Mortalitätsanalyse mit den Daten der Deutschen Rentenversicherung – Methodische Überlegungen zum Stichprobencharakter

Thomas Salzmann / Martin Kohls
Universität Rostock

1 Einleitung

Für Sterblichkeitsuntersuchungen bieten sich in Deutschland zwei methodisch unterschiedliche Herangehensweisen an. Zum einen liefert die amtliche Statistik stichtagsbezogene Bevölkerungsbestände und Sterbefälle aufgeschlüsselt nach Alter, Geschlecht, Staatsangehörigkeit, Todesursachen, usw. Mit diesen Angaben lassen sich Mortalitätsmaße (z. B. altersstandardisierte Sterberaten oder Sterbetafeln) für vorher klassifizierte Personengruppen berechnen. Problematisch bei dieser Vorgehensweise ist das Querschnittsdesign, das Sterblichkeitsanalysen nur durch Korrelationszuweisungen auf der Makroebene erlaubt. Eine andere Möglichkeit stellen bevölkerungsrepräsentative Längsschnittanalysen dar. Hier können „reale" Lebensläufe von Personen verfolgt werden. Dieses Untersuchungsdesign ist allerdings sehr aufwändig und dementsprechend selten. Die Fallzahlen zu bestimmten Merkmalen sind in diesen Studien oft sehr gering und können die Repräsentativität der Ergebnisse in Frage stellen.

Durch den Aufbau des Forschungsdatenzentrums der Rentenversicherung ist es möglich geworden, die prozessproduzierten Daten der Rentenversicherung für Sterblichkeitsuntersuchungen und andere wissenschaftliche Fragestellungen zu nutzen (vgl. *Scholz* 2006, *Gaudecker* 2006). Zudem besitzen die Daten der Deutschen Rentenversicherung eine hohe Validität. Während zwischen der amtlichen Statistik und den tatsächlichen (nicht messbaren) Beständen, hauptsächlich in den oberen Altersstufen und bei der nicht-deutschen Bevölkerung, erhebliche Abweichungen aufgrund der kumulierten Fortschreibungsfehler bestehen dürften, sind für diese Personengruppen in den Datensätzen der Rentenversicherung keine Verzerrungen zu erwarten.

Um Sterblichkeitsanalysen mit den Daten der Deutschen Rentenversicherung vornehmen zu können, bedarf es der Datensätze Rentenbestand und Rentenwegfall. Das Forschungsdatenzentrum der Rentenversicherung hat entsprechende Scientific Use Files (SUFs) aus den prozessproduzierten Daten der Deutschen Rentenversicherung gezogen. Dies sind zum einen die 1% Stichprobe aus dem gesamten Rentenbestand zum 31.12. sowie die 10%ige Stichprobe aus den Rentenwegfällen von Januar bis November eines Jahres. Aufgrund der nicht verbundenen Stichprobenziehung können beide SUF-Datensätze als voneinander unabhängige Stichproben angesehen werden.[1] Um sie für Mortalitätsanalysen nutzen zu können, erfordert es vorab einiger methodischer Überlegungen, die den vorliegenden Stichprobencharakter der beiden Datensätze berücksichtigen. In der vorliegenden Arbeit wird aufgezeigt, welche methodi-

1 Es ist unbestritten möglich, dass z. B. ein Rentner gleichzeitig in der SUF-Stichprobe Rentenwegfall 2002 und in der Stichprobe Rentenbestand 2001 vertreten ist. Allerdings beträgt die Wahrscheinlichkeit hierfür lediglich 0,1%.

schen Probleme mit den beiden Datensätzen auftreten, wenn Signifikanzaussagen getätigt werden sollen. Wir geben zunächst einen kurzen Überblick über die Möglichkeiten von Mortalitätsanalysen mit Hilfe der amtlichen Statistik in den USA. Dabei zeigt sich, dass in den Vereinigten Staaten bereits seit längerer Zeit Erfahrungen mit der Verwendung von Stichproben für bevölkerungsrepräsentative Sterblichkeitsuntersuchungen bestehen. Eine Möglichkeit, um Analysen unter diesen Voraussetzungen vornehmen zu können, wird von uns auf die Daten der Deutschen Rentenversicherung übertragen. Aussagen zu differentieller Sterblichkeit unter der Beachtung der üblichen statistischen Testverfahren sind dadurch noch nicht möglich. Mit Hilfe einer Monte-Carlo-Simulation lassen sich aber Konfidenzintervalle ermitteln, die den Stichprobencharakter entsprechend berücksichtigen. Da längst nicht alle wissenschaftlichen Sterblichkeitsuntersuchungen mit statistischen Testverfahren arbeiten, wird zunächst beschrieben, wie diese Testverfahren in Mortalitätsanalysen integriert werden können.

2 Statistische Testverfahren für signifikante Aussagen in Sterblichkeitsuntersuchungen

Das Statistische Bundesamt ermittelt mit Hilfe des Sterbeziffernverfahrens auf der Basis von Jahresendbeständen und unter der Annahme gleichverteilter Sterbefälle im Kalenderjahr jährlich Sterbewahrscheinlichkeiten nach Einzelalter. Die Sterbewahrscheinlichkeiten bilden die Grundlage für die Berechnung der jeweiligen vollständigen oder abgekürzten Sterbetafel. Neben dem Sterbeziffernverfahren werden in der Literatur auch andere Möglichkeiten zur Berechnung von Sterbewahrscheinlichkeiten diskutiert (vgl. *Flaskämper* 1962; *Esenwein-Rothe* 1981; *Dinkel* 2006).

Allgemein werden Sterbewahrscheinlichkeiten in einem Alter x im Jahr t folgendermaßen berechnet:

Formel 1:

$$ {}_1q_x = \frac{Sterbefälle\ im\ Alter\ x\ im\ Jahr\ t}{Personen, die\ im\ Jahr\ t\ das\ exakte\ Alter\ x\ erreicht\ haben} $$

Der heutige Bestand des Nenners (in einem bestimmten Alter x) ist für die gesamte Bundesrepublik durch die Bevölkerungsfortschreibung seit der letzten Volkszählung 1987 bekannt und kann als Grundgesamtheit angesehen werden. Der Zähler, die Sterbefälle der Risikopopulation im Alter x, kann nur aus dem Bevölkerungsbestand in diesem Alter stammen, wenn von dem Sonderproblem der Einbeziehung von Migration abstrahiert wird. Der Bestand des Nenners kann als nicht zufallsschwankende Größe angesehen werden. Der Bestand des Zählers dagegen ist das Ergebnis einer ganz bestimmten Zufallsrealisation von Sterbefällen im Alter x im Jahr t, die sich aus der altersspezifischen Risikopopulation im Jahr t ergibt. Eine zufällige andere Realisation wäre ebenso möglich.

