

Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR WORKING PAPER WP 2006-018
JULY 2006

**Nutzungsmöglichkeiten von Daten
der gesetzlichen Rentenversicherung
über das Forschungsdatenzentrum
der Rentenversicherung (FDZ-RV)**

Ralf K. Himmelreicher
Hans-Martin v. Gaudecker
Rembrandt D. Scholz (scholz@demogr.mpg.de)

This working paper has been approved for release by: Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de),
Head of the Laboratory for Demographic Data.

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review.
Views or opinions expressed in working papers are attributable to the authors and do not necessarily
reflect those of the Institute.

Nutzungsmöglichkeiten von Daten der gesetzlichen Rentenversicherung über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV)

Eine Beschreibung von Nutzungsmöglichkeiten an Hand des Projektes

'Differentielle Sterblichkeit'

Dr. Ralf K. Himmelreicher, Forschungsdatenzentrum der Rentenversicherung, Berlin

Hans-Martin v. Gaudecker, Mannheimer Forschungsinstitut Ökonomie und demographischer Wandel (MEA)

Dr. Rembrandt D. Scholz, Max-Planck-Institut für demografische Forschung Rostock

Dieser Beitrag beschreibt an Hand des Projektes ‚Differentielle Sterblichkeit‘, in welcher Form die Daten der Rentenversicherung über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) genutzt werden können: Zunächst wurde der Scientific Use File (SUF) Demografie genutzt. An Hand dieses kommentierten Datensatzes aus dem Datenangebot des FDZ-RV ist es möglich, die Charakteristika der Datensätze aus dem Bereich Rentenstatistik kennen zu lernen. Sind insbesondere höhere Fallzahlen und spezielle Differenzierungen von Merkmalen wichtig, dann können Analysen auf Gastwissenschaftler-Arbeitsplätzen mit speziellen Datensätzen eine Alternative darstellen. Als dritte Form der Datennutzung wurde hier das kontrollierte Fernrechnen, bei dem Wissenschaftler keinen Kontakt mit den Daten haben, praktiziert.

Die Befunde, die das Projekt ‚Differentielle Sterblichkeit‘ ermitteln konnte, verweisen darauf, dass auf Sterbetafeln basierende Sterblichkeitsanalysen sehr hohe Fallzahlen benötigen, die von Bevölkerungsumfragen nicht erreicht werden können. Im Ergebnis zeigen die Berechnungen für die fernere Lebenserwartung im Alter von 65 Jahren für in Deutschland lebende Männer, dass die mittlere Lebenserwartung bei 15,7 Jahren und somit auf einem Niveau mit den Berechnungen des Statistischen Bundesamtes liegt. Bezieht man das Lebensarbeitseinkommen in die Analysen ein, dann zeigt sich eine positive statistisch signifikante Assoziation zwischen Lebenseinkommen und Lebenserwartung. Die Lebenserwartungen variieren allerdings bei der überwiegenden Mehrheit der Männer lediglich plus minus eineinhalb Jahre um die durchschnittliche Lebenserwartung.

1. Einleitung

Das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) wurde im Januar 2004 gegründet. Es wird vom Bundesministerium für Bildung und Forschung (BMBF) und dem Forschungsnetzwerk Alterssicherung (FNA) der Deutschen Rentenversicherung Bund im Rahmen einer Aufbauphase bis August 2006 gefördert.

Die wichtigsten Aufgaben des FDZ-RV lassen sich in vier Kernzielen bündeln: Das Forschungsdatenzentrum möchte durch die Bereitstellung von Mikrodaten im Sinne eines übergeordneten Zieles zur Versachlichung der sozialpolitischen Diskussion beitragen. Hierzu gibt es zweitens methodische Hinweise und Anmerkungen, um Analysen mit den prozessproduzierten Daten der Rentenversicherung zu vereinfachen. Drittens soll durch Information und Dokumentation unter Anderem über das Internetportal <http://www.fdz-rv.de> ein Überblick über das Analysepotenzial der über das FDZ-RV beziehbaren Daten gegeben werden. Viertens möchte das Forschungsdatenzentrum der Rentenversicherung einen Beitrag dazu leisten, Diskussionen unter und zwischen Datennutzern sowie Beschäftigten der Deutschen Rentenversicherung zu intensivieren, um durch diesen Diskurs wissenschaftliche Erkenntnisse – zum Beispiel über die Ausgestaltung der Alterssicherung - auf ein solides empirisches Fundament zu stellen (vgl. Rische 2006).

Die in diesem Beitrag dargestellten Nutzungsmöglichkeiten von Daten der Rentenversicherung über das Forschungsdatenzentrum beleuchten zum einen die genannten vier Kernziele des FDZ-RV. Zum anderen werden die drei verschiedenen Formen der Datennutzung, die der nicht kommerziellen Wissenschaft für Forschungsvorhaben im Bereich der sozialen Sicherung zur Verfügung stehen, dokumentiert.¹ Dabei werden die Nutzungsmöglichkeiten an Hand eines konkreten Projektes beschrieben; diese Vorgehensweise ist geeignet zu zeigen, in welchem Zusammenhang die Befunde mit der jeweils gewählten Form der Datennutzung stehen können. Als konkretes Forschungsvorhaben wurde das Projekt „Differentielle Sterblichkeit“ gewählt, weil im Rahmen dieser Forschungs Kooperation erstmals alle drei Nutzungsformen der Daten des FDZ-RV praktiziert wurden. Im Hinblick auf die Nutzungsformen ist anzumerken, dass für das Gros der wissenschaftlichen Fragestellungen Scientific Use Files (SUFs) in Form von Basisfiles, vorhandenen oder gegebenenfalls noch zu erstellenden Themenfiles eine hinreichende Operationalisierung der Fragestellung ermöglichen sollten.

