

# Wie „repräsentativ“ sind GKV-Daten?

## Demografische und soziale Unterschiede und Ähnlichkeiten zwischen einer GKV-Versichertenpopulation, der Bevölkerung Niedersachsens sowie der Bundesrepublik am Beispiel der AOK Niedersachsen

In einer 2008 veröffentlichten Studie zur Mitgliederstruktur von Krankenversicherungen wurde gezeigt, dass sich die Sozialstruktur, Geschlechterverhältnisse sowie die Morbiditäten zwischen unterschiedlichen Versicherungen deutlich voneinander unterscheiden [1, 2]. Die Versicherten der Allgemeinen Ortskrankenkassen (AOK) und der Betriebskrankenkassen (BKK) wiesen in diesem Vergleich am häufigsten Gesundheitsbeeinträchtigungen auf. Grundlage der Analysen waren Surveydaten aus dem Bertelsmann-Gesundheitsmonitor 2007. Unter Verwendung der gleichen Datenquelle wurden Untersuchungen zur Prävalenz von Diabetes je nach Krankenversicherungszugehörigkeit durchgeführt [3]. Wie oben [1, 2] für allgemeine Krankheitsindikatoren berichtet, unterschieden sich die diesbezüglichen Prävalenzraten bei den Versicherten der unterschiedlichen Kassen deutlich; wiederum wiesen die BKK- und die AOK-Versicherten die höchsten Raten auf. Bei Versicherten von Ersatzkassen waren die Prävalenzen für Diabetes hingegen um mehr als die Hälfte niedriger, die bei privat Versicherten lagen dazwischen. Insbesondere das letztgenannte Ergebnis ist überraschend, da Erkrankungsrisiken im Allgemeinen mit

steigendem Einkommen, zunehmender Qualifikation und steigender beruflicher Position sinken [4].

Beiden Studien ist gemeinsam, dass sie nur Unterschiede zwischen den Versicherten unterschiedlicher Kassen untersuchten, dennoch zieht Schnee [1] daraus den Schluss: „... die Allgemeinen Ortskrankenkassen versichern einkommensschwache und bildungsferne Schichten und haben damit eine völlig andere Klientel als beispielsweise die Techniker Krankenkasse“ (S. 100). Aus dieser Aussage könnte der Eindruck entstehen, dass Analysen mit GKV-Daten grundsätzlich zu verzerrten Ergebnissen führen. Tatsächlich wurde aber in beiden Studien die Frage nach der Übereinstimmung der Verteilung relevanter Merkmale zwischen den Kassenversicherten und der Gesamtbevölkerung nicht behandelt. Es bleibt deshalb offen, ob und in welchem Ausmaß sich die unterschiedlichen Versichertenpopulationen von der Allgemeinbevölkerung unterscheiden. Die Qualität von GKV-Daten wird bisher – außer in wenigen Studien – zu selten oder zu wenig differenziert diskutiert. Die Verwendung von GKV-Daten als Basis für wissenschaftliche Analysen wird häufig wegen fehlender Kenntnisse ihrer konkre-

ten Möglichkeiten und Grenzen abgelehnt [5] oder wegen ihres als begrenzt unterstellten Geltungsanspruchs kritisiert [6]. Andererseits werden aber auch Studien unter der Annahme durchgeführt, dass GKV-Daten für die Beantwortung einer breiten Variation von Fragestellungen verwendbar sind [7, 8].

Eine Beurteilung der wissenschaftlichen Brauchbarkeit von GKV-Daten ist jedoch notwendig, da ihre Verwendung weit verbreitet ist und zukünftig noch weiter zunehmen wird [9, 10, 11]. Auch der Bericht des Sachverständigenrats für das Gesundheitswesen aus dem Jahr 2007 [12] empfiehlt eine stärkere Nutzung von Routinedaten für die Gesundheitsforschung, da sie es ermöglichen, auch seltene Erkrankungen in ihren Verläufen und sozialstrukturellen Verteilungen auf der Basis großer Fallzahlen zu untersuchen.

Vor diesem Hintergrund soll im vorliegenden Beitrag untersucht werden, ob die Verteilung zentraler sozialstruktureller Merkmale in GKV-Daten von der in der allgemeinen Bevölkerung abweicht, und wenn ja, wie groß diese Unterschiede sind.

Diese beiden Fragen werden in den folgenden Analysen aufgenommen und für einen Datensatz empirisch unter-

**Tab. 1** Geschlechterverteilung unter den Versicherten der AOK Niedersachsen sowie in der Bevölkerung Niedersachsens und der Bundesrepublik in den Jahren 2004 und 2009

	AOK Niedersachsen			Bevölkerung Niedersachsen			Bevölkerung Bundesrepublik		
	Männlich	Weiblich	Gesamt	Männlich	Weiblich	Gesamt	Männlich	Weiblich	Gesamt
2004	46,8%	53,2%	100,0%	49,0%	51,0%	100,0%	48,9%	51,1%	100,0%
	994.559	1.129.997	2.124.556	3.919.375	4.081.534	8.000.909	40.353.627	42.147.222	82.500.849
2009	47,4%	52,6%	100,0%	49,1%	50,9%	100,0%	49,0%	51,0%	100,0%
	978.832	1.087.998	2.066.830	3.894.627	4.034.188	7.928.815	40.103.606	41.698.651	81.802.257

Stand: 31.12. des Jahres.

sucht. Es werden Zahlen der statistischen Ämter verwendet, um Variablen aus einem aktuellen GKV-Datenbestand mit den gleichen Variablen auf der Ebene eines Bundeslands (Niedersachsen) sowie für Deutschland insgesamt zu vergleichen. Als vorteilhaft erweist es sich, dass die Qualifikationen und Berufe der sozialversicherungspflichtig Beschäftigten in den GKV-Daten und in den Daten der Statistischen Ämter des Bundes sowie der Länder in gleicher Weise kodiert [13]. Diese Klassifikationen basieren auf dem „Schlüsselverzeichnis für Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung“ [14]. Im vorliegenden Fall werden die Daten der AOK Niedersachsen aus den Jahren 2004 und 2009 mit den entsprechenden Informationen für Niedersachsen und für die Bundesrepublik verglichen.

