

Andreas Wolik

## Empirische Zusammenhangsanalyse von Arzneimitteln und Diagnosen sowie deren Berücksichtigung innerhalb des morbiditätsorientierten Risikostrukturausgleichs

Im morbiditätsorientierten Risikostrukturausgleich wurde vom Gesetzgeber nicht nur die Berücksichtigung von Diagnosen, sondern auch von Arzneimitteln gefordert. In der bisherigen Umsetzung wurde dieser Forderung dadurch entsprochen, dass eine Reihe von Arzneimittelinformationen ausschließlich als Prüfkriterium zur Validierung einer Auswahl von Diagnosen genutzt wird. Der Artikel zeigt Ansätze, wie medizinisch definierte Zusammenhänge zwischen Arzneimittelwirkstoffen und Diagnosen durch den Einsatz eines datengetriebenen Verfahrens (Assoziationsanalyse) ergänzt werden können. Neben dem Nachweis der grundsätzlichen Anwendbarkeit dieser Methode ergeben sich auch erste Hinweise auf mögliche Erweiterungen der im Kontext des morbiditätsorientierten Risikostrukturausgleichs verwendeten Zuordnungen.

### Problembeschreibung

Innerhalb des morbiditätsorientierten Risikostrukturausgleichs (RSA) sind Diagnosen zu identifizieren, die jeweils auf eine von maximal 80 Krankheiten mit einer hinreichenden Schwere oder Kostenintensität hinweisen. Über die Berücksichtigung der Diagnosen hinaus sollten auch die Arzneimittelverordnungen Eingang in den RSA finden. Dieser Forderung des Gesetzgebers hat das Bundesversicherungsamt (BVA) dadurch entsprochen, dass neben der Identifikation von Diagnosen und den daraus ableitbaren Hierarchisierten Morbiditätsgruppen (HMG) Arzneimittel zur Validierung einer Menge von ambulanten Diagnosen herangezogen werden. Das Aufgreifkriterium „Arzneimitteltherapie“ prüft zudem eine hinreichend lange Behandlungsdauer in Abhängigkeit von der betrachteten Krankheit. Bei akutem Behandlungsbedarf wird eine Dosierung von zehn Tagesdosen und in allen anderen Fällen eine Dosierung von mindestens einem halben Jahr, sprich 183 Tagesdosen, gefordert. Somit ist der Zweck der Berücksichtigung der Arzneimittel-ATC (Anatomisch-Therapeutisch-Chemisches Klassifikationssystem)

ausschließlich die Validierung vorliegender Diagnoseinformationen. Es soll aus medizinischer Sicht sichergestellt werden, dass eine hinreichende Schwere der Erkrankung vorliegt und der Morbiditätszuschlag gerechtfertigt ist.

Bei der im morbiditätsorientierten RSA umgesetzten Berücksichtigung von Arzneimittelverordnungen zur Diagnosevalidierung stellt sich die Frage, ob die Liste der für die jeweiligen Diagnosen berücksichtigten Arzneimittel vollständig ist. So ist die Vollständigkeit der Validierungsliste zum einen durch das schlichte Vergessen von Arzneimitteln gefährdet, zum anderen aber auch durch systembedingte fehlende Berücksichtigung von Innovationen. Die Innovationsproblematik wird verständlicherweise insbesondere von der Pharmaindustrie gesehen, da aus der prospektiven Gestaltung des morbiditätsorientierten RSA eine um mindestens ein Jahr verzögerte Berücksichtigung von neuen, im Markt vertretenen Produkten folgt. So entsteht eine weitere Schwelle, neue pharmazeutische Produkte innerhalb der GKV-Versorgung zu etablieren (Jaeckel 2008). Aber auch die Krankenkassen sind von diesem „Innovation Gap“ betroffen, da sie erst einmal in Vorlage für diese neuen Medikamente gehen müssen. Allerdings ist dieses Problem systemimmanent und kann einzig durch frühzeitige Berücksichtigung neuer Arzneimittel minimiert werden.

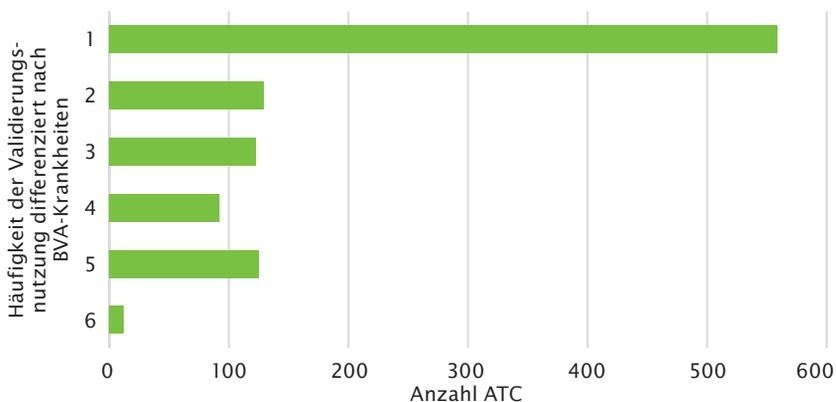
Darüber hinaus können Arzneimittel nicht berücksichtigt worden sein, weil sie nicht ausdrücklich für die Behandlung bestimmter Diagnosen zugelassen sind. Die Nutzung eines Medikaments außerhalb seines in der Zulassung definierten Gebrauchs, der sogenannte „Off-Label-Use“, gehört zur alltäglichen Praxis, daher ist die Berücksichtigung eines solchen Ordnungsverhaltens innerhalb der morbiditätsorientierten RSA-Validierungskriterien sinnvoll und würde innerhalb des stetigen Anpassungsprozesses des morbiditätsorientierten RSA helfen, den medizinischen Fortschritt abzubilden.