Im zweiten Abschnitt wird das Projekt „Differentielle Sterblichkeit“ vorgestellt. Der dritte Abschnitt ist dreigeteilt: er verfolgt in chronologischer Reihenfolge wesentliche Schritte der Analysen zur differentiellen Sterblichkeit und stellt diese sowohl in inhaltlicher als auch in organisatorischer Perspektive heraus. Ein besonderes Augenmerk liegt dabei auf den drei verschiedenen Formen des Datenzugangs (Scientific Use Files, Gastwissenschaftler-Arbeitsplatz und kontrollierter Datenfernverarbeitung) sowie den Erfahrungen, die mit ihnen gesammelt wurden. Es wird deutlich, dass eine stufenweise Nutzung der Daten erfolgen

¹ In diesem Beitrag werden jene Nutzungsmöglichkeiten, die für Studierende und die Fachöffentlichkeit anhand absolut anonymisierter Public Use Files angeboten werden, nicht thematisiert.

sollte. Zu Beginn einer wissenschaftlichen Datenauswertung ist in jedem Fall die Nutzung eines Scientific Use Files (SUF) zu empfehlen, um sich mit den Besonderheiten der prozessproduzierten Daten der Rentenversicherung vertraut machen zu können. Im vierten Abschnitt werden die mit den jeweiligen Daten berechneten empirischen Befunde und Erfahrungen zusammengefasst sowie ein kurzer Ausblick auf das Datenangebot im FDZ-RV der Zukunft gegeben.

2. Zum Projekt ‚Differentielle Sterblichkeit‘

Das Projekt ‚Differentielle Sterblichkeit‘ ist am Mannheimer Forschungsinstitut Ökonomie und demographischer Wandel (MEA) angesiedelt. Es wird unter der Leitung von Prof. Axel Börsch-Supan, Ph.D. von Hans-Martin von Gaudecker bearbeitet. Institutionelle Verbindungen zwischen dem MEA und dem FDZ-RV bestehen seit der Gründung des Forschungsdatenzentrums. Sowohl Mitarbeiter des MEA als auch des MPI für Demografie waren beim ersten (Juni 2004) und zweiten (Juni 2005) Workshop des FDZ-RV sowohl als Teilnehmer als auch als Referenten vertreten.² Insofern kann inzwischen auf eine mehrjährige und produktive Zusammenarbeit zwischen MEA, MPI für Demografie und FDZ-RV zurückgeblendet werden. Im Rahmen dieser Zusammenarbeit konnten im Bereich von Datenbedarf und Zugangswegen Lösungen gefunden werden, die datenschutzrechtlichen Bestimmungen genügen und den zu Projektbeginn formulierten Datenbedarf weitgehend zur Verfügung stellen (vgl. Gaudecker 2004: 128 und Scholz 2004: 145). Allerdings sind solche Lösungen für das FDZ-RV ressourcenintensiv und benötigen einen längeren zeitlichen Vorlauf.

Die Grundfragestellung des Projektes ‚Differentielle Sterblichkeit‘ lautet: Gibt es zwischen verschiedenen Einkommensgruppen Unterschiede bei der ferneren Lebenserwartung im Alter ab 65 Jahren? Und falls ja, wie hoch sind diese? Die Mikrodaten der Deutschen Rentenversicherung eignen sich für solche Untersuchungen besonders gut, da sie eine Messgröße für das Lebensarbeitseinkommen von rentenversicherten Personen beinhalten. Aufgrund von erstaunlich hohen jährlichen Schwankungen des Einkommens in Umfragedaten sind diese für Untersuchungen zum Lebensarbeitseinkommen nicht geeignet (vgl. zum Beispiel Haider und Solon 2006). Ferner ermöglichen die prozessproduzierten Daten Analysen mit sehr hohen Fallzahlen. Wie sich im weiteren Verlauf des Artikels zeigen wird, sind diese für die statistische Aussagekraft der Ergebnisse von besonderer Bedeutung.

² Siehe hierzu Gaudecker (2004, 2006), Berkel und Börsch-Supan (2006) sowie Scholz (2004, 2006).

Im Hinblick auf die Interpretation der hier vorgestellten Befunde ist Folgendes zu beachten: Die Analysen dokumentieren lediglich die statistische Assoziation der Größen „Lebensarbeitseinkommen“ und „Lebenserwartung“. Sie lassen keinen Rückschluss darauf zu, ob ein höheres Einkommen kausal zu einer höheren Lebenserwartung führt. Dafür gibt es zwei wesentliche Gründe. Zum einen sind Faktoren zu benennen, die sowohl das Einkommen als auch die Sterblichkeit beeinflussen. Ein Beispiel hierfür ist die Bildung, für die in internationalen Studien ein kausaler Effekt auf die Sterblichkeit nachgewiesen wurde.³ Ein höheres Bildungsniveau führt demnach sowohl zu geringerer Sterblichkeit als auch zu einem höheren Einkommen. Allein aufgrund dieser Effekte kann in Studien zu Einkommen und Sterblichkeit eine positive Assoziation der beiden Größen auftreten. Der zweite Grund liegt darin, dass Menschen, die von Geburt an oder zu Beginn ihrer Erwerbsphase gesundheitlich beeinträchtigt sind, tendenziell ein geringeres Einkommen erzielen werden als solche mit einer robusten Konstitution. Aufgrund ihres schlechteren Gesundheitszustands weisen sie jedoch auch eine geringere Lebenserwartung auf. Wiederum muss dies nichts mit dem Einkommen zu tun haben. Anders ausgedrückt lässt sich die Variable „Renteneinkommen“ als ein Konglomerat aus verschiedenen Variablen verstehen, das sich unter anderem aus schulischer Bildung, beruflicher Ausbildung, Gesundheitszustand und aus dem Einkommen aus rentenversicherungspflichtiger Tätigkeit zusammensetzt. Im Rahmen der hier vorgelegten Analysen wird nicht untersucht, ob ein Einfluss einzelner Faktoren auf die Sterblichkeit vorhanden ist, bzw. wie stark er ist. Untersucht wird die Assoziation von „Lebensarbeitseinkommen“ und „Lebenserwartung“.