## Material und Methoden

### Datengrundlage

Zur Beantwortung der aufgestellten Frage wurde ein pseudonymisierter Datensatz der AOK Niedersachsen für die Jahre 2004 und 2009 verwendet. Er enthält Angaben zum Geschlecht, Geburtsmonat und -jahr, zum Wohnort und zum Versicherungsverlauf. Insgesamt liegen Daten für über 3,3 Mio. Versicherte vor.

Als Vergleichsbasis wurden Daten der Statistischen Ämter des Bundes und der Länder herangezogen [13]. Für die allgemeinen Bevölkerungsvergleiche wurden die Geschlechter- und Altersangaben (15 Kategorien) für die Einwohner Niedersachsens und der Bundesrepublik verwendet. Aus dem AOK-Datenbestand wurden alle Versicherten entfernt, deren Wohnort außerhalb Niedersachsens liegt (ca. 6,5% der Versicherten).

Für diesen Bereinigungsprozess wurden die Wohnortsangaben als zweistellige Postleitzahlangaben genutzt. Die personenbezogenen Daten liegen nicht als Prozessdaten, sondern als Querschnittsdaten zum 31.12.2010 vor. Die Unschärfen einer zweistelligen Postleitzahlzuordnung zum Bundesland Niedersachsen und auch des Querschnittscharakters der Wohnortdaten wurden mangels genauerer Informationen in Kauf genommen.

Des Weiteren werden AOK-Versicherte mit der Bevölkerung Niedersachsens und der Bundesrepublik hinsichtlich ihrer Qualifikation und der ausgeübten Tätigkeiten (Berufe) verglichen. Entsprechende Informationen werden vom Arbeitgeber im Rahmen der sog. Jahresmeldung (§ 28a des SGB V) an die Krankenkasse übermittelt, die diese Daten entsprechend der Datenerfassungs- und -übermittlungsverordnung [15] an die Sozialversicherungsträger weiterleitet. Der Natur dieser Daten entsprechend können in diese Vergleiche nur Personen einbezogen werden, die am Stichtag sozialversicherungspflichtig beschäftigt sind.

Aus den Daten der Statistischen Ämter des Bundes und der Länder konnte eine entsprechende Datenquelle herangezogen werden, die nur Daten zu sozialversicherungspflichtig Beschäftigten enthält, die in Niedersachsen bzw. in der Bundesrepublik wohnhaft sind. Ein Vergleich anhand der in den Daten der Statistischen Ämter des Bundes und der Länder ansonsten üblichen Gliederung nach Arbeitsortprinzip (Daten für Beschäftigte in Unternehmen, die im jeweiligen Bundesland ihren Hauptsitz haben) war nicht möglich, da die AOK-Daten keine entsprechenden Informationen enthalten.

### In den Vergleichen verwendete Variablen

Neben den oben genannten demografischen Informationen ist für die AOK-Versicherten auch die Versichertenart verfügbar, die eine Identifikation der sozialversicherungspflichtig Beschäftigten ermöglicht. Für eine eindeutige Identifikation der Versichertenart wurde die Expertise von AOK-Fachexperten eingeholt.

In den Verlaufsdaten sind für sozialversicherungspflichtig Beschäftigte der amtliche dreistellige Tätigkeitsschlüssel mit Angaben zum Beruf sowie die nach 8 Kategorien gegliederte amtliche Kategorisierung zur Stellung im Beruf und zur Berufsausbildung nach 7 Kategorien verfügbar. In den amtlichen Statistiken über sozialversicherungspflichtig Beschäftigte werden die Daten zur beruflichen Tätigkeit in zusammengefasster Form berichtet: Qualifikation wird in 3 Kategorien dargestellt (mit/ohne Berufsausbildung und mit Hochschulabschluss), und aus den ursprünglich dreistelligen detaillierten Berufsbezeichnungen wurden für die amtlichen Berichte 10 Berufsgruppen gebildet (s. dazu [Tab. 5](#)). Die AOK-Daten ließen sich nach einer gründlichen Plausibilitätsprüfung in gleicher Weise wie in den statistischen Daten zusammenfassen.

## Ergebnisse

### Geschlecht

Die Geschlechterverteilung der Versicherten der AOK Niedersachsen weist für das Jahr 2004 mit 53,2% einen etwas höheren Frauenanteil auf; im Land Niedersachsen ist der Frauenanteil mit 51% geringfügig niedriger, das Gleiche gilt für die Bundesrepublik (Frauenanteil 51,1%). Im Jahr 2009 sinkt der Frauenanteil unter den

AOK-Versicherten auf 52,6%, in Niedersachsen beträgt er 50,9% und in der Bundesrepublik 51% (■ **Tab. 1**).

