

**Management von Leistungsrisiken einer Krankenkasse  
mit Predictive Modeling**

Diplomarbeit

zur Erlangung des Grades eines Diplom-Ökonomen der  
Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von:

Name:

Töller

Vorname:

Jos



Erstprüfer:

Prof. Dr. Michael H. Breitner

Hannover, den 22.02.2007

## Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>III</b>
<b>Tabellenverzeichnis .....</b>	<b>IV</b>
<b>Abkürzungsverzeichnis .....</b>	<b>V</b>
<b>1 Einleitung .....</b>	<b>1</b>
1.1 Problemstellung und Zielsetzung .....	1
1.2 Gang der Untersuchung .....	3
<b>2 Begriffsbestimmungen und Grundlagen .....</b>	<b>4</b>
2.1 Leistungsrisiken .....	4
2.1.1 Definition .....	4
2.1.2 Struktur der Leistungsausgaben .....	6
2.1.3 Steuerungsmöglichkeiten .....	8
2.2 Predictive Modeling .....	10
2.2.1 Definition .....	10
2.2.2 Abgrenzung zu Prior-Expense-Analysen .....	12
2.2.3 Einsatzmöglichkeiten zum Management der Leistungsrisiken .....	14
<b>3 Theoretische Ausgestaltung von Predictive Models .....</b>	<b>18</b>
3.1 Modellierungsansätze .....	18
3.2 Risk Grouper basierte Modelle .....	21
3.2.1 Grundlegende Konzeption .....	21
3.2.2 Datenbasis .....	23
3.2.2.1 Morbiditätsindikatoren .....	23
3.2.2.2 Codierungen .....	25
3.2.3 Risiko- und Ausgabenschätzung .....	27
3.2.3.1 Zellenansatz .....	27
3.2.3.2 Zuschlagsansatz .....	28
3.2.3.3 Vergleich der Ansätze .....	30
3.2.4 Zeitgleiche und prospektive Anwendung .....	31
3.2.5 Modelloutput .....	32
3.2.6 Kritische Würdigung .....	34
<b>4 Vergleich bestehender Predictive Modeling Systeme .....</b>	<b>37</b>
4.1 Vergleichskriterien .....	37

4.2	Übersicht ausgewählter Systeme .....	39
4.2.1	ACG Case-Mix System .....	39
4.2.2	IHCIS Impact Pro .....	41
4.2.3	DxCG Risk Smart Global .....	43
4.3	Empirische Testergebnisse .....	45
4.3.1	Maße zur Schätzgenauigkeit.....	45
4.3.2	SOA Risk Adjuster Study.....	47
4.3.3	IGES/Lauterbach/Wasem-Studie .....	53
4.4	Anpassungsmöglichkeit an GKV-Strukturen .....	57
4.5	Kritische Würdigung und Systemauswahl .....	60
<b>5</b>	<b>Implementierung eines Risk Grouper basierten Predictive Models.....</b>	<b>64</b>
5.1	Vorgehensmodell.....	64
5.1.1	Ableitung eines Vorgehensmodells.....	64
5.1.2	Implementierungsphasen .....	67
5.2	Kritische Erfolgsfaktoren bei der Implementierung.....	69
5.2.1	Risiko-Framework .....	69
5.2.2	Systemauswahl .....	71
5.2.3	Datenqualität.....	73
5.2.4	Klassifikation und Kostengewichte .....	74
5.2.5	Risikomanagement-Prozess.....	76
5.2.6	Risiko-Berichtswesen .....	77
<b>6</b>	<b>Schluss .....</b>	<b>79</b>
6.1	Fazit .....	79
6.2	Ausblick.....	81
	<b>Literaturverzeichnis .....</b>	<b>83</b>
	<b>Ehrenwörtliche Erklärung.....</b>	<b>92</b>