Betrachten wir beispielhaft die insgesamt 11.847 männlichen Sterbefälle im Alter 75 im Jahr 2002: Ohne Weiteres hätte auch eine Zahl von 11.712 oder 12.020 Sterbefällen realisiert werden können. Aufgrund der Zufälligkeit sollte für die Sterbefälle (im Alter x im Jahr t), in Abhängigkeit der Fallzahl, ein Konfidenzintervall für die konkret beobachteten Sterbefälle bestimmt werden, welches mit einer bestimmten Wahrscheinlichkeit angibt, in welchen Grenzen der „wahre" Wert des Beobachtungswertes liegt. Da aber die Grundgesamtheit meist nicht bekannt ist, muss die Testsituation umgedreht werden. Dabei wird unterstellt, dass die beobachtete Stichprobenrealisation (der z. B. die 11.847 männlichen Sterbefälle im Jahr 2002) den wahren Erwartungswert abbildet. Es wird nun gefragt, in welchen Grenzen alle zufälligen Stichproben liegen müssten, um anzunehmen, dass das Resultat der Grundgesamtheit entsprechen könnte.

Konfidenzintervalle können, in Abhängigkeit vom konkreten Untersuchungsgegenstand, für Sterbefälle oder Sterbewahrscheinlichkeiten gebildet werden. Um Konfidenzintervalle ermitteln zu können, ist die Berechnung der Varianz und der Standardabweichung des konkret beobachteten Wertes für die Sterbefälle bzw. der Sterbewahrscheinlichkeit ${}_1q_x$ notwendig. Da eine Sterbewahrscheinlichkeit einen Anteilswert[2] darstellt, ist die in der Statistik bekannte Varianzformel anzuwenden. Für die Varianz der Sterbefälle in einer binomialverteilten[3] Grundgesamtheit gilt, wobei ${}_1D_x$ die Sterbefälle in einem Alter x und P_x den Anfangsbestand im Alter x darstellt:

Formel 2:

$$Var({}_1D_x) = P_x \cdot {}_1q_x \cdot (1 - {}_1q_x)$$

Daraus ergibt sich die Standardabweichung:

Formel 3:

$$SE({}_1D_x) = \sqrt{Var({}_1D_x)}$$

Woraus sich folgendes Konfidenzintervall (für 1- $\alpha = 0{,}05$)[4] ableitet:

$${}_1D_x - 1{,}96 \bullet SE({}_1D_x) \leq {}_1D_x \leq {}_1D_x + 1{,}96 \bullet SE({}_1D_x)$$

Die Varianz und Standardabweichung für die beobachtete Sterbewahrscheinlichkeit ${}_1q_x$ kann folgendermaßen berechnet werden:

Formel 4:

$$Var({}_1q_x) = {}_1q_x \cdot (1 - {}_1q_x)/P_x$$

$$SE({}_1q_x) = \sqrt{Var({}_1q_x)}$$

Als Konfidenzintervall ergibt sich:

$${}_1q_x - 1{,}96 \bullet SE({}_1q_x) \leq {}_1q_x \leq {}_1q_x + 1{,}96 \bullet SE({}_1q_x)$$

2 Anteilswert bedeutet hier, dass von einer konkret beobachteten Zahl von Personen, die im Jahr t das Alter x erreicht haben, ein bestimmter Anteil von Personen im Verlauf des Jahres sterben.

3 Exakt definiert ist die zugrunde liegende Verteilung eine hypergeometrische Verteilung, da eine verstorbene Person nicht wieder in die Grundgesamtheit zurückgelangen kann. Es ist jedoch zulässig, ab einem bestimmten Stichprobenumfang n in Abhängigkeit von der Größe des Anteilwertes (hier ${}_1q_x$) die hypergeometrische Verteilung durch die Binomialverteilung zu approximieren.

4 Der Wert von 1,96 ergibt sich aus der Standardnormalverteilung aus der bekannt ist, dass 95 Prozent der möglichen Stichprobenrealisationen in einem Bereich des 1,96fachen der Standardabweichung nach beiden Seiten um den Erwartungswert der Grundgesamtheit verteilt liegen.

Die Formeln 3 und 4 zeigen, dass die Varianz der Sterbefälle bzw. der altersspezifischen Sterbewahrscheinlichkeit neben dem Wert der Sterbewahrscheinlichkeit allein von der beobachteten Fallzahl in der Stichprobe abhängt. Gerade in differentiellen Sterblichkeitsanalysen ist dies oftmals problematisch. Daher muss bei jeder Analyse zwischen der Differenziertheit der Stichprobe und der statistischen Aussagefähigkeit abgewogen werden.

Wir wollen das Vorgehen an unserem Anfangsbeispiel verdeutlichen und jeweils ein Konfidenzintervall der 95%-wahrscheinlichen Realisation der männlichen Sterbefälle und der altersspezifischen Sterbewahrscheinlichkeit für die Männer im Alter 75 bestimmen. Bei insgesamt 11.847 Sterbefällen im Alter 75 und 258.625 Männern, die dieses Alter erreichten, ergibt sich daraus eine beobachtete Sterbewahrscheinlichkeit von ${}_1q_{75} = 0{,}0458076$. Mit den Formeln 3 und 4 können die Konfidenzintervalle für ${}_1D_{75}$ und ${}_1q_{75}$ berechnet werden.

Konfidenzintervall (1- $\alpha = 0{,}05$) für die männlichen Sterbefälle im Alter 75:

$${}_1D_{75} - 1{,}96 \cdot SE\,({}_1D_{75}) \leq {}_1D_{75} \leq {}_1D_{75} + 1{,}96 \cdot SE\,({}_1D_{75})$$

$$11.639 \leq 11.847 \leq 12.055$$

Konfidenzintervall (1- $\alpha = 0{,}05$) für die männliche Sterbewahrscheinlichkeit im Alter 75:

$${}_1q_{75} - 1{,}96 \cdot SE\,({}_1q_{75}) \leq {}_1q_{75} \leq {}_1q_{75} + 1{,}96 \cdot SE\,({}_1q_{75})$$

$$0{,}0450019 \leq 0{,}0458076 \leq 0{,}0466134$$

Deutlich wird, dass selbst für die Werte der Bundesrepublik ein relativ breites Konfidenzintervall der 75-jährigen männlichen Sterbefälle bzw. der Sterbewahrscheinlichkeit ${}_1q_{75}$ entsteht. Werden bei differentiellen Mortalitätsanalysen die Fallzahlen immer weiter reduziert, ergibt sich jeweils ein noch breiteres Konfidenzintervall. Für die Bundesrepublik sind – wie für die meisten Industriestaaten – breite Konfidenzintervalle gerade in den jüngsten und in den obersten Altersstufen zu beobachten. Grundsätzlich kann gesagt werden: Je kleiner die Stichprobe ausfällt und/oder je weniger Sterbefälle zu beobachten sind, desto größer ist die Unsicherheit über die beobachteten Werte und umso breiter wird das Konfidenzintervall.

Sind die altersspezifischen Sterbewahrscheinlichkeiten ${}_1q_x$ für Alter $x = 0$ bis $x = n$ bekannt, können daraus Sterbetafeln und die daraus abgeleiteten Parameter e_0 bzw. e_x (die durchschnittliche Lebenserwartung im Alter 0 bzw. x) berechnet werden. Wenn die der Berechnung von e_x zu Grunde liegenden ${}_1q_x$ jeweils ein Konfidenzintervall aufweisen, dann muss gleiches auch für e_0 bzw. e_x gelten, in welche die einzelnen ${}_1q_x$-Werte gewichtet eingehen. Stammen die Sterbefälle aus der Grundgesamtheit der jeweiligen Risikopopulation, kann auch ein Konfidenzintervall für den Parameterwert e_0 bzw. e_x angegeben werden, die entweder mit Hilfe numerischer Näherungslösungen (*Chiang* 1984) oder durch Simulationsverfahren (z. B. Bootstrapping) ermittelt werden können.