## Alter

Die Altersverteilungen der AOK-Versicherten, der Bevölkerung Niedersachsens sowie der Bundesrepublik gleichen sich in den Jahren 2004 und 2009 (■ **Abb. 1, 2**); diese Verteilungen nähern sich zwischen 2004 und 2009 weiter an. Die mittleren Altersgruppen sind unter den AOK-Versicherten etwas schwächer vertreten als in der Bevölkerung Niedersachsens und der Bundesrepublik. Diese Unterschiede sind jedoch gering und überschreiten bei Personen unterhalb des 65. Lebensjahrs in keinem Fall 2 Prozentpunkte der jeweiligen Gesamtgruppe; dabei zeigen sich bei den männlichen Versicherten geringere Unterschiede als bei den Frauen (die Geschlechterunterschiede sind nicht in der Grafik dargestellt). Keine Unterschiede zeigen sich für die Altersgruppen unterhalb des 20. Lebensjahrs sowie zwischen dem 50. und dem 65. Lebensjahr; in den darüber liegenden Altersgruppen unterscheiden sich die Verteilungen wiederum nur gering.

## Erwerbstätigkeit

Der Anteil an sozialversicherungspflichtig beschäftigten Männern unter den AOK-Versicherten lag im Jahr 2004 bei 35,1%, bei Frauen bei 21,7%. Beim Vergleich mit den Bevölkerungsdaten fällt der gleiche Anteil sozialversicherungspflichtig beschäftigter Männer in Niedersachsen im Jahr 2004 auf, aber auch die geringe Differenz zu den Relationen in der gesamten Bundesrepublik (■ **Tab. 2**). Bei Frauen lagen die Unterschiede bei ca. 5–7%, damit sind sozialversicherungspflichtig beschäftigte Frauen unter den AOK-Versicherten unterrepräsentiert.

Im Jahr 2009 steigt der Anteil an sozialversicherungspflichtig Beschäftigten unter den AOK-Versicherten auf 23,2% (Frauen) bzw. auf 35,8% (Männer), was die zunehmende Zahl von Personen in bezahlten Tätigkeiten und die abnehmende Arbeitslosigkeit abbildet. Unter der Bevölkerung Niedersachsens (28,1% bzw. 36,2%) bzw. in der Bundesrepublik (30%

Bundesgesundheitsbl 2013 · 56:447–454 DOI 10.1007/s00103-012-1626-9  
© Springer-Verlag Berlin Heidelberg 2013

J. Jaunzeme · S. Eberhard · S. Geyer

## Wie „repräsentativ“ sind GKV-Daten? Demografische und soziale Unterschiede und Ähnlichkeiten zwischen einer GKV-Versichertenpopulation, der Bevölkerung Niedersachsens sowie der Bundesrepublik am Beispiel der AOK Niedersachsen

### Zusammenfassung

Anhand von Daten der AOK Niedersachsen wird im vorliegenden Beitrag untersucht, ob sich die Versicherten hinsichtlich sozialstruktureller Merkmale von der Bevölkerung Niedersachsens und der Bundesrepublik unterscheiden. Die Vergleichsdaten stammen von den statistischen Ämtern; sie sind nach den gleichen Kriterien kodiert und können daher direkt für Vergleiche verwendet werden. Die Geschlechterverteilung unter den AOK-Versicherten unterscheidet sich von der in der Bevölkerung Niedersachsens und in der gesamten Bundesrepublik kaum. Gering sind auch die Unterschiede bei der Altersverteilung. Unter den AOK-Versicherten ist der Anteil der Beschäftigten bei den männlichen Versicherten etwa gleich groß wie in Niedersachsen und in Deutschland, bei Frauen ist er geringer. Die Beschäftigten unter den AOK-Versicherten sind im Durchschnitt niedriger qualifiziert als in Niedersachsen und in Deutsch-

land, die höheren Qualifikationsgruppen sind jedoch ausreichend stark vertreten. Wesentliche sozialstrukturelle Merkmale sind bei den AOK-Versicherten teilweise anders verteilt als in der Gesamtbevölkerung Niedersachsens bzw. in Gesamtdeutschland. Die diesbezügliche Repräsentativität ist jedoch keine Voraussetzung dafür, um Zusammenhänge zwischen sozialstrukturellen Merkmalen und gesundheitlichen Beeinträchtigungen bzw. Mustern der Inanspruchnahme medizinischer Leistungen untersuchen zu können. Entscheidend ist, dass in allen Subgruppen genügend Personen vorhanden sind. Aufgrund der hohen Fallzahl ist dies in den AOK-Daten gegeben.

### Schlüsselwörter

Routinedaten · Bevölkerungsvergleiche · Repräsentativität · Gesetzliche Krankenversicherung · Sozialstruktur

## How “representative” are SHI (statutory health insurance) data? Demographic and social differences and similarities between an SHI-insured population, the population of Lower Saxony, and that of the Federal Republic of Germany using the example of the AOK in Lower Saxony

### Abstract

Using data obtained from a statutory health insurance (AOK) in the federal state of Lower Saxony, this study examined whether there were differences between the insured population compared with that of Lower Saxony (Niedersachsen) and of Germany with respect to social structural characteristics. Data for the comparisons were provided by the statistical office of Germany, and all datasets were coded according to the same criteria. The differences in gender distribution and age distribution between the AOK, Lower Saxony, and Germany were small. The share of employed individuals among the insured compared with those of Lower Saxony and Germany did not differ for males, but it was lower in women. In the insured population a higher proportion of individuals had lower qualification