## 1 Einleitung

### 1.1 Problemstellung und Zielsetzung

Die Motivation für diese Arbeit entstand vor dem Hintergrund des Vorhabens einer großen gesetzlichen Krankenkasse, aufbauend auf der Analyse ihrer Versichertenbestandsrisiken ein Risikomanagement zu implementieren. Ziel der Krankenkasse ist es, mittels einer (morbidityorientierten) Klassifizierung des Versichertenbestandes durch zuverlässige Prognosemodelle den erwarteten aktuellen und zukünftigen Ressourceneinsatz planbar zu machen und so zu einer Erhöhung der Wirtschaftlichkeit und der Qualität der Gesundheitsversorgung zu gelangen.<sup>1</sup> Darüber hinaus soll die „strategische Analyse- und Prognosefähigkeit“<sup>2</sup> der Kasse ausgebaut werden, indem z. B. Prognosen über risikobehaftete Entwicklungen im Versichertenbestand abgeleitet werden.<sup>3</sup> Die Krankenkasse entgegnet somit dem wachsenden Wettbewerbsdruck und der zunehmenden Dynamik im Umfeld der gesetzlichen Krankenversicherung, u. a. in Folge kassenartenübergreifenden Fusionen und der für 2009 geplanten Einführung des morbidityorientierten Risikostrukturausgleichs (Morbi-RSA).

Die Suche nach etablierten Ansätzen zur Prognose des versichertenbezogenen Ressourcenbedarfes im Vorfeld dieser Arbeit führte zu der Erkenntnis, dass in den USA seit längerem Ansätze bestehen, in denen aufbauend auf der Morbidität als erklärende Variable der zukünftige Ressourcenbedarf versichertenbezogen prognostiziert wird. Die dazu notwendigen Methoden entstammen einer Disziplin, die als „*Predictive Modeling*“<sup>4</sup> bezeichnet wird. Zentrale Fragestellungen des Predictive Modeling - Ansatzes sind u. a.:

- Wie entwickeln sich die Leistungsausgaben für eine bestimmte Altersgruppe?
- Wer sind die zukünftigen Hochkosten-Versicherten und welche Merkmale kennzeichnen diese Versicherten?

---

<sup>1</sup> Vgl. *Nicht öffentl. Verf.* (2006), S. 3.

<sup>2</sup> *Nicht öffentl. Verf.* (2006), S. 3.

<sup>3</sup> Vgl. *Nicht öffentl. Verf.* (2006), S. 3.

<sup>4</sup> Zu Deutsch: prädiktive Modellierung.

In Deutschland war der Einsatz von Predictive Modeling im Gesundheitswesen bis 2004 noch weitgehend unerforscht.<sup>5</sup> Mit dem geplanten Einsatz von Predictive Models im morbiditätsorientierten Risikostrukturausgleich<sup>6</sup> und in der ärztlichen Vergütung<sup>7</sup> steigt jedoch das Interesse im deutschen Gesundheitswesen, so dass mittlerweile auch Krankenkassen Interesse am Einsatz von Predictive Modeling zeigen<sup>8</sup>. Die dazu notwendige Datengrundlage ist bei den Kassen durch die elektronische Abrechnung mit dem Leistungserbringer prinzipiell vorhanden. Inzwischen haben einige US-amerikanische Anbieter ihre Predictive Modeling - Systeme an die Bedingungen des deutschen Gesundheitswesens angepasst, so dass der Anpassungsaufwand für diese Systeme geringer denn je erscheint.

Die Darstellung und kritische Würdigung der Potentiale von Predictive Modeling zum Management der Leistungsrisiken einer Krankenkasse bilden den Gegenstand dieser Untersuchung. Ziel der vorliegenden Arbeit ist es, Antworten auf folgende, für Entscheidungsträger einer Krankenkasse aus fachlicher Sicht relevante, W-Fragen zu geben:

1. Was ist Predictive Modeling und wie kann es zum Management der Leistungsrisiken der Krankenkasse eingesetzt werden?
2. Welche verschiedenen methodischen Predictive Modeling - Ansätze bestehen und welche Vor-/ Nachteile sind mit diesen verbunden?
3. Welche Predictive Modeling - Systeme sind am Markt verfügbar und wie hoch ist der Anpassungsbedarf an die Strukturen einer deutschen Krankenkasse?
4. Wie sollte ein solches System sinnvoller Weise implementiert werden?

