

BIG DATA IM KRANKENVERSICHERUNGSMARKT

Relevanz, Anwendungen, Chancen und Hindernisse

Prof. Dr. Dubravko Radić
Dr. Marija Radić
Nadine Metzger
Anne-Sophie Pohl
Nicole Schuldt

Leipzig, 28. Oktober 2016

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	I
Abbildungsverzeichnis.....	II
1 Big Data: Trend oder gängige Praxis?.....	1
2 Executive Summary.....	2
3 Ergebnisse der Befragung.....	4
3.1 <i>Vorgehensweise</i>	4
3.2 <i>Studienergebnisse</i>	6
3.2.1 Begriffsverständnis	6
3.2.2 Strategische Einsatzsituation	7
3.2.3 Daten- und Analysesituation.....	14
4 Fazit und Ausblick	20
5 Quellenverzeichnis.....	21
6 Elemente einer Big Data-Strategie.....	22

Abkürzungsverzeichnis

AOK	Allgemeine Ortskrankenkasse
BMG	Bundesministerium für Gesundheit
GKV	Gesetzliche Krankenversicherung
PKV	Private Krankenversicherung
vdek	Verband der Ersatzkassen e. V.

Abbildungsverzeichnis

Abb. 01:	Grundgesamtheit und Stichprobe der Onlinebefragung im Vergleich.....	4
Abb. 02:	Teilnehmende gesetzliche Krankenkassen nach Größe.....	4
Abb. 03:	Aktueller Einsatz von Big Data-Technologien in deutschen Krankenversicherungen	5
Abb. 04:	Kenntnisse im Hinblick auf Big Data-Anwendungen	6
Abb. 05:	Verständnis des Begriffs Big Data nach Art der Krankenversicherung.....	6
Abb. 06:	Einsatz von Big Data-Technologien nach Krankenversicherungsart und -größe	7
Abb. 07:	Einführungsphasen von Big Data-Lösungen von Early Adopters und Early / Late Majority	7
Abb. 08:	Existenz einer Big Data-Strategie bei Early Adopters und Early / Late Majority.....	8
Abb. 09:	Verantwortlichkeit von Big Data-Anwendungen bei Early Adopters und Early / Late Majority	9
Abb. 10:	Ziele des Einsatzes von Big Data-Technologien von Early Adopters und Early / Late Majority	10
Abb. 11:	Big Data-Anwendungsbereiche von Early Adopters und Early / Late Majority.....	11
Abb. 12:	Hemmnisse bei der Big Data-Umsetzung von Early Adopters und Early / Late Majority	12
Abb. 13:	Eingeschätzte Hemmnisse bei der Big Data-Umsetzung von Laggards und Lack of Information	13
Abb. 14:	Analytische Nutzung der festgelegten Datenquellen der Early Adopters und Early / Late Majority	14
Abb. 15:	Datenquellen und analytisch genutzte Datenquellen der Early Adopters und Early / Late Majority	14
Abb. 16:	Treiber für große Datenmengen von Early Adopters und Early / Late Majority.....	16
Abb. 17:	Spezielle Formen der Datenanalyse bei Early Adopters und Early / Late Majority.....	16
Abb. 18:	Fragestellungen von Datenanalysen bei Early Adopters und Early / Late Majority.....	17
Abb. 19:	Vorstellbare Änderungen in der Datenanalyse durch Big Data	18
Abb. 20:	Änderungen in der Datenanalyse durch das Thema Big Data	19
Abb. 21:	Die wichtigsten Ergebnisse im Überblick	20

1 Big Data: Trend oder gängige Praxis?

Unser Gesundheitssystem steht vor enormen Herausforderungen wie dem demographischen Wandel und einer steigenden Anzahl an chronisch Kranken (Zillmann 2013). Infolgedessen werden die Gesundheitsausgaben auch in den kommenden Jahren weiter wachsen (Statistisches Bundesamt 2016). Verstärkt durch zahlreiche Reformen im Gesundheitsmarkt verschärft sich die Finanzlage der Kostenträger stetig (Bundesregierung 2015; vdek 2016). Der Einsatz von technologischen Konzepten und Analysetechniken soll dabei helfen, diese Herausforderungen zu bewältigen. Aufgrund der digitalen Erfassung und Speicherung der Patienten- sowie Behandlungsdaten stehen bereits heute gewaltige Datenvolumina zur Verfügung. Bis zum Jahr 2020 wird ein jährlich generiertes digitales Datenvolumen von 40 Zettabytes (40×10^{21} Bytes) weltweit prognostiziert (EMC Corporation 2012). Ein Drittel dieser Daten werden aus dem Gesundheitsmarkt stammen. Diese sind in unverarbeiteter Form relativ nutzlos. Oberstes Ziel sollte es deshalb sein, die erzeugten Daten auszuwerten und deren enormes Potenzial intelligent zu erschließen. Da sich jedoch das Datenspektrum aufgrund von Größe und Struktur oft vom bisherigen Bestand unterscheidet, kann es mit herkömmlichen Ressourcen nicht nutzenbringend ausgewertet werden, weshalb nach neuen Big Data-Konzepten verlangt wird.

Was versteht man unter Big Data?

Unter Big Data versteht man die Gewinnung und Nutzung großer Datenmengen aus vielfältigen Quellen mit einer hohen Verarbeitungsgeschwindigkeit zur Erzeugung wirtschaftlichen Nutzens (Diessner et al. 2015).

Wesentliche Charakteristika von Big Data sind die sogenannten 3 Vs:

- Datenmenge (**Volume**): Terabytes oder mehr
- Datenvielfalt (**Variety**): Verarbeitung auch unstrukturierter Daten
- Geschwindigkeit (**Velocity**): oft Verarbeitung in Echtzeit

Aufgrund der erhöhten Geschwindigkeit und größeren Speicherkapazität der Rechen-technik können heute Datenmengen analysiert werden, die vor wenigen Jahren noch nicht zugänglich waren.

Zwar gibt es zahlreiche Studien zum Thema Big Data, aber nur wenige beziehen sich auf den Krankenversicherungsmarkt. Zillmann (2013) kommt in seinem Trendpapier zu der Erkenntnis, dass bei Krankenkassen diesbezüglich Aufholbedarf besteht und Lösungen gefunden werden müssen, um die wachsenden Datenmengen zu bewältigen. Brunner (2014) untersucht dieses Thema am Beispiel einer einzelnen Krankenkasse. Die vorliegende Studie des Fraunhofer IMW, der Gesundheitsforen Leipzig GmbH und der Universität Leipzig verfolgt die Zielsetzung, diese Forschungslücke zu schließen. Im Rahmen der vorliegenden Studie soll empirisch untersucht werden, welche Relevanz das Thema Big Data für den Krankenversicherungsmarkt hat. Auf Basis einer umfassenden Befragung unter Vertretern von privaten und gesetzlichen Krankenversicherungen soll insbesondere geklärt werden, welches Verständnis diese vom Begriff Big Data haben, inwieweit der Einsatz von Big Data-Technologien bzw. -Anwendungen von strategischer Bedeutung ist und welche tatsächliche Daten- und Analysesituation aktuell vorherrscht.

2 Executive Summary

Begriffsverständnis

- Der Begriff Big Data wird relativ ähnlich wahrgenommen
- Big Data wird meist anhand der drei Eigenschaften definiert:
 - Datenmenge (Volume): Terabytes oder mehr
 - Datenvielfalt (Variety): Verarbeitung auch unstrukturierter Daten
 - Geschwindigkeit (Velocity): oft Verarbeitung in Echtzeit

Strategische Einsatzsituation

- Krankenversicherungen setzen bereits heute größtenteils Big Data-Technologien ein oder haben es geplant und informieren sich derzeit.
- In vielen Unternehmen besteht eine Big Data-Strategie. Zwar ist eine solche von Vorteil, jedoch nicht zwingend für die Umsetzung notwendig.
- Die Verantwortung für Betrieb und Umsetzung der Technologien liegen bei der IT. Auftraggeber der Analysen ist überwiegend das Versorgungs- und Leistungsmanagement. Für das Reporting ist die Analytik oder der Bereich Finanzen und Controlling zuständig.
- Mit den Ergebnissen des Einsatzes von Big Data-Technologien werden Ziele aus den Bereichen Finanzen, Steuerung sowie zum Teil Marketing und Vertrieb verfolgt.
- Primäre Big Data-Anwendungsgebiete finden sich vor allem in den Bereichen Markt und Versorgung. Das Gebiet Personal, Recht und Finanzen ist weniger relevant.
- Hemmnisse bei der Umsetzung von Big Data-Anwendungen stellen hauptsächlich Datenschutzbestimmungen, mangelnde zeitliche und personelle Ressourcen dar.

Daten- und Analysesituation

- Es liegen mehrheitlich unternehmensinterne Datensätze vor.
- Die Daten stammen insbesondere aus Versorgungs- und Leistungsabrechnungen. Weitere Daten stammen aus gesetzlichen Quellen wie amtlichen Statistiken und dem Morbi-RSA-Verfahren.
- Treiber für zunehmende Verwendung von Big Data-Analysen stellt der Unternehmensbereich Versorgung dar, der zudem die größten Datenvolumina stellt. Der Unternehmensbereich Markt ist aufgrund von Datenschutzbestimmungen weniger relevant für zukünftige Big Data-Anwendungen.
- In der GKV liegen systembedingt größere Datenvolumina als in der PKV vor.
- Der Datenbestand wird größtenteils kontinuierlich aktualisiert.
- Die relevanten Datenquellen werden von mehr als der Hälfte der Krankenversicherungen für Analysezwecke genutzt.
- Die Datenanalysen werden sowohl manuell als auch automatisch initiiert und durchgeführt.
- Häufige Fragestellungen, die mittels Datenanalysen untersucht werden, lauten: „Was ist geschehen?“ und „Warum ist es geschehen?“ Der Fragestellung „Was könnte geschehen?“ wird bislang nur teilweise nachgegangen. Analysen in der PKV haben einen stärkeren Vergangenheitsbezug als in der GKV.
- Neben den klassischen multivariaten Verfahren werden von einigen Krankenversicherungen auch Geodaten- und Text-Analysen durchgeführt.

Fazit und Ausblick

- Die Datenvielfalt wird heute noch nicht genügend ausgeschöpft: Krankenversicherungen verwenden für Big Data-Analysen fast ausschließlich strukturierte Daten, da unstrukturierte Daten häufig nicht systematisch erfasst werden können.
- Die Krankenversicherungen verfügen über große Datenmengen.
- Die notwendige Technologie zur Durchführung von Big Data Analysen ist vorhanden und stellt kein Hemmnis dar.
- Hemmnisse stellen vor allen Dingen fehlende personelle Ressourcen dar.
- Das Potenzial von Big Data wurde erkannt. Es gilt nun, Big Data-Anwendungen in das strategische und operative Geschäft zu integrieren.
- Big Data ist kein Zukunftstrend mehr, sondern wird in der Praxis bereits angewendet.

