

1 Einleitung und deskriptive Datenanalyse

In der Sozialversicherung treffen zwei grundlegend verschiedene Ansichten aufeinander. Auf der einen Seite stehen Patienten und Ärzte für die bestmögliche Behandlung von Krankheiten im Zentrum. Der Kostenfaktor spielt durch den praktisch nicht vorhandenen Selbstbehalt keine Rolle, natürlich unter der Annahme, dass nur Sozialversicherungsleistungen konsumiert werden. Für das Management der Sozialversicherung ist hingegen der Weg der Behandlung nicht relevant, sondern rein die Finanzierung. Da das Gesamtbudget im Allgemeinen nie ausreichen kann, um jedem Patienten eine optimale Behandlung zu bieten, muss ein Gleichgewicht zwischen den Kosten und der Effektivität einer Behandlung gefunden werden.

Aus diesem Grund ist es sinnvoll, die Variablen der Behandlung und die Kostenvariablen mit Hilfe von Regressionsrechnungen in einen Zusammenhang zu bringen. Im einfachsten Fall bedeutet dies, die Leistungen mit ihren Preisen zu gewichten, um die Kosten eindeutig bestimmen zu können. Dadurch ist es auch möglich, den verwendeten Trainingsdatensatz zu erstellen.

Es gibt jedoch mehrere Probleme. Die Preise von Leistungen sind nicht konstant. Sie ändern sich über die Zeit bzw. werden durch Deckelungen beeinflusst. Darüber hinaus sind die ausgewählten Leistungen nicht konstant. Kleine Wechsel von Wirkstoffstärken bei Medikamenten oder ein Wechsel von einem Arztbesuch zu einem Hausbesuch würden zu einer praktisch unüberschaubaren Menge an beschreibenden Variablen führen, die insbesondere in einem starken Zusammenhang stehen. Auch wenn die Berechnung eines solchen Modells keine Probleme bereitet, so können gewonnene Rückschlüsse kaum interpretiert werden. Daher macht es Sinn, diese verwandten Variablen zusammenzufassen und einen Fehler in der Beschreibung zu akzeptieren. Darüber hinaus ermöglicht der Fehler innerhalb geeignet gewählter Gruppen Rückschlüsse über die Konsistenz der Behandlungen der gewählten Gruppe.

Ziel dieser Arbeit ist es, zu prüfen, ob einfache Regressionsmodelle allgemeine Aussagen über Zusammenhänge zulassen. Insbesondere soll aus Basis des Beispiels von Verordnungen von Antipsychotika zu prüfen, ob einfache, in der Sozialversicherung häufig verwendete Regressionsmodelle überhaupt zulässig sind, bzw. welche Adaptionen vorgenommen werden müssen und welche Risiken im Rahmen der Berechnungen existieren.

1.1 Datenstruktur

Insgesamt drei österreichische Gebietskrankenkassen aus dem Burgenland, Kärnten und Salzburg haben sich bereit erklärt, Daten zum Thema Antipsychotika für Analysen zur Verfügung zu stellen. Die zu analysierenden Bewegungsdaten stammen aus den hauseigenen FOKO Datenbanken. Die gelieferten Bewegungsdatensätze beinhalten Arztkosten, Heilmittelverordnungen und Krankenhausaufenthalte.

Arztkosten:	Krankenhausaufenthalte:	Heilmittelverordnungen:
<ul style="list-style-type: none"> • Patient • Quartal • Arzt • Fachgruppe Arzt • Arztbezirk • Anzahl an Konsultationen • erzeugte Kosten 	<ul style="list-style-type: none"> • Patient • Leistungsdatum • Krankenhausdiag. (ICD-10) • Aufenthaltstage 	<ul style="list-style-type: none"> • Patient • Arzt • Fach • Verordnungsjahr • Rezeptmonat • Einreichdatum des Rezeptes • Verordnungen • Kosten • Pharmanummer

Die Ausprägungen Patient und Arzt wurden mit einem md5 Schlüssel mit trägerspezifischem Passwort verschlüsselt. Der Bezirk wurde ebenfalls mit einer Laufnummer versehen anstelle des echten Bezirks. Diese Bewegungsdaten stehen für die Jahre 2007 bis 2010 zur Verfügung. Diese Daten stehen für zwei Gruppen zur Verfügung. Eine Gruppe beinhaltet sämtliche Patienten die zwischen 2007 und 2010 Antipsychotika erhalten haben. Die zweite Gruppe ist als Kontrollgruppe gedacht und ist eine Stichprobe von 24 893 Patienten. Da diese Analysen keine statistischen Test zu Prävalenzen beinhalten, wurde die Kontrollgruppe nicht in die Untersuchungen inkludiert. Angereichert wurden die Bewegungsdaten mit Stammdaten zu Patienten, Versicherungsträgern und Pharmanummern.