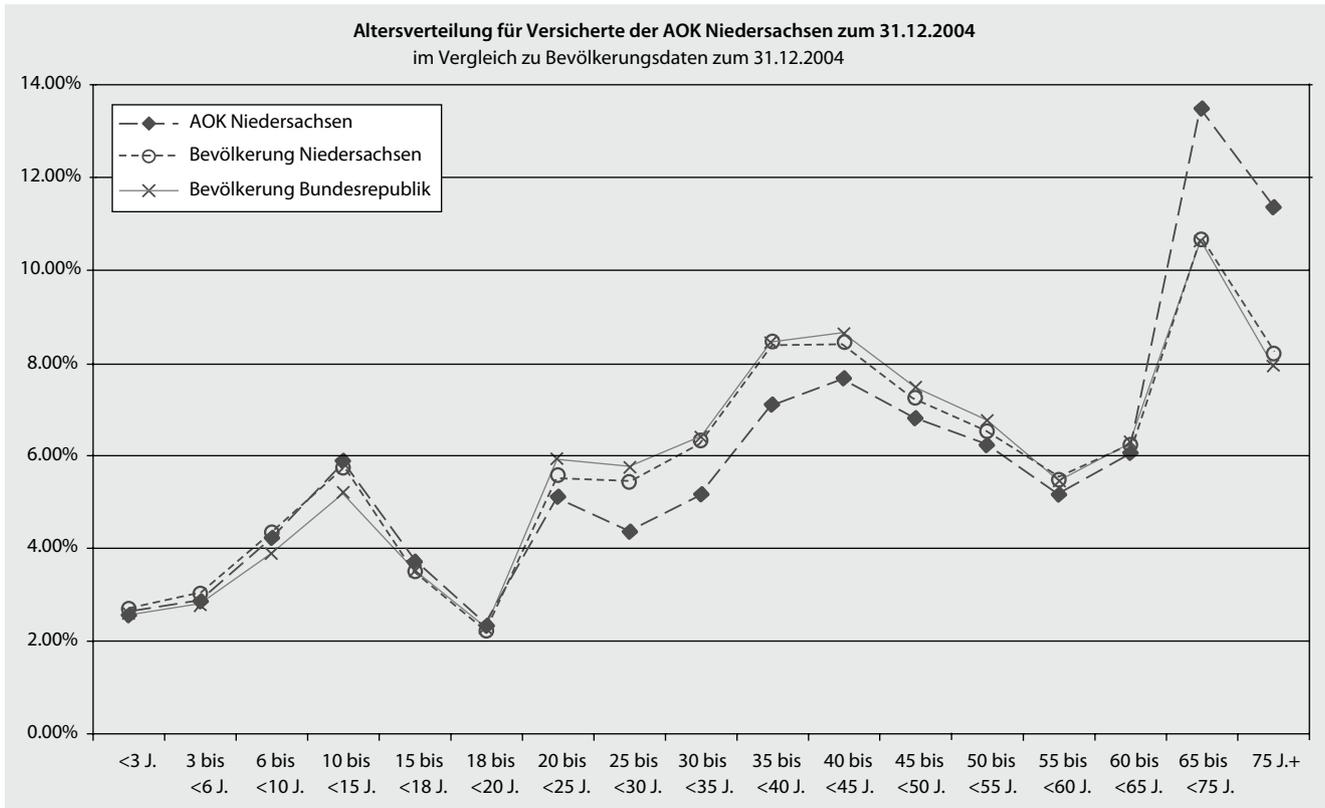
levels than in Lower Saxony or in Germany; the number of individuals with higher qualifications was, however, sufficient to permit statistical analyses. There were differences in the distributions of social structural characteristics between the health insurance population on the one hand and the populations of Lower Saxony and of Germany on the other. Due to the high number of cases, it is nevertheless possible to analyze associations between social structural variables, health impairments, and patterns of health care utilization.

### Keywords

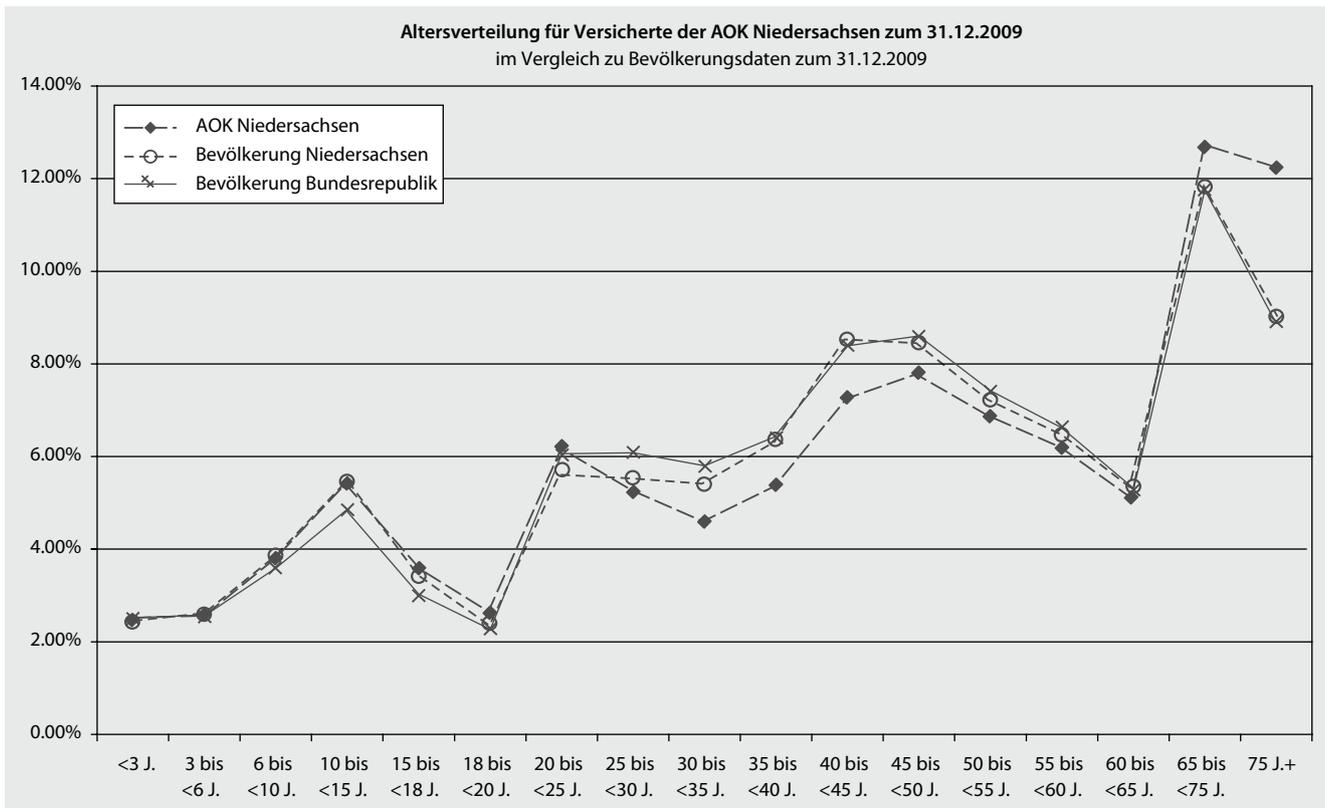
Routine data · Population-based comparisons · Representativity · Statutory health insurance · Social structure