Neben dem Aspekt der Vollständigkeit der RSA-Validierungsliste stellt sich die Frage, ob und in welchem Maße es möglich ist, Arzneimittelverordnungen als Ersatz für Diagnosen zu verwenden. Dieser Ansatz ginge dann weit über die bisherige Berücksichtigung der Arzneimittel innerhalb des morbiditätsorientierten RSA hinaus. Die Motivation für eine derartige Nutzung von Arzneimittelinformationen ergibt sich unter anderem aus Studien, die eine kritische Diagnosequalität im ambulanten Bereich feststellen (siehe hierzu Giersiepen et al. 2007). Daraus lässt sich ableiten, dass die Morbidität eines Versicherten durch ambulante Diagnosen nur unzureichend beschrieben werden kann.

Eine Hürde bei der Nutzung von Arzneimitteln als Diagnosesubstitut ist die in den meisten Fällen nicht gegebene eindeutige Zuordnung eines Medikamentes zu genau einer Diagnose und damit zu einer einzigen Krankheit. So zeigt sich auch bei in den Validierungskriterien berücksichtigten Arzneimitteln eine Mehrfachnutzung einzelner Medikamente. In Abbildung 1 ist dieser Zusammenhang innerhalb der BVA-Referenztabellen zwischen ATC und Krankheit aufgeführt.

**Abbildung 1:** Mehrfachzuweisungen der zur Validierung genutzten ATC zu Krankheiten

---



---

Datenbasis: BVA Änderungsbekanntgabe vom 2. März 2009

Bei der mehrfachen Nutzung der zur Validierung genutzten Arzneimittel ist es sehr wahrscheinlich, dass nicht alle Zusammenhänge gleich stark sind. Der Grad des Zusammenhangs zwischen den Arzneimitteln und Diagnosen kann ermittelt werden und lässt sich am besten mit Wahrscheinlichkeiten bewerten. Die Methodik der Assoziationsanalyse, mit der sowohl die Zusammenhänge selbst als auch ihre Wahrscheinlichkeiten bestimmt werden können, ist hierzu ein adäquater Ansatz.

### Methodik

Die Assoziationsanalyse ist eine Methode aus den Verfahren des maschinellen Lernens und arbeitet autonom mit den zur Verfügung gestellten Daten. Sie ist besonders dazu geeignet, Zusammenhänge in komplexen Datenbeständen zwischen den Eigenschaften beziehungsweise Zuständen untersuchter Objekte, wie Kunden oder Patienten, zu identifizieren.

Die Assoziationsanalyse findet ihre Anwendung insbesondere in der Analyse von großen Transaktionsdatenbeständen, wie sie bei Handelstransaktionen, als Warenkörbe von gemeinsam gekauften Produkten, auftreten. Daher wird die Assoziationsanalyse auch oft Warenkorbanalyse genannt. Die Ergebnisse von Assoziationsanalysen begegnen uns im Alltag ständig. So wird man auf interessante Produkte im Online-Buchhandel aufmerksam gemacht oder das Warensortiment im Supermarkt ist so sortiert, dass die gemeinsam konsumierten Dinge dicht beieinander stehen. Ein weiteres Anwendungsfeld für die Assoziationsanalyse ist die Betrugsidentifikation (Fraud Detection) in sehr großen Datenbeständen, wie sie im Wertpapierhandel oder auch Kreditkartengeschäft vorliegen.

Bei Bollinger (1996: 257) wird folgende kurze Definition gegeben: „Assoziationsregeln beschreiben Korrelation zwischen gemeinsam auftretenden Dingen“. Im Allgemeinen ist das Ziel einer Assoziationsanalyse die Identifizierung einer Menge von Eigenschaften beziehungsweise Zuständen von Untersuchungsobjekten, die gemeinsam mit einer disjunkten

(unterschiedlichen) Menge auftreten. Ein solcher Zusammenhang ist dann als Regel formulierbar (vergleiche Hettich und Hippner 2009). So ist die Interpretation der versichertenbezogenen Informationen von Arzneimitteln und Diagnosen als „Warenkorb“ ein adäquater Ansatz zur Bildung einer Analysegrundlage und der Schritt von „Kunden, die x gekauft haben, haben auch y gekauft“ zu „Patienten, bei denen a verschrieben wurde, ist b diagnostiziert worden“ sehr naheliegend. Die einzelnen Kombinationspaare zwischen Arzneimittel und Diagnose werden im Kontext der Assoziationsanalyse als Transaktionen bezeichnet.

Die Algorithmen der Assoziationsanalyse suchen in den gegebenen Datenbeständen nach Regeln, die mithilfe der Kriterien Relevanz (Support) und der bedingten Wahrscheinlichkeit (Confidence) bewertet werden. Je höher eine Mindestrelevanz als Schwellwert gesetzt wird, desto häufiger ist im oben dargestellten Beispiel ein Arzneimittel-Diagnose-Zusammenhang vertreten. Die Mindestrelevanz wird im Verhältnis zu allen Transaktionen gemessen. Dies bedeutet im umgekehrten Fall, dass ein niedriger Mindestsupport zwar dazu führt, dass mehr Zusammenhänge identifiziert werden können, aber mehr exotische beziehungsweise seltene Zusammenhänge aufgezeigt werden. Da solche seltenen Zusammenhänge oftmals Artefakte darstellen, erhöht sich die Wahrscheinlichkeit, dass falsche Zusammenhänge identifiziert werden.

Je höher die Confidence ist, desto häufiger ist in den Fällen, in denen ein bestimmtes Arzneimittel verordnet wurde, eine bestimmte Diagnose ebenfalls zu finden. Die Wahrscheinlichkeit wird also im Verhältnis zu den Transaktionen gemessen, an denen das untersuchte Arzneimittel beteiligt ist.