3 Ergebnisse der Befragung

3.1 Vorgehensweise

Das Ziel der Studie war es, herauszufinden, welche Relevanz Big Data im Krankenversicherungsmarkt hat. Hierfür wurde ein zweistufiges Verfahren gewählt: In einem ersten Schritt wurde eine Onlineumfrage unter privaten und gesetzlichen Krankenversicherungen (PKV und GKV) durchgeführt, deren Ergebnisse in einem zweiten Schritt von Fachexperten validiert wurden.

Onlinebefragung

Im Rahmen der Onlinebefragung wurden 2.502 Probanden mit unterschiedlichen Funktionen aus 157 Krankenversicherungen angeschrieben. Für die Ergebnisauswertung wurden nur die Antworten derjenigen Teilnehmer berücksichtigt, die eine Vorstellung vom Begriff Big Data hatten, was auf 96 Probanden zutraf. Die Verteilung von GKV und PKV ist im Vergleich zum gesamten deutschen Krankenversicherungsmarkt annähernd repräsentativ (vgl. Abbildung 1).

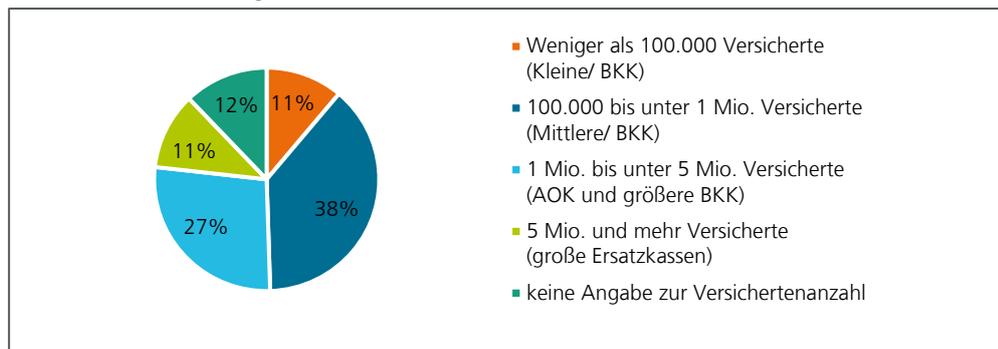
Abb. 01: Grundgesamtheit und Stichprobe der Onlinebefragung im Vergleich



n = 96; Quelle: Eigene Darstellung; BMG (2016), S. 37; Verband der privaten Krankenversicherung (2016).

Die Mehrheit der Befragten ist in der GKV tätig, die im Weiteren deshalb nach Versichertenanzahl differenziert wird (Abbildung 2). An der Befragung nahmen vor allem Führungskräfte sowie Mitarbeiter aus den Bereichen Finanzen und Controlling (29 Prozent), Analytik (27 Prozent) sowie Versorgungs- und Leistungsmanagement (23 Prozent) teil.

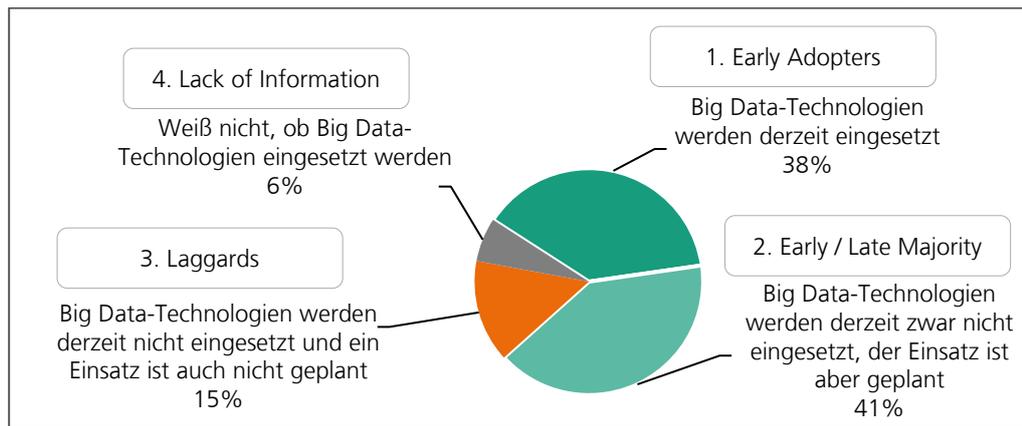
Abb. 02: Teilnehmende gesetzliche Krankenkassen nach Größe



n = 73; Quelle: Eigene Darstellung.

In der Ergebnisauswertung wird zudem nach der Einsatzsituation von Big Data-Technologien unterschieden. Angelehnt an den Innovationszyklus von Rogers (1995, S. 22) lassen sich dabei die folgenden Gruppen unterscheiden: 1) Early Adopters, 2) Early / Late Majority, 3) Laggards und 4) Lack of Information.

Abb. 03: Aktueller Einsatz von Big Data-Technologien in deutschen Krankenversicherungen



n = 96; Quelle: Eigene Darstellung.

Die Gruppen 1 und 2 setzen in ihren Unternehmen aktuell Big Data-Technologien ein oder planen den Einsatz. Die Gruppen 3 und 4 gehören insgesamt zu den Nachzüglern. In ihren Antworten werden daher nicht bestehende Unternehmensrealitäten, sondern vielmehr Erwartungen widerspiegelt.

Experteninterviews

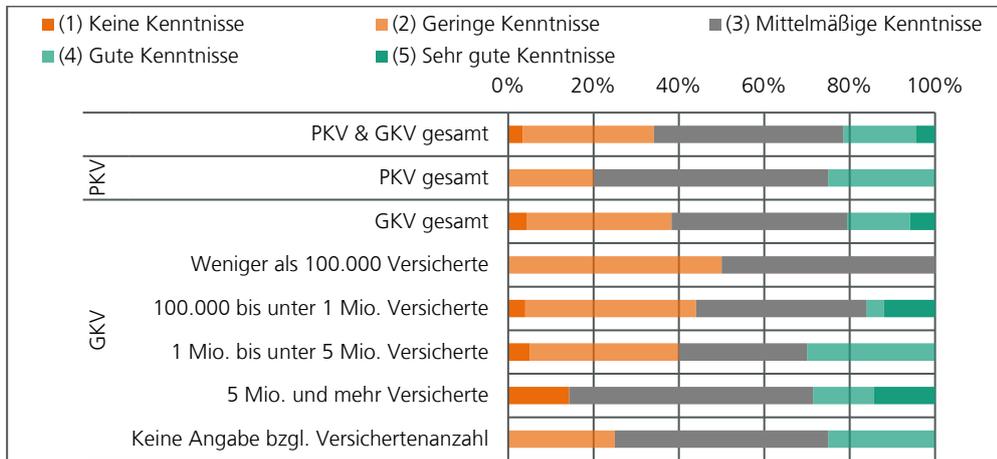
Im Anschluss an die Online-Befragung wurden Experteninterviews mit fünf ausgewählten Vertretern aus GKV und PKV durchgeführt. Die wichtigsten Erkenntnisse werden in Auszügen in Form von Expertenkommentaren präsentiert.

3.2 Studienergebnisse

3.2.1 Begriffsverständnis

Die meisten Probanden geben an, Kenntnisse im Hinblick auf Big Data-Anwendungen vorweisen zu können (Abbildung 4). Vertreter der PKV und größerer GKVen schätzen ihre Kenntnisse diesbezüglich höher ein.

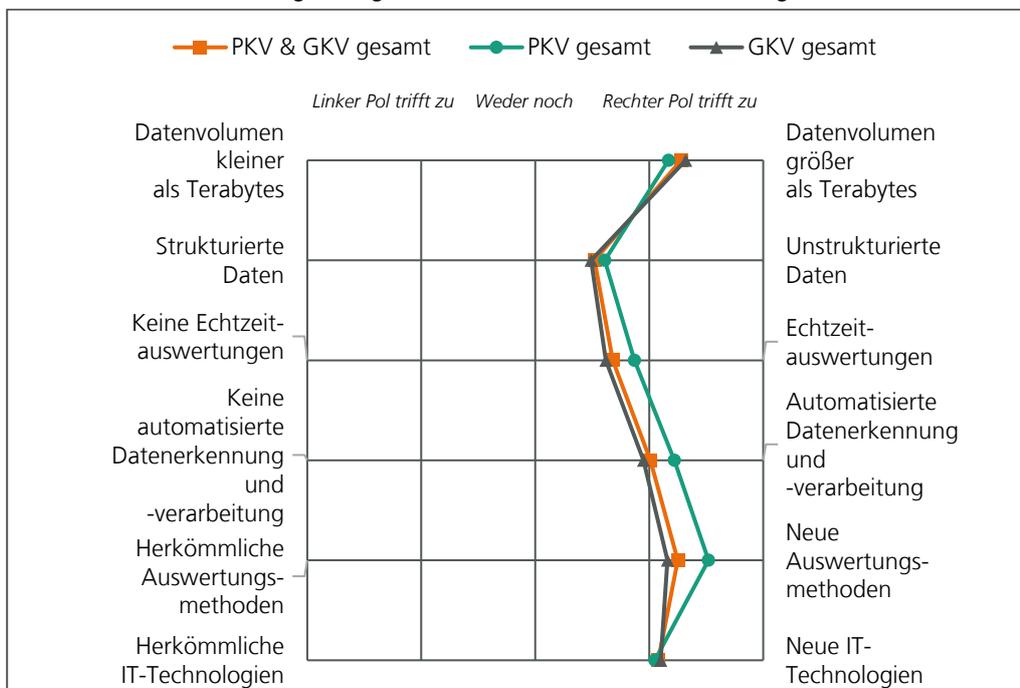
Abb. 04: Kenntnisse im Hinblick auf Big Data-Anwendungen



n = 88; Quelle: Eigene Darstellung.

Die guten Kenntnisse dieser Befragten spiegeln zudem ein relativ einheitliches Verständnis des Begriffs Big Data wider (Abbildung 5). Im Rahmen der Interviews zeigte sich, dass der Begriff häufig mithilfe des 3-V-Modells eingegrenzt wird. Rund ein Drittel der Befragten gab an, unter dem Terminus alle drei Dimensionen (Datenmenge, -vielfalt und Verarbeitungsgeschwindigkeit) zu verstehen. 97 Prozent der Befragten verbinden mindestens eines der Kriterien mit Big Data.

Abb. 05: Verständnis des Begriffs Big Data nach Art der Krankenversicherung



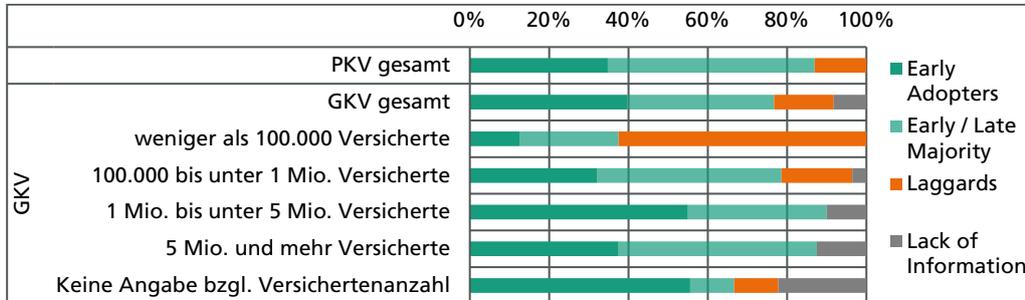
$n_1/n_3/n_4/n_5 = 96$, $n_1/n_6 = 95$; Quelle: Eigene Darstellung.