3. Nutzung von FDZ-RV-Daten durch das Projekt ‚Differentielle Sterblichkeit‘

Im Rahmen des Projektes ‚Differentielle Sterblichkeit‘ sind die Daten des FDZ-RV aus drei Gründen von besonderer Bedeutung. Dies ist einerseits auf das in den Daten enthaltene Merkmal ‚Entgeltpunkte‘ bzw. die ‚Summe persönlicher Entgeltpunkte‘ zurückzuführen. Das Merkmal ‚Summe persönlicher Entgeltpunkte‘ wird hier analog zum Lebensarbeitseinkommen aus versicherter Beschäftigung (bis zur Beitragsbemessungsgrenze) benutzt. Wegen der Beitragsbemessungsgrenze unterliegen die beobachteten Entgeltpunkte und Rentenzahlbeträge einer Rechtszensierung, weil über dieser liegende Arbeitseinkommen die Ansprüche an die Rentenversicherung nicht erhöhen. Zudem besteht eine Linkszensierung des Lebensarbeitseinkommens, weil die Entgeltpunkte durch umverteilende Maßnahmen in der Rentenversicherung zum Teil erhöht wurden, wie z.B. im Fall einer Rentenerhöhung wegen ‚Rente nach Mindesteinkommen‘, die bis 1992 bezogen werden konnte (vgl. Himmelreicher 2006: 60f.). Beide Zensierungen werden in der

³ Lleras-Muney, Adriana: The Relationship Between Education and Adult Mortality in the United States, Review of Economic Studies, 72, S. 189-221, Januar 2005.

vorliegenden Analyse nicht beachtet. Dennoch stellt die ‚Summe persönlicher Entgeltpunkte‘ eine valide Messgröße für das Lebensarbeitseinkommen dar, welches seinerseits eine theoretisch relevante Größe für den Zusammenhang zwischen Einkommen und Mortalität ist. Während das jährliche Arbeitseinkommen in Abhängigkeit von der individuellen Position im Lebenszyklus, Arbeitslosigkeit usw. oftmals starken Schwankungen unterliegt, insbesondere in Umfragedaten (vgl. Himmelreicher 2001: 77ff.), bildet die Gesamtzahl der in der Erwerbsbiografie erarbeiteten Entgeltpunkte ein adäquates Maß für die tatsächliche Einkommensposition von überwiegend abhängig Beschäftigten (vgl. Unger 2006). Die Arbeitseinkünfte von Personen, die nur einen Teil ihres Erwerbslebens abhängig beschäftigt und ansonsten selbständig oder verbeamtet waren, einem freien Beruf nachgegangen sind oder länger im Ausland gearbeitet haben, werden von den Daten nur ungenügend repräsentiert. Trotz dieser Einschränkung werden im weiteren Verlauf dieser Arbeit die Begriffe (Lebensarbeits-) Einkommen und Summe der Entgeltpunkte synonym verwendet.

Der zweite wichtige Grund zur Nutzung von FDZ-RV-Daten ist auf die hohen Fallzahlen zurückzuführen. Für die Berechnung der Sterblichkeit werden die Daten des Rentenbestandes und die des Rentenwegfalls benutzt. Beide Statistiken der Deutschen Rentenversicherung sind Vollerhebungen, die je nach Nutzungsart im FDZ-RV unterschiedlich aufbereitet werden.

Drittens sind die FDZ-RV-Daten prozessproduziert. D.h., die in den Datensätzen enthaltenen versicherungsrechtlich relevanten Merkmale, wie z.B. die Summe der Entgeltpunkte, sind von sehr hoher Qualität. Im Unterschied zu Befragungsdaten treten bei prozessproduzierten Daten keine Erinnerungsfehler, keine Antwortverweigerungen und keine Panelmortalität auf (vgl. Wübbecke 2006: 160).

Im Hinblick auf den untersuchten Personenkreis ist darauf hinzuweisen, dass die Analyse auf Männer begrenzt wurde. Dies liegt daran, dass die Erwerbsquote von Frauen in den betrachteten Kohorten vergleichsweise gering ist und damit die Annahme, dass die Zahl der individuellen Entgeltpunkte ein hinreichendes Maß für das dem Haushalt zur Verfügung stehende Lebensarbeitseinkommen ist, nicht plausibel erscheint. Im Hinblick auf eine adäquate Erfassung des Haushaltskontextes wäre eine Messgröße, die das gesamte Haushaltseinkommen erfassen kann, erforderlich. Die hier zu Grunde liegenden Mikrodaten beziehen sich auf Renten und haben weder individuelle Verknüpfungsmöglichkeiten noch solche im Rahmen von Ehepartnern. Die Zahl der Entgeltpunkte von Männern dürfte jedoch wegen ihren überwiegend langjährigen Erwerbsbiografien eine hinreichende Approximation darstellen (vgl. Himmelreicher und Frommert 2006). Ferner werden die Analysen auf jene Männer in den alten und neuen Bundesländern begrenzt, die eigene Versichertenrenten (Renten wegen Alters und wegen Erwerbsminderung) beziehen; ausgeschlossen werden dadurch Männer mit Hinterbliebenenrenten.

Damit ist bereits das Hauptmerkmal der Analyse, die Summe der persönlichen Entgeltpunkte, genannt. Aufgrund des Systems der Rentenberechnung bieten diese eine Messgröße für das Einkommen aus rentenversicherungspflichtiger Arbeit im gesamten Lebenszyklus. Es werden nur Personen mit mindestens 20 persönlichen Entgeltpunkten berücksichtigt, da bei niedrigeren Werten davon auszugehen ist, dass der überwiegende Teil des Lebensarbeitseinkommens nicht aus rentenversicherungspflichtiger Beschäftigung stammte. Ausgehend von diesen Prämissen wurde wie folgt vorgegangen: Das Lebensarbeitseinkommen wird in 11 Klassen à 5 Entgeltpunkte eingeteilt, für jede Klasse wird separat die fernere Lebenserwartung im Alter 65 berechnet (im weiteren Verlauf dieser Arbeit wird der Begriff Lebenserwartung synonym für fernere Lebenserwartung im Alter 65 verwendet).

Im Rahmen des Projektes ‚Differentielle Sterblichkeit‘ wurde in der Zeit zwischen März und Dezember 2005 zunächst der SUF Themendatensatz „Demographie 2003“ beantragt und genutzt (siehe 3.1). Im nächsten Schritt wurde ein spezieller Datensatz für die Nutzung an einem Gastwissenschaftler-Arbeitsplatz in den Räumen des Forschungsdatenzentrums in Berlin erstellt (siehe 3.2). Basierend auf diesen Erfahrungen konnte schließlich das Verfahren zur kontrollierten Datenfernverarbeitung zur Anwendung kommen (siehe 3.3).

3.1 Nutzung des SUF Demographie 2003

Der Themendatensatz SUF Demographie war neben dem Basisfile Versichertenrentenzugang 2003 einer der ersten vom FDZ zur Verfügung gestellten Scientific Use Files (SUFs). Der Themendatensatz Demographie besteht aus mehreren SUFs, die vom FDZ-RV aus den Statistiken des Rentenbestands und Rentenwegfalls mehrerer Berichtsjahre erstellt wurden. Eine Dokumentation zu den Scientific Use Files Rentenzugang und Rentenbestand („Themenfile Demografie“) kann über das Internetportal unter <http://www.fdz-rv.de> eingesehen werden. Detaillierte Informationen zum Themenfile SUF Demografie sind in Luckert (2006: 30-33) dokumentiert. Insofern wird an dieser Stelle lediglich auf einige wichtige Eigenschaften und Merkmale der Datensätze eingegangen.