Weitaus schwieriger ist die Situation, wenn für die Berechnung einer Sterbewahrscheinlichkeit in einem Alter beide Komponenten der Formel 1 Zufallsrealisationen darstellen, wie es für die SUFs „Rentenwegfall“ und „Rentenbestand“ des Forschungsdatenzentrums der Fall ist. Das gleiche Problem stellt sich in US-amerikanischen Studien schon seit längerer Zeit und soll im folgenden Abschnitt kurz umrissen werden.

3 Datensätze für Sterblichkeitsuntersuchungen in den USA

Die Vereinigten Staaten besitzen kein Melderegister, wie es z. B. viele europäische Staaten vorweisen können. Demografisch relevante Daten für die Gesamtbevölkerung werden seit 1790 alle zehn Jahre durch eine (klassische) Volkszählung erfasst. Jährliche Informationen für die demografischen Ereignisse Fertilität, Mortalität und Migration werden dagegen durch zahlreiche Stichproben erhoben oder mit Hilfe indirekter Messverfahren berechnet. Das Streben der US-amerikanischen Behörden neben dem „decennial census" umfangreichere Informationen für bevölkerungsrelevante Fragestellungen anzubieten, führte im Verlauf der letzten Jahrzehnte zu einer Vielzahl an umfangreichen Datensätzen.

Bis in die sechziger Jahre des 20. Jahrhunderts beschränkten sich demografische Analysen eher auf kleinräumige Gebiete, wie z. B. die Großstädte der USA. Vor allem für die Stadt Chicago liegen zahlreiche Untersuchungen vor. Unter anderem errechneten *Coombs* (1941) und *Kitagawa* und *Hauser* (1964) für den Großraum Chicago z. B. altersstandardisierte Sterberaten für bestimmte Einkommensklassen nach ethnischer Zugehörigkeit und Geschlecht. Um eine Zuordnung der verstorbenen Personen (Zählerpopulation) mit deren sozioökonomischen Merkmalen zu erhalten, sortierten sie die Sterbfälle den „census tracts" zu, in denen diese erfasst worden waren. Census tracts sind möglichst homogene geografische Einheiten mit jeweils ca. 4.000 Einwohnern, in welche die Vereinigten Staaten für Volkszählungszwecke unterteilt werden. Ein wesentlicher Vorteil dieser kleinräumigen Gebiete ist die Verfügbarkeit der demografischen Merkmale ihrer Wohnbevölkerung, welche *Coombs* und *Kitagawa* und *Hauser* gleichzeitig als Nennerpopulation für die Berechnung von altersstandardisierten Sterberaten dienten. Beide Untersuchungen basieren aber streng genommen ausschließlich auf einer Korrelation zwischen Zähler (Sterbefall) und der Ausprägung des Nenners (Risikopopulation).

Eine ebenfalls von *Kitagawa* und *Hauser* im Jahr 1973 als „Matched Record Study" veröffentlichte Untersuchung, kann trotz methodischer Kritik (vgl. *Pamuk* 1985, *Preston* und *Elo* 1994, *Schalick* 2000) als der Beginn der systematischen Mortalitätsanalyse auf nationaler Ebene für die Vereinigten Staaten angesehen werden. Die Datenbasis bildete eine Stichprobe aus allen zwischen Mai und August 1960 auf dem Gebiet der USA registrierten Sterbefällen im Alter 25 und älter. Für diese Sterbefälle wurde rückwirkend versucht, Personenangaben aus der Volkszählung 1960 (mit dem Stichtag 1. April) zuzuweisen. Dies gelang mit Hilfe einer zusätzlichen Follow-up Befragung der nächsten Angehörigen für rund 263.000 Verstorbene. Hier konnten zum ersten Mal sozioökonomische Ausprägungen dem entsprechenden Sterbefall zugeordnet werden.

Als eine Reaktion auf die „Matched Record Study" wurde der National Mortality Followback Survey (NMFS) konzipiert, der ausschließlich Personenangaben über Verstorbene bereitstellt. Der erste NMFS wurde 1961 erhoben. In einer mehr oder weniger zwanglosen Folge wurden bis heute fünf weitere Erhebungen seiner Art durchgeführt. Das Anliegen des NMFS war und ist es, mehr über die Sterblichkeit in Bezug auf die Ursachen und die Entwicklung bestimmter Krankheiten zu erfahren, als es über die alleinige Auswertung der Totenscheine (U.S. Standard Certificate of Death) möglich ist. Zu diesem Zweck wird eine geschichtete Stichprobe aus dem Current Mortality Sample (CMS) – ein Datensatz mit zehn Prozent aller auf dem Gebiet der USA ausgestellten Totenscheine – entnommen. Über einen Fragebogen an die nächsten Angehörigen und Telefoninterviews werden weitere Informationen über den Verstorbenen gesammelt. Neben einem feststehenden Frageteil, der demografische Merkmale erfasst, werden variable Themenschwerpunkte integriert, die je nach Aktualität neu hinzu- oder wieder herausgenommen werden.

Für Mortalitätsanalysen ist der NMFS im Grunde nicht geeignet, da hier ausschließlich Personenangaben für Verstorbene vorliegen. Um den Survey dennoch für Sterblichkeitsuntersuchungen nutzen zu können, müssten Angaben über die Risikopopulation vorliegen. Diese werden unter anderem mit dem National Health Interview Survey (NHIS) erhoben. Im NHIS wird der Gesundheitsstatus der "U.S. civilian non-institutionalized" Bevölkerung durch ein freiwilliges Einzelinterview pro Haushalt abgefragt. Der Survey wird seit 1957 jährlich durchgeführt und umfasst mittlerweile 49.000 Haushalte mit ca. 127.000 Personen und ist damit eines der größten Datenerhebungsprogramme in den USA. Der NHIS setzt sich aus vier einzelnen Panels zusammen, die jeweils für sich allein repräsentative Stichproben der nicht institutionalisierten amerikanischen Bevölkerung abbilden. Inhaltlich ist der Survey wie der NMFS in zwei Abschnitte gegliedert. Ein Teil befasst sich mit dem Erfragen von demografischen und sozioökonomischen Merkmalen sowie Basisinformationen zum Gesundheitszustand der im Haushalt lebenden Personen. Ein zweiter Teil beinhaltet je nach Relevanz und Wichtigkeit unterschiedliche Fragenkomplexe zu gesundheitsrelevanten Themen, welche es zusätzlich erlauben für bestimmte Bevölkerungsminderheiten gesonderte und detaillierte Angaben zu erhalten. Der NHIS ist für sich allein ebenfalls ein ungeeigneter Datensatz für Mortalitätsuntersuchungen, da der in der Zukunft eintretende Todesfall zum Interviewzeitpunkt unbekannt ist. Erst die Kombination von NMFS (Zähler) und NHIS (Nenner) ermöglicht Mortalitätsanalysen (vgl. *Pappas et al.* 1993, *Schalick et al.* 2000).