3.2.2 Strategische Einsatzsituation

Nutzung von Big Data-Technologien

Der Großteil der Befragten gibt an, Big Data-Technologien bereits einzusetzen. Dabei hat die PKV die Nase vorn. Bei der GKV zeigen sich deutliche Unterschiede zwischen größeren und kleineren Krankenkassen.

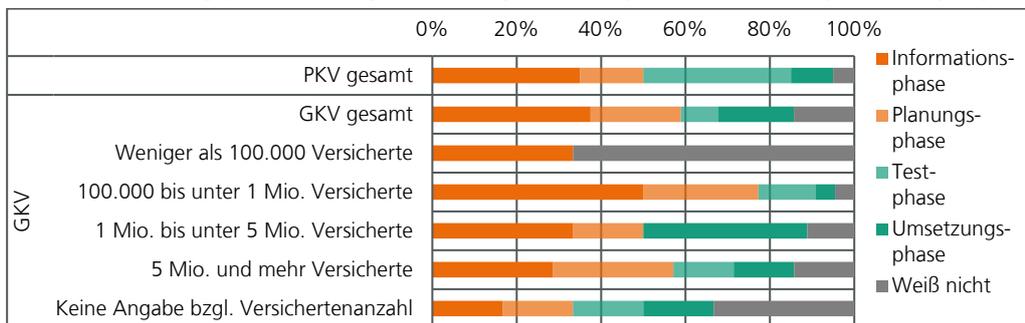
Abb. 06: Einsatz von Big Data-Technologien nach Krankenversicherungsart und -größe



$n_{PKV} = 23$, $n_{GKV} = 73$; Quelle: Eigene Darstellung.

Um den Entwicklungsstand von Big Data-Aktivitäten im eigenen Unternehmen heraus zu arbeiten, sollten die Teilnehmer aus den Gruppen der Early Adopters und Early / Late Majority angeben, in welcher Phase sich ihr Unternehmen aktuell befindet (Abbildung 7). Auch hier lässt sich ablesen, dass die PKV im Gegensatz zur GKV im Durchschnitt bereits häufiger die Informations- und Planungsphase durchlaufen hat. Die großen GKV liegen mindestens gleichauf, während kleinere Kassen bei dem Thema noch zurückhaltend sind.

Abb. 07: Einführungsphasen von Big Data-Lösungen von Early Adopters und Early / Late Majority



$n_{PKV} = 20$, $n_{GKV} = 56$; Quelle: Eigene Darstellung.

Dass sich eine klare unternehmensübergreifende Einordnung in eine der Phasen als relativ schwierig gestaltet, zeigte sich im Interview mit den Experten. Problematisch hierbei sei vor allem die themen- oder projektabhängige Einschätzung des Fortschritts.

Technikaffine Bereiche – wie z. B. IT, Analytik oder Controlling – arbeiten oft unmittelbar mit Big Data-Instrumenten. Die in Voranalysen gebündelten Daten bilden den Ausgangspunkt zur Weiterverarbeitung in anderen Abteilungen. Versorgungsbezogene Fachabteilungen nutzen hierfür eher einfache Analysewerkzeuge oder führen eigene Analysen durch, um konkrete Ergebnisse zu erhalten.

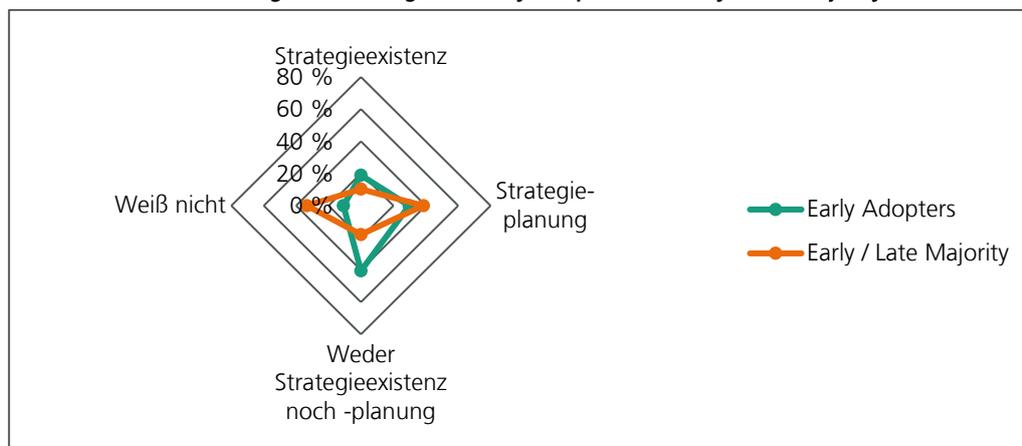
Matthias Waack, Gesundheitsforen Leipzig

Big Data-Strategie

Obwohl bei einem Großteil der Krankenversicherungen Big Data-Technologien eingesetzt werden oder der Einsatz geplant ist, verfügt nur etwa ein Viertel über eine explizite Big Data-Strategie (Abbildung 8). 40 Prozent der Early Adopters verfolgen weder eine Strategie noch planen sie eine solche. Weitere rund 40 Prozent der Early Adopters und der Early / Late Majority befinden sich aktuell in der Strategieplanung. Die Ergebnisse unterscheiden sich dabei kaum hinsichtlich der Art der Krankenversicherung.

Diese widersprüchlich erscheinenden Ergebnisse wurden im Rahmen der Experteninterviews näher beleuchtet: Im Gegensatz zu einem strategisch angestoßenen Top-Down-Ansatz wird die Umsetzung von Big Data-Anwendungen häufig ohne eine explizite Strategie durch einen agilen iterativen Bottom-Up-Prozess vorangetrieben. Einig waren sich die Experten darin, dass eine klare Strategie viele Vorteile mit sich bringt. Wie eine solche aussehen könnte, wird kurz im Anhang diskutiert.

Abb. 08: Existenz einer Big Data-Strategie bei Early Adopters und Early / Late Majority



$n_{\text{Early Adopters}} = 37$, $n_{\text{Early / Late Majority}} = 39$; Quelle: Eigene Darstellung.

Was beinhaltet für Sie eine gute Big Data-Strategie?

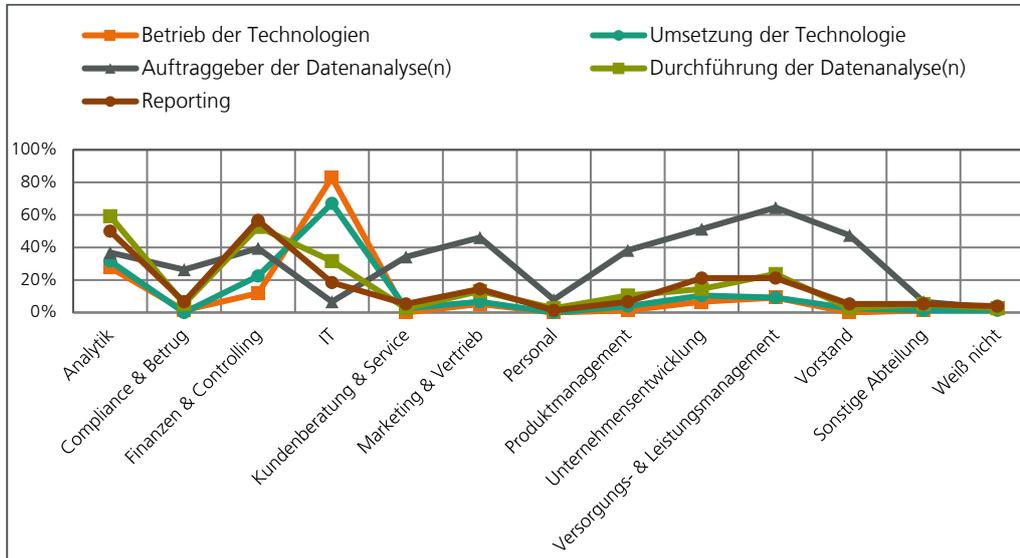
Die Grundlage für eine gute Big Data-Strategie bildet ein ausgereiftes Government-Konzept. Dies beinhaltet die zentralen Fragestellungen: Welche Entscheidungen sind durch Wen in welcher Rolle bezüglich Was (Information) Wann (Zeitpunkt) Wie (Prozess) zu treffen. Dadurch werden die einzelnen Stakeholder innerhalb des Unternehmens identifiziert. Innerhalb der Gesetzlichen Krankenversicherung besitzt der Datenschutz einen enormen Stellenwert. Aus diesem Grund ist es unerlässlich, den Einsatzzweck der Daten zu definieren und den Datenfluss abzubilden. Einen Zugriff auf die Daten erhält in der Folge ein bestimmter Anwenderkreis für einen definierten Anwendungsfall.

Stefan Baßler, AOK PLUS

Verantwortlichkeit bei Big Data-Anwendungen

Im Bereich der Verantwortlichkeiten von Big Data-Anwendungen zeigt sich in den Unternehmen ein relativ einheitliches Bild (Abbildung 9): Der Betrieb, die Umsetzung der Infrastruktur und die Technologie liegen in der Verantwortung der IT.

Abb. 09: Verantwortlichkeit von Big Data-Anwendungen bei Early Adopters und Early / Late Majority



n = 76; Mehrfachauswahl; Quelle: Eigene Darstellung.

Auftraggeber der Datenanalysen sind verschiedene Fachbereiche: Allen voran das Versorgungs- und Leistungsmanagement, der Vorstand, die Unternehmensentwicklung sowie Marketing und Vertrieb. Die Durchführung der Datenanalysen und das Reporting übernehmen am häufigsten die Bereiche Finanzen und Controlling oder Analytik. Die Ergebnisse zwischen GKV und PKV weichen hierbei kaum ab.

Interdisziplinarität der Teamkompetenzen und Transparenz der Prozesse spielen bei der Implementierung von Big Data-Lösungen eine entscheidende Rolle. Die effiziente Nutzung von Big Data setzt voraus, dass sie bereichsübergreifend im Unternehmen angesiedelt wird.

