Langwirksame  $\beta$ -2-Mimetika: R03AC12, R03AC13, R03CC12, R03CC13

Kombination langwirksame  $\beta$ -2-Mimetika + inhalative Corticoide: R03AK06, R03AK07

Dem könnte man ein typisches Asthma Medikament gegenüberstellen, also:

### Asthma:

Leukotrienrezeptor-Antagonisten: R03DC03

Abbildung 18 zeigt die Gegenüberstellungen dieser beiden ATC Gruppen. Jeder Punkt entspricht einer Person, und die x- und y-Koordinate in den Plots gibt die Summe der ATC codes für alle COPD bzw. für alle Asthma ATCs an (hier nur ein ATC Code). Dies ist sowohl links als auch rechts gezeigt. Die Farben hingegen markieren Personen, bei denen zumindest eine Diagnose J44 (links, rot) bzw. J45 (rechts, grün) vergeben wurde. Nicht-färbige (schwarze) Punkte entsprechen also Personen, die keine J44 oder J45 Diagnose im betrachteten Zeitraum erhalten haben, was aber nicht heißen muss, dass sie dennoch COPD- oder Asthma-Patienten sind.

Man erkennt in Abbildung 18 einen klaren Unterschied: Im Vergleich zu “rot” nimmt “grün” im Plot wesentlich weniger Platz ein, vor allem bezüglich der x-Koordinate. Anders ausgedrückt ist die Diagnose J45 (Asthma) sehr untypisch bei hohen Anzahlen von ATC codes für COPD. Man könnte also eine vertikale Schranke finden, die “grün” nach rechts hin begrenzt. Personen mit größeren x-Werten wären somit mit hoher Wahrscheinlichkeit COPD-Patienten, nicht aber Asthma-Patienten. Dieser Ansatz wird nachfolgend gemacht. Man könnte eventuell auch eine komplexere Funktion als eine vertikale Gerade finden, um die “grünen” Punkte einzugrenzen. Im Sinne der einfacheren Interpretierbarkeit wird dies aber nicht vorgenommen.

Man bemerke, dass der umgekehrte Schluss nicht möglich wäre: Im linken Plot ist “rot” nach oben hin begrenzt. Allerdings ist im rechten Plot auch “grün” nach oben begrenzt. Das heißt, man kann keine y-Schranke finden, die Asthma isolieren könnte.

Für eine Bestimmung eines **cut-off Wertes** müssen folgende Dinge bedacht werden:

- Der cut-off Wert soll einer einfachen Regel folgen. Wir werden folgende Regel wählen: *Wenn die Summe der für COPD typischen ATC codes unseren zu bestimmenden cut-off Wert überschreitet, wird es wahrscheinlicher, dass die Person COPD als Asthma hat.*