Eine weitere Kombinationsmöglichkeit des NHIS bietet sich mit dem National Death Index (NDI) an. Der NDI erfasst seit 1972 für jeden verstorbenen US-Bürger über 17 Jahren alle auf dem Totenschein eingetragenen Informationen. Zwar ist das Nichtberücksichtigen der „institutionalisierten" Personen ein Nachteil des NHIS, dennoch wird er (seit seiner konzeptionellen Umstellung im Jahr 1986) in Verbindung mit den NDI als die beste Datengrundlage für Mortalitätsanalysen auf dem Gebiet der USA angesehen (vgl. *Preston* 1994). Um vorhandene Angaben über eine Person aus dem NHIS mit dem NDI abgleichen zu können, wurden zwölf Kriterien entwickelt, über die eine im NHIS erfasste Person im NDI identifiziert werden kann. *Rogers et al.* (2000) nutzten diese Möglichkeit für umfangreiche Sterblichkeitsanalysen in den Jahren 1991 bis 1995.

Weitere Möglichkeiten für Mortalitätsuntersuchungen in Kombination mit dem NDI sind der National Longitudinal Mortality Survey (vgl. *Rogot et al.* 1992, *Preston und Elo* 1994) und die Longitudinal Studies of Aging (LSOA I+II). Während der NMLS ca. 1,3 Mio. Personenangaben aus 11 Stichproben des Current Population Survey[5] im Zeitraum von 1973 bis 1985 und einer Volkszählungsstichprobe aus dem Jahr 1980 umfasst, stehen im LSOA I+II ausgewählte Personen des NHIS 1994 unter Beobachtung.

5 Der Current Population Survey ist eine monatliche Stichprobe von ca. 50.000 Haushalten auf dem Gebiet der USA.

3.1 Eine mögliche Kombination von zwei voneinander unabhängigen Stichproben für Mortalitätsanalysen in den USA

Liegen keine Längsschnittdaten vor, müssen für Mortalitätsuntersuchungen zwei unterschiedliche Datenquellen für den Zähler (Sterbefälle) und für den Nenner (Risikopopulation) herangezogen werden. Am konkreten Beispiel des NMFS 1986 und des NHIS 1986 und deren Verwendung als Datenbasis für Mortalitätsanalysen durch zwei Autorengruppen soll gezeigt werden, wie bevölkerungsrepräsentative Aussagen zur Sterblichkeit gewonnen werden können und welche bestehende Restriktionen entsprechend berücksichtigt werden müssen.

In der Stichprobe des NMFS 1986 sind insgesamt 18.733 Sterbefälle US-amerikanischer Staatsangehöriger ab Alter 25 enthalten. Der Datensatz des NHIS 1986 umfasst 23.838 Haushalte mit insgesamt 62.052 Haushaltsmitgliedern. *Pappas et al.* (1993) und *Schalick et al.* (2000) verwendeten beide Datensätze für bevölkerungsrepräsentative Sterblichkeitsanalysen von Personen zwischen Alter 25 - 64 und kamen verallgemeinernd zu dem Ergebnis, dass eine zunehmende Diskrepanz zwischen den Sterblichkeitsrisiken bestimmter sozioökonomischer Bevölkerungsgruppen besteht. Aufgrund der unterschiedlichen Personengesamtheiten in beiden Stichproben mussten beide Datensätze zunächst aneinander angepasst werden, indem die „institutionalisierten" Personen aus dem NMFS sowie die „Hispanics" aus beiden Datensätzen ausgeschlossen wurden. Dadurch reduzierte sich vor allem im NMFS 1986 die ursprüngliche Fallzahl von 18.733 auf 6.985 (vgl. *Pappas et al.* 1993) bzw. 6.259 (vgl. *Schalick et al.* 2000).

Die Tatsache, dass beide Datensätze zwei voneinander unabhängige Stichproben darstellen, schränkt die Zahl möglicher Maßzahlen zur Darstellung von Sterblichkeitsunterschieden zwischen zwei Personengruppen stark ein. So konnten *Pappas et al.* (1993) bzw. *Schalick et al.* (2000) nur relationale Mortalitätsmaße verwenden, wie z. B. den Index of Dissimilarity, den Slope Index of Inequality, den Relative Index of Inequality oder das Relative Risiko. Mit dem letztgenannten Maß „Relatives Risiko" (RR) konnten die Autoren das (Sterbe)Risiko zwischen zwei Bevölkerungsgruppen mit bestimmten Merkmalsausprägungen vergleichen.

Formel 5: $$RR = \frac{\frac{D_{x,i}}{P_{x,i}}}{\frac{D_{x,\neq i}}{P_{x,\neq i}}} = \frac{d_{x,i}}{d_{x,\neq i}}$$

In differentiellen Sterblichkeitsuntersuchungen drückt RR aus, dass die mit einem bestimmten Merkmal versehene Bevölkerungsgruppe (z. B. sozioökonomischer Status i) ein *x*-Prozent höheres Sterberisiko aufweist als die zum Vergleich herangezogene Bevölkerungsgruppe ohne die betrachtete Merkmalsausprägung. Ist RR = 1, sind die Sterberaten der vergleichenden Bevölkerungsgruppen identisch und ein erhöhtes Sterberisiko kann ausgeschlossen werden. Ist RR > 1, ist jene mit der bestimmten Disposition versehene Bevölkerungsgruppe einem höheren Sterberisiko ausgesetzt als ihre Referenzpopulation. Ein Kausalbeweis zwischen der erhöhten Sterblichkeit und der betrachteten Merkmalsausprägung ist damit aber noch nicht hergestellt. Hierzu müsste bei Kontrolle aller weiteren in Frage kommenden Ursachen eine statistisch robuste Beziehung zwischen der Risikoausprägung *i* und der dadurch resultierenden Sterblichkeit hergestellt werden.

Ist wie im konkreten Fall das RR von den Autoren für eine bestimmte Personengruppe aus der Stichprobe berechnet worden, stellt sich die Frage nach der Übertragbarkeit auf die Allgemeinbevölkerung. Hierfür verwendeten sowohl *Pappas et al.* (1993) als auch *Schalick et al.* (2000) Gewichtungsfaktoren, mit deren Hilfe die Fallzahlen des NMFS und NHIS auf bevölkerungsrepräsentative Werte des Jahres 1986 umgerechnet werden können. Bei den Gewichtungsfaktoren handelt es sich jedoch um feststehende Faktoren, die sich für bestimmte Personengruppen je nach Merkmalsausprägung in ihrer Höhe unterscheiden. Mit den hoch gewichteten Fallzahlen konnte wiederum ein RR berechnet werden.

Unter der Annahme, dass hier zwei zufällige und voneinander unabhängige Stichprobenrealisationen vorliegen, kann diese Vorgehensweise bei Wiederholung der Stichprobenziehung dazu führen, dass man für bestimmte Alters- bzw. Bevölkerungsgruppen andere gewichtete Fallzahlen erhält als jene im Originaldatensatz. Verzerrungen bei der Berechnung der Sterberaten sind dadurch nicht ausgeschlossen. Ein Konfidenzintervall, welches die Unsicherheit durch die Zufälligkeit der Stichprobenfallzahlen und damit der Ergebnisse ausdrückt, gaben die Autoren aber nicht an.