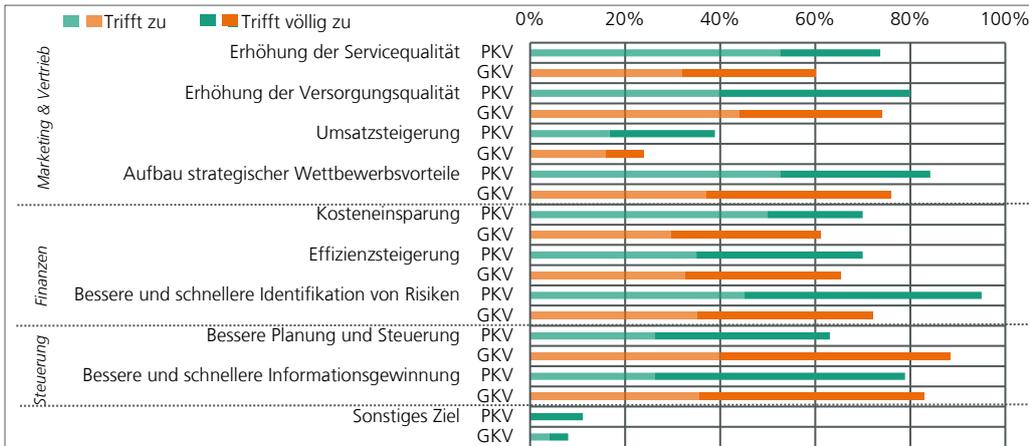
Matthias Waack, Gesundheitsforen Leipzig

Ziele von Big Data-Anwendungen

Alle vier Gruppen wurden zu den Zielen bei der Anwendung von Big Data befragt. In Abbildung 10 sind die Ergebnisse der Unternehmen, die derzeit bereits Big Data-Technologien einsetzen oder den Einsatz planen, zu sehen. Die Ziele wurden mittels einer Faktorenanalyse kategorisiert. Diese ergibt drei Zielkategorien: Marketing und Vertrieb, Finanzen sowie Steuerung. Die Ergebnisse zeigen, dass die meisten abgefragten Ziele vergleichsweise hoch bewertet wurden.

Aus Sicht der PKV sind vor allem die bessere und schnellere Identifikation von Risiken, der Aufbau strategischer Wettbewerbsvorteile und die Erhöhung der Service- und Versorgungsqualität sehr wichtig. Für die GKV stehen die bessere Planung, Steuerung und Informationsgewinnung sowie der Aufbau strategischer Wettbewerbsvorteile im Vordergrund.

Abb. 10: Ziele des Einsatzes von Big Data-Technologien bei Early Adopters und Early / Late Majority



$n_{PKV} = 18-20$, $n_{GKV} = 50-55$; Quelle: Eigene Darstellung.

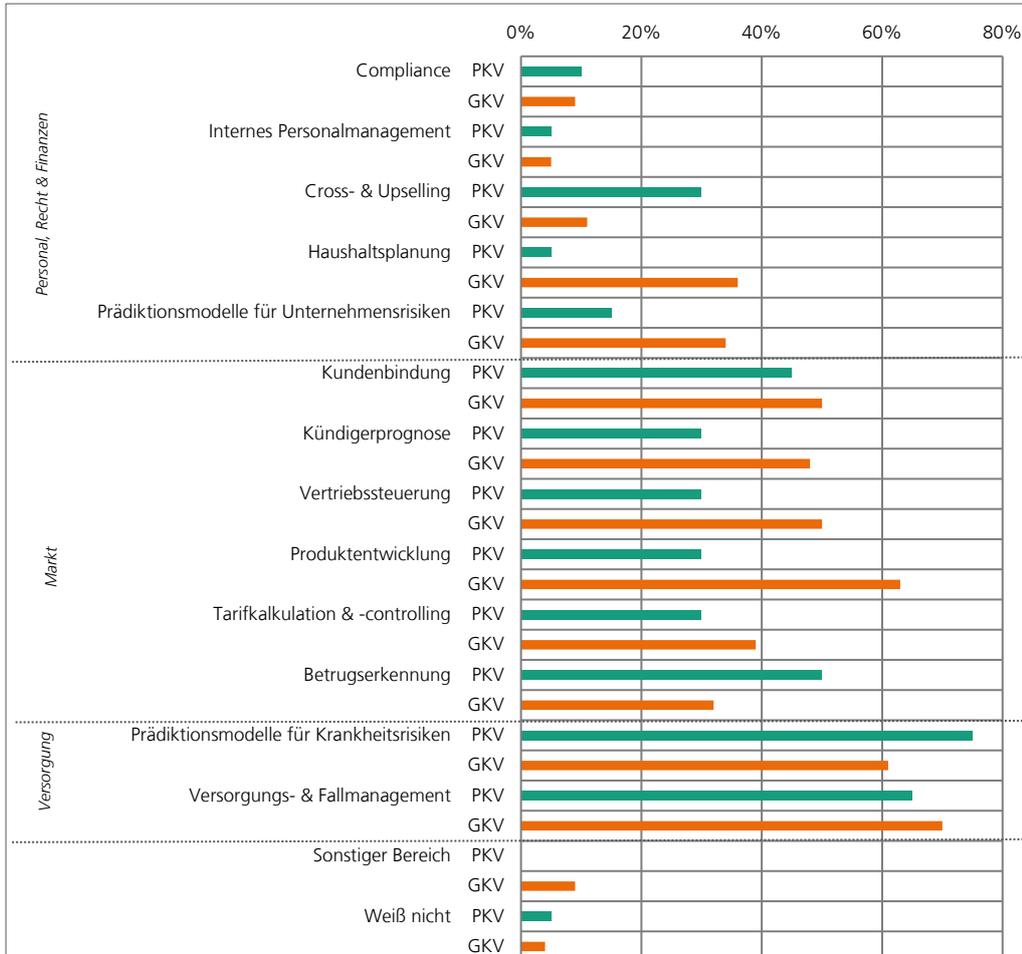
Es entspricht dem GKV-Wirtschaftlichkeitsgebot, Krankheiten frühzeitig zu behandeln. Dies ist durchaus auch im Sinne der Versicherten. Das setzt allerdings voraus, diese Krankheiten auch frühzeitig zu erkennen, um ihnen nicht nur kurativ, sondern auch präventiv entgegenzuwirken. Datenbasierte Prognosen helfen, betroffene Versicherte zu identifizieren, ihnen Angebote für spezielle Vorsorgeuntersuchungen zu machen sowie in passenden Präventions- oder Versorgungsprogrammen unterzubringen. Im Idealfall profitieren Kassen und Patienten davon, weil auf Basis gesicherter Diagnosen Krankheitsverläufe zu einem möglichst frühen Zeitpunkt positiv beeinflusst werden können, ohne dabei die Refinanzierung der notwendigen Ausgaben aus dem Gesundheitsfonds aus dem Auge zu verlieren.

Ronald Schwarz, BIG direkt gesund

Anwendungsbereiche

Die Gruppierung der Anwendungsbereiche von Big Data mittels einer Clusteranalyse (Abbildung 11) ergibt drei Bereiche: Personal, Recht und Finanzen, Markt sowie Versorgung. Big Data wird aktuell vor allem in den Bereichen Versorgung und Markt eingesetzt, weniger in den administrativen Bereichen. Sowohl GKV als auch PKV nutzen Big Data für Prädiktionsmodelle von Krankheitsrisiken und das Versorgungs- und Fallmanagement.

Abb. 11: Big Data-Anwendungsbereiche bei Early Adopters und Early / Late Majority



$n_{PKV} = 20$, $n_{GKV} = 56$; Mehrfachauswahl; Quelle: Eigene Darstellung.

Anwendungsbereich Versorgungs- und Fallmanagement

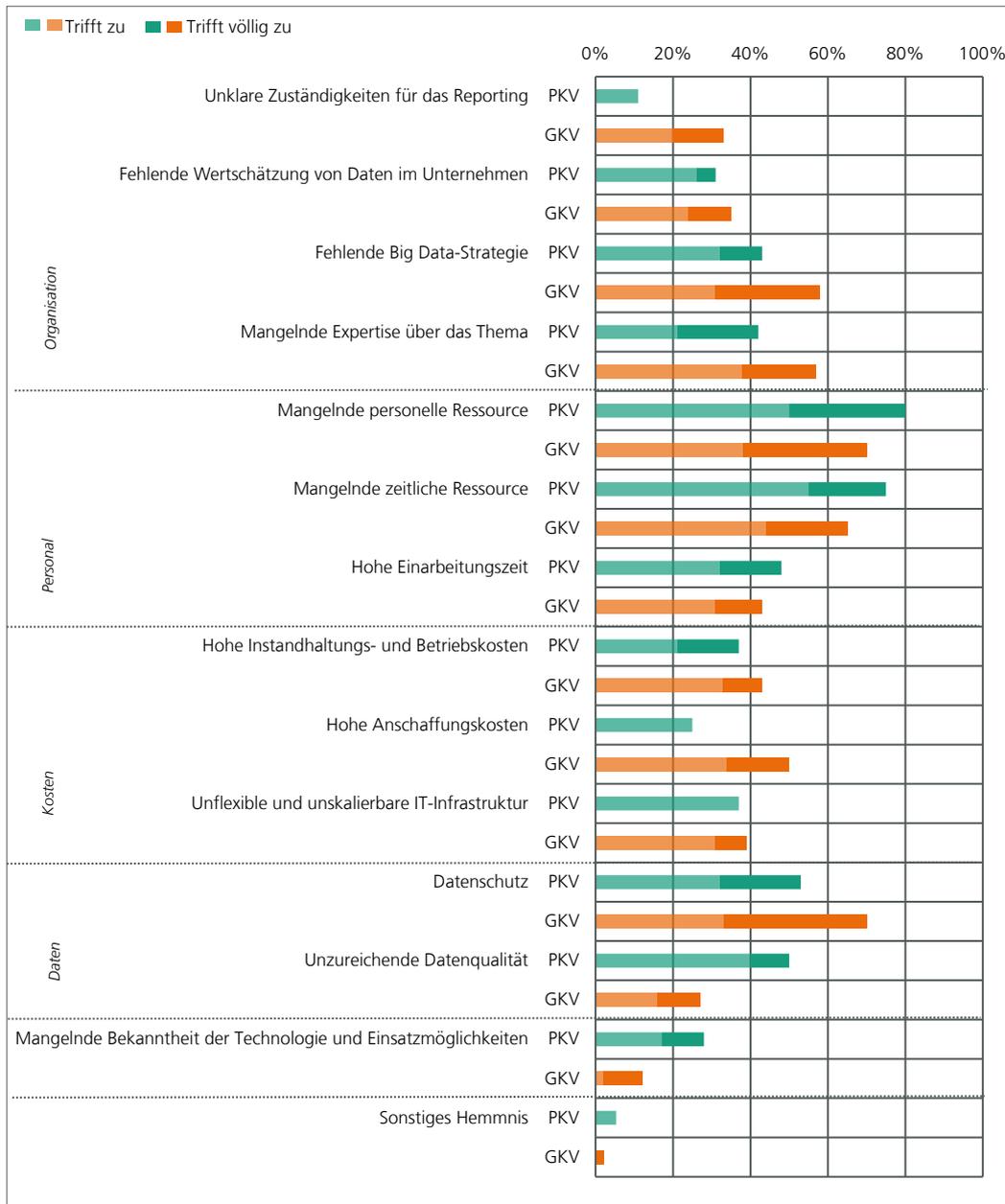
Je mehr Informationen über einen Versicherten und einen Leistungserbringer vorliegen, desto besser können die Gesundheitsversorgung sowie die damit verbundenen Leistungsausgaben durch innovative (Vergütungs-)Modelle beeinflusst und die Behandlungsqualität verbessert werden. Einfaches Beispiel: Ein Versicherter mit Rückenschmerzen möchte gerne eine orthopädische Rehabilitation durchführen. Aufgrund einer Niereninsuffizienz ist der Versicherte dialysepflichtig, was er seiner Krankenkasse bei der Antragstellung mitteilt. Mit diesem Wissen kann die Krankenkasse dem Versicherten vorschlagen, die Rehabilitation in einer bestimmten Rehabilitationseinrichtung zu absolvieren, in der neben der Therapie der Rückenschmerzen auch die Dialysebehandlung sichergestellt werden kann.