Die Assoziationsanalyse ist bei den gegebenen großen Datenvolumina von Arzneimittel- und Diagnosedaten ein vielversprechendes Verfahren. Um Zusammenhänge zwischen Verordnungen und Diagnosen zu finden, ist man gezwungen, jedes Arzneimittel eines Versicherten mit all seinen Diagnosen in Verbindung zu setzen. Wegen der großen

Kombinationsvielfalt bedarf es intelligenter Verfahren, um den Analyse-  
raum rechnerisch bewältigen zu können. Dies ist bei den Assoziations-  
algorithmen der Fall.

### Zusammenhangsanalyse und Beschreibung der Datengrundlage

Auf Basis der vorgestellten Methodik lässt sich die bestehende RSA-  
Systematik sinnvoll erweitern, indem bestehende Zusammenhänge zwi-  
schen Arzneimitteln und Diagnosen, mit einer Wahrscheinlichkeit ver-  
sehen, genutzt werden. Im Folgenden wird die für die späteren Analysen  
genutzte Datenbasis vorgestellt und die notwendigen Vorverarbeitungss-  
chritte dargelegt.

### Datenbasis

Erste Untersuchungsergebnisse in (Weisser et al. 2008) deuten darauf  
hin, dass datengetriebene Ansätze zur Untersuchung der Arzneimittel-  
Diagnose-Zusammenhänge insbesondere bei chronischen Erkrankungen  
Erfolg versprechen. Da die in der RSA-Systematik genutzten hierarchi-  
sierten Morbiditätsgruppen (HMG) zum größten Teil so definiert sind,  
dass sie chronische Erkrankungen repräsentieren, wurde die Analyse-  
basis auf Versicherte mit mindestens einer HMG eingeschränkt.

Die genutzten Einzelfalldaten bestehen aus den Quartalsmeldungen  
der ambulanten Diagnose- und Verordnungsdaten des Jahres 2007. Ver-  
wendet werden nur die Diagnosen, die vom behandelnden Arzt als ge-  
sichert gekennzeichnet wurden. Diagnosen und Verordnungen eines  
Versicherten sind aber anhand der Daten nicht kausal zueinander in  
Beziehung zu bringen, da es im bestehenden Abrechnungssystem nicht  
gefordert ist, diesen Zusammenhang zu dokumentieren. Daher kann  
man anhand der Daten eines Versicherten a priori nur einen vollstän-  
digen kombinatorischen Zusammenhang annehmen, wodurch sich eine  
extreme Vervielfachung der Daten ergibt. So ergeben sich beispiels-  
weise aus fünf Diagnosen und drei Arzneimittelverordnungen bereits  
 $3 * 5 = 15$  zu berücksichtigende Transaktionen.

Um die Anzahl der Kombinationsmöglichkeiten zwischen Diagnosen und Arzneimitteln zu reduzieren, werden die Informationen auf verschiedene Arten zusammengeführt. Zunächst werden die Diagnosen eines Versicherten nur mit den Verordnungen desselben Arztes kombiniert. Diese Vorgehensweise bewirkt die Bildung von homogeneren Untergruppen der Zusammenhänge ATC-ICD auf Arzt-Versicherten-Ebene und schafft damit eine Verminderung der Bildung unnützer beziehungsweise falscher Informationen, das sogenannte Datenrauschen. Erst durch die zusätzliche Berücksichtigung des Arztes ist es möglich, bestimmte ATC-ICD-Kombinationen eines Versicherten nicht zu berücksichtigen. So wird ein Orthopäde andere Medikamente verordnen als ein Augenarzt. Als Resultat dieser Vorgehensweise vermindern sich die kombinatorischen Möglichkeiten drastisch.

Wegen der Betrachtung auf Arzt-Patienten-Ebene entsteht durch Überweisungen oftmals eine Aufteilung zwischen Diagnosen und Verordnungen auf verschiedene Ärzte. Um diese Arzneimittel- und Diagnoseinformationen trotzdem zusammenzuführen, werden in einem weiteren Schritt die Informationen eines zwischenzeitlich hinzugezogenen Arztes besonders berücksichtigt und die durch eine Überweisung generierten Diagnosen und Arzneimittelverordnungen dem überweisenden Arzt zugeordnet. Das Beispiel des Hausarztes macht diese Problematik sofort deutlich. Wenn ein Hausarzt seinen Patienten zu einem Facharzt überweist (etwa zum Orthopäden), so wird der Hausarzt die Behandlungsempfehlung des Facharztes bei der Weiterbehandlung in den meisten Fällen übernehmen, aber die Codierung der Diagnosen nicht in allen Fällen wiederholend durchführen. Die medikamentöse Dauerbehandlung wird allerdings in den meisten Fällen beim Hausarzt erfolgen. Daher werden die Facharzt-Diagnosen dem weiterbehandelnden Hausarzt zugeordnet. Die gegebenen Daten lassen diese Zuordnung zwischen überweisendem und behandelndem beziehungsweise diagnostizierendem Arzt zu und bewirken eine weitere Verbesserung der Analysebasis. Zudem werden nur die Daten eines Quartals miteinander kombiniert, um so die kombinatorischen

Möglichkeiten weiter zu vermindern und damit homogenere Subgruppen zu bilden.