bzw. 36,8%) steigen diese Anteile noch etwas deutlicher als unter den AOK-Versicherten.

Wird der Anteil der Erwerbstätigen nach Altersgruppen betrachtet, zeigen sich zwischen den 3 miteinander vergli-



**Abb. 1** Altersverteilung für Versicherte der AOK Niedersachsen zum 31.12.2004 im Vergleich zu Bevölkerungsdaten zum 31.12.2004



**Abb. 2** Altersverteilung für Versicherte der AOK Niedersachsen zum 31.12.2009 im Vergleich zu Bevölkerungsdaten zum 31.12.2009

**Tab. 2** Anteile der sozialversicherungspflichtig Beschäftigten unter den Versicherten der AOK Niedersachsen bzw. in der Gesamtbevölkerung Niedersachsens und der Bundesrepublik nach Geschlecht in den Jahren 2004 und 2009

	AOK-Niedersachsen		Niedersachsen		Bundesrepublik	
	Männlich	Weiblich	Männlich	Weiblich	Männlich	Weiblich
2004	35,1%	21,7%	35,1%	26,5%	35,8%	28,3%
	349.358	245.708	1.374.766	1.079.979	14.462.484	11.942.805
2009	35,8%	23,2%	36,2%	28,1%	36,8%	30,0%
	350.060	252.359	1.410.297	1.135.102	14.750.589	12.512.666

Fälle nach Wohnort des Beschäftigten ausgewählt.

**Tab. 3** Anteil der sozialversicherungspflichtig Beschäftigten unter den Versicherten der AOK Niedersachsen bzw. in der Gesamtbevölkerung Niedersachsens und der Bundesrepublik nach Altersgruppen in den Jahren 2004 und 2009

Altersgruppen	2004			2009		
	AOK Niedersachsen	Niedersachsen	Bundesrepublik	AOK Niedersachsen	Niedersachsen	Bundesrepublik
15 bis <20 Jahre	2,7%	2,8%	3,3%	2,6%	2,7%	2,9%
	16.230	68.955	875.848	15.449	67.485	789.757
20 bis <25 Jahre	9,3%	8,6%	8,9%	11,0%	8,8%	8,8%
	55.136	211.317	2.340.434	66.023	222.378	2.385.488
25 bis <30 Jahre	8,3%	9,8%	10,0%	9,7%	9,9%	10,6%
	49.071	240.682	2.637.864	58.583	249.763	2.867.126
30 bis <50 Jahre	55,3%	57,5%	56,4%	48,4%	53,4%	52,3%
	328.893	1.406.299	14.834.740	290.949	1.352.108	14.187.081
50 bis <60 Jahre	20,7%	18,2%	18,4%	23,5%	21,6%	21,8%
	122.965	445.702	4.829.272	141.154	546.693	5.908.902
60 bis <65 Jahre	3,7%	2,9%	3,0%	4,8%	3,7%	3,7%
	22.142	71.397	778.164	29.133	94.064	991.031
<b>Gesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
	<b>594.437</b>	<b>2.444.352</b>	<b>26.296.322</b>	<b>601.291</b>	<b>2.532.491</b>	<b>27.129.385</b>

Stand: 30.6. des jeweiligen Jahres.

**Tab. 4** Anteile der sozialversicherungspflichtig Beschäftigten unter den Versicherten der AOK Niedersachsen bzw. in der Gesamtbevölkerung Niedersachsens und der Bundesrepublik nach Berufsausbildung in den Jahren 2004 und 2009

	2004			2009		
	AOK Niedersachsen	Niedersachsen	Bundesrepublik	AOK Niedersachsen	Niedersachsen	Bundesrepublik
Ohne Berufsausbildung	31,5%	18,0%	18,8%	28,5%	16,0%	17,2%
	149.575	382.337	4.299.444	129.634	338.084	3.883.938
Mit Berufsausbildung	66,8%	73,4%	70,5%	68,7%	73,6%	70,3%
	317.358	1.562.283	16.168.691	312.896	1.554.945	15.915.146
Mit Hochschulabschluss	1,7%	8,7%	10,7%	2,8%	10,4%	12,6%
	8205	184.647	2.455.125	12.795	218.926	2.844.266
<b>Gesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
	<b>475.138</b>	<b>2.129.267</b>	<b>22.923.260</b>	<b>455.325</b>	<b>2.111.955</b>	<b>22.643.350</b>

Fälle nach Wohnort des Beschäftigten ausgewählt. Nur Fälle berücksichtigt, die eine gültige Angabe zur Qualifikation haben. Stand: 30.06. des jeweiligen Jahres.

chenen Populationen geringe Unterschiede; die größten finden sich in den mittleren Altersgruppen (■ Tab. 3).