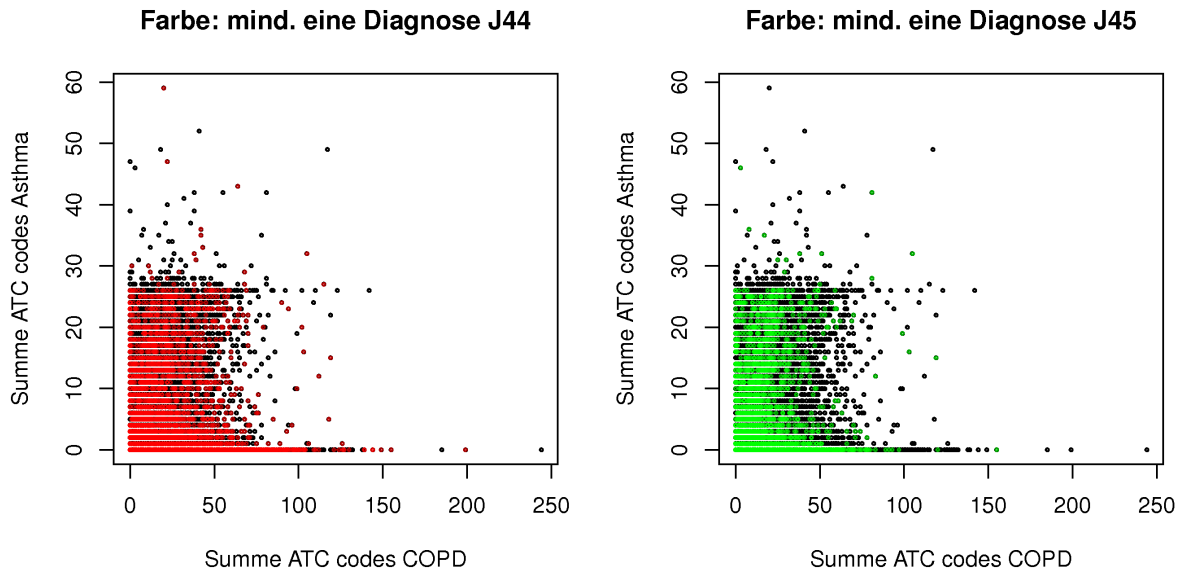


Abbildung 18: Summe der typischen ATC codes für COPD (horizontal) bzw. für Asthma (vertikal) pro Person. Die Farben entsprechen mindestens einer Diagnose J44 (links, rot) bzw. J45 (rechts, grün).

- Die Anzahl der Diagnosen J44 ist in der Population wesentlich höher als jene für J45. Wenn diese Anzahlen pro Person direkt verglichen werden, muss man mit der Gesamtsumme der codes normieren.

### Vorgangsweise:

- (1) Schränke die Population auf jene Personen ein, die mehr als 40 Jahre alt sind, und für die zumindest eine Diagnose J44 oder J45 gemacht wurde (70604 Personen).
- (2) Bezeichne mit  $x_c$  einen cut-off Wert bezogen auf die x-Achse von Abbildung 18, also auf die Summe der für COPD typischen ATC codes. Es wird im Folgenden  $x_c$  schrittweise im Intervall  $[1, 70]$  variiert. Mit  $x$  bezeichnen wir einen beliebigen Wert auf dieser Achse.
- (3) Zähle die Anzahl der Diagnosen J44 bzw. J45 von jenen Personen, für die  $x \geq x_c$  ist, und normiere mit der Gesamtanzahl von Diagnosen J44 bzw. J45.
- (4) Zähle die Anzahl der Personen mit Diagnose J44 bzw. J45, eingeschränkt auf jene Personen, für die  $x \geq x_c$  ist, und normiere mit der Gesamtanzahl von Personen mit Diagnose J44 bzw. J45.

Die resultierenden Zahlen nehmen natürlich mit größer werdendem Wert  $x_c$  rasch ab, weil nicht mehr viele Personen das Kriterium erfüllen. Uns interessieren allerdings nur relative Unterschiede bezüglich J44 und J45.

Abbildung 19 zeigt die Resultate, links für die Diagnosen laut Punkt (3), rechts für die Personen laut Punkt (4). Auf der x-Achse findet man wie gewohnt die Summe der ATC codes, die für COPD typisch sind, hier liegt auch der potentielle cut-off Wert  $x_c$ . Auf der y-Achse sind die relativen Unterschiede für die Häufigkeit der vorkommenden Diagnosen

J44 bzw. J45. In der linken Grafik erkennt man, dass für alle Grenzwerte  $x_c$  größer als 10 der code J44 in Relation häufiger als J45 vorkommt. Ab einem Wert von etwa 60 scheint der Unterschied wieder geringer zu werden, was aber nur auf die Instabilität der kleiner werdenden Anzahl von Personen zurückzuführen ist. Die Aussagen bezogen auf die Personen (rechts) ist ähnlich, hier überwiegt aber J44 erst bei einem etwas höheren Grenzwert.