Der Untersuchungsgegenstand beschränkt sich auf die gesetzliche Krankenversicherung. Zwar ist der Einsatz von Predictive Modeling ebenso bei privaten Krankenversicherern denkbar, aufgrund systematischer Unterschiede<sup>9</sup> zwischen GKV und PKV ist es jedoch notwendig, die Betrachtung auf eine Art der

---

<sup>5</sup> Vgl. *Elsner, C. et al.* (2004), S. 1724.

<sup>6</sup> Der Einsatz von Predictive Models im morbiditätsorientierten Risikostrukturausgleich wurde im Rahmen der *IGES/Lauterbach/Wasem-Studie* analysiert. Vgl. *IGES et al.* (2004a).

<sup>7</sup> Vgl. *Stillfried, D. Graf v.* (2006).

<sup>8</sup> Vgl. *Reschke, P. / Sehlen, S.* (2005), S. 19; *ACG* (2007).

<sup>9</sup> Diese Unterschiede betreffen sowohl die Informations- und Abrechnungsprozesse der Krankenkassen als auch die Steuerungsmöglichkeit der Leistungsrisiken, die für gesetzliche Krankenkassen durch das 5. Sozialgesetzbuch (SGB V) reguliert werden.

Krankenversicherung einzuschränken um Untersuchungen von praktischer Relevanz durchführen zu können.

## **1.2 Gang der Untersuchung**

Der Untersuchungsaufbau der Arbeit orientiert sich an den eingangs in der Problemstellung formulierten W-Fragen.

Im zweiten Kapitel werden die für diese Arbeit grundlegenden Begriffe Leistungsrisiken und Predictive Modeling erläutert und Einsatzmöglichkeiten von Predictive Modeling zur Steuerung der Leistungsrisiken aufgezeigt. Damit wird die erste W-Frage aufgegriffen: Was ist Predictive Modeling und wie kann es zum Management der Leistungsrisiken der Krankenkasse eingesetzt werden?

Ziel des dritten Kapitels ist es, Antworten auf die Fragen nach den methodisch-technischen Predictive Modeling - Ansätzen und deren Vor-/ Nachteilen (zweite W-Frage) zu geben. Dazu werden alternative Ansätze innerhalb der Predictive Modeling - Konzeption beschrieben und ein Vergleich sowie eine kritische Würdigung der theoretischen Ausgestaltung von Predictive Models vorgenommen.

Im vierten und fünften Teil dieser Arbeit sind Antworten auf die Frage „Welche Systeme sind am Markt verfügbar und wie hoch ist der Anpassungsbedarf an die Strukturen einer deutschen Krankenkasse? und „Wie sollte ein System sinnvoller Weise implementiert werden?“ zu erarbeiten. Aufbauend auf den vorangegangenen theoretischen Untersuchungen werden hierzu alternative Predictive Modeling - Systeme und deren Anpassungsbedarf an die GKV-Strukturen evaluiert (viertes Kapitel) und das Vorgehen zur Implementierung eines Predictive Modeling - Systems in einer Krankenkasse abgeleitet (fünftes Kapitel).

Das sechste Kapitel fasst die Erkenntnisse der Arbeit abschließend zusammen und nimmt eine kritische Gesamtbetrachtung des Einsatzes von Predictive Modeling zum Management der Leistungsrisiken vor.