4 Mortalitätsanalysen mit den SUF-Datensätzen der Deutschen Rentenversicherung unter Berücksichtigung des Stichprobencharakters

Wenn Mortalitätsanalysen mit den SUF-Datensätzen der Deutschen Rentenversicherung vorgenommen werden, ergeben sich die gleichen methodischen Restriktionen, wie für die oben vorgestellten Studien.[6]

Wir wollen beispielhaft für das Jahr 2002 die Sterblichkeit der Männer mit deutscher Staatsangehörigkeit gegenüber den Männern mit nicht-deutscher Staatsangehörigkeit im Alter 75 vergleichen. Dazu benötigen wir die Sterbefälle im Alter 75, die wir aus dem Datensatz Rentenwegfall 2002 herausfiltern können. Die (eigentlich) dazugehörige Risikopopulation erhalten wir aus den Daten des Rentenbestandes zum 31.12. des Jahres 2001 im Alter 75. In unserem konkreten Beispiel ist die Wahrscheinlichkeit mit 0,1% jedoch sehr gering, dass die beobachteten Sterbefälle aus dem Datensatz Rentenwegfall 2002 tatsächlich aus der Risikopopulation des Rentenbestandes zum 31.12. des Jahres 2001 stammen. Beide Datensätze sind quasi einzelne, voneinander unabhängige (Querschnitts-) Stichproben, die jeweils für sich Zufallsschwankungen ausgesetzt sind. Im vorliegenden Fall kann deshalb die Varianzformel (Formel 4) zur Berechnung eines Konfidenzintervalls nicht verwendet werden, da sie nur zulässig ist, wenn die beobachteten Ereignisse (Rentenwegfälle) eindeutig einer Personengesamtheit (Rentenbestand) zuzuordnen sind.

Ein weiteres Problem stellt der Gewichtungsfaktor (=Hochrechnungsfaktor) für die jeweilige Stichprobenziehung dar. Um bevölkerungsrepräsentative Aussagen zur Sterblichkeit in einem

6 Inzwischen ist es für Wissenschaftler möglich, in Abstimmung mit dem Forschungsdatenzentrum der Rentenversicherung im Rahmen von Gastwissenschaftler-Arbeitsplätzen bzw. kontrolliertem Fernrechnen eine höhere Ausschöpfungsquote der Stichprobe bzw. sogar eine Totalauswertung zu realisieren (vgl. *Himmelreicher, Gaudecker* und *Scholz* 2006).

Alter x zu erhalten, müssen beide Datensätze (in unserem Bsp.: Rentenwegfall 2002 und Rentenbestand 2001) mit den jeweils vorgegebenen und feststehenden Gewichtungsfaktoren korrigiert werden. Für die gesamte Stichprobe Rentenwegfall 2002 beträgt die Ausschöpfungsquote ca. 10% (vgl. DRV Bund 2005), was einen Gewichtungsfaktor von ca. 10 ergibt. Dagegen kann für die einzelnen Altersstufen nie eine genaue Ausschöpfungsquote von 10% realisiert werden. Auch der Rentenbestand muss gewichtet werden (bei einer Ausschöpfungsquote von 1% mit dem Faktor 100), weist aber alters- bzw. merkmalsspezifisch ebenfalls unterschiedliche Stichprobenrealisationen auf. Die Berechnung von altersspezifischer Sterblichkeit wird damit durch die zufällige Stichprobenauswahl verzerrt.

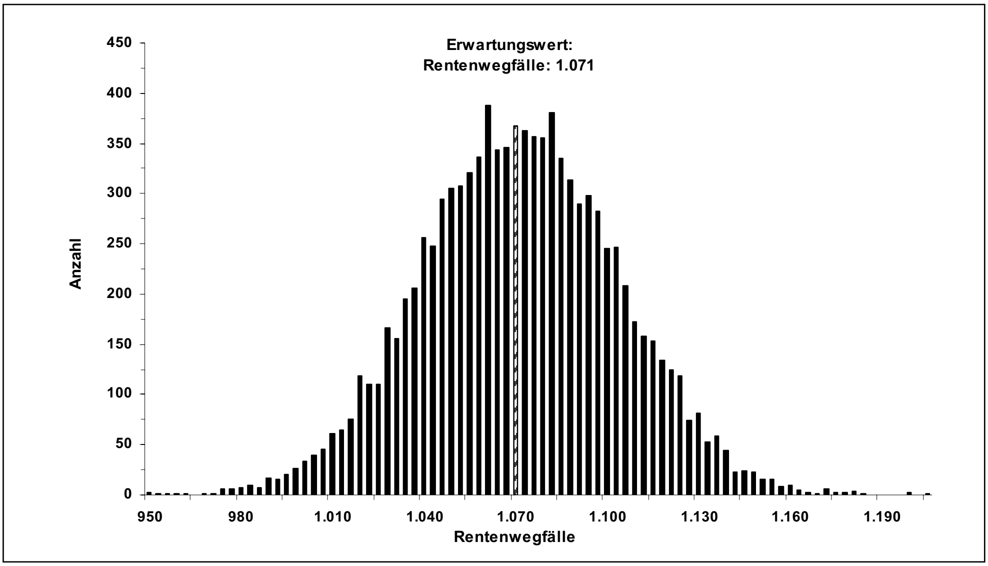
Wie schon die Autoren der US-amerikanischen Studien wollen wir für unser Beispiel das Maß Relatives Risiko für die Darstellung von möglichen Sterblichkeitsunterschieden zwischen den Männern mit deutscher und nicht-deutscher Staatsangehörigkeit verwenden. Um RR berechnen zu können, wird zunächst eine Häufigkeitsauszählung durchgeführt. Für die Männer mit deutscher Staatsangehörigkeit werden 1.071 Rentenwegfälle und 2.435 Rentenbestände gezählt. Für die Männer mit nicht-deutscher Staatsangehörigkeit ergeben sich 104 Rentenwegfälle und 259 Personen im Rentenbestand. Mit Hilfe der Formel 5 und unter der Berücksichtigung der Gewichtungsfaktoren ergibt sich ein RR = 1,09536, welches zunächst als ein höheres Sterberisiko der Männer mit deutscher gegenüber den Männer mit nicht-deutscher Staatsangehörigkeit interpretiert werden kann. Um dem Stichprobencharakter und damit der möglichen Zufälligkeit und Unsicherheit des Ergebnisses gerecht zu werden, muss für das RR in Abhängigkeit der beobachteten Fallzahlen und der Gewichtungsfaktoren ein Konfidenzintervall um den berechneten Wert der Stichprobenrealisierung angegeben werden. Da die statistische Voraussetzung, dass die Sterbefälle ein Bestandteil der zugehörigen Risikopopulation sein müssen nicht gegeben ist, darf die von *Woolf* (1955) vorgeschlagenen Methode in diesem Fall nicht angewandt werden. Die Schätzung eines Konfidenzintervalls um die „doppelt" zufallsschwankende Größe (RR) ist aber z. B. durch ein Simulationsverfahren (im konkreten Fall mittels einer Monte-Carlo-Simulation[7]) möglich.