Ein weiterer Ansatz zur Verringerung der Zuordnungsmöglichkeiten wäre die Berücksichtigung eines Diagnosedatums. Da die Apothekendaten mit einem Verordnungsdatum versehen sind, ließe sich hierdurch eine schärfere Zuordnung erreichen. Die Diagnosedaten enthalten jedoch kein Datum der Diagnosestellung, sondern lediglich Behandlungsbeginn und Behandlungsende eines Behandlungsfalles im gegebenen Quartal. Es handelt sich um das Datum des ersten und letzten Arzt-Patienten-Kontaktes im Quartal. Betrachtet man jedoch die besondere Konstellation eines einmaligen Kontaktes im Quartal (Behandlungsbeginn = Behandlungsende), so ist man in dieser besonderen Konstellation in der Lage, eine tagesgenaue Zuordnung von Arzneimittel und Diagnose durchführen zu können. Die zusätzliche Berücksichtigung des Behandlungsdatums in der oben beschriebenen Weise stellt bei der vorliegenden Datengrundlage die Methode mit dem geringsten „Rauschen“ dar. Eine Analyse der sich ergebenden Transaktionen nach Facharztgruppen zeigt jedoch, dass diese Methode ein verzerrtes Bild erzeugt und die Daten daher nicht als repräsentative Teilmenge betrachtet werden können. Insbesondere die Menge der Transaktionen, die aufgrund von allgemeinärztlicher oder internistischer Behandlung entstehen, wäre stark reduziert, da diese wenige eintägige Behandlungsfälle aufweisen.

### Problem der Multimorbidität

Die Multimorbidität vieler Versicherter in der Untersuchungsstichprobe bewirkt ein Verrauschen der Daten, da in dieser Untergruppe ein besonders hohes Maß an kombinatorischen Möglichkeiten zwischen den Arzneimitteln und den Diagnosen vorliegt. Dennoch zeigt sich in den Analysen, dass eine Beschränkung auf Versicherte mit einer einzigen HMG als Indikator für eine „Monomorbidität“ die Ergebnisqualität nicht besonders stark verbessert, aber in der Folge bestimmte Diagnosen nicht mehr durch das Regelwerk betrachtet werden. Ein Grund hierfür ist das häufige gemeinsame Auftreten bestimmter Diagnosen. Ein Beispiel ist

der Diabetes Mellitus und die Hypertonie. Eine Einschränkung auf die Versicherten mit genau einer HMG erhöht die Wahrscheinlichkeit, dass durch das Kriterium der Mindestrelevanz Regeln mit solchen Diagnosen ausgefiltert werden, da sie eher selten allein auftreten.

Nach Abwägung der beschriebenen Effekte wurden Versicherte mit mindestens einer HMG ausgewählt. Aus dieser Grundgesamtheit wurde wegen der höheren Praktikabilität eine in etwa fünfprozentige Zufallsstichprobe gezogen und dabei wurden etwa 139.000 Versicherte ausgewählt. Diesen konnten für das betrachtete Jahr 8.571.294 Arzneimittelverordnungen und 7.290.840 Diagnosen zugewiesen werden. Durch die oben beschriebene Zusammenführung von Diagnosen und Arzneimitteln eines Versicherten bei einem Arzt in einem Quartal ergaben sich 82.481.606 Arzneimittel-Diagnosepaare, die in der Nomenklatur der Assoziationsanalyse den Transaktionen entsprechen.

### Empirische Analysen

Die oben definierte Datenbasis wurde mithilfe der Assoziationsanalyse mittels eines Assoziations-Regel-Algorithmus bearbeitet. Dabei wurde der im SQL-Server 2005 implementierte Algorithmus „Microsoft Association Rules“ genutzt (vergleiche Tang und McLennan 2005). Das Ergebnis eines Analyselaufes ist sowohl eine Regelliste als auch eine Visualisierung durch ein Abhängigkeitsnetz. Die Berücksichtigung von Alter und Geschlecht erfolgte aus zweierlei Gründen. Zum einen wird auch durch das BVA eine differenzierte Betrachtung nach Alter und Geschlecht durchgeführt und zum anderen kann gezeigt werden, dass sich durch die vorgestellte Methodik auch eine medizinisch sinnvolle Differenzierung der Medikamentengabe nach dem Geschlecht ergeben kann. So können dieselben Medikamente in unter Umständen völlig unterschiedlichen und geschlechtsspezifischen Diagnosen Anwendung finden. Ein Beispiel hierfür ist Buserelin (ATC L02AE01). Die Anwendung bei Männern erfolgt zur Behandlung eines Prostatakarzinoms. Die durch das Verfahren generierten Regelzusammenhänge mit dem Arzneimittel Buserelin sind in Tabelle 1 aufgeführt.

**Tabelle 1:** Beispiel einer geschlechtsspezifischen Regel für Männer

Wahrscheinlichkeit	ATC-Bez.	Geschlecht	Alter	ICD-Bezeichnung
65%	Buserelin	männlich	älter als 77	Bösartige Neubildung der Prostata
64%	Buserelin	männlich		Bösartige Neubildung der Prostata
56%	Buserelin	männlich	zwischen 66 und 77	Bösartige Neubildung der Prostata

Datenbasis: BARMER-Daten 2007

Bei Frauen hingegen wird Buserelin für die Reifung der Ovarialfollikel und damit für die Vorbereitung und Unterstützung einer künstlichen Befruchtung verordnet.

**Tabelle 2:** Beispiel einer geschlechtsspezifischen Regel für Frauen

Wahrscheinlichkeit	ATC-Bez.	Geschlecht	Alter	ICD-Bezeichnung
12,5%	Buserelin	weiblich		Fertilisationsfördernde Maßnahme, nicht näher bezeichnet

Datenbasis: BARMER-Daten 2007

In der zuerst generierten Regelmenge verhinderte allerdings eine zu hoch gesetzte Mindestrelevanz (Support), dass Regeln, die Buserelin bei Frauen beinhalten, aufgeführt wurden. Daher wurde ein Analyselauf mit herabgesetzter Mindestrelevanz durchgeführt, indem die Anwendung von Buserelin bei Frauen mit einer Wahrscheinlichkeit von 12,5 Prozent bewertet wurde.

Die geringe Wahrscheinlichkeit liegt insbesondere in den Komorbiditäten dieser kleinen Subgruppe begründet. Eine Detailanalyse zeigt, dass in der untersuchten Gruppe der weiblichen Buserelin-Anwender eine Reihe

von Schwangerschaftsuntersuchungen mit den zugehörigen Diagnosen durchgeführt wird. Da die Daten miteinander kombiniert werden, ergibt sich eine Wahrscheinlichkeit von 12,5 Prozent.