## 6 Schluss

### 6.1 Fazit

Predictive Models in der hier untersuchten Ausführung als *Risk Grouper basierte Predictive Models* greifen mit der Morbidität des Versicherten eine wesentliche Determinante zur Prognose der Leistungsausgaben auf. Die im 4. Kapitel der Arbeit vorgestellten Testergebnisse zeigen jedoch, dass bei prospektiver Anwendung des Modells und ohne Begrenzung der Schadensdaten lediglich ein  $R^2$  von 24%, bei zeitgleicher Anwendung ein  $R^2$  von 48% erreicht wird.<sup>201</sup> D. h. der Anteil der erklärten Varianz an der Gesamtvarianz der Ausgabenunterschiede zwischen den Versicherten eines Versichertenkollektives beträgt lediglich 24% bzw. 48%. Dies ist zum einen darauf zurückzuführen, dass ein großer Teil der Ausgaben aus prospektiver Sicht rein zufällig anfällt und nicht durch die aktuell für den Versicherten vorliegenden Morbiditätsindikatoren prognostizierbar ist. Darüber hinaus ist jedoch davon auszugehen, dass die in den Risk Groupen verwendeten Morbiditätsindikatoren allein nicht sämtliche Ausgabenunterschiede zwischen den Versicherten erklären können. Die Vernachlässigung komplexer Merkmale wie der Anbieterstruktur am Wohnort des Versicherten und des Versichertenverhaltens sind weitere Erklärungsansätze für den hohen Anteil der nicht erklärten Varianz der Ausgabenunterschiede. Ungeklärt bleibt, ob die in den angeführten Studien erzielte Schätzgüte im praktischen Einsatz bei einer Krankenkasse überhaupt erreicht werden kann, da aus den Stichproben nachträglich Versicherte mit unvollständigen und fehlerhaften Daten von der Prognose ausgeschlossen wurden.

Die Ausführungen zur theoretischen Ausgestaltung von Risk Grouper basierten Predictive Models zeigten, dass die Risk Grouper aufgrund ihrer Modellstruktur mit der zusätzlichen Berücksichtigung komplexer Merkmale überfordert wären, da der Klassifikationsalgorithmus sämtliche Merkmale des Versicherten zu berücksichtigen hätte. Die Modellentwickler müssten darüber hinaus die Zusammenhänge zwischen komplexen Merkmalen und der Morbidität kennen, um sie in den Klassifikationsalgorithmus einpflegen zu können. An dieser Stelle wird das, dem Modellierungsansatz der Risk Grouper immanente, Problem deutlich: Um mit Risk

---

<sup>201</sup> Diese Werte beziehen sich auf das IPHCC+RxGroups Modell in der IGES/Lauterbach/Wasem-Studie. In der SOA-Studie lag der höchste  $R^2$ -Wert bei 15,4% in der prospektiven Anwendung und 46,6% in der zeitgleichen Anwendung, jeweils erzielt vom DCG/HCC-Modell.

Groupern Hypothesen aus Daten zu gewinnen, müssen im ersten Schritt selbst umfangreiche Hypothesen angenommen werden. Aus diesem Grund arbeiten Risk Grouper mit nur wenigen, aber aussagekräftigen Prädiktoren. Künstliche Neuronale Netze – die sogenannte dritte Generation von Predictive Models – könnten sich als Alternative erweisen, da die aufwendige Modellierung der Abhängigkeiten vom Modellentwickler nicht „per Hand“ programmiert werden muss, sondern eine Anpassung des KNN durch den iterativen Lernprozess anhand von Beispielen erfolgt. Bei entsprechender Generalisierungsfähigkeit des Netzes kann dieses durch Training an neue Daten angepasst werden. Im Gegensatz zu den hier untersuchten Risk Groupern werden von MLPs zudem auch nicht-lineare Abhängigkeiten zwischen den Prädiktoren und den Output-Variablen berücksichtigt.

Dennoch sind mit den etablierten Risk Grouper basierten Predictive Models bessere Ergebnisse zu erzielen als unter alleiniger Verwendung der soziodemographischen Variablen Alter und Geschlecht oder auf Basis der Ausgaben vergangener Jahre. Für den Einsatz von Risk Groupern spricht zudem, dass die notwendige Datengrundlage zumindest für den stationären Bereich und für Arzneimitteldaten in ausreichender Qualität zeitnah verfügbar ist. Für den künftigen Morbi-RSA werden die Krankenkassen ohnehin Personenstammdaten und Morbiditätsinformationen ihrer Versicherten liefern müssen und dabei die notwendige Datenqualität sicherzustellen haben, da jeder Datenausfall in den Morbiditätsinformationen mit einem Verlust an Ausgleichszahlungen einhergeht und zur Folge hätte, dass nur alters- und geschlechtsspezifische Basisbeträge angesetzt werden. Es stellt sich also die Frage, warum die Krankenkasse das Predictive Model nicht selbst einsetzt, wenn sie bereits für den Morbi-RSA die notwendige Datengrundlage unter der Nebenbedingung einer entsprechenden Qualitätssicherung bereitstellen muss. Vor dem Hintergrund der Tatsache, dass einige Anbieter von Risk Grouper basierten Predictive Models ihre Systeme bereits an die im deutschen Gesundheitswesen verwendeten Codierungsstandards angepasst haben, scheint die Implementierung für große Einzelkassen tragbar.