Die Besonderheit der Monte-Carlo-Simulation liegt in ihrer Einfachheit des Algorithmus. Unter der Berücksichtigung des Stichprobencharakters wird der Erwartungswert beliebig oft simuliert. Jede Wiederholung (k) ist dabei von der vorangegangenen unabhängig. In unserem Beispiel werden die vier benötigten Stichprobenrealisationen der Formel 5 (die jeweiligen Rentenwegfälle und die Rentenbestände der Männer mit deutscher bzw. nicht-deutscher Staatsangehörigkeit im Alter 75) simuliert. Da der eigentliche Erwartungswert für die Sterbefälle und die Risikopopulation für unser Beispiel in der Grundgesamtheit nicht bekannt ist, werden die beiden Stichproben (Rentenwegfall und Rentenbestand) als Grundgesamtheit angesehen und die jeweilige Stichprobenrealisation als Erwartungswert unterstellt. Die Stichprobenrealisationen stellen einen Anteilswert am Stichprobenumfang dar. Um eine robuste Schätzung zu erhalten, simulieren wir k = 10.000 binomialverteilte Zufallsrealisationen.

In **Abbildung 1** ist die Häufigkeitsverteilung der simulierten Stichprobenfallzahl der Rentenwegfälle für die Männer (im Alter 75) mit deutscher Staatsangehörigkeit (für k = 10.000) dargestellt. Die einzelnen Ausprägungen sind entsprechend ihrer Häufigkeit auf der Abszisse abgetragen und ordnen sich normal verteilt um den Erwartungswert von 1.071. Das Ergebnis der Si-

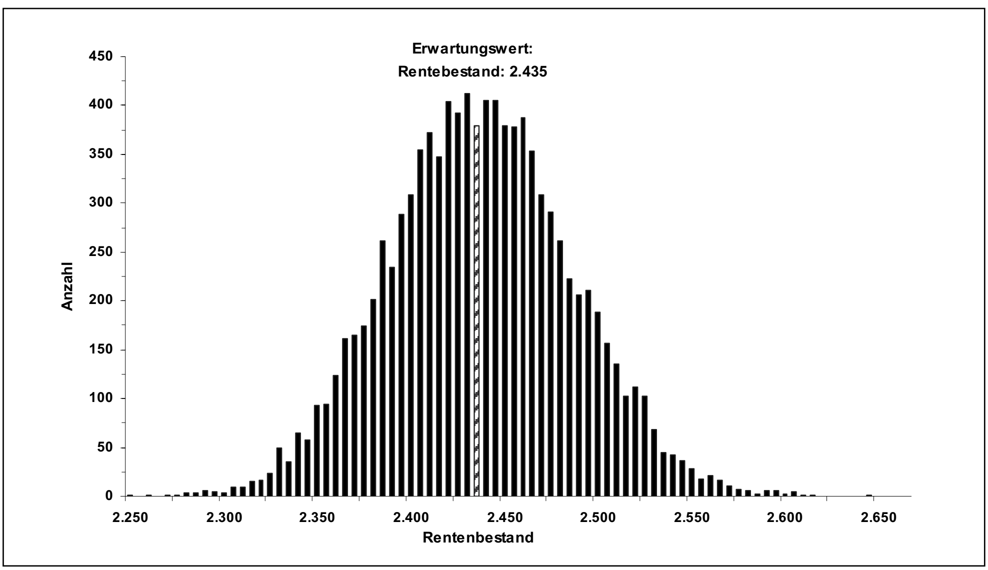
7 Für die Simulation wurde ein Makro unter Verwendung von VBA für Microsoft Excel programmiert.

Abbildung 1: Die Häufigkeitsverteilung der Rentenwegfälle mit dem Erwartungswert der Männer im Alter 75 mit deutscher Staatsangehörigkeit für 10.000 Simulationen



Quelle: Eigene Berechnung mit den Daten des FDZ-RV – SUFRTBN01VWITD und FDZ-RV – SUFRTWF02VWITD.

Abbildung 2: Die Häufigkeitsverteilung der Rentenbestände mit dem Erwartungswert der Männer im Alter 75 mit deutscher Staatsangehörigkeit für 10.000 Simulationen



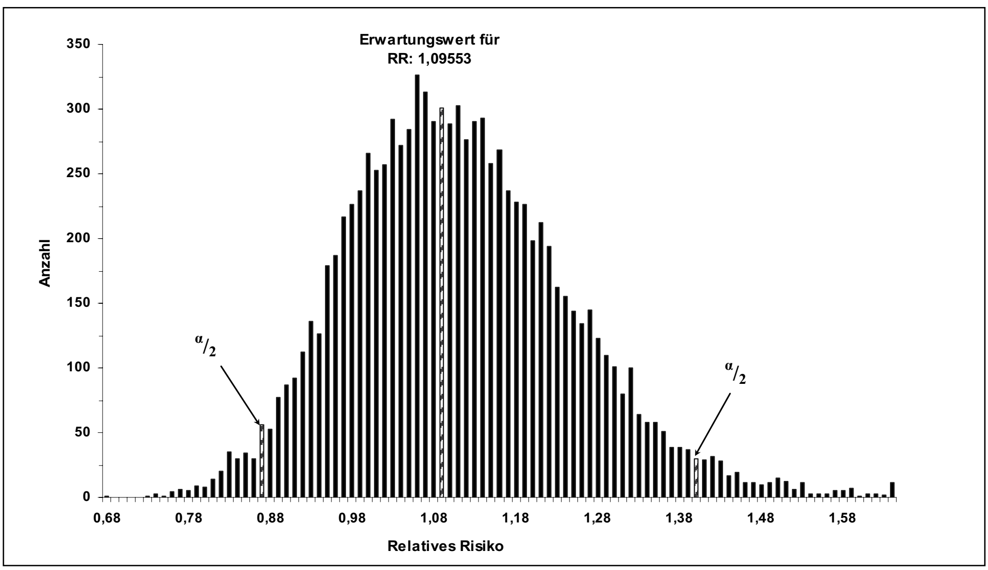
Quelle: Eigene Berechnung mit den Daten des FDZ-RV – SUFRTBN01VWITD und FDZ-RV – SUFRTWF02VWITD.

mulation für die Rentenbestände der gleichen Personengruppe ist in **Abbildung 2** zu sehen. Auch hier ergibt sich eine Anordnung der simulierten Ausprägungen für die Rentenbestände um den tatsächlichen Beobachtungswert herum. Für die Männer mit nicht-deutscher Staatsangehörigkeit (im Alter 75) wird die Simulation der Rentenwegfälle und Rentenbestände mit k = 10.000 ebenfalls durchgeführt (ohne Abb.).

Jede simulierte Stichprobenrealisation wird mit ihrem jeweiligen feststehenden Gewichtungsfaktor multipliziert. Mit den gewichteten Fallzahlen erfolgt die Berechnung der Maßzahl RR. Durch die Simulation des RR auf der Basis der gewichteten Stichprobenrealisationen ist zum einen die Bedingung der zufälligen Stichprobenziehung um den ursprünglichen Erwartungswert der Stichprobe gegeben und zum anderen wird die Repräsentativität der Ergebnisse für die Grundgesamtheit gewahrt.