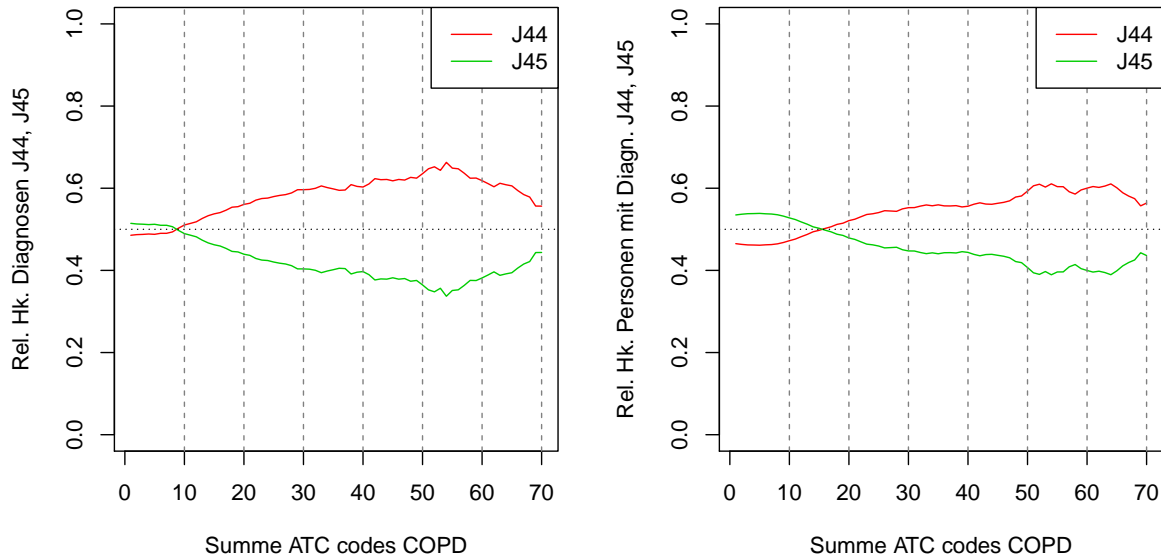


Abbildung 19: Relative Häufigkeit von Diagnosen J44/J45 (links) bzw. Personen mit Diagnosen J44/J45 (rechts) von Personen mit einer Mindestsumme des x-Wertes.

Obige Vorgangsweise könnte eventuell verbessert werden, indem man bei jeder Person betrachtet, ob J44 und J45 gleichzeitig auftreten, bzw. wie unterschiedlich diese sind. Wir gehen also zunächst wie oben vor, modifizieren aber die Punkte (3) und (4):

- (3\*) Normiere die Anzahlen von Diagnosen J44 bzw. J45 mit den jeweiligen Gesamtanzahlen von Diagnosen unserer Population. Zähle, wie oft die normierten Diagnosen J44 größer sind als jene von J45, und zwar von jenen Personen, für die  $x \geq x_c$  ist.
- (4\*) Normiere die Anzahlen von Personen mit Diagnose J44 bzw. J45 mit den jeweiligen Gesamtanzahlen von Personen unserer Population. Zähle, wie oft die normierten Anzahlen für J44 größer sind als jene von J45, und zwar von jenen Personen, für die  $x \geq x_c$  ist.

Die Resultate werden in [Abbildung 20](#) gezeigt, links für Diagnosen, rechts für Personen mit den Diagnosen. Trotz unterschiedlicher Vorgangsweise gleichen sich die Bilder. Ein Grenzwert bezogen auf die Diagnose-Häufigkeiten erscheint sicherer und besser als einer bezogen auf die Personen-Häufigkeiten. Auch dies ist plausibel, weil für Personen-Häufigkeiten nur 0/1 Information eingeht, ob die Diagnose auftritt, bei Diagnose-Häufigkeiten aber die tatsächlichen Anzahlen. Ein Grenzwert von  $x_c = 20$  könnte sinnvoll sein.

Eine abschließende Fragestellung wird sein, wie viele Personen nun insgesamt an COPD leiden. Wir schränken dies auf Personen ein mit Alter höher als 40 Jahre. [Abbildung 21](#) zeigt den Zusammenhang mit der Wahl des Grenzwertes (der sich auf die Summe der für COPD typischen ATC codes bezieht), und der Anzahl von Personen  $>40$