Patientenstammdaten:	Versicherungsträger (Anspruchsberechtigte):	DEKO:
<ul style="list-style-type: none"> • Patient • Geschlecht • Geburtsjahr • Todesdatum • Bezirk des Hauptsitzes 	<ul style="list-style-type: none"> • Versicherungs- träger • Jahr • Geschlecht • Alter • Anspruchs- berechtigte 	Beinhaltet alle gängigen Ausprägungen und insbesondere DDDs.

1.2 Patientenstruktur

Um die Patientenstruktur genauer analysieren zu können, wurde eine Tabelle angelegt, die sämtliche Antipsychotikapatienten sowie einige charakterisierende Ausprägungen beinhaltet. Diese Struktur stellt gleichzeitig, getrennt nach spezifischen Unterkriterien, auf die erst später eingegangen werden soll, die Datenbasis für die Anwendung der multivariaten Analyse dar.

- Patient
- Geburtsjahr
- Kosten Allgemeinmediziner
- Kosten Facharzt
- Kosten Antipsychotika
- Verordnungen Antipsychotika
- Kosten Heilmittel Sonstiges
- Verordnungen Heilmittel Sonstiges
- Verordnungszeitraum
- Anzahl Pharmanummern Antipsychotika
- Anzahl Wirkstoffe Antipsychotika

Ausprägungen wie die Kosten für Allgemeinmediziner weisen extreme Ausreißer auf, die genauer betrachtet werden müssen. Eine kurze Zusammenfassung der Werte zeigt den großen Abstand zwischen den normalen Bereichen und dem maximalen Wert. Diese Zusammenfassung wird in Tabelle 1.1 für die Kosten für Allgemeinmediziner dargestellt.

Tabelle 1.1: Zusammenfassung von Kosten für Allgemeinmediziner

Min.	1st Qu.	Median	Mean	3rd Qu.	Max
1.53	947.2	1827	2333	3096	137800

Vermutlich handelt es sich hier um Sammelnummern.

1.3 Verordnungszeitraum

Der Verordnungszeitraum ist definiert als die Anzahl an Tagen zwischen der ersten und der letzten Antipsychotikaverordnung eines Patienten im Zeitraum zwischen 2007 und 2010, berechnet aus der Ausprägung *eindat*, dem Einreichdatum des Rezeptes. Er ist eines der wichtigsten Charakteristika der Antipsychotikapatienten und gibt Aufschluss über die Therapietreue der Patienten. Bei perfekter Versorgung der Patienten würden in jedem

Zeitintervall die gleiche Menge neuer Patienten hinzukommen (unter der Annahme dass sich die Altersverteilung nicht über die Zeit ändert) und bis zu ihrer Heilung bzw. ihrem Tod im System verbleiben. Betrachtet man in einem derartigen System die Verteilung der Patienten auf den Verordnungszeitraum, so würde es um den maximal möglichen Verordnungszeitraum eine Anhäufung von Patienten geben, die bereits vor 2007 diagnostiziert wurden. Über den restlichen Zeitraum wären die Patienten gleichverteilt, da nach Annahme zu jedem Zeitpunkt gleich viele Patienten diagnostiziert werden. Die reale Verteilung ist in Abbildung 1.1 dargestellt.