Dirk Engst, AOK Nordost

Hemmnisse bei der Anwendung von Big Data

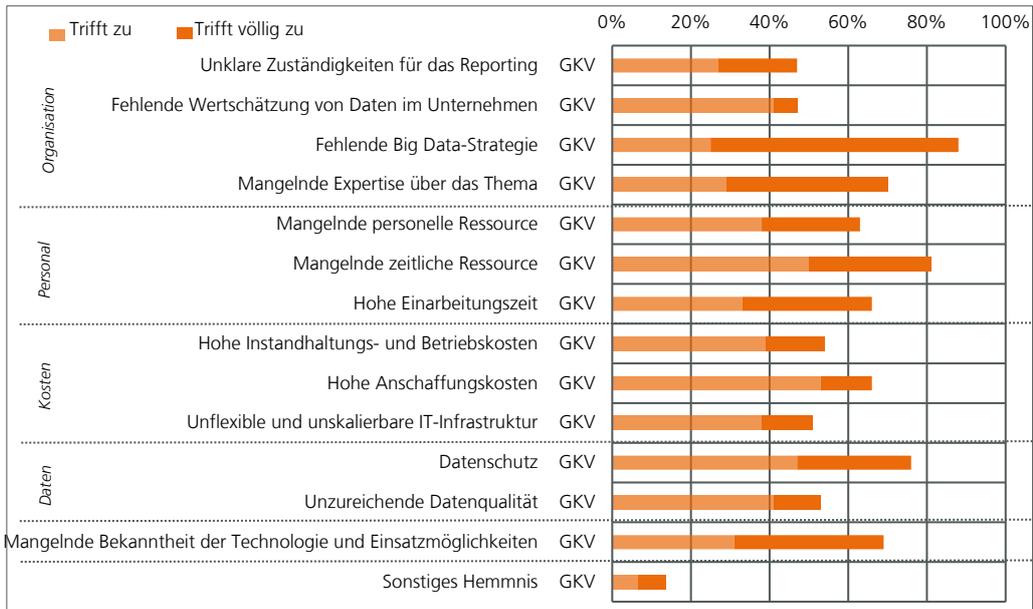
Welche Hemmnisse bei der Anwendung von Big Data bestehen, zeigt Abbildung 12. Bei der PKV wird Big Data bereits eingesetzt, jedoch fehlen hier vor allem personelle und zeitliche Ressourcen. Bei der GKV bestehen die Barrieren in Datenschutz, mangelnder Expertise, zusätzlichen Kosten und einer fehlenden Strategie. In der Gruppe der Laggards und Lack of Information sind aufgrund zu geringer Fallzahlen der PKV nur die Ergebnisse der GKV ausgewertet worden. In diesen Gruppen kommt zu den genannten Hemmnissen unzureichendes Wissen über die Technologie und deren Anwendungsmöglichkeiten hinzu (Abbildung 13). Im Anhang greifen wir insbesondere das Thema fehlender personeller Ressourcen auf und diskutieren mögliche Ansätze.

Abb. 12: Hemmnisse bei der Umsetzung von Big Data durch Early Adopters und Early / Late Majority



n_{PKV} = 19-20, n_{GKV} = 48-56; Quelle: Eigene Darstellung.

Abb. 13: Eingeschätzte Hemmnisse bei der Big Data-Umsetzung von Laggards und Lack of Information



n_{GKV} = 13-17; Quelle: Eigene Darstellung.

Die Befragten fürchten zu hohe Betriebskosten und eine unflexible IT-Infrastruktur im Zusammenhang mit Big Data. Diese Einschätzung zeigt, dass die Auseinandersetzung mit dem Thema Big Data in den Häusern noch nicht umfassend erfolgt. Denn es sind performante, anwenderfreundliche und auch individuelle Software- und Beratungs-Lösungen verfügbar. Auch kleine und mittlere Krankenkassen können so langfristig von Big Data-Technologien profitieren.

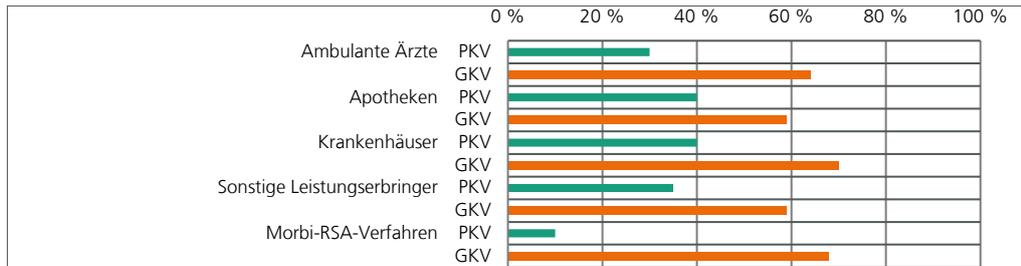
Matthias Waack, Gesundheitsforen Leipzig

3.2.3 Daten- und Analysesituation

Datenquellen

Bei der GKV ist der Datenaustausch insbesondere im Versorgungsbereich gesetzlich vorgeschrieben. Das Datenvolumen fällt umso größer aus, je mehr Versicherte eine Kasse hat. Die Antworten von GKV und PKV unterscheiden sich hinsichtlich der analytischen Nutzung der leistungsbezogenen Datenquellen deshalb deutlich voneinander (Abbildung 14).

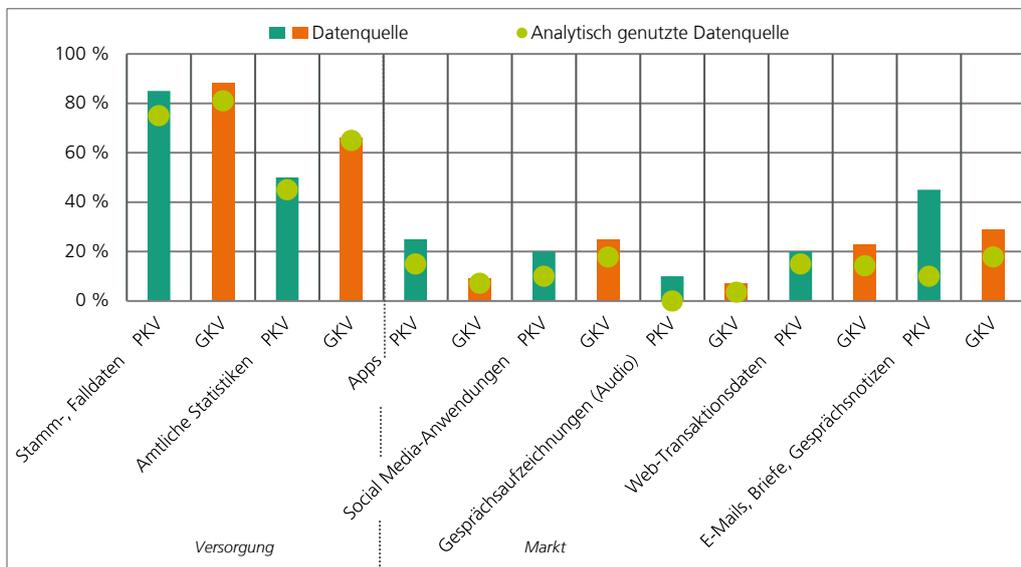
Abb. 14: Analytische Nutzung der festgelegten Datenquellen durch Early Adopters und Early / Late Majority



$n_{PKV} = 20$, $n_{GKV} = 56$; Mehrfachauswahl; Quelle: Eigene Darstellung.

Neben diesen vom jeweiligen Krankenversicherungssystem abhängigen Datenquellen wurden weitere Quellen abgefragt. Diese lassen sich in zwei Bereiche gliedern: Schnittstelle Versorgung und Schnittstelle Markt. Während die Datenquellen im Bereich Versorgung gleichermaßen genutzt werden, findet die Nutzung der Marktdaten nur sehr rudimentär statt (Abbildungen 15 und 16).

Abb. 15: Datenquellen und analytisch genutzte Datenquellen der Early Adopters und Early / Late Majority



$n_{PKV} = 20$, $n_{GKV} = 56$; Mehrfachauswahl; Quelle: Eigene Darstellung.

Allgemein formuliert bedeutet Big Data heute nicht nur über viele Daten zu verfügen, sondern Daten aus unterschiedlichen Quellen intelligent zu nutzen. Ich möchte diese einmal in drei Klassen einteilen:

1. Interne Daten, d.h. Daten aus operativen Datenbanken, z. B. dem CRM-, dem Marketing- oder auch dem Leistungsbereich.
2. Daten an der Schnittstelle zum Kunden bzw. Versicherten, z. B. von eigenen Web-Plattformen, Gesundheitsportalen, Apps oder Wearables.
3. Daten aus dem Internet, wie z. B. Web-Services, Social Media oder Adserver.

Ronald Schwarz, BIG direkt gesund

Treiber für große Datenmengen

Als Treiber für große Datenmengen werden insgesamt von allen Probanden die Stamm- und Falldaten, Abrechnungsdaten von Leistungserbringern und das damit einhergehende Morbi-RSA-Verfahren gesehen. Abbildung 16 zeigt die Ergebnisse der Early Adopters und Early / Late Majority.