Die weiteren Analysen verfolgen zwei Fragestellungen. Zuerst wird der Kontext der Arzneimittel zur Diagnosevalidierung betrachtet und geprüft, ob es auf Basis der gegebenen Daten möglich ist, die im morbiditätsorientierten RSA genutzte ATC-Liste zur Validierung bestimmter Diagnosen mit Arzneimitteln zu erweitern. Dies wäre sinnvoll, wenn signifikante und starke Regeln für die zu validierenden Diagnosen gefunden werden würden und die verwendeten Arzneimittel in der Validierungsliste nicht aufgeführt wären. Zusätzlich wird geprüft, ob bereits zur Validierung bestimmter Diagnosen aufgeführte Arzneimittel nicht auch für andere Diagnosen zur Validierung herangezogen werden sollten.

In der zweiten Fragestellung wird untersucht, in welchem Umfang die Assoziationsanalyse für die Identifizierung und Gewichtung der Zusammenhänge zwischen Arzneimitteln und Diagnosen mit hohen Wahrscheinlichkeiten genutzt werden kann. Das Ziel ist es, auf diese Weise Zusammenhänge mit besonders hoher Wahrscheinlichkeit zu identifizieren, um die Wahrscheinlichkeiten der Regeln als Maß für die Substituierbarkeit der jeweiligen Diagnosen durch die Arzneimittelverordnungen zu nutzen.

Bei beiden Fragestellungen sollen die ausgewiesenen Zusammenhänge nur für die im morbiditätsorientierten-RSA relevanten Diagnosen angegeben werden. Daher wird für die Analysen eine Filterung genutzt, die die Diagnosen ausschließlich auf die RSA-relevanten Diagnosen einschränkt. Bei den Analysen zeigt sich jedoch, dass es die Regelqualität verschlechtert, wenn man die Filterung vor der Regelgenerierung durchführt. Daher werden alle Regeln erst nachträglich auf die im morbiditätsorientierten-RSA relevante Menge gefiltert. Die Ursache hierfür ist dieselbe wie bei einer Einschränkung der Arzneimittelverordnungen und wird im Folgenden diskutiert.

### Analyse einer möglichen Ergänzung bestehender ATC-Validierungskriterien

Für diese Untersuchung wird die Datenbasis in einem ersten Analyseansatz auf die Arzneimittel-ATC eingeschränkt, die nicht Teil der Arzneimittel-Validierungsreferenz des BVA sind. Daher werden alle innerhalb der RSA-Prüfkriterien genutzten ATC ausgefiltert. Dadurch können sich allerdings auch keine Regeln der Art: „Insulin -> Diabetes Mellitus“ bilden, da Insulin bereits durch die BVA-Validierungsreferenz genutzt wird.

Bei der Regelgenerierung zeigt sich, dass erst bei einem niedrig gesetzten Schwellwert der Mindestrelevanz eine Regelmenge generiert wurde. Daher entstehen sehr spezifische Regeln, die eine Tendenz zu Artefakten aufweisen. Die stärksten Regeln der unter den beschriebenen Einschränkungen generierten Regelmenge sind in Tabelle 3 dargestellt. Der stärkste Zusammenhang zwischen Glucagon und Diabetes Mellitus ist allerdings auf eine kleine Subgruppe von Notfallversorgungen zurückzuführen. Es fällt ferner auf, dass es wenige starke Regeln gibt. So wird ein Zusammenhang zwischen dem H1-Antihistaminikum Mizolastin und der Hypertonie als Regel dargestellt. Diese Regel ist eine Folge der Komorbiditäten in einer kleinen Subgruppe von Hypertonikern, da jedes Arzneimittel mit allen Diagnosen dieser Patienten in Verbindung gesetzt wird.

Dieser Effekt wird allerdings durch die Beschränkung auf die nicht zur Validierung genutzten Arzneimittel verstärkt. Ein Grund für diesen Effekt ist die Verminderung der Transaktionsmenge, die dazu führt, dass Artefakte beziehungsweise spezifische Zusammenhänge auf kleinen Subgruppen eine hohe Wahrscheinlichkeit erhalten und zu starken Regeln werden. Daher hat es sich als sinnvoll erwiesen, keine der Regelgenerierung vorgelagerte Filterung durchzuführen und die Mindestrelevanz möglichst hoch zu setzen.

**Tabelle 3:** Regeln mit neuen Arzneimittel-Diagnose-Zusammenhängen für nicht bereits berücksichtigte Arzneimittel

Wahrscheinlichkeit	ATC-Bezeichnung	Geschlecht	Alter	ICD-Bezeichnung
56%	Glucagon		jünger als 31	Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
45%	Glucagon		zwischen 31 und 51	Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
41%	Glucagon	männlich		Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
36%	Glucagon			Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
29%	Glucagon	weiblich		Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
27%	Mizolastin	männlich		Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
23%	Mizolastin		zwischen 51 und 66	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
17%	Hydroxyzin		älter als 77	Nicht näher bezeichnete Demenz
15%	Detajmium			Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
14%	Glucagon	männlich		Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes], ohne Komplikationen, nicht als entgleist bezeichnet
13%	Sulfasalazin		zwischen 66 und 78	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise

Datenbasis: BARMER-Daten 2007

Daher wird die Analyse mit einer Filterung der ATC nach der erfolgten Regelbildung und der Berücksichtigung der ATC-ICD-Relationen aus den BVA-Validierungslisten erneut durchgeführt. Das bedeutet, dass die in den Referenztabellen definierten Zusammenhänge ignoriert werden und nur die Zusammenhänge ausgewiesen werden, die nicht Bestandteil der Validierungsreferenz sind. Hierdurch ist es möglich, Medikamente zur Validierung von Diagnosen aufzuzeigen, für die sie nicht vordringlich definiert sind, aber in der medizinischen Versorgung verordnet werden.