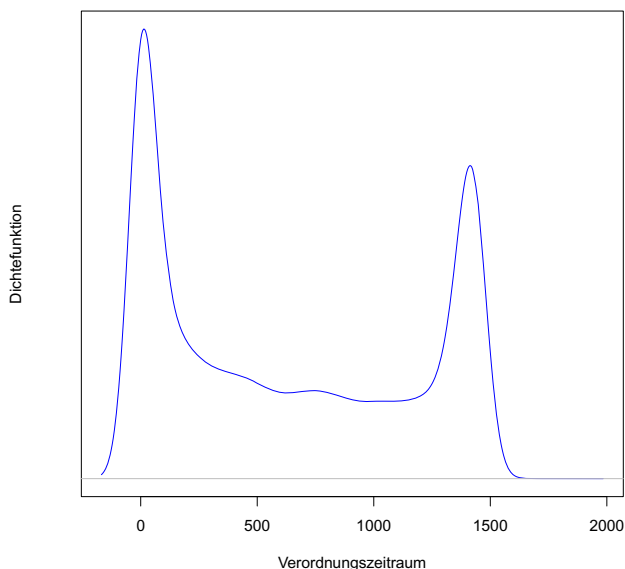


Abbildung 1.1: Verteilung des Verordnungszeitraumes

Die Ausreißer ganz rechts sind wahrscheinlich durch Fehler in der Datierung der Verordnungen zu erklären. In der gelieferten Datei "BGKK APHM.txt", etwa in Zeile 74 661, findet sich ein Patient mit Einreichdatum 1.1.2006 und zugehörigem Jahr 2008. Die gleiche Situation tritt auch bei den anderen Versicherungsträgern auf. Insgesamt treten pro Versicherungsträger zwischen 20 und 30 Verordnungen auf, welche für das erste Halbjahr 2006 oder früher datiert sind, aber in der Ausprägung *jahr* mit 2007 oder 2008 markiert sind.

Der Peak auf der rechten Seite der Abbildung 1.1 entspricht dem Peak des perfekten Modells. Der linke wesentlich größere Peak jedoch ist eine Art Gegenstück. Diese Patienten haben ein bis zwei Verordnungen erhalten und danach ihre Therapie wieder abgebrochen.

Die Patienten zwischen den Peaks können sowohl zur Gruppe der Therapietreuen gehören, die im Zeitraum 2007 bis 2010 ihre Therapie begonnen haben, als auch zur Gruppe der Therapieabbrecher, die mehr als eine Verordnung konsumiert haben.

Die Patienten werden folgendermaßen in die Gruppen Therapietreue und Therapieabbrecher getrennt. Da Therapietreue regelmäßig Verordnungen konsumieren müssen werden all jene Patienten mit mehr als einem Verordnungszeitpunkt und der letzten Verordnung nach dem 1.11.2010 als Langzeittherapiepatienten betrachtet. Der Grund, warum Patienten mit einer Verordnung hier ausgeschlossen werden, ist dass die Wahrscheinlichkeit eines Therapieabbruchs höher ist, als jene einer Dauertherapie. Das Grenzdatum wurde dabei so gewählt, dass die resultierenden Gruppen möglichst keinen Peak an den Rändern haben, gemäß den Voraussetzungen des Modells der optimalen Versorgung. Diese Trennung führt zu Abbildung 1.2.

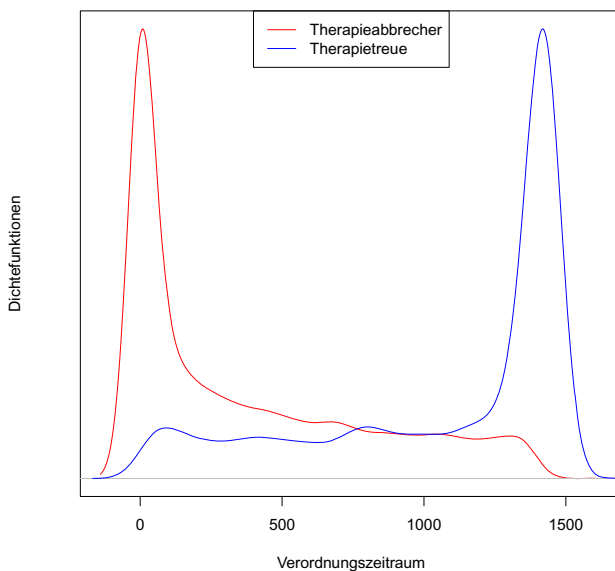


Abbildung 1.2: Gruppierung nach Therapietreue

Die Höhe der Peaks in Abbildung 1.2 darf nicht interpretiert werden, da es sich um voneinander unabhängige Dichtefunktionen handelt. Das Bild der einzelnen Kurven zeigt jedoch, dass die Trennungskriterien durchaus sinnvoll gewählt sind. Die Gruppe der Therapietreuen, bleibt über den betrachteten Zeitraum annähernd konstant, während die Therapieabbrecher nach ihrem Maximum bei dem Abbruch nach der ersten Verordnung