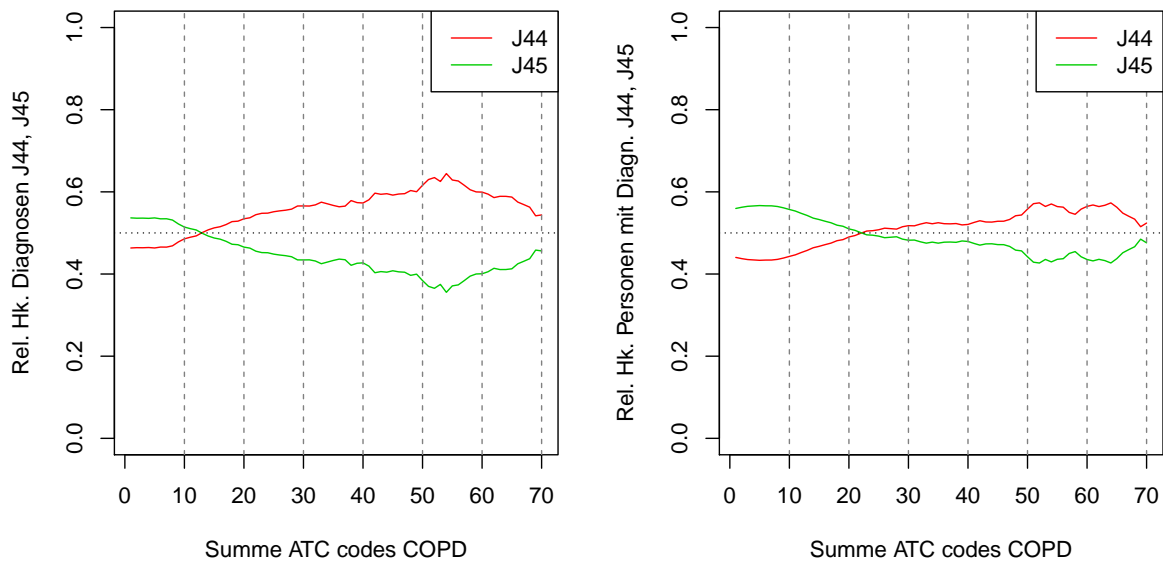


Abbildung 20: Relative Häufigkeit von Diagnosen J44>J45 (links) bzw. Personen mit Diagnosen J44>J45 (rechts) von Personen mit einer Mindestsumme des x-Wertes.

Jahre. Bei einer Wahl des Grenzwertes von  $x_c = 20$  würde man **40000 Personen als COPD-Patienten** ausweisen.

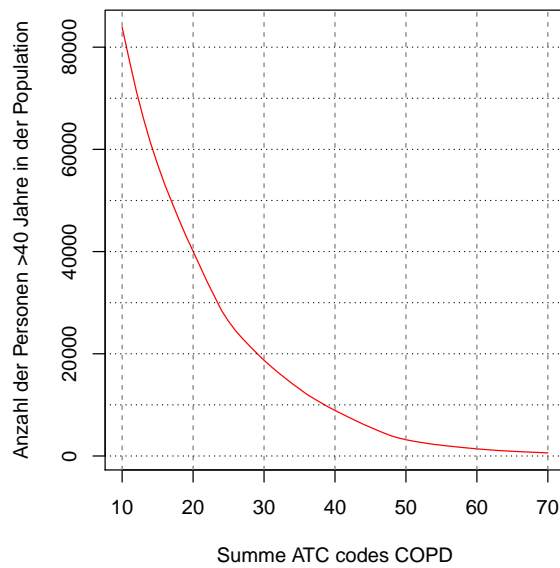


Abbildung 21: Anzahl von Personen >40 Jahre gegen die Summe der für COPD typischen ATC codes.

## 7 Zusammenfassung der Ergebnisse

Die wichtigste Aussage dieser Studie ist, dass COPD isoliert werden kann, nicht aber Asthma. Für die Isolation von COPD kann ein Grenzwert angegeben werden, der nicht die Information der Diagnosen J44 oder J45 benötigt, sondern die Information der Summe der ATC codes

R03BB04,  
R03BA01, R03BA02, R03BA05,  
R03AC12, R03AC13 R03CC12, R03CC13,  
R03AK06, R03AK07

für die betrachtete Person. Liegt diese Summe im 2-Jahreszeitraum über 20 (im Jahreszeitraum über 10), dann ist das ein starkes Indiz für COPD. Aus den Daten für 2006 und 2007 konnten auf diese Weise 40000 Personen identifiziert werden. Diese Schranke bzw. diese Aussagen sind nur gültig für Personen über 40 Jahre.

Die getroffenen Aussagen sind abhängig von den zugrunde liegenden Daten, also von der Qualität der ATC Kodierung, und vor allem von der Zuverlässigkeit der ICD codes J44 und J45. Bei vielen Personen treten beide codes auf, was eine Isolation von COPD erschwert. Der gewonnene Grenzwert ist bewusst relativ hoch gewählt, sodass die Wahrscheinlichkeit, eine Person mit COPD zu detektieren, deutlich höher ist als die Wahrscheinlichkeit, dass diese Person Asthma hat. Würde der Grenzwert verringert werden, dann könnten zwar mehr Personen isoliert werden, aber das Risiko von "false positives", also Personen mit Asthma, wird höher.