Die Datensätze Rentenbestand 1993-2003 enthalten eine Einprozentstichprobe der Personen, die im Dezember des jeweiligen Berichtsjahres eine Rente bezogen haben (Stichprobenumfang pro Jahr ca. 220.000 Fälle). Die für die Zwecke dieser Arbeit wichtigsten enthaltenen Charakteristika sind Geburtsjahr und Monat sowie das Geschlecht des Versicherten, Staatsangehörigkeit und Wohnsitz, die Gesamtzahl der Entgeltpunkte und die Art der Rente (Alters- und Erwerbsminderungsrenten sowie Witwen- und Witwerrenten).

Korrespondierend dazu enthalten die Datensätze Rentenwegfall 1993-2003 eine Zehnprozentstichprobe derjenigen Renten, die im Berichtsjahr – in den Wegfallsmonaten

Januar bis November - aufgrund des Todes des Berechtigten weggefallen sind (Stichprobenumfang ca. 95.000 bis 125.000 Fälle pro Jahr).

Im Rahmen des Projektes Differentielle Sterblichkeit wurden ausschließlich die SUFs Rentenbestand (FDZ-RV - SUFRTBN03VWITD) und Rentenwegfall (FDZ-RV - SUFRTWF03VWITD) des Jahres 2003 benutzt. In allen folgenden Analysen wurden die beiden Datensätze in einem Datensatz zusammengefasst. Dieser neue Datensatz wird FDZ-RV – SUFRTBNWF03DEM genannt; er hat einen Stichprobenumfang von 81.610 Fällen (Männer mit Versichertenrenten).

Zusätzlich zu den oben genannten Vorzügen der Entgeltpunkte als Messgröße des Lebenseinkommens haben diese Datensätze den Vorteil von sehr großen Fallzahlen sowie die hohe Verlässlichkeit der administrativen Daten gegenüber Eigenangaben zum Einkommen in Umfragen. Im vorliegenden Beitrag, der stärker die Nutzungsmöglichkeiten im Rahmen des FDZ-RV als Inhalte fokussiert, wird stärker auf die den Analysen zugrunde liegenden Daten als auf die Befunde eingegangen.⁴

Die Daten des Themendatensatzes SUF Demographie wurden nach Antrag und Vertrag zur Datennutzung auf CD-ROM bereitgestellt und inklusive der Dokumentation versandt. Zusammen mit den roten Statistikbänden „Rentenwegfall“ und „Rentenbestand“⁵ konnten die Daten innerhalb kurzer Zeit für erste Auswertungen aufbereitet werden. Die Ergebnisse auf Basis des speziell aufbereiteten SUF Demografie 2003 (SUFRTBNWF03DEM) finden sich in Abbildung 1. Die fernere Lebenserwartung der Männer im Alter von 65 Jahren wurde berechnet nach dem klassischen Verfahren, das beispielsweise in Chiang (1984) dargestellt ist. In der Sterblichkeitsberechnung wird zunächst auf Monatsebene für jede Alters- und Einkommensklasse die Mortalitätsrate als Quotient der Gesamtzahl der Gestorbenen und der Gesamtzahl der während der Periode gelebten Jahre berechnet. Das Problem unterschiedlicher Stichprobenumfänge (10% im Wegfall vs. 1% im Bestand) wurde durch entsprechende Gewichtung gelöst, aufgrund der fehlenden Dezember-Wegfälle wurden alle Wegfälle zusätzlich mit 12/11 gewichtet. Die analytischen Konfidenzintervalle basieren ebenfalls auf den in Chiang (1984) aufgeführten Formeln.

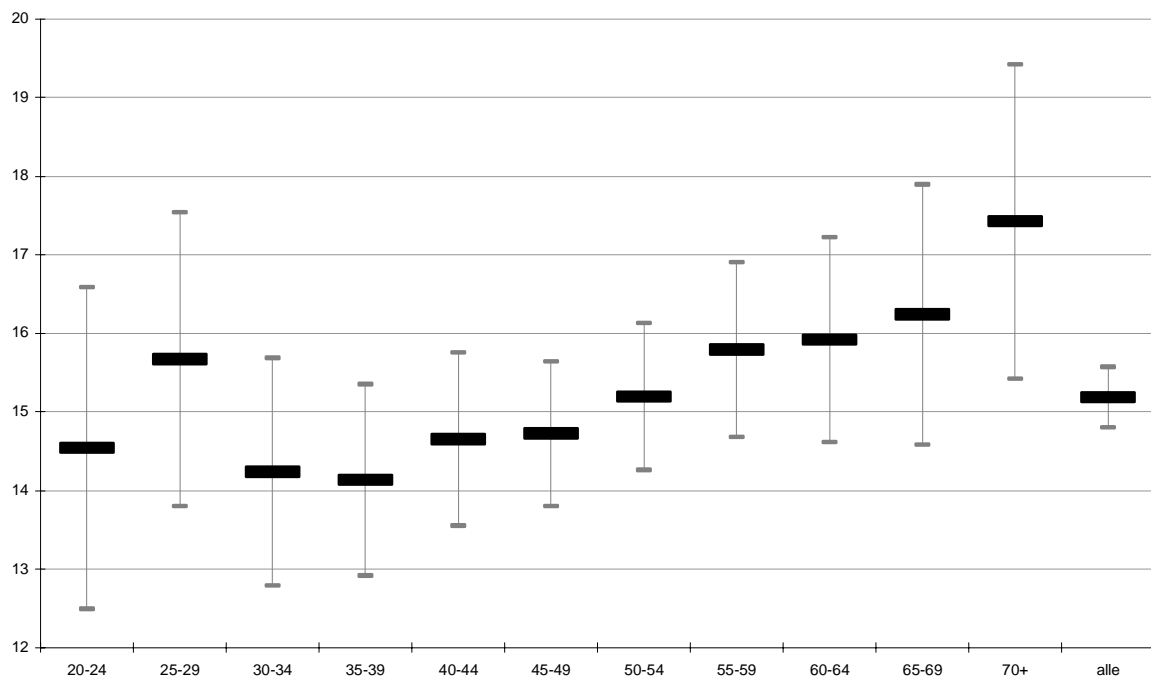
Es wird deutlich, dass die Lebenserwartung von Männern mit den Entgeltpunkten ansteigt, wenn sie 35 und mehr Entgeltpunkte haben. Allerdings sind die Unterschiede nicht statistisch signifikant – man kann nicht mit hinreichender Sicherheit sagen, ob die Unterschiede tatsächlich vorhanden sind, oder ob sie lediglich auf Zufallsschwankungen in „kleinen“ Datensätzen beruhen. Auf den ersten Blick erstaunlich ist die ebenfalls zu beobachtende höhere durchschnittliche Lebenserwartung in den unteren Einkommensklassen. Dies liegt

⁴ Zu den Befunden siehe Gaudecker und Scholz (2006).