In der neu generierten Regelmenge werden dann auch neue Regelzusammenhänge ausgewiesen (vergleiche Tabelle 4). Von den identifizierten Zusammenhängen könnten sich einige für eine Erweiterung der Validierungsliste – auch aus medizinischer Sicht – als sinnvoll erweisen. Dies ist zum einen das Arzneimittel Risperidon, das in der Regelmenge einen Zusammenhang zur Demenz aufzeigt. In den Fachinformationen zu Risperidon wird der Wirkstoff für das Behandlungsgebiet der Schizophrenie ausgewiesen. Dennoch ist das Medikament auch zur Therapie gegen schwere Verhaltensstörungen bei Demenz zugelassen.

Analog wird das Medikament Melperon sowohl zur Behandlung von Verwirrheitszuständen und psychomotorischer Unruhe bei Psychosen als auch für die im Regelwerk identifizierte Nutzung von Erregungszuständen bei organisch bedingter Demenz genutzt. Die ausgewiesenen Zusammenhänge zwischen dem Arzneimittelwirkstoff Biperiden und der Schizophrenie sind auf einen indirekten Zusammenhang zurückzuführen. Biperiden wird zur Behandlung von Bewegungsstörungen im Rahmen einer Parkinson-Erkrankung genutzt. Ähnliche Symptome werden durch Neuroleptika hervorgerufen, die zur Therapie der Schizophrenie eingesetzt werden. Biperiden lindert die Nebenwirkungen der Neuroleptika im Rahmen der Schizophreniebehandlung.

**Tabelle 4:** Regeln mit neuen Arzneimittel-Diagnose-Zusammenhängen für bereits berücksichtigte Arzneimittel

Wahrscheinlichkeit	ATC-Bezeichnung	Geschlecht	Alter	ICD-Bezeichnung
53%	Biperiden		jünger als 36	Paranoide Schizophrenie
36%	Biperiden		zwischen 51 und 66	Paranoide Schizophrenie
34%	Biperiden		zwischen 36 und 51	Paranoide Schizophrenie
32%	Biperiden	weiblich		Paranoide Schizophrenie
30%	Biperiden			Paranoide Schizophrenie
28%	Biperiden	männlich		Paranoide Schizophrenie
23%	Melperon		älter als 77	Nicht näher bezeichnete Demenz
19%	Risperidon		älter als 77	Nicht näher bezeichnete Demenz
16%	Pirenzepin			Paranoide Schizophrenie
16%	Pipamperon		älter als 77	Demenz bei Alzheimer-Krankheit, mit spätem Beginn (Typ 1)
15%	Melperon	weiblich		Nicht näher bezeichnete Demenz
15%	Kombinationen	weiblich		Nicht näher bezeichnete Demenz
13%	Biperiden		zwischen 36 und 51	Schizophrenes Residuum
13%	Melperon			Nicht näher bezeichnete Demenz
13%	Biperiden	männlich		Schizophrenes Residuum
12%	Propafenon		älter als 77	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
12%	Kombinationen			Nicht näher bezeichnete Demenz
10%	Folsäure		zwischen 66 und 78	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
10%	Pentaerythryltetranitrat		zwischen 51 und 66	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
10%	Biperiden			Schizophrenes Residuum
10%	Hydrocortison	männlich		Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise

Wahrscheinlichkeit	ATC-Bezeichnung	Geschlecht	Alter	ICD-Bezeichnung
10%	Levodopa in Kombination mit Carbidopa	männlich		Nicht näher bezeichnete Demenz
10%	Propafenon	weiblich		Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
10%	Acetyldigoxin		zwischen 66 und 78	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
10%	Physiologische Kochsalzlösung		älter als 77	Nicht näher bezeichnete Demenz
9%	Triamcinolon-Depot		zwischen 66 und 78	Gonarthrose, nicht näher bezeichnet
9%	Methylprednisolonacetat		jünger als 36	Asthma bronchiale, nicht näher bezeichnet
9%	Isosorbidmononitrat	männlich		Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
9%	Kombinationen		zwischen 51 und 66	Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise
8%	Warfarin			Essentielle Hypertonie, nicht näher bezeichnet, ohne Angabe einer hypertensiven Krise

Datenbasis: BARMER-Daten 2007

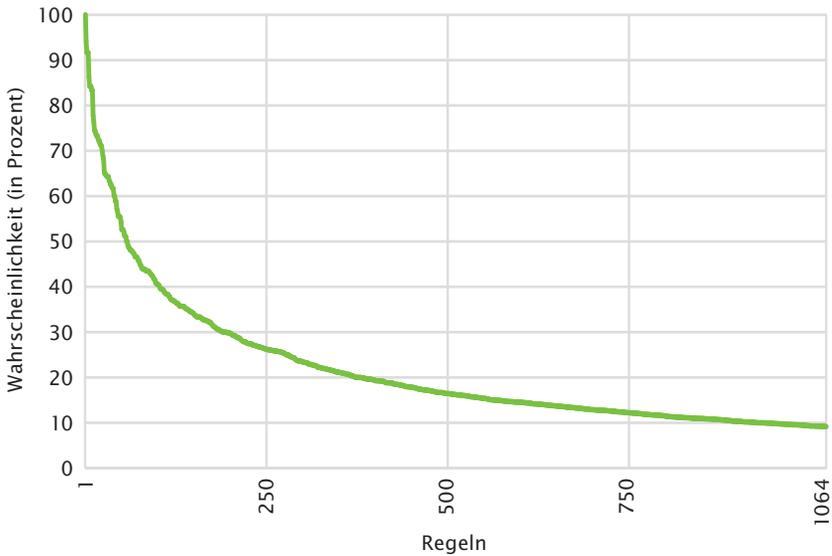
Neben diesen neurologischen Diagnosen zeigten sich auch in den Validierungslisten ungenutzte Zusammenhänge zur Hypertonie. Dies gilt sowohl für Isosorbidmononitrat als auch für Warfarin, die beide für die Behandlung der Hypertonie mit ihren Folgeerscheinungen genutzt werden. Beide Arzneimittel helfen, eine Hypertonie-Diagnose zu validieren.