Vor dem Hintergrund der im 2. Kapitel aufgezeigten Einsatzbereiche von Predictive Modeling in Krankenkassen ist festzuhalten, dass Nutzenpotentiale für den Einsatz von Predictive Modeling vor allem in den Bereichen Leistung und Vertrag bestehen. Für den Vertriebsbereich ist die fehlende Einbindung von Finanzierungsdaten ein potentiell



Hindernis, da sich gute Risiken in der GKV auch an den einkommensabhängigen Beiträgen messen.<sup>202</sup> Doch selbst bei Einbindung von Finanzierungsinformationen wäre die Aussagekraft gering, da für Neukunden mangels historischer Morbiditätsinformationen nur ein alters- und geschlechtsspezifischer Basisbetrag ermittelt werden kann. Aufgrund des Prognosehorizontes von 12 bis 18 Monaten sind die Risk Grouper zudem nur für die kurzfristige Anwendung geeignet. Die strategische Ausrichtung der Krankenkasse auf Basis der Predictive Modeling - Prognosen ist nicht möglich.

Bei der Implementierung eines Predictive Modeling - Systems sollten neben den technischen Fragen der Auswahl und Anpassung des Systems sowie der Bereitstellung der notwendigen Daten in ausreichender Datenqualität auch der Einfluss auf die bestehenden Prozesse und das Berichtswesen analysiert werden. Entsprechend liegt im aufgezeigten Vorgehensmodell ein Großteil der organisatorischen Implementierungsmaßnahmen noch vor der Systemauswahl, damit ein frühzeitiger organisatorischer Wandel im Sinne des Change Managements vollzogen wird.

## 6.2 Ausblick

Zukünftig ist davon auszugehen, dass Anbieter von Predictive Modeling - Systemen verstärkt verschiedene Modellierungsmethoden kombinieren werden, um die Nachteile bestehender Lösungen zu verringern. Die Lösung könnte in der Kombination der verschiedenen Predictive Modeling - Methoden liegen, wobei zunächst eine Aufspaltung der Daten nach festen Regeln in Untergruppen und darauf aufbauend die Analyse der Daten mit statistischen Methoden und Künstlichen Neuronalen Netzen erfolgt.<sup>203</sup> Die Integration weiterer Datenquellen wie genetischer Daten und Laborwerte könnte darüber hinaus die Prognosegüte erhöhen.<sup>204</sup> Jedoch entscheidet nicht nur der Umfang und die Vielfalt der Daten über die Prognosegüte, sondern auch die Qualität der verwendeten Daten. Von zukünftigen Systemen wird erwartet, dass sie eine Empfindlichkeit gegenüber geringer Datenqualität aufweisen und in Form von Plausibilitätschecks (Coding Audit) selbst erkennen, dass eine vom Leistungserbringer erstellte Diagnose nicht zutreffen kann.<sup>205</sup> Ob sich die Datenqualität für den aktuell

---

<sup>202</sup> Vgl. *Wasem, J. et al.* (2005), S. 8.

<sup>203</sup> Vgl. *Elsner, C. et al.* (2004), S. 1725.

<sup>204</sup> Vgl. *Mazzi, C. et al.* (2004).

<sup>205</sup> Vgl. *Mazzi, C. et al.* (2004).

kritischen Bereich der ambulanten Versorgung in Folge der für 2007 geplanten Umstellung auf die morbiditätsorientierte Vergütung der Leistungserbringer verbessert, bleibt abzuwarten.