⁵ vgl. VDR (2004a) Statistik Band 148 und 149 (2004b), siehe auch <http://www.fdz-rv.de>, Link: Datenangebot.

vermutlich daran, dass ein Teil der dort enthaltenen Rentner höhere Einkommen außerhalb der Rentenversicherung erzielt hat. Zu denken ist insbesondere an Beamte, Freiberufler und Selbständige, die im Laufe ihres Erwerbslebens die allgemeine Wartezeit von fünf Jahren erfüllt haben, jedoch den überwiegenden Teil ihrer Erwerbstätigkeit nicht sozialversicherungspflichtig beschäftigt waren. Häufig erzielen diese Personengruppen höhere Einkommen und sind somit im Hinblick auf ihre Lebenslage eher mit denjenigen in den höheren Einkommensklassen vergleichbar und dürften eine ähnlich hohe Lebenserwartung aufweisen. Aus diesen Überlegungen ergibt sich, dass vor allem in den Einkommensklassen unter 35 Entgeltpunkten Personen enthalten sein dürften, deren Lebenseinkünfte untererfasst sind.

Abbildung 1: Fernere Lebenserwartung von Männern im Alter 65 nach Zahl der persönlichen Entgeltpunkte, Versicherte mit mindestens 20 persönlichen Entgeltpunkten, (1993-2003)



Quelle: FDZ-RV – SUFRTBNWF03DEM, eigene Berechnungen. Die dünnen vertikalen Linien repräsentieren 99-prozentige Konfidenzintervalle.

Der Themendatensatz SUF Demografie spiegelt die Pluralität der Forschungsinteressen im Bereich der Demografie wider. So enthält er zahlreiche Merkmale und Personenkreise, die für viele Fragestellungen benötigt werden, jedoch im vorliegenden Projekt keine Rolle spielen (Geschlecht, Witwerrenten, Berufsklassenschlüssel, Bundesländer, Zahl der Kinder u.a.). Mit anderen Worten, der Kompromiss zwischen Detailliertheit und Datenschutz war für die spezifischen Anforderungen des Projektes „Differenzielle Sterblichkeit“ nicht hinreichend; deshalb wurde die Erstellung eines neuen Themenfiles vorbereitet. Die Überlegungen dazu

begannen im Rahmen des 2. Workshops des FDZ-RV im Juni 2005, bei dem auch die Kooperation von Rembrandt D. Scholz und Hans-Martin von Gaudecker für Projekte im Bereich der differentiellen Sterblichkeit zustande gekommen ist (vgl. die Ergebnisse in Gaudecker und Scholz (2006)). Aus beiden Institutionen wurden die Prioritäten formuliert, den der spezielle Themendatensatz „Differentielle Sterblichkeit“, der ausschließlich auf einem Gastwissenschaftler-Arbeitsplatz bearbeitet werden kann, enthalten soll.

Bei der Erstellung des Gastwissenschaftler-Themenfiles „Differentielle Sterblichkeit“ waren drei wesentliche Punkte von Bedeutung: Erstens, wie lassen sich die Personen in den unteren Einkommensklassen unterscheiden nach denjenigen mit einem niedrigen und denjenigen mit einem hohen Lebensarbeitseinkommen? Zweitens, lässt sich eine noch bessere Messgröße als die persönlichen Entgeltpunkte für das Lebensarbeitseinkommen finden? Drittens, kann die Stichprobengröße erhöht werden?

3.2 Gastwissenschaftler-Themenfile zur Analyse der ‚Differentiellen Sterblichkeit‘ auf einem Gastwissenschaftler-Arbeitsplatz

Im Juli 2005 wurde im Rahmen der Diskussionen auf dem zweiten Workshops des Forschungsdatenzentrums der Rentenversicherung ein Themenfile zur Analyse differentieller Sterblichkeitsprozesse skizziert. Wichtig war dabei, dass sowohl auf Seiten der Datennehmer als auch auf der Seite der Datengeber datenschutzrechtliche Kenntnisse sowie eine diesbezügliche Sensibilität herrschten.⁶ Grundsätzlich bewirkt die Anonymisierung eine Reduktion der in den Daten enthaltenen Information und schränkt dadurch das Analysepotenzial ein. Insofern ist eine Balance zwischen Einhaltung der datenschutzrechtlichen Regelungen und Analysemöglichkeiten zu finden. Die zu Grunde liegenden Kriterien für die Umsetzung der faktischen Anonymisierung basieren auf folgenden Standards, die sich an Empfehlungen, die in einem gemeinsamen Forschungsprojekt des Statistischen Bundesamtes und ZUMA Mannheim zu Beginn der 90er Jahre entwickelt wurden - ergänzt um einige Spezifika der GRV -, orientieren (vgl. Stegmann u.a. 2005: 210f.):

(1) Direkte Identifikatoren (wie Personen- und sonstige Ordnungsnummern) dürfen nicht weitergegeben werden. (2) Die Anordnung der Datensätze im zu übermittelnden Datenbestand muss systemfrei sein. (3) Es ist eine Stichprobe zu ziehen. (4) Als unmittelbare Regionalangabe wird in der Regel das Bundesland übermittelt. Die maximale

⁶ In diesem Zusammenhang sind die datenschutzrechtlichen Ausführungen von Heese (2004) von Bedeutung.