## Qualifikation der sozialversicherungspflichtig Beschäftigten

Bei Betrachtung der Qualifikation der Beschäftigten zeigen sich – anders als bei den bisher dargelegten Merkmalen – zwischen AOK-Versicherten und den diesbezüglichen landes- bzw. bundesweiten Daten deutliche Unterschiede (■ Tab. 4). Für 2004 ist der Anteil der Beschäftigten ohne Berufsausbildung unter den AOK-Versicherten mit 31,5% deutlich höher als in der Bevölkerung Niedersachsens (18%) bzw. der bundesweiten Bevölkerung (18,8%). Ähnlich ungleichgewichtig ist der Anteil der Beschäftigten mit Berufsausbildung (66,8% bei den AOK-Versicherten vs. 73,4% in Niedersachsen bzw. 70,5% in der Bundesrepublik). Der Anteil an Hochschulabsolventinnen und -absolventen unter den AOK-Versicherten ist mit 1,7% nochmals geringer als unter der Bevölkerung Niedersachsens (8,7%) und der gesamten Bundesrepublik (10,7%). Bei einer getrennten Betrachtung nach Geschlechtern zeigt sich, dass diese Ungleichgewichte bei Männern stärker ausgeprägt sind als bei Frauen, was sich aus den geringeren Durchschnittseinkommen bei Frauen erklären könnte.

Für 2009 reproduzieren sich diese Relationen mit leicht veränderten Anteilen und Zahlen. Die Anteile an Hochschulabsolventen steigen insgesamt zwischen 2004 bis 2009 sowohl in Niedersachsen als auch in Deutschland an; diese Entwicklung zeigt sich aber unter den AOK-Versicherten noch deutlicher.

## Berufsgruppen

Unter den Versicherten der AOK finden sich als häufigste Berufsgruppen die Kategorien der Fertigungsberufe und der Dienstleistungsbereich (■ Tab. 5). Entsprechendes gilt, wenn man die Bevölkerung Niedersachsens bzw. der Bundesrepublik betrachtet. Allerdings sind die Fertigungsberufe in der AOK im Vergleich überrepräsentiert, während der Dienstleistungsbereich einen wesentlich kleineren Anteil einnimmt. Diese Unterschiede

**Tab. 5** Anteile und Anzahl der sozialversicherungspflichtig Beschäftigten unter den Versicherten der AOK Niedersachsen bzw. in der Gesamtbevölkerung Niedersachsens und der Bundesrepublik nach Berufsgruppen in den Jahren 2004 und 2009

	2004		2009			
	AOK Niedersachsen	Niedersachsen	Bundesrepublik	AOK Niedersachsen	Niedersachsen	Bundesrepublik
Landwirte u. Ä.	3,0%	1,8%	1,5%	3,0%	2,0%	1,5%
	17.385	43.133	399.289	18.087	47.162	407.399
Bergbau	0,2%	0,2%	0,1%	0,2%	0,1%	0,1%
	1322	3672	37.981	1082	3476	29.016
Fertigung	41,3%	29,8%	27,5%	38,5%	28,1%	25,8%
	242.980	697.888	7.290.477	229.398	678.684	7.053.352
Technische	1,6%	6,2%	7,0%	1,9%	6,2%	6,9%
	9645	144.299	1.848.798	11.523	148.892	1.893.073
Dienstleistungen, darunter	53,8%	62,0%	63,9%	56,4%	63,7%	65,7%
	316.295	1.451.743	16.947.437	336.495	1.538.068	17.997.256
Waren- und Dienstleistungskaufleute	7,7%	12,0%	12,1%	8,8%	12,0%	12,2%
	45.152	280.193	3.200.423	52.480	289.511	3.338.105
Verkehr	12,4%	7,8%	7,3%	11,8%	7,8%	7,2%
	73.033	182.247	1.933.724	70.488	188.828	1.984.463
Organisation. Büro	8,8%	19,3%	21,6%	9,8%	19,4%	22,0%
	51.868	451.868	5.737.595	58.524	468.206	6.015.189
Gesundheit	5,2%	7,9%	7,3%	5,8%	8,2%	7,6%
	30.745	184.755	1.940.443	34.470	198.297	2.085.841
Soziale und wissenschaftliche	4,9%	6,1%	6,1%	5,7%	6,9%	6,9%
	28.890	141.718	1.624.256	34.241	167.420	1.884.649
Allgemeine Dienstleistung	14,7%	9,0%	9,5%	14,5%	9,3%	9,8%
	86.607	210.962	2.510.996	86.292	225.806	2.689.009
Gesamt	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Stand: 30.06. des jeweiligen Jahres.

sind deutlich und insbesondere vor dem Hintergrund der Verteilung in der Bevölkerung Niedersachsens auffällig, da die betrachteten AOK-Versicherten eine Untergruppe dieser Population sind.

Wird der Dienstleistungsbereich weiter ausdifferenziert, zeigt sich, dass die Bereiche „Verkehr“ und „allgemeine Dienstleistung“ unter den AOK-Versicherten stärker vertreten sind als erwartet.

## Diskussion

Die Untersuchungen haben gezeigt, dass sich die Geschlechterverteilung unter den AOK-Versicherten kaum von der in der Bevölkerung Niedersachsens und der gesamten Bundesrepublik unterscheidet. Auch bei der Altersverteilung finden sich zwischen den AOK-Versicherten und den

beiden anderen Populationen nur geringe Unterschiede. Entsprechende Differenzen werden erst in den mittleren Altersgruppen deutlich; sie überschreiten nur bei über 65-Jährigen 2 Prozentpunkte.