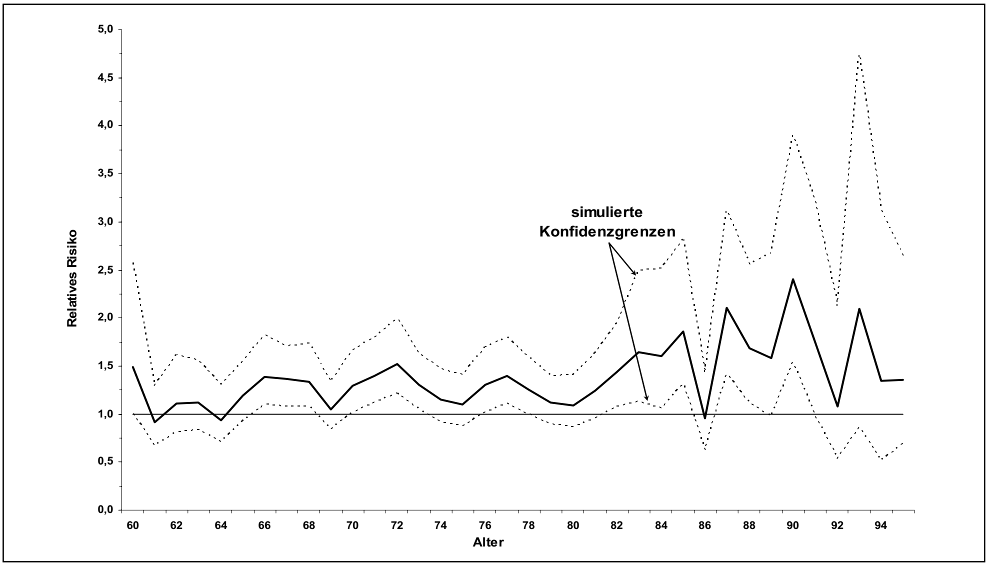
In **Abbildung 3** ist zu erkennen, dass auch die Streuung um den Erwartungswert von RR (von 1,09536) annähernd normal verteilt ist. Bei einem unterstelltem Signifikanzniveau von 1 – α = 0,05 kann angegeben werden, wie groß die Abweichungen vom Erwartungswert sein dürfen, um das RR der Stichprobenrealisationen mit einem Sicherheitsniveau von 95 Prozent annehmen zu können. Im konkreten Fall erstreckt sich das Konfidenzintervall von 0,87 bis 1,40. Ein Unsicherheitsbereich der tatsächlichen Stichprobenrealisation ist damit angegeben, für die Interpretation des Relativen (Sterbe)Risikos ergibt sich jedoch eine weitere Restriktion. Für unser Beispiel schließen die Konfidenzgrenzen des RR den Wert Eins ein; was bedeutet, dass ein erhöhtes Sterberisiko der deutschen gegenüber den nicht-deutschen Männern von RR = 1,09536

Abbildung 3: Die Häufigkeitsverteilung für das Relative (Sterbe)Risiko der Männer im Alter 75 mit deutscher Staatsangehörigkeit gegenüber den Männern im Alter 75 mit nicht-deutscher Staatsangehörigkeit für 10.000 Simulationen



Quelle: Eigene Berechnung mit den Daten des FDZ-RV – SUFRTBN01VWITD und FDZ-RV – SUFRTWF02VWITD.

Abbildung 4: Das Relative (Sterbe)Risiko der Männer mit deutscher Staatsangehörigkeit gegenüber den Männern mit nicht-deutscher Staatsangehörigkeit für das jeweilige Alter x = 60 bis 95 sowie die simulierten Konfidenzgrenzen für $1 - \alpha = 0{,}95$



Quelle: Eigene Berechnung mit den Daten des FDZ-RV – SUFRTBN01VWITD und FDZ-RV – SUFRTWF02VWITD.

nicht als statistisch signifikant angesehen werden kann. Die korrekte Interpretation muss lauten: Mit einer Irrtumswahrscheinlichkeit von $\alpha = 0{,}05$ ist für die Männer im Alter 75 im Jahr 2002 kein relativer Mortalitätsunterschied zwischen deutscher und nicht-deutscher Staatsangehörigkeit festzustellen.

In **Abbildung 4** ist das Relative Risiko jeweils für die Altersstufen 60 bis 95 dargestellt. Dabei zeigt sich, dass das Sterberisiko bei den deutschen Männern in nahezu allen Altersstufen über dem der nicht-deutschen Männer liegt. Ein möglicher Erklärungsansatz hierfür könnte die selektive Sterblichkeit der betrachteten Kohorten während des 2. Weltkrieges sein. Die möglichen Ursachen sollen an dieser Stelle aber nicht weiter interessieren. Die Schlussfolgerung, dass die Männer mit deutscher Staatsangehörigkeit im Alter zwischen 60 und 95 eine höhere Sterblichkeit aufweisen als die Männer mit nicht-deutscher Staatsangehörigkeit ist deshalb aber noch nicht zulässig. Dazu dürften die Grenzen des jeweiligen Konfidenzintervalls den Wert Eins nicht mit einschließen, was aber auf ca. 70% der betrachteten Altersstufen zwischen 60 und 95 Jahren zutrifft. Trotz der Gewichtungsfaktoren sind die Fallzahlen der tatsächlichen Stichprobenrealisationen für viele Altersstufen zu klein. Eine Möglichkeit die Fallzahlen zu erhöhen wäre die Betrachtung von zusammengefassten Altersklassen. Eine andere Vorgehensweise statistisch gesicherte Ergebnisse zu erhalten, kann über die Erhöhung der Stichprobenfallzahlen erreicht werden (vgl. *Himmelreicher, Gaudecker* und *Scholz* 2006).

5 Zusammenfassung

Wenn die Sterbefälle und die Risikopopulation, wie bei den prozessproduzierten Daten der Deutschen Rentenversicherung, aus zwei voneinander unabhängigen Querschnittsstichproben stammen, ist eine Berechnung von altersspezifischen Sterbewahrscheinlichkeiten und weiteren darauf basierenden Parametern nicht möglich. Eine Alternative ist die Anwendung relationaler Mortalitätsmaße wie z. B. das Relative Risiko. Um den Stichprobencharakter der Datensätze bei der Interpretation der Ergebnisse zu berücksichtigen, müssen Konfidenzintervalle angegeben werden.

Eine Möglichkeit der Konfidenzintervallschätzung für das Mortalitätsmaß Relatives Risiko wurde von uns mittels einer Monte-Carlo-Simulation aufgezeigt. Dabei werden die jeweiligen Stichprobenrealisationen der Sterbefälle und der Risikopopulation der betrachteten Merkmalsgruppen mit einer vorgegebenen Wahrscheinlichkeitsverteilung beliebig oft simuliert. Mit Hilfe der simulierten Häufigkeitsverteilung um den Erwartungswert können dann Konfidenzgrenzen für das Relative Risiko bestimmt werden.