Tiefe der regionalen Gliederung bilden Kreise und kreisfreie Städte, wobei die Punkte 1 bis 3 und 5 bis 7 zu beachten sind. (5) Staatsangehörigkeiten oder Gruppen von Nationalitäten, denen weniger als ca. 50.000 Personen in der Bundesrepublik Deutschland angehören, dürfen nicht identifizierbar sein. (6) Jede ausgewiesene Merkmalsausprägung diskreter Variablen muss in ihrer univariaten Verteilung in der Grundgesamtheit mindestens 5.000 Fälle umfassen. Ist diese Voraussetzung nicht erfüllt, ist eine sachgerechte Vergrößerung bzw. Zusammenfassung vorzunehmen, bis die vorgegebene Fallzahl erreicht wird. (7) Angaben zu Entgeltpunkten, z. B. aus Beitragszeiten, beitragsgeminderten Zeiten und beitragsfreien Zeiten, werden ganzzahlig gerundet oder bis zur ersten Nachkommastelle weitergegeben. Zusätzlich wird durch Zusammenfassung der Werte am oberen Rand der Verteilung eine Begrenzung vorgenommen. (8) Werte zur Rentenberechnung (Entgeltpunkte, Angaben zur Dauer rentenrechtlicher Zeiten) werden nicht differenziert nach Versicherungszweig und Ost/West weitergegeben und nach oben begrenzt (vgl. Pkt. 7.). (10) Datumsangaben werden aufgeteilt in Jahres- und Monatsangaben. Die Jahresangaben werden bei einer Besetzung von 5.000 Fällen und weniger in der Grundgesamtheit dementsprechend gruppiert.

Neben den Grundsätzen zur faktischen Anonymisierung sind bei der Einrichtung von Gastwissenschaftler-Arbeitsplätzen nach § 78a Satz 1 SGB X einige Vorkehrungen zu treffen, damit an solchen Arbeitsplätzen personenbezogene Daten bearbeitet werden können:⁷ (1) Zutritts- und Zugangskontrolle: Unbefugten ist der Zutritt zu Datenverarbeitungsanlagen, mit denen personenbezogene Daten verarbeitet oder genutzt werden, zu verwehren. Es ist zu verhindern, dass Unbefugte diese Datenverarbeitungssysteme nutzen. (2) Zugriffskontrolle: Es ist sicherzustellen, dass nur auf Daten zugegriffen werden kann, für die eine Zugriffsberechtigung besteht. (3) Weitergabekontrolle: Es muss gewährleistet werden, dass der Datenträger nicht unbefugt gelesen, kopiert, verändert oder entfernt wird. Die Wege der Datenweitergabe müssen nachvollziehbar sein. (4) Eingabe- und Verfügbarkeitskontrolle: Es muss feststellbar sein, von wem personenbezogene Daten in DV-Systeme eingegeben, verändert oder entfernt wurden. Die Möglichkeit der zufälligen Zerstörung ist auszuschließen.

Vor dem Hintergrund dieser Regelungen konnte der spezifische Datenbedarf und die Merkmalszusammenstellung konzipiert werden. Hierbei ist zu beachten, dass zur Wahrung der faktischen Anonymität der an einem Gastwissenschaftler-Arbeitsplatz verwendeten Daten eine Balance aus feinerer Differenzierung und der Löschung von Merkmalen zu finden ist, auch wenn eine Hinzuziehung externer Information an einem nicht vernetzten Arbeitsplatzrechner nicht möglich ist.

⁷ Die Arbeitsbedingungen an Gastwissenschaftler-Arbeitsplätzen in öffentlich geförderten Forschungsdatenzentren werden im Rahmen des Arbeitskreises öffentlicher Forschungsdatenzentren (AKöffFDZ) harmonisiert.

Im Unterschied zu den SUF Themendatensätzen Demografie wurden bei dem Themendatensatz Differentielle Sterblichkeit 2003 zahlreiche Merkmale gelöscht. Zunächst wurden, weil die Analysen zur Sterblichkeit auf Männer begrenzt wurden, nur Männer im Gastwissenschaftler-Themenfile abgespeichert. Da die Analysen sich auf männliche Versichertenrentner beziehen, wurden jene Merkmale, die in Zusammenhang mit Hinterbliebenenrenten stehen gelöscht. Des Weiteren wurden einige sozio-demografische Variable aus dem Datensatz entfernt, wie die Anzahl der Kinder, Familienstand und der Beruf. Die datenschutzrechtlich sensiblen Merkmale zur Staatsangehörigkeit und zum Wohnort, differenziert nach Bundesländern, wurden in Deutsche und Ausländer und der Wohnort in alte und neue Bundesländer sowie Ausland vergrößert.

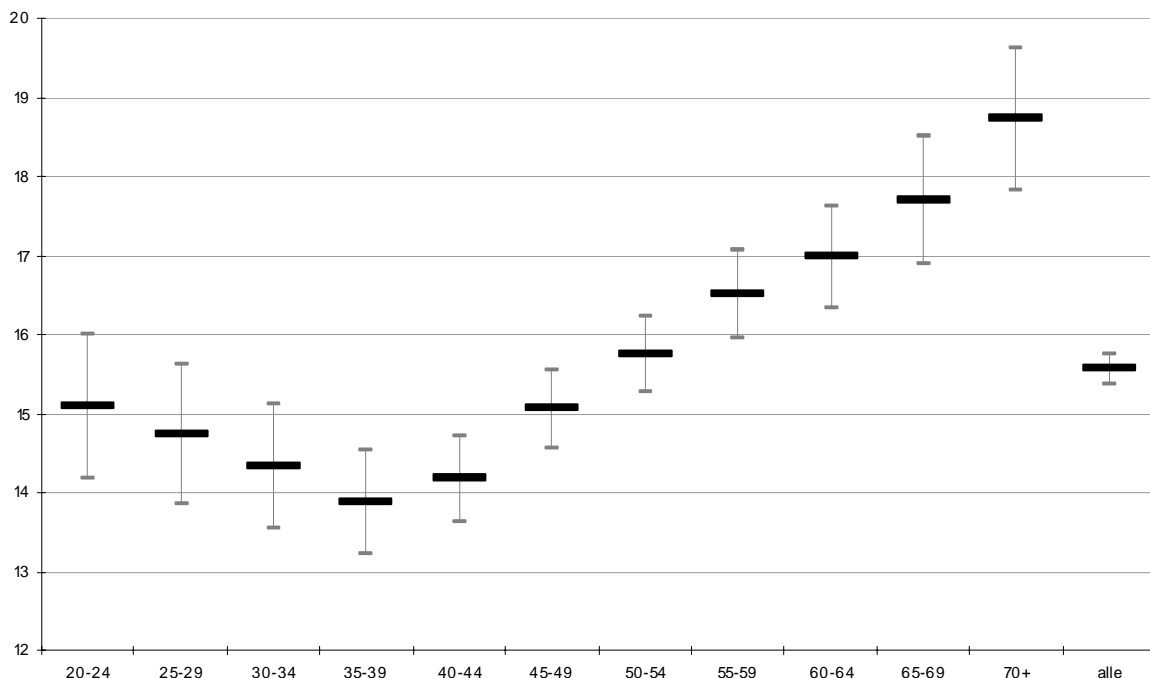
Auf Grund der Eingrenzung auf Versichertenrenten, des Verzichts insbesondere auf differenzierte regionale Angaben, solche zur Staatsbürgerschaft und Herkunftsland bei Renten, die Zeiten nach dem Fremdrentengesetz aufweisen, und der ausschließlichen Nutzung der Daten auf einem kontrollierten Gastwissenschaftler-Arbeitsplatz, konnte der Stichprobenumfang beim Rentenwegfall von 10 auf 25 Prozent und beim Rentenbestand von einem auf vier Prozent erhöht werden. Die unterschiedlichen Stichprobenumfänge wurden wie zuvor bei der Analyse auf Basis des SUF Demografie durch entsprechende Gewichtung angepasst. Die unterschiedlichen Bezugsperioden (31. Dezember im Bestand vs. Anfang Dezember bis Ende November im Wegfall) wurden dadurch angeglichen, dass alle Wegfälle einen Monat später gezählt wurden. Der Datensatz Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit‘ hat einen Stichprobenumfang von 295.894 männlichen Versichertenrentnern.