Unter den AOK-Versicherten ist der Anteil der Beschäftigten bei den Männern in etwa gleich groß, bei den Frauen aber geringer als in Niedersachsen bzw. in Deutschland insgesamt. Ein höherer Anteil AOK-Versicherter findet sich in den geringer qualifizierten Positionen, ohne dass die höheren Qualifikationsgruppen fehlen. Deren Anteil steigt über die Jahre insbesondere bei den Frauen an, und es sind in allen Kategorien ausreichend große Fallzahlen vorhanden, um statistische Analysen durchführen zu können. Bei den Berufsgruppen sind unter den AOK-Versicherten im Vergleich zur Bevölkerung in

Niedersachsen und im Bund Fertigungsberufe überrepräsentiert, der Dienstleistungsbereich ist hingegen schwächer vertreten. Bei den Dienstleistungsberufen findet sich wiederum ein Übergewicht der einfacheren Tätigkeiten.

Vor diesem Hintergrund lässt sich die eingangs formulierte Forschungsfrage wie folgt beantworten: Im untersuchten GKV-Datensatz liegen fast alle Qualifikations- und Berufsgruppen vor, die auch in den Daten der statistischen Ämter enthalten sind. Es finden sich im Vergleich der 3 Populationen zwar diesbezüglich leichte Verschiebungen, aber nicht so stark, dass der Schluss gerechtfertigt ist, dass es sich bei den AOK-Versicherten nur oder auch überwiegend um einkommensschwache und bildungsferne soziale Gruppen handelt. Anhand der vorliegenden Daten können zwar keine Aussagen über die Gesamtbevölkerung gemacht werden, aber es können Zusammenhänge zwischen sozialstrukturellen, demografischen Merkmalen und Erkrankungen bzw. Mustern der Inanspruchnahme untersucht werden. Dies gilt auch für Versicherte mit einem hohen Qualifikationsniveau oder mit technischen Berufen, da sie unter den Versicherten der AOK Niedersachsen in ausreichend großer Zahl vertreten sind. Damit eignen sich diese Daten, um eine Vielzahl von Fragen der Gesundheitsberichterstattung zu behandeln, insbesondere zu sozialen Unterschieden bei Morbiditäten oder zur Versorgung mit medizinischen Leistungen.

Der vorliegende Aufsatz argumentiert, dass die Aussagekraft von GKV-Daten nicht primär auf deren Repräsentativität beruht, sondern vielmehr, dass die relevanten Merkmale ausreichend häufig in diesen Daten vertreten sein müssen, um eine Stratifizierung vornehmen zu können. Dieses Argument gilt jedoch nicht uneingeschränkt. So liegen sozialstrukturelle Informationen nur für berufstätige Versicherte vor, d. h., die für gesundheitsbezogene Fragestellungen sehr relevante Gruppe der über 65-Jährigen ist in GKV-Daten sozialstrukturell nicht klassifiziert. Ein weiteres Problem ergibt sich aus den Prävalenzen unterschiedlicher Erkrankungen: Während relativ häufige Krankheitsbilder nach Stratifizierung statistisch

untersucht werden können, kommt man bei eher seltenen Erkrankungen schnell an eine Grenze.

Eine Generalisierung der obigen Schlussfolgerung auf Datensätze anderer Krankenversicherungen ist nicht zulässig, weil dazu für jeden einzelnen Datensatz Informationen zur Größe der Versicherungsschaft sowie zu ihrer sozialstrukturellen Zusammensetzung erforderlich sind [3]. Die Chancen, krankheits- und sozialstrukturbezogene Analysen durchführen zu können, steigen mit zunehmender Heterogenität und Größe der Versichertenpopulation. Es sollte daher nicht gefragt werden, ob eine Versichertengruppe im Hinblick auf definierte Merkmale repräsentativ ist, sondern ob die interessierenden Merkmale bzw. Merkmalskombinationen ausreichend häufig vertreten sind, um die interessierenden Zusammenhänge statistisch sichtbar werden zu lassen. Trotzdem werden in den Daten der gesetzlichen Krankenversicherungen bestimmte Personengruppen nicht oder zu selten vertreten sein. Dies bezieht sich z. B. auf typischerweise in der PKV versicherte Personen, insbesondere Beamte, Pensionäre und Selbstständige [16].

Der folgende Punkt betrifft die Verwendbarkeit der Daten; der Vollständigkeit halber muss darauf hingewiesen werden, dass sie auch mit Blick auf ihre Validität untersucht werden müssen. Validierungsstudien wurden unter anderem für ambulante und stationäre Diagnosen [17] sowie für auf Krankenscheinen angegebene psychiatrische Diagnosen [18] durchgeführt. Es liegen damit Informationen zur Brauchbarkeit und Verwendbarkeit der Daten vor. Bei den hier verwendeten sozialstrukturellen Merkmalen betrifft dies die Korrektheit der Jahresmeldungen zur Qualifikation sowie die Berufsgruppenzugehörigkeit, die im Falle der GKV-Zugehörigkeit jährlich vom Arbeitgeber an die Krankenversicherung erfolgen muss [19]. Auch wenn erste eigene Untersuchungen diesbezüglich auf eine recht hohe Validität schließen lassen, steht eine systematische Untersuchung aus. Mit den hier präsentierten Ergebnissen wurde aber ein Anfang gemacht, und der verwendete Datensatz hat sich als tauglich erwiesen, um

soziale Differenzierungen im Hinblick auf Erkrankungen und Inanspruchnahmen zu untersuchen. Die hier praktizierte Verfahrensweise sollte auch auf andere GKV-Datensätze übertragen werden können, weil die erforderlichen statistischen Daten auf Landes- und Bundesebene allgemein verfügbar sind.