Als Resultat beider Analysen ergaben sich zwei unterschiedliche Ergebnisse. In der ersten Analyse zeigte sich, dass die ATC der Validierungsliste auf Basis der gegebenen Daten und der vorgestellten Methodik nicht um weitere Arzneimittel erweitert werden kann. Dies ist sicherlich auch der umfangreichen Berücksichtigung von 1.040 ATC-Codes für die zu validierenden 728 Diagnosen zu verdanken.

Hingegen zeigte sich in der zweiten Analyse, dass die bereits verwendeten Arzneimittel auch zur Validierung weiterer Krankheiten verwendet werden sollten, da nachvollziehbare Zusammenhänge extrahiert werden konnten. So ergeben sich neue Arzneimittel-Diagnose-Kombinationen, die zusätzliche Validierungsmöglichkeiten für Diagnosen darstellen.

### Bestimmung der Substituierbarkeit von morbiditätsorientierten RSA-Diagnosen durch Arzneimittelverordnungen

In einer weiteren Analyse erfolgt die Untersuchung der Arzneimittel- und Diagnosedaten ohne Einschränkung auf eine Teilmenge der Arzneimittel und die zu validierenden Diagnosen. Ziel ist es, möglichst viele signifikante Zusammenhänge zu identifizieren, die für eine mit Wahrscheinlichkeiten gewichtete Substitutionsliste genutzt werden können. Durch die Berücksichtigung aller Arzneimittel ergibt sich ein deutlich stärker ausdifferenziertes Regelwerk. So werden 1.064 Regeln generiert, die die Zusammenhänge von Arzneimittel, Alter und Geschlecht und RSA-relevanten Diagnosen darstellen. Diese Regeln differenzieren dabei 322 identifizierte Arzneimittel-Diagnose-Zusammenhänge nach den zusätzlichen Kriterien Alter und Geschlecht. Die generierten Regeln implizieren ausschließlich die für den morbiditätsorientierten RSA-relevanten Diagnosen und werden, aus den oben beschriebenen Gründen, erst nach der erfolgten Regelgenerierung auf die RSA-Diagnosen eingeschränkt. In Abbildung 2 ist der Verlauf der Wahrscheinlichkeiten dieser Regelmenge bis zur geringsten vorliegenden Regelwahrscheinlichkeit von neun Prozent abgebildet.

**Abbildung 2:** Wahrscheinlichkeitsverlauf des Regelwerks ohne Filter

Datenbasis: BARMER-Daten 2007

So wurden auf der Basis von Regeln mit einer mindestens 50-prozentigen Wahrscheinlichkeit 32 Arzneimittel-Diagnose-Zusammenhänge identifiziert. Zur einfacheren Darstellung sind diese in Tabelle 5 als Zusammenhänge von Arzneimitteln und Diagnosen ohne Differenzierung nach Alter und Geschlecht in kompakter Form dargestellt. Die in Tabelle 5 aufgeführten Arzneimittelverordnungen zeigen somit die in den zugrunde liegenden Daten in besonders hohem Maß zur Substitution geeigneten Diagnosen auf.

**Tabelle 5:** Auswahl der stärksten ATC-ICD-Zusammenhänge mit Wahrscheinlichkeit größer 50 Prozent

ATC-Bezeichnung	ICD-Bezeichnung	Validierungs- Diagnose	ATC in Liste
Triptorelin	Bösartige Neubildung der Prostata	Nein	Nein
Flutamid	Bösartige Neubildung der Prostata	Nein	Nein
Mitomycin	Bösartige Neubildung: Harnblase, nicht näher bezeichnet	Nein	Nein
Imatinib	Chronische myeloische Leukämie; ohne Angabe einer kompl. Remission	Nein	Nein
Gliquidon	Nicht primär insulinabhängiger Diabetes	Nein	Ja
Estramustin	Bösartige Neubildung der Prostata	Nein	Nein
Goserelin	Bösartige Neubildung der Prostata	Nein	Nein
Leuprorelin	Bösartige Neubildung der Prostata	Nein	Nein
Bicalutamid	Bösartige Neubildung der Prostata	Nein	Nein
Ribavirin	Chronische Virushepatitis C	Nein	Ja
Leuprorelin	Bösartige Neubildung: Brustdrüse nicht näher bezeichnet	Nein	Nein
Cyproteron	Bösartige Neubildung der Prostata	Nein	Nein
Nevirapin	Nicht näher bezeichnete HIV-Krankheit	Nein	Ja
Buserelin	Bösartige Neubildung der Prostata	Nein	Nein
Lopinavir	Nicht näher bezeichnete HIV-Krankheit	Nein	Ja
Benzylpenicillin- Benzathin	Erysipel (Wundrose)	Ja	Ja
Phenobarbital	Epilepsie, nicht näher bezeichnet	Nein	Ja
Aripiprazol	Paranoide Schizophrenie	Ja	Ja
Metformin und Rosiglitazon	Nicht primär insulinabhängiger Diabetes	Nein	Ja
Miglitol	Nicht primär insulinabhängiger Diabetes	Nein	Ja
Zidovudin, Lamivudin und Abacavir	Nicht näher bezeichnete HIV- Krankheit	Nein	Nein
Goserelin	Bösartige Neubildung: Brustdrüse nicht näher bezeichnet	Nein	Nein
Ziprasidon	Paranoide Schizophrenie	Ja	Ja