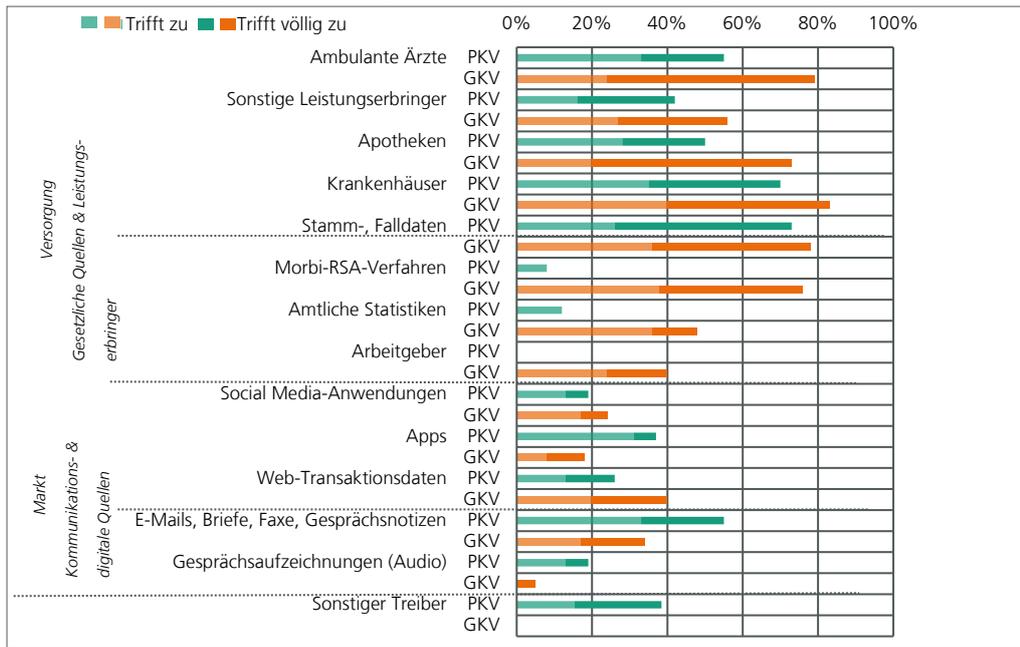
Die Antwortkategorien lassen sich erneut gliedern in: „Versorgung“ (Leistungserbringer und gesetzlich festgelegte Quellen) sowie „Markt“ (Kommunikations- und weitere digitalen Quellen). Wie in Abbildung 16 zu sehen ist, stammen die größten Datenmengen aus der Versorgung. Der Einsatzzweck dieser (Abrechnungs-)Daten ist gesetzlich stark reglementiert, weshalb eine willkürliche oder missbräuchliche Verwendung ausgeschlossen ist. Die Versorgungsquellen als „klassischer“ Treiber von Big Data werden auch weiterhin ihre Relevanz behalten. Wie jedoch in den Interviews deutlich wurde, werden digital getriebene Marktdaten (wie bspw. Geodaten oder Apps) für die Versicherer zukünftig immer wichtiger. Aber auch hier wurden datenschutzrechtliche Bedenken geäußert.

Die Kenntnis über Nutzungsprofile oder Konsumverhalten der Versicherten würde den Krankenkassen die Möglichkeit eröffnen, ihre Versorgungsangebote und Programme zielgruppenorientiert anzupassen.

Eine gesundheitsspezifische App, in der bspw. spielerische Elemente im Sinne des Gamification-Ansatzes aufgegriffen werden, könnte motivierend auf den Versicherten wirken, die Kundenbindung stärken und gleichzeitig relevante Daten erfassen. Um datenschutzrechtliche Bedenken auszuräumen zu können, bedarf es an dieser Stelle jedoch einer staatlichen Regulierung der Nutzungsmöglichkeiten dieser Daten.

Matthias Waack, Gesundheitsforen Leipzig

Abb. 16: Treiber für große Datenmengen von Early Adopters und Early / Late Majority



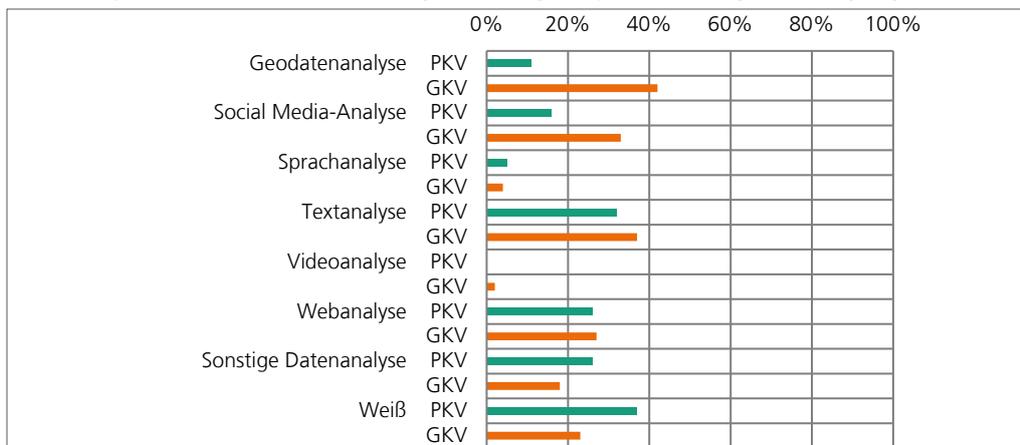
n_{PKV} = 13-19, n_{GKV} = 35-53; Quelle: Eigene Darstellung.

Formen der Datenanalyse

Insgesamt werden die Analysen in Krankenversicherungen sowohl manuell als auch automatisch initiiert und durchgeführt. Dabei zeigt sich, dass die Early Adopters und Early / Late Majority insgesamt tendenziell mehr zur automatischen Variante tendieren als die Laggards, jedoch im Mittel genauso oft angeben, dass sowohl die manuelle als auch die automatische Situation vorliegt. In den Interviews wurde dies bestätigt und die Probanden gaben an, dass zwar teilweise automatische Analysen gefahren werden, die Ergebnisse jedoch stets von Experten auf fachliche Richtigkeit überprüft werden.

Hinsichtlich verschiedener Analyseverfahren bestätigten die Experten, dass in Krankenversicherungen klassische multivariate Verfahren der Datenanalyse durchgeführt werden. In der Onlinebefragung wurden insbesondere Analysen für unstrukturierte Daten fokussiert (Abbildung 17 und 18).

Abb. 17: Spezielle Formen der Datenanalyse bei Early Adopters und Early / Late Majority



n_{PKV} = 19, n_{GKV} = 52; Mehrfachauswahl; Quelle: Eigene Darstellung.

Es zeigte sich, dass spezielle Analyseformen für unstrukturierte Daten in der GKV und PKV verhältnismäßig wenig durchgeführt werden. Dies kann jedoch auch daran liegen,

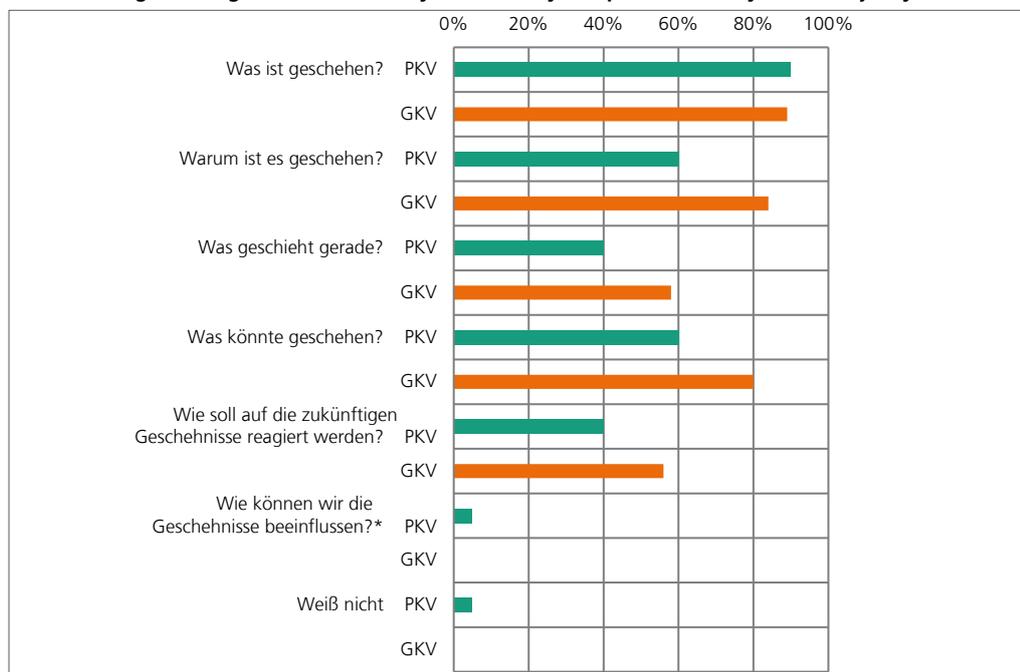
dass die Daten in den Unternehmen weitgehend strukturiert vorliegen, z. B. Daten aus Krankenhäusern und Apotheken.

Im Gegensatz dazu sind bspw. die Abrechnungen von ambulanten Ärzten, zumindest in der PKV, eher unstrukturiert. Problem hierbei stellen vor allem fehlende Standards für die ambulante Diagnose auf Rechnungen dar. Jeweils mehr als 50 Prozent der Probanden aus der Gruppe der Early Adopters, Early / Late Majority und Laggards gaben an, dass die Daten aus diesen drei Leistungsbereichen analytisch genutzt werden. Hierin kann der Grund für die relativ seltene Durchführung dieser speziellen Formen der Datenanalyse bei unstrukturierten Daten gesehen werden.

Verfolgte Fragestellungen von Datenanalysen

Im Rahmen der Analysesituation wurden die verfolgten Fragestellungen von Datenanalysen abgefragt. Diese Fragestellungen unterscheiden sich dabei im Zeitbezug (Abbildung 18 und 19).

Abb. 18: Fragestellungen von Datenanalysen bei Early Adopters und Early / Late Majority

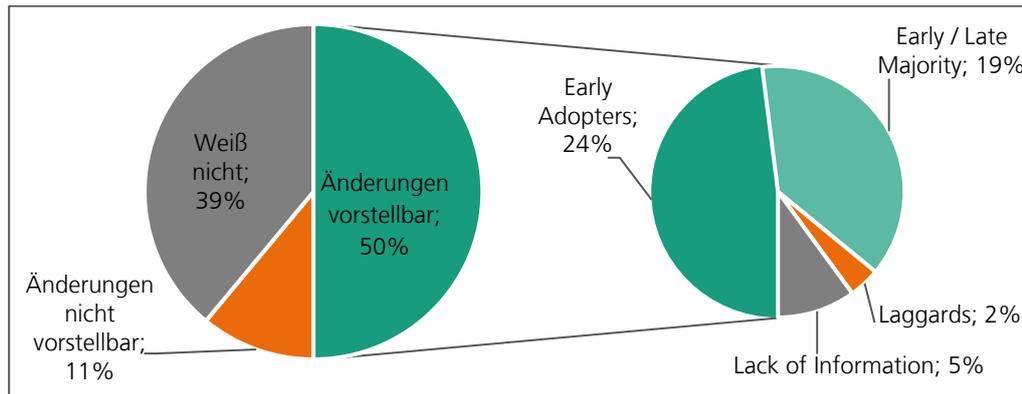


$n_{PKV} = 20$, $n_{GKV} = 55$; Mehrfachauswahl; *Kategorie Sonstige; Quelle: Eigene Darstellung.

Änderungen durch Big Data

Insgesamt geben ungefähr 50 Prozent der Probanden an, dass sie sich Änderungen der Datenanalyse vorstellen können (Abbildung 19): Sowohl in der PKV (55 Prozent) als auch in der GKV (48 Prozent) wird diese Auffassung geteilt.

Abb. 19: Vorstellbare Änderungen in der Datenanalyse durch Big Data



n = 89; Quelle: Eigene Darstellung.