## Korrespondenzadresse

**Prof. Dr. S. Geyer**  
Medizinische Soziologie OE 5420,  
Medizinische Hochschule Hannover  
30625 Hannover  
Geyer.siegfried@mh-hannover.de

**Danksagung.** Wir danken Ingrid Schubert, Peter Ihle, Sascha Abbas und Rebecca Hein (Forschungsgruppe Primärmedizinische Versorgung, Universität Köln) für eine konstruktive Diskussion des Textes und den beiden anonymen Gutachtern für ihre sehr hilfreichen Kommentare.

**Interessenkonflikt.** Der korrespondierende Autor weist für sich und seine Koautoren auf folgende Beziehungen hin: Sveja Eberhard ist Mitarbeiterin der AOK Niedersachsen, bei Siegfried Geyer und Jelena Jaunzeme bestehen keine Interessenskonflikte.

## Literatur

- Schnee M (2008) Sozioökonomische Strukturen und Morbidität in den gesetzlichen Krankenkassen. In: Böcken J, Braun B, Amhof R (Hrsg) Gesundheitsmonitor 2008. Gesundheitsversorgung und Gestaltungsoptionen aus der Perspektive der Bevölkerung. Verlag Bertelsmann Stiftung, Gütersloh, S 88–104
- Hoffmann F, Icks A (2012) Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen für die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. Gesundheitswesen 74:291–297
- Hoffmann F, Icks A (2011) Diabetes prevalence based on health insurance claims: large differences between companies. Diabet Med 28:919–923
- Mielck A (2000) Soziale Ungleichheit und Gesundheit. Huber, Bern
- Braun B, Müller R (2006) Gesundheitsberichterstattung mit GKV-Daten. In: Müller R, Braun B (Hrsg) Vom Quer- zum Längsschnitt mit GKV-Daten. Schwäbisch Gmünd: GEK, Gmünder Ersatzkasse: 6–22
- Bowles D, Damm O, Greiner W (2011) Gesundheitsbezogene Versorgungsforschung mit GKV-Routinedaten – Grenzen am Beispiel der Prophylaxe venöser Thromboembolien in der Hüft- und Kniegelenkendoprothetik. Gesundh Ökon Qual Manag 16:96–107
- Geyer S (2008) Sozialstruktur und Krankheit: Analysen mit Daten der gesetzlichen Krankenversicherung. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 51:1164–1172
- Huber J, Mielck A (2010) Morbidität und Gesundheitsversorgung bei GKV- und PKV-Versicherten Forschungsstand empirischer Studien. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 53:925–938
- Swart E, Ihle P (2005) Routinedaten im Gesundheitswesen. Hans Huber, Bern
- Schubert I, Köster I, Küpper-Nybelen J, Ihle P (2008) Versorgungsforschung nach mit GKV-Routinedaten. Nutzungsmöglichkeiten versichertenbezogener Krankenkassendaten für Fragestellungen der Versorgungsforschung. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 51:1095–1105
- Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) (2008) Gute Praxis Sekundärdatenanalyse, Revision nach Überarbeitung. [http://www.gesundheitsforschung-bmbf.de/\\_media/GPS.pdf](http://www.gesundheitsforschung-bmbf.de/_media/GPS.pdf). URL: [http://www.gesundheitsforschung-bmbf.de/\\_media/GPS.pdf](http://www.gesundheitsforschung-bmbf.de/_media/GPS.pdf)
- (o A) (2007) Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen. Gutachten 2007 des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen: Kooperation und Verantwortung. Voraussetzungen einer zielorientierten Gesundheitsversorgung. Drucksache 16/6339. Deutscher Bundestag, Berlin
- Statistische Ämter des Bundes und der Länder. Regionaldatenbank Deutschland. <http://www.regionalstatistik.de>
- Bundesagentur für Arbeit (2012) Schlüsselverzeichnis für Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung, Version 2003. [http://www.arbeitsagentur.de/nn\\_27830/Navigation/zentral/Veroeffentlichungen/Schlüsselverzeichnis/Schlüsselverzeichnis-Nav.html](http://www.arbeitsagentur.de/nn_27830/Navigation/zentral/Veroeffentlichungen/Schlüsselverzeichnis/Schlüsselverzeichnis-Nav.html). URL: [http://www.arbeitsagentur.de/nn\\_27830/Navigation/zentral/Veroeffentlichungen/Schlüsselverzeichnis/Schlüsselverzeichnis-Nav.html](http://www.arbeitsagentur.de/nn_27830/Navigation/zentral/Veroeffentlichungen/Schlüsselverzeichnis/Schlüsselverzeichnis-Nav.html)
- Bundesministerium für Arbeit und Sozialordnung (2006) Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung. Bundesgesetzblatt 1:152
- Finkenstädt V, Keßler T (2012) Die sozioökonomische Struktur der PKV-Versicherten – Ergebnisse der Einkommens- und Verbrauchsstichprobe 2008. Wissenschaftliches Institut der PKV, Köln
- Schubert I, Ihle P, Köster I (2010) Interne Validierung von Diagnosen in GKV Routinedaten: Konzeption mit Beispielen und Falldefinition. Gesundheitswesen 72:316–322
- Hach I, Rentsch A, Ruhl U, Becker E et al (2003) Validität von Krankenscheindiagnosen psychischer Störungen. Gesundheitswesen 65:359–364
- Geyer S (2005) Die Bestimmung der sozioökonomischen Position in Prozessdaten und ihre Verwendung in Sekundärdatenanalysen. In: Swart E, Ihle P (Hrsg) Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Huber, Bern, S 203–214

Hier steht eine Anzeige.