Im Unterschied zum SUF Themenfile Demografie ermöglicht der Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit‘ die Identifikation von Personengruppen, die vermutlich einen größeren Teil ihres Lebensarbeitseinkommens außerhalb von sozialversicherungspflichtigen Beschäftigungsverhältnissen erzielten. Dies soll an Hand der Merkmale Art des Krankenversicherungsverhältnisses, Summe der vollwertigen Beitragszeiten sowie der Summe der rentenrechtlichen Zeiten erfolgen. Da Selbständige, Freiberufler und Beamte in der Regel nicht in der gesetzlichen Krankenversicherung pflichtversichert sind, lassen sich diejenigen von ihnen identifizieren, die am Ende ihres Berufslebens nicht wieder in einem sozialversicherungspflichtigen Beschäftigungsverhältnis standen. Weiterhin kann man für Personen mit kurzen rentenrechtlichen Zeiten bzw. Beitragszeiten davon ausgehen, dass sie in den nicht belegten Zeiträumen andere Einkünfte bezogen haben. Bezüglich des zweiten Punktes fand die Summe der Entgeltpunkte aus Beitragszeiten Eingang in den Datensatz. Diese Variable ist näher an der theoretisch interessanten Größe „Lebensarbeitseinkommen“, weil Zugangsfaktoren, Entgeltpunkte für

beitragsfreie Zeiten und weitere Elemente der Rentenberechnung nicht berücksichtigt werden.

Die Ergebnisse der Berechnungen zur ferneren Lebenserwartung von Männern im Alter von 65 Jahren nach Entgeltpunktklassen auf Basis des Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit‘ sind in Abbildung 2 zu sehen. Die Ergebnisse wurden analog zu den in Abbildung 1 vorgenommenen Spezifikation berechnet, lediglich die zu Grunde liegende Datenbasis wurde verändert. Vergleicht man Abbildung 1 mit 2, so sieht man, dass die allgemeine fernere Lebenserwartung nahezu identisch ist: 15,56 Jahren im Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit‘ stehen 15,43 Jahre im SUF Demografie gegenüber. Bricht man die Analyse jedoch auf einzelne Entgeltpunktklassen herunter, so lassen sich teilweise große Unterschiede in den Punktschätzungen feststellen. Diese Unterschiede sind jedoch statistisch nicht signifikant: Die Unterschiede in der Lebenserwartung bei benachbarten Entgeltpunktklassen liegen stets innerhalb der 99-prozentigen Konfidenzintervalle. Obwohl die arithmetischen Sterblichkeitsmittelwerte auf den ersten Blick stark voneinander abweichen, sind sie unter Berücksichtigung der Stichprobengröße nicht signifikant.

Abbildung 2: Fernere Lebenserwartung von Männern im Alter 65 nach Zahl der persönlichen Entgeltpunkte, Versicherte mit mindestens 20 persönlichen Entgeltpunkten, Jahr 2003



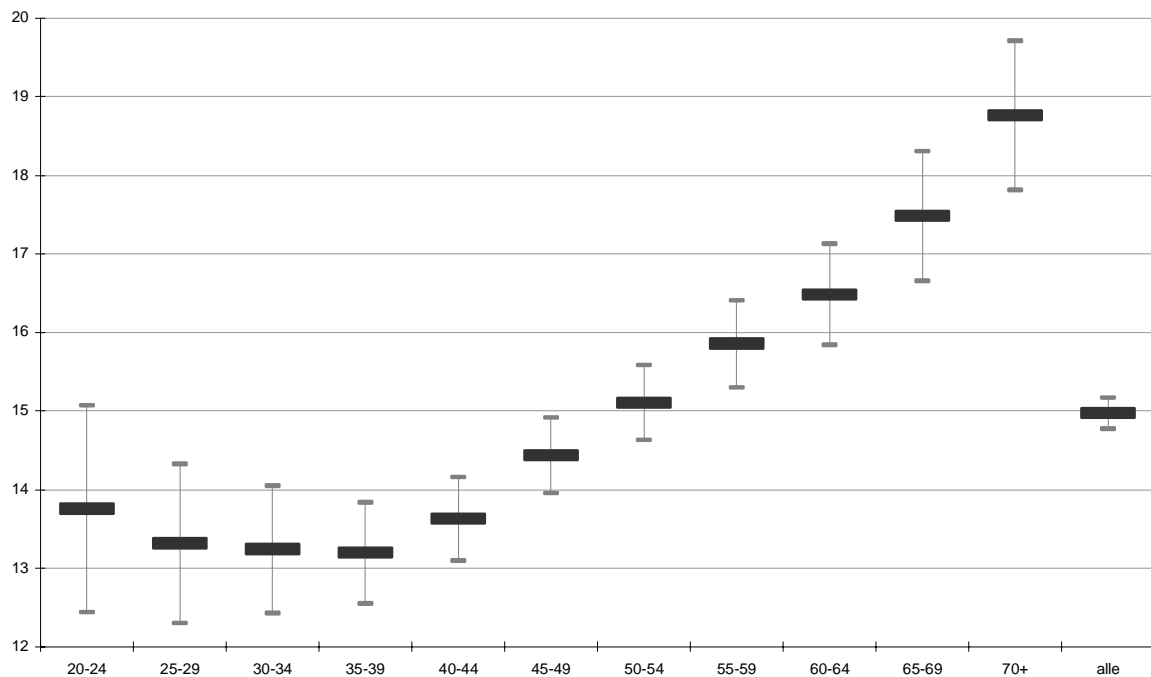
Quelle: FDZ-RV, Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘, eigene Berechnungen. Die dünnen vertikalen Linien repräsentieren 99-prozentige Konfidenzintervalle.