Die Ergebnisse der Onlinebefragung verdeutlichen, dass die Änderungen in der Datenanalyse durch das Thema Big Data vielfältig wahrgenommen werden. Bei einer inhaltlichen Kategorisierung der offenen Antworten zeigt sich, dass hauptsächlich Änderungen der Auswertungsmöglichkeiten erkannt werden (Abbildung 20).

Wenn auch ohne Big Data-Anwendungen schon eine ganze Menge möglich war, stellte sich insbesondere die Analyse (leistungsbereichs-)übergreifender Daten durch viele kleine Einzelschritte manchmal sehr mühselig dar. Big Data eröffnet neue Möglichkeiten, so dass vieles einfacher und schneller geht. Früher hatte man verschiedene Excel-Tabellen, die bei Bedarf in eine Access-Datenbank importiert wurden, um sie für die Nutzung in einer Statistiksoftware aufzubereiten etc. Alle relevanten Daten stehen jetzt in einem Data-Warehouse an einer Stelle pseudonymisiert für Analysen zur Verfügung.

Dirk Enge, AOK Nordost

Die Verfügbarkeit und der Zugriff auf die Daten haben sich über die letzten Jahre stark verändert. In der Vergangenheit wurden die einzelnen Leistungssektoren in separaten Datensilos vorgehalten. Inzwischen wurden die Datenstruktur und der Datenzugang derart verändert, dass eine Kombination verschiedener Leistungsbereiche unkompliziert durchgeführt werden kann.

Stefan Baßler, AOK PLUS

Ich sehe zwei grundsätzliche Veränderungen, die mit Big Data und Digitalisierung einhergehen: Die erste Veränderung ist Geschwindigkeit und die zweite Veränderung ist Transparenz. Alle Dinge werden schneller und transparenter – und wie immer hat dies Vor- und Nachteile.

Wenn meine GKV mir frühzeitig ein Herz-Kreislauf-Training anbietet, weil meine Vorerkrankungen oder in Zukunft zudem vielleicht die Daten aus meinem Wearable ein entsprechendes Risikomuster erkennen lassen, ist dies zunächst einmal eine gute Sache. Es erfordert jedoch auch von mir eine gewisse Mitwirkung und frühzeitige Auseinandersetzung mit einer Erkrankung, die subjektiv vielleicht gar nicht wahrnehmbar ist und nur mit Verfahren wie Predictive Analytics als wahrscheinlich vorhergesagt wird. Die Frage ist, wollen wir das?

Das andere Thema Transparenz ist sicherlich noch ambivalenter. Die Datenquellen, welche eine hohe Komplexität aufweisen, müssen vor Missbrauch und Zugriffen Dritter geschützt werden. Die Transparenz entsteht erst, wenn Mustererkennungsverfahren, also Vorhersagealgorithmen, auf diese Datenquellen angewendet werden. Die Frage, die sich hier stellt, ist: Tragen wir das dazugehörige Know-how in die Kassen

hinein oder unterstützen wir mittelfristig eine Entwicklung, bei der diese Daten von externen Anbietern, die heute bereits über entsprechende Algorithmen verfügen, auf Wunsch der Versicherten analysiert werden? Erste Studien zeigen, dass das Vertrauen der Versicherten in ihre Kasse vergleichsweise groß ist.

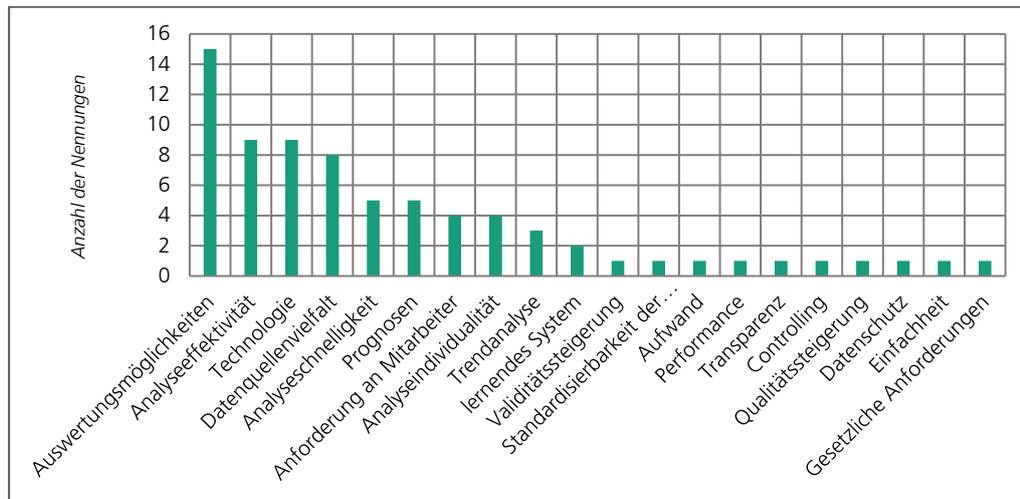
Ronald Schwarz, BIG direkt gesund

Big Data bietet die Möglichkeit, Komplexität und komplexe Sachverhalte, die durch einen Menschen nicht einzel-analytisch lösbar sind, zu erkennen und Zusammenhänge herzustellen. Der Vorteil von Big Data mit neuen Auswertungsmethoden liegt darin, solche Zusammenhänge, die nicht offenkundig sind, überhaupt ermitteln zu können. Dabei werden auch Ergebnisse ermittelt, die zuvor nicht vermutet worden wären oder hätten herausgearbeitet werden können.

Volker Wagner, AOK Hessen

Im Rahmen der Experteninterviews zeigte sich, dass Verfahren oder Methoden der Datenanalyse durch das Thema Big Data nicht unbedingt adaptiert wurden. Des Weiteren werden die Änderungen meist nicht in spezifischen Aspekten festgemacht. Zwar wurden im Experteninterview konkrete Analyseverfahren angesprochen, die im Unternehmen unter dem Stichwort Big Data-Analysen laufen. Jedoch stellte sich mehrheitlich heraus, dass sich die neue Qualität durch Big Data vielmehr in der Änderung des gesamten Unternehmensprozesses, insbesondere bezüglich der Daten- und Analysesituation, wahrnehmen lässt.

Abb. 20: Änderungen in der Datenanalyse durch das Thema Big Data

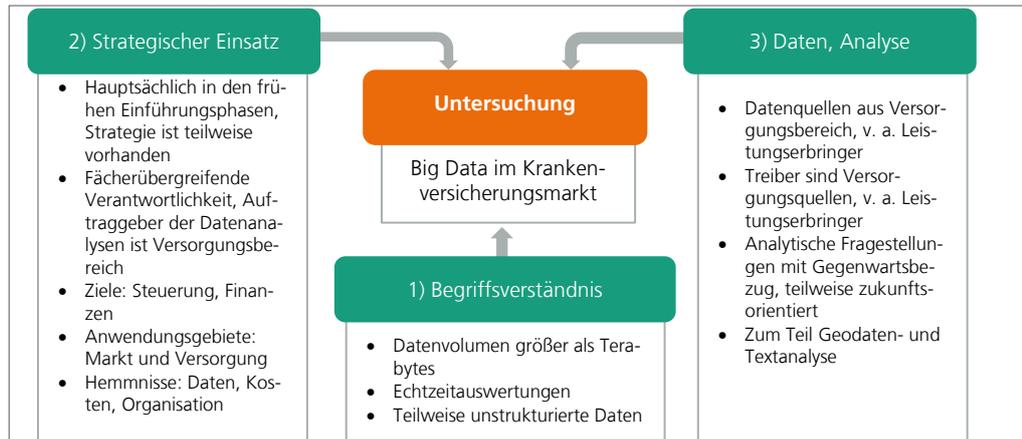


n = 44; Quelle: Eigene Darstellung.

4 Fazit und Ausblick

Big Data spielt im gesamten Krankenversicherungsmarkt eine relevante Rolle. Sowohl die GKV als auch die PKV zeigen sich dem Thema gegenüber aufgeschlossen: 80 Prozent der Unternehmensvertreter geben an, dass Big Data-Technologien bereits eingesetzt werden oder der Einsatz geplant ist.

Abb. 21: Die wichtigsten Ergebnisse im Überblick



Quelle: Eigene Darstellung.

Wie ist das Verständnis des Begriffs Big Data? Im Krankenversicherungsmarkt besteht eine relativ ähnliche Auffassung von Big Data. Die Interviews bestätigten, dass der Begriff nicht nur an der stark wachsenden Datenmenge und Verarbeitungsgeschwindigkeit festgemacht wird, sondern vielmehr an der Datenvielfalt.

Wie wird der Einsatz von Big Data-Technologien/-Anwendungen wahrgenommen? Der Markt ist bereits heute mehrheitlich auf den Umgang mit großen Datenmengen vorbereitet. Der Einsatz von Big Data-Technologien ist aufgrund von Synergieeffekten stark von der Unternehmensgröße abhängig. Entsprechend stellen fehlende personelle Ressourcen Hemmnisse bei der Umsetzung dar. Als weitere große Barriere gilt der Datenschutz. Big Data-Anwendungen werden vor allem im Versorgungs- und Marktbereich durchgeführt, um Ziele des Finanz-, Marketing-, Vertriebs- und Steuerungsbereichs zu erreichen.

Wie ist die aktuelle Daten- und Analysesituation? Systembedingt verfügen gesetzliche Versicherer über größere Datenmenge als private. Die Daten kommen hauptsächlich aus dem Versorgungsbereich und werden teilweise analytisch genutzt. Die Wichtigkeit dieser treibenden Datenquellen wird sich auch zukünftig nicht ändern. Die Fragestellungen in der Analytik beziehen sich größtenteils auf die Gegenwart. Bei der GKV spielt des Weiteren u. a. auch Frage „Was könnte geschehen?“ eine wichtige Rolle.

Das Schlagwort Big Data ist im Krankenversicherungsmarkt nicht nur ein Trend, sondern bereits heute gängige Praxis. Zwar bringt Big Data viele Risiken mit sich, wie den Missbrauch von Daten, aber auch enorme Potenziale. Meist sind die gigantischen Datenmengen, die stark vom Versorgungsbereich getrieben werden, bereits in den Unternehmen vorhanden. Auch die Technologie, um diese Datenmengen zu bewältigen, ist auf dem Markt. Des Weiteren gibt es viele externe Datenquellen, die oft nur noch erschlossen werden müssen. Solche typischerweise unstrukturierten Daten, wie etwa Social Media-Daten, sind prädestiniert für Big Data-Anwendungen. Somit besteht bereits heute ein enormes Potenzial für Krankenversicherungen, das oft nur noch genutzt werden muss. Viele Versicherer haben die Chance bereits erkannt und ihre intern vorhandenen Datentöpfe kombiniert, um sie auf sinnvolle Muster zu untersuchen. Big Data sollte von allen als Chance gesehen werden, um bestehende Daten in einen konkreten Nutzen umzuwandeln und Wettbewerbsvorteile zu erzielen.