Würden bei Mortalitätsanalysen mit den Datensätzen der Deutschen Rentenversicherung Vollerhebungen für die jeweiligen Rentenwegfälle und Rentenbestände durchgeführt (vgl. *Rehfeld* und *Scheitl* 1986, 1991), würde die Limitation der zugrunde liegenden Daten wesentlich geringer ausfallen. Eine weitere Alternative, hinsichtlich differenzierterer Mortalitätsanalysen, wäre die Verwendung von Längsschnittdaten, wie dies z. B. in der Versichertenkontenstichprobe der Fall ist (allerdings nur bis Alter 65), in der die Rentenwegfälle eindeutig dem vorherigen Rentenbestand zugeordnet werden können.

Ein weiterer denkbarer Schritt wäre, zukünftig die sehr validen Daten der Deutschen Rentenversicherung mit Daten wie z. B. der Bundesagentur für Arbeit oder der Krankenversicherungen unter der Maßgabe eines hinreichenden Datenschutzes zu verknüpfen. Dass dies methodisch ohne Weiteres möglich ist, beweisen die Erfahrungen in den USA. Mit Hilfe dieser Verknüpfung(en) wären umfangreiche Längsschnittanalysen – nicht nur für Mortalitätsuntersuchungen – möglich.

Literaturverzeichnis

Chiang, L. C. (1984): The Life Table and its Applications. R. E. Krieger Publishing: Malabar, Florida.

Coombs, L. C. (1941): Economic Differentials in Causes of Death. In: Medical Care 1, 246-255.

Deutsche Rentenversicherung Bund (2005): Codeplan Demographiedatensatz: Rentenwegfall und Rentenbestand 1993-2003, FDZ-RV – SUFRTBNXXVWITD und FDZ-RV – SUFRTWFXXVWITD.

Dinkel, R. H. (2007): Demographie Band II - Mortalität. Universität Rostock. Im Erscheinen.

Esenwein-Rothe, I. (1982): Einführung in die Demographie. Franz Steiner: Wiesbaden.

Flaskämper, P. (1962): Bevölkerungsstatistik. Richard Meiner: Hamburg.

Gaudecker, H.-M. (2006): Differentielle Sterblichkeit in der GRV: Problemaufriss und erste Berechnungen. In: DRV-Schriften Band 55/2005, 242-252.

Himmelreicher, R. K., Gaudecker, H.-M. & Scholz, R. D. (2006): Nutzungsmöglichkeiten von Daten der Deutschen Rentenversicherung über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). In: MPIDR Working Paper WP 2006-018 July 2006.

Horm, J. (1996): The National Health Interview Survey and the National Death Index. National Center for Health Statistics: Maryland.

Kitagawa, E. M. & Hauser, P. M. (1964): Trends in differential fertility and mortality in a metropolis - Chicago. In: Burgess, E. & Bogue, D. (Hrsg.), Contributions to urban sociology. University of Chicago Press, Chicago, 59-85.

Kitagawa, E. M. & Hauser, P. M. (1973): Differential Mortality in the United States – A Study in Socioeconomic Epidemiology. Harvard University Press: Cambridge, Massachusetts.

Pamuk, R. E. (1985): Social Class Inequality in Mortality from 1921 to 1972 in England and Wales. In: Population Studies, Vol. 39, 17-31.

Pappas, G., Queen, S., Hadden, W. C. & Fisher, G. (1993): The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986. In: New England Journal of Medicine, Vol. 329, 103-109.

Preston, S. H. & Elo, I. T. (1994): Are Educational Differentials in Mortality Increasing in the United States? In: Working Paper Series No. 95-01, Population Studies Center, University of Pennsylvania.

Rehfeld, U. & Scheitl, O. (1986): Die Rentnersterblichkeit 1985: Aktuelle Ergebnisse für Altersrentner und einige spezielle Fallgruppen. Deutsche Rentenversicherung, 11-12, S.729-737.

Rehfeld, U. & Scheitl, O. (1991): Sterblichkeit und fernere Lebenserwartung von Rentnern der gesetzlichen Rentenversicherung – aktuelle Ergebnisse für 1986/1988 und Bilanz zum bisherigen Untersuchungsstand. Deutsche Rentenversicherung, 4-5, 289-320.

Rogers, R. G., Hummer R. A. & Nam, C. B. (2000): Living and Dying in the USA – Behavioural, Health, and Social Differentials of Adult Mortality. Academic Press: San Diego, California.

Rogot, E., Sorlie, P. D., Johnson, N. J. & Schmitt, C. (1992): A mortality study of 1.3 million persons by demographic, social, and economic factors: 1979-1985 Follow up. NIH: Bethesda, Maryland.

Schalick, L., Hadden, W. C., Pamuk, E., Pappas, G. & Navaro, V. (2000): The Widening Gap in Death Rates among Income Groups in the United States from 1967 to 1986. In: International Journal of Health Services, Vol. 30, Nr. 1, 13-26.

Scholz, R. D. (2006): Differentielle Sterblichkeitsanalyse mit den Daten der Deutschen Rentenversicherung. In: DRV-Schriften Band 55/2005, 253-266.

Woolf, B. D. (1955): On Estimating the Relation Between Blood Group and Disease. In: Annals of Human Genetics, Vol. 10, 251-253.

Diplom-Demograph Thomas Salzmann studierte Volkswirtschaftslehre an der Humboldt Universität zu Berlin und Demographie an der Universität Rostock. Er ist zur Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Demographie und Ökonometrie der Universität Rostock beschäftigt. Seine Arbeitsgebiete erstrecken sich auf die Mortalitäts- und Migrationsforschung. Zudem arbeitet er auf dem Gebiet der methodischen Verfahren für Bevölkerungsprognosen. *thom.salzmann@uni-rostock.de*

Diplom-Demograph Martin Kohls studierte Demographie an der Universität Rostock. Zurzeit ist er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Demographie und Ökonometrie tätig. Neben umfangreichen Lehraufgaben widmet er sich vor allem demografisch relevanten Fragestellungen der Mortalitätsforschung. Dazu zählen unter anderem differentielle Sterblichkeitsanalysen mit verschiedenen Querschnitts- und Längsschnittdatensätzen sowie die Erstellung und Verbesserung von Pflegebedarfsprognosen für Deutschland und dem Bundesland Mecklenburg-Vorpommern unter Berücksichtigung eines Mortalitätsfortschritts. *martin.kohls@uni-rostock.de*

Die Forschungsschwerpunkte des **Lehrstuhls Demographie und Ökonometrie der Universität Rostock** unter Leitung von *Professor Dr. R. H. Dinkel* liegen im Bereich der Mortalitäts- und Migrationsforschung. Dabei wird das Hauptaugenmerk auf die Analyse und Generierung geeigneter quantitativ orientierter Instrumente zur Mortalitäts- und Migrationsmessung gelegt. Ziel der Arbeit ist es, die Ursachen und Folgen demografischer Entwicklungen unter Berücksichtigung sämtlicher demografischer Parameter mit geeigneten quantitativen Messinstrumenten zu erforschen und mögliche Zukunftsszenarien abzuschätzen. Ein weiteres Forschungsgebiet stellt die Erstellung neuer Verfahren für Bevölkerungsprognosen dar. *www.wiwi.uni-rostock.de/~demograf.*