ATC-Bezeichnung	ICD-Bezeichnung	Validierungs- Diagnose	ATC in Liste
Acebutolol	Essentielle Hypertonie, nicht näher bezeichnet	Ja	Ja
Glatirameracetat	Multiple Sklerose, nicht näher bezeichnet	Nein	Nein
Metformin und Pioglitazon	Nicht primär insulinabhängiger Diabetes	Nein	Ja
Fludrocortison	Adrenogenitale Störung, nicht näher bezeichnet	Nein	Nein
Levomethadon	Chronische Virushepatitis C	Nein	Nein
Biperiden	Paranoide Schizophrenie	Ja	Ja
Efavirenz	Nicht näher bezeichnete HIV-Krankheit	Nein	Ja
Buprenorphin	Psychische und Verhaltens- störungen durch Opiode	Nein	Nein
Exemestan	Bösartige Neubildung: Brustdrüse nicht näher bezeichnet	Nein	Nein

Datenbasis: BARMER-Daten 2007

### Fazit und Ausblick

Die Ergebnisse zeigen die grundsätzliche Anwendbarkeit der Methodik der Assoziationsanalyse auf Arzneimittel und Diagnosedaten. Es lässt sich nachweisen, dass sich in den Daten Zusammenhänge bei einzelnen Arzneimitteln widerspiegeln, die innerhalb der BVA-Referenztabellen zur Validierung einiger Diagnosen nicht betrachtet werden und somit Vorschläge für die Erweiterung der bestehenden Diagnosevalidierungsliste darstellen.

Darüber hinaus zeigte sich, dass für den morbiditätsorientierten RSA relevante Diagnosen durch hinreichend starke Regelzusammenhänge prognostiziert werden können. Diese Regeln bewerten die Zusammenhänge zwischen Arzneimittel und Diagnosen und geben mit ihrer Wahrscheinlichkeit ein Maß für die Substituierbarkeit der Diagnose durch ein Arzneimittel an und stellen damit eine sinnvolle Erweiterung der RSA-Systematik dar, indem die Arzneimittel nicht nur zur Validierung

genutzt werden, sondern in einigen Fällen auch Diagnosen beziehungsweise Krankheiten graduell repräsentieren können. Diese Anwendungsmöglichkeit erscheint besonders zu einem Zeitpunkt vor oder während der Einführungsphase von Codierrichtlinien sinnvoll, da in dieser Phase ein besonders hoher Nutzen einer erweiterten Diagnosevalidierung vorliegt, man aber annehmen darf, dass die Codierrichtlinien die Diagnosequalität weitestgehend verbessern werden.

Die Analysen machten aber auch deutlich, dass die Grenzen des Verfahrens durch den groben zeitlichen Kontext von Arznei- und Diagnose Information auf Quartalsbasis gesetzt sind. Daher sollte man bei der Einführung von Codierrichtlinien im ambulanten Bereich dringend eine möglichst gute Zuordnung von Arzneimittelverordnung und Diagnosen fordern. Sollte sich eine Umsetzung in der Praxis als zu komplex erweisen, so wäre es bereits ein großer Schritt hin zu einer besseren Zuordnung, wenn die Diagnosen mit einem Datum versehen werden müssten. So ließe sich unter Nutzung der hier beschriebenen Methodik und der neu zur Verfügung stehenden zeitlichen Referenz ein signifikant besseres Ergebnis erwarten.

Perspektivisch kann eine Anwendungsmöglichkeit der hier vorgestellten Methodik in der Identifizierung von statistischen Auffälligkeiten liegen und darin, Neben- und Wechselwirkungen zwischen Arzneimitteln zu indizieren. In Chen et al. (2005) ist beispielsweise eine Anwendung der Assoziationsanalyse zur Identifikation von medikamentösen Nebenwirkungen vorgestellt worden. Eine solche Analyse würde insbesondere im Kontext der elektronischen Gesundheitskarte Sinn machen und deren Ziele unterstützen.

## Literatur

- Bollinger, T. (1996): Assoziationsregeln – Analyse eines Data Mining Verfahrens. In: Zeitschrift Informatik Spektrum. Ausgabe 19. S. 257-261.
- Chen, J., H. He, J. Li, H. Jin, D. McAullay, G. Williams, R. Sparks und C. Kelman (2005). Representing Association Classification Rules Mined from Health Data. Lecture Notes in Computer Science. In: Knowledge Based Intelligent Information and Engineering Systems. Berlin und Heidelberg. S. 1.225-1.231.
- Giersiepen, K., H. Pohlabein, G. Egidi und I. Pigeot (2007): Die ICD-Kodierqualität für Diagnosen in der ambulanten Versorgung. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz. Nummer 8. Berlin und Heidelberg.
- Hettich, S., und H. Hippner (2009): [www.ku-eichstaett.de/Fakultaeten/WWF/Lehrstuehle/WI/Lehre/dm\\_v/HF\\_sections/content/DM%205.pdf](http://www.ku-eichstaett.de/Fakultaeten/WWF/Lehrstuehle/WI/Lehre/dm_v/HF_sections/content/DM%205.pdf) (Download 22. Juli 2009).
- Jaeckel, R. (2008). Einfluss des „Morbi-RSA“ auf die Versorgung mit innovativen Arzneimitteln. In: Gesellschaftspolitische Kommentare. Nummer 9/08. S. 6-8.
- Tang Z. H., und J. Mac Lennan (2005): Data Mining with SQL Server 2005. Wiley.
- Weisser, A., G. Endel, P. Filzmoser und M. Gyimesi (2008): ATC -> ICD – evaluating the reliability of prognoses for ICD-10 diagnoses derived from the ATC-Code of prescriptions. 24th Patient Classification Systems International (PCSI) Working Conference Lisbon, Portugal. 8.-11. October 2008. BMC Health Services Research 2008. 8 (Suppl 1): A10.