Diese Problematik tritt trotz der vergleichsweise großen Stichproben von annähernd 300.000 Fällen auf. An dieser Stelle wird der große Datenbedarf demographischer Techniken

deutlich. Man muss sich das Verfahren dahingehend vorstellen, dass für jede der 35 Altersklassen (vom 65. bis zum 100. Lebensjahr) multipliziert mit 11 Entgeltpunktklassen eine Tabelle mit 385 Zellen entsteht, bei der für jede Zelle die Sterbewahrscheinlichkeit geschätzt wird. Im Durchschnitt macht dies 212 (780) Beobachtungen pro Klasse im SUF Demografie (Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit‘). Diese Zahl erscheint zunächst recht hoch, jedoch konzentriert sich die Verteilung der Personen stark auf die mittleren Entgeltpunkt- und niedrigen Altersklassen. So finden sich zum Beispiel für die unterste Einkommensklasse im SUF Demografie durchschnittlich nur 97 Personen in jeder Altersstufe. Daraus resultieren die hohen Standardfehler. Im Unterschied dazu konzentrieren sich die Rentner auf den Bereich zwischen 40 und 59 Entgeltpunkte, weshalb in diesem Bereich vergleichsweise präzise Aussagen möglich sind.

In Abbildung 3 werden Ergebnisse dargestellt, die ebenfalls auf Datenbasis des Gastwissenschaftler-Themenfiles ‚Differentielle Sterblichkeit‘ berechnet wurden. Jedoch wurde hier versucht, den Personenkreis auf diejenigen einzugrenzen, die den überwiegenden Teil ihres Erwerbslebens sozialversicherungspflichtig beschäftigt waren. Dazu wurden die Lebenserwartungen für diejenigen Personengruppen berechnet, die in der gesetzlichen Krankenversicherung pflichtversichert waren und mindestens 25 Jahre an rentenrechtlichen Zeiten aufwiesen. Der Stichprobenumfang reduziert sich durch diese Selektion von 295.894 auf 210.482 männliche Versichertenrentner. Ein Vergleich mit Abbildung 2 zeigt, dass die Punktschätzungen für die Gruppen mit mindestens 50 persönlichen Entgeltpunkten in beiden Spezifikationen nahezu identisch sind. In den niedrigen Entgeltpunktklassen sinken die durchschnittlichen Lebenserwartungen deutlich, insbesondere steigen sie in den untersten Klassen nicht mehr so stark an. Die allgemeine Lebenserwartung sinkt um rund sieben Monate auf 14,97 Jahre. Dieser Rückgang ist statistisch signifikant; es wurde also eine Personengruppe mit einer überdurchschnittlich hohen Lebenserwartung ausgeschlossen. Es ist davon auszugehen, dass sich trotz dieser Maßnahmen - nur GKV-Pflichtversicherte mit mindestens 25 Jahren rentenrechtlicher Zeiten - noch immer Personen mit höheren Lebenseinkommen in den unteren Entgeltpunktklassen befinden und der leichte Anstieg der durchschnittlichen Lebenserwartung in den niedrigsten Gruppen eine Folge dieser Annahme ist.

Abbildung 3: Fernere Lebenserwartung von Männern im Alter 65 nach Zahl der persönlichen Entgeltpunkte, nur GKV-Pflichtversicherte mit mindestens 25 Jahren rentenrechtlicher Zeiten und mindestens 20 persönlichen Entgeltpunkten



Quelle: FDZ-RV, Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘, eigene Berechnungen. Die dünnen vertikalen Linien repräsentieren 99-prozentige Konfidenzintervalle.

An dieser Stelle wird deutlich, wie voraussetzungsreich Analysen zum Thema Einkommen und Sterblichkeit sind, weil es aus methodischer Perspektive anspruchsvoll ist, das gesamte Lebenseinkommen (bis zur Verrentung) zu ermitteln. Bei den hier verwendeten Daten des FDZ-RV trifft dies in besonderem Maße für solche Einkünfte zu, die nicht in der Rentenversicherung verbeitragt wurden.

Im Rahmen der Analysen mit dem Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ stellte sich heraus, dass die detaillierteren Variablen zur Rentenberechnung (Entgeltpunkte aus Beitragszeiten, vollwertige Beitragszeiten, beitragsgeminderte Zeiten) lediglich für diejenigen Personen vorlagen, die wegen der Konsequenzen des In-Kraft-Tretens des Rentenreformgesetzes 1992 (RRG '92) für die statistische Berichterstattung nach 1992 in Rente gegangen sind. Damit konnten nur die 1928 und später geborenen Kohorten analysiert werden. Die ältesten Individuen sind somit im Berichtsjahr 2003 fünfundsiebzig Jahre alt; eine Berechnung der Lebenserwartung ist somit nicht möglich. Stattdessen wurde die Wahrscheinlichkeit, das 74. Lebensjahr zu erreichen, berechnet. Analoge Ergebnisse für 2002 sind in Gaudecker und Scholz (2006) abgedruckt. Für die Interpretation der Sterblichkeitsbefunde in diesem Beitrag ist anzumerken, dass die Unterschiede zwischen den beiden Merkmalen „persönliche Entgeltpunkte“ und „Entgeltpunkte aus Beitragszeiten“ nicht so groß sind, als dass sie die

Validität des Merkmals persönliche