tendenziell früher abbrechen. Vordergründig sieht es so aus, als ob es unwahrscheinlicher wird, dass ein Patient seine Therapie abbricht, je länger er sie angenommen hat. Zu einem späteren Zeitpunkt wird ausgeführt, dass dies nicht der Fall ist. Tabelle 1.2 gibt einen Überblick über die Patienten die hinter den Gruppen stehen. Um robustere Aussagen treffen zu können werden die Quartile, der Median sowie die Anzahl an Patienten die keinen Eintrag für diese Daten haben dargestellt. Die Kosten bzw. Verordnungen wurden immer aus Einzeldatensätzen aggregiert. Kein Eintrag ist also gleichwertig mit keine Kosten. Trotzdem werden die Patienten im Rahmen der Quartilsberechnungen einmal weggelassen und einmal mit 0 gerechnet. Man kann einmal die Kosten betrachten die eine durchschnittliche Person der Population erzeugt und einmal jene die eine Person der Population erzeugt, die entsprechende Leistungen in Anspruch nimmt.

Tabelle 1.2: Vergleich Therapieabbrecher Therapietreue

	Therapieabbrecher*				Therapietreue**			
	1st Qu.	Med.	3rd Qu.	NA	1st Qu.	Med.	3rd Qu.	NA
Geb.jahr	1928	1947	1964	15	1932	1949	1964	0
Kost. AM ₁	788	1613	2829	1045	1258	2183	3506	148
Kost. FA ₁	116	438	550	21916	158	413	928	9823
Kost. AM ₂	719	1553	2778	0	1241	2167	3495	0
Kost. FA ₂	0	0	144	0	0	47	353	0
Kost. AP	23	76	255	0	181	576	2221	0
Vord. AP	1	4	13	0	14	35	74	0
Kost. sonst. HM ₁	64	369	1543	4478	601	1933	4510	284
Vord. sonst. HM ₁	4	24	91	4478	44	127	267	284
# Pharnr. AP	1	1	2	0	1	2	5	0
# Wirkstoffe AP	1	1	2	0	1	2	3	0

1 Patienten ohne Kosten werden nicht berücksichtigt.

2 Patienten ohne Kosten werden mit Kosten von 0 gewertet.

* Gesamtpopulation von 35 034 Patienten

** Gesamtpopulation von 20 511 Patienten

Dadurch ist es etwa möglich zu sehen, dass zumindest der Median der Kosten für Fachärzte der Therapieabbrecher höher ist als der Median der Therapietreuen. Berücksichtigt man aber, dass 2/3 der Therapieabbrecher Kosten von 0 erzeugen, so fällt der Median der Therapieabbrecher offensichtlich auf 0.

Der Mittelwert, sowie die Extremwerte werden bewusst nicht dargestellt, da sie durch die Ausreißer, welche vermutlich Sammelnummern sind das Bild stark verzerren. Aus diesem Grund müssen die Aussagen die aus den Quartilswerten abgeleitet werden auch vorsichtig betrachtet werden. Im Fall der Kosten für Allgemeinmediziner etwa fallen knapp 20% der Kosten auf die teuersten 5% der Patienten, unabhängig davon ob es sich um Therapieabbrecher oder Therapietreue handelt. Ein Entfernen der Extremwerte, die die Hauptverantwortung für diese Aussage tragen, ist nicht sinnvoll, da hier ein wesentlicher Teil der realen Kosten verloren gehen würde.

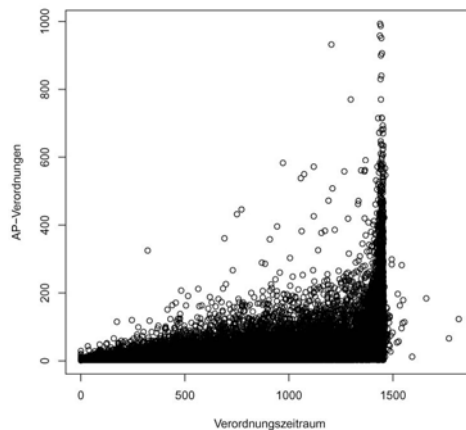
Bei Personen die einmal eine Therapie begonnen haben ist es, sofern die richtige Diagnose erstellt wurde, aus medizinischer Sicht, mit hoher Wahrscheinlichkeit empfehlenswert die

Tabelle 1.3: Verhältnis zwischen Therapieabbrecher und Therapietreuen.

	BGKK	KGKK	SGKK
Weiblich	1.52	1.4	1.78
Männlich	1.99	1.79	2.09
Gesamt	1.69	1.55	1.91

Therapie fortzusetzen. Wie aber gezeigt wurde, bricht eine absolute Mehrheit ihre Therapie wieder ab.