5 Quellenverzeichnis

BMG1 (2016) – Bundesministerium für Gesundheit, Gesetzliche Krankenversicherung, Mitglieder, mitversicherte Angehörige und Krankenstand, Monatswerte Januar-Juli 2016, URL: http://www.bmg.bund.de/fileadmin/dateien/Downloads/Statistiken/GKV/Mitglieder_Versicherte/KM1_Januar_bis_Juli_2016.pdf, Stand: 13.08.2016.

Brunner, T. (2014), Big Data aus Sicht einer Krankenkasse, in: Langkafel, P. (Hrsg.), Big Data in Medizin und Gesundheitswirtschaft, Heidelberg.

Bundesregierung (2015), Neue Regelungen ab Januar 2016, Presse- und Informationsamt der Bundesregierung (Hrsg.), URL: <https://www.bundesregierung.de/Content/DE/Artikel/ArtikelNeuregelungen/2016/neuregelungen-januar/2015-12-21-neuregelungen-erster-januar-2016.html?nn=694676#doc1677570bodyText2>, Stand: 16.08.2016.

Chute, C. G.; Ullman-Cullere, M.; Wood, G. M.; Lin, S. M. & Pathak J. (2013), Some experiences and opportunities for big data in translational research, Genetics in Medicine, Vol. 15, No. 10, S. 802-809.

Diessner, H.; Früh, F.; Gantner, M.; Henrichsen, F.; Kaiser, S.; Roos, W.; Sontow, K.; Wolf, G. & Zöllner, B. (2015), ECM Trend-Check 2015, Publikation der BITKOM, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (Hrsg.), Berlin.

EMC Corporation (2012), Prognose zum Volumen der jährlich generierten digitalen Datenmenge weltweit in den Jahren 2005 bis 2020 (in Exabyte), URL: <http://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/>, Stand 15.08.2016.

Meinel, C.; Koppenhagen, N.; Bildmayer, R.; Falkenberg, G.; Greger, H.; Hrkalovic, N.; Juffernbruch, K.; Laskowski, K.; Lenk, A.; Löffler, S.; Lott, K.; Mentzinis, P.; Peuker, M.; Rabe, A.; Saurer, B. R.; Schwaderer H.; Schwalb, D.; Stürner, G.; Vaillant, S.; Wagner, M.; Weber, M.; Widmer, M. & Zünkeler, M. (2015), Thesenpapier zum Schwerpunktthema Smart Data im Gesundheitswesen, Publikation der Plattform „Innovative Digitalisierung der Wirtschaft“ im Nationalen IT-Gipfel, Fokusgruppe Intelligente Vernetzung/ Projektgruppe Smart Data (Hrsg.), URL: <https://www.bitkom.org/Publikationen/2016/Positionspapiere/Thesenpapier-zum-Schwerpunktthema-Smart-Data-im-Gesundheitswesen/it-gipfel-2015-thesenpapier-zum-schwerpunktthema-smart-data-im-gesundheitswesen.pdf>, Stand: 24.08.2016.

Provost, F.; Fawcett, T. (2013), Data Science for Business, 1. Aufl., Cambridge et al.

vdek (2016) – Verband der Ersatzkassen e. V., Daten zum Gesundheitswesen: Finanzierung, URL: https://www.vdek.com/presse/daten/c_einnahmen-ausgaben.html, Stand: 16.08.2016.

Verband der Privaten Krankenversicherung (2016), URL: <http://www.pkv.de/verband/ueber-uns/>, Stand: 16.08.2016.

Rogers, E. M. (1995), Diffusion of Innovations, 4. Aufl., New York.

Sabherwal, R.; Becerra-Ferdandez, I. (2011), Business Intelligence, 1. Aufl., New Jersey et al.

Statistisches Bundesamt (2016), Jährliche Gesundheitsausgaben in Deutschland in den Jahren von 1992 bis 2014 (in Millionen Euro), URL: <http://de.statista.com/statistik/daten/studie/5463/umfrage/gesundheitswesen-in-deutschland---ausgaben-seit-1992/>, Stand: 16.08.2016.

Zillmann, M. (2013), Big Data bei Krankenversicherungen, Publikation der Lündendonk GmbH (Hrsg.), in fachlicher Zusammenarbeit mit der SAS Institute GmbH, Kaufbeuren.

6 Elemente einer Big Data-Strategie

Eine Big Data-Strategie aufzubauen ist eine kostspielige Investition, die jedoch großes Potenzial hat. Beispiele hierfür gibt es zuhauf. Neben den berühmten wie Google, Amazon und Netflix gibt es viele kleinere, etwa die Signet Bank aus Virginia, USA (vgl. Provost & Fawcett 2013). Wie viele andere Banken litt auch Signet an industrietypischen hohen "Charge Off-Rates" (Anteil an ausgefallenen Krediten) im Kreditkartengeschäft. Es folgte der Entschluss, diese Raten mit Big Data-Ansätzen besser zu prognostizieren. Mittels realer Feldexperimente wurden Daten gesammelt, die überhaupt erst eine bessere Prognose und Steuerung ermöglichten – und das, obwohl diese Experimente, bei denen verschiedene Kreditkonditionen an Kunden getestet wurden, über Jahre die Kosten in die Höhe trieben. Das Ergebnis: Im Jahr 2000 wurde aus der kleinen unbekannteren Signet Bank die hochprofitable Capital One ausgegründet mit der industrieweit niedrigsten Charge-Off-Rate überhaupt.

Eine Big Data-Strategie bedeutet also nicht nur, in Technologie, Software und Humankapital zu investieren oder in die Erhebung und Qualität der notwendigen Daten, sondern auch die Bereitschaft zu entwickeln, datengetrieben und experimentell zu denken und eine Unternehmenskultur zu schaffen, in der sich dieses Denken entwickeln kann. Charakteristisch für erfolgreiche Big Data-Strategien ist dabei ein „Catch22-Effekt“: je mehr Daten, desto höher das Potenzial für Big Data-Anwendungen, desto zufriedener die Kunden, desto mehr Kunden, desto mehr Kundendaten, usw. In diesem sich verstärkenden Kreislauf gibt es dabei viele kleine "Tipping Points", die man erkennen und ausnutzen muss, da man sonst in einen strategischen Nachteil gegenüber der Konkurrenz gerät.

Wann ist also "Big Data" eine strategische Option für Unternehmen? „Big Data“ als Strategie muss zunächst zu dem jeweiligen Unternehmenskontext passen. Beispielsweise war es für Dell von Vorteil, mehr Informationen über seine Kunden zu sammeln, weil Dell über ein Direkt-Vertriebsnetz verfügte und so diese Informationen ausnutzen konnte. Für den Konkurrenten Compaq, der seine Computer vornehmlich indirekt vertrieb, galt das nicht. Nur wenn Konkurrenten nicht über die gleichen strategischen „Big Data“-Vorteile verfügen oder diese ohne Weiteres imitieren können, besteht die Chance auf einen nachhaltigen Wettbewerbsvorteil.

Mögliche Elemente einer "Big Data"-Strategie, neben dem separat behandelten Punkt Personal, sind dabei (vgl. Sabherwal & Becerra-Fernandez 2011):

- **Infrastruktur:** IT, Tools, Daten & Qualität der Daten.
- **Strukturen:** Big Data-Governance, d. h. Festlegung, wer in einem "Big Data-Steering Committee" über Big Data-Strategien mitentscheiden darf (typischerweise Mitglieder aus der IT-Abteilung, aber auch anderen funktionalen Abteilungen, die Big Data anwenden). Zusätzlich kann ein Big Data-Kompetenz-Center geschaffen werden, in dem alle Experten zu Big Data gebündelt werden, um die verstreut vorhandene Expertise dem gesamten Unternehmen zur Verfügung zu stellen. Auf diese Weise werden Redundanzen verhindert und gemeinsame Standards etabliert, z. B. im Hinblick auf vorhandene Daten, Metadaten und deren Qualität. Aber auch Training, Ausbildung und Entwicklung von Tools können hier gebündelt werden.
- **Prozesse:** Z. B. nach welchen Kriterien werden Big Data-Anwendungen genehmigt? Braucht man einen Business Case? Welche Zielgrößen möchte man verbessern? Soll ein ROI gemessen werden? Oder betrachtet man Big Data eher als F&E?
- **Kultur:** Erfolg von Big Data-Strategien hängt nicht nur von der Technologie ab, sondern davon, dass alle Verantwortlichen im Unternehmen die Relevanz und Wichtigkeit von Daten und datenbasierten Entscheidungen teilen.

Big Data-Hemmnis: Personelle Ressourcen

Während sehr viele und sehr gute Big Data-Tools umsonst verfügbar sind (Python und R z. B. sind freie Programmiersprachen), gilt das gleiche leider nicht für Big Data-Spezialisten. Und das, obwohl in der Zwischenzeit Big Data-Studiengänge und Fortbildungsangebote geradezu aus dem Boden schießen.

Gute Big Data-Teams sind ein wesentlicher strategischer Wettbewerbsvorteil. „Big Data-Talent“ ist jedoch ungleich verteilt. Wie im Fußball gibt es nur sehr wenige „Big Data-Stars“, die sich ihre Arbeitgeber und die Konditionen aussuchen können. Das sieht man z. B. daran, dass immer wieder die gleichen Teams Big Data-Wettbewerbe, wie etwa den KDD Cup oder den Netflix-Wettbewerb, gewinnen (vgl. Provost & Fawcett 2013).

Big Data-Spezialisten kann man nicht ohne Weiteres selbst kurzfristig ausbilden, da einen solchen ebenso sehr Erfahrung ausmacht wie technisches Wissen. Man braucht also bereits Big Data-Spezialisten im Haus, die aus Talenten selbst wieder Spezialisten formen. Man braucht diese Spezialisten außerdem im Haus, um überhaupt erst die Qualität von Bewerbern ex ante einschätzen zu können.

Noch kritischer ist das Big Data-Management, das von Individuen an der Schnittstelle zwischen Big Data-Technologien und den betriebswirtschaftlichen Anwendungen übernommen werden muss. Diese Individuen müssen sowohl technisch mithalten als auch zukünftige betriebswirtschaftliche Anwendungen antizipieren können.

Wie kann man nun Data Science-Talente attrahieren? Es gibt auf diese Frage keine einfache Antwort, jedoch die tröstliche Erkenntnis, dass die Motive von Big Data-Spezialisten oftmals sehr heterogen sind und somit kreativ ausgenutzt werden können. Die einen wollen schnell in Führungspositionen aufsteigen, die anderen einen sicheren Job, wiederum andere mit Top-Talenten zusammenarbeiten oder „an etwas Großem“ wirken.

Was tun, wenn man in der kurzen Frist keine Big Data-Spezialisten für das eigene Unternehmen findet? Eine Möglichkeit besteht in der Kooperation mit der Wissenschaft, z. B. in Form gemeinsam finanzierter Doktorarbeiten. Und die letzte Alternative ist schließlich, das Thema Big Data nach außen zu verlagern und abzuwarten, bis sich der Markt für Big Data-Spezialisten wieder „abgekühlt“ hat (eine Liste mit internationalen Big Data-Consultants findet sich z. B. auf KDnuggets).