1.3.1 Anzahl an Verordnungen

**Abbildung 1.3:** Verordnungszeitraum gegen Antipsychotika Verordnungen

Es klingt plausibel, dass die Anzahl der Verordnungen die ein Patient konsumiert mit der Dauer seiner Therapie, dem Verordnungszeitraum, linear zusammenhängt. Diese Annahme ist für die Masse der Patienten gültig. Es gibt jedoch eine Untergruppe, deren Behandlung abweicht. Je nachdem wie die Untergruppe der auffälligen Patienten definiert wird, erkennt man schnell, dass die Konzentration der Verordnungen und Kosten in dieser Gruppe sehr hoch ist.

Die Konzentration wird in Abbildung 1.4 dargestellt. Die Kostenfunktion entsteht folgendermaßen. Die Patienten werden nach Höhe der Kosten aufsteigend geordnet. Auf der y-Achse wird die normierte kumulative Summe der Kosten aufgetragen, auf der x-Achse der Anteil, den diese Patienten an der Gesamtmenge ausmachen. Dadurch kann man schnell

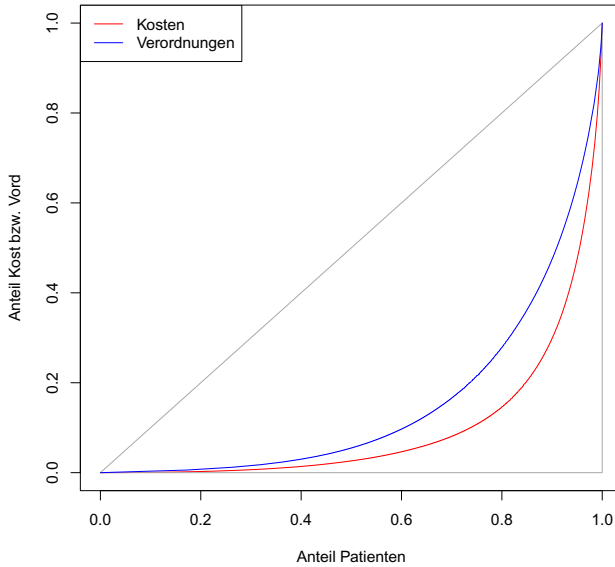


Abbildung 1.4: Konzentration der Kosten und Verordnungen

sehen, wie stark sich die Kosten bzw. Verordnungen auf eine kleine Gruppe von Patienten konzentriert. In Abbildung 1.4 sind zusätzlich die beiden Extremkurven in grau eingezeichnet. Einmal die Diagonale, welche einer Gleichverteilung der Kosten auf alle Patienten entspricht, andererseits fast entlang der Achsen, wo die gesamten Kosten auf einen Patienten entfallen.

Die Verteilung kann auf Basis dieses Diagramms beispielsweise mit Hilfe der Lorenzkonzentration gegeben werden. Das Konzept wird etwa in Assenmacher (2010) vorgestellt. Die Lorenzkonzentration oder auch Gini Koeffizient L ist definiert als das Verhältnis zwischen der echten Fläche A zwischen der Kurve und der Diagonalen und der maximal möglichen Fläche. Der Gini Koeffizient variiert dabei zwischen 0 im Fall einer Gleichverteilung, da hier die Fläche A verschwindet, und 1 im Fall einer Konzentration auf einen einzigen Patienten.

$$L = \frac{2nA}{n-1} \quad n \dots \text{Anzahl an Beobachtungen}$$

$$L_{\text{Kost}} = 0,80$$

$$L_{\text{Vord}} = 0,69$$

Im Allgemeinen gilt, dass die TOP 5% der Patienten bezogen auf konsumierte Verordnungen bzw. erzeugte Heilmittelkosten für Antipsychotika in Summe 36% der Verordnungen und 52,5 % der Kosten ausmachen. Die TOP 2,5% der Patienten im Vergleich konsumieren noch immer 23,5% der Verordnungen und 36% der Kosten. Einerseits ist der Unterschied zwischen den Anteilen an den Verordnungen und den Kosten unerwartet groß, andererseits ist die Konzentration durchwegs sehr hoch.

Multivariate statistische Analyse von Gesundheitsdaten
österreichischer Sozialversicherungsträger

Ortner, Th.

2015, XI, 67 S. 26 Abb., Softcover

ISBN: 978-3-658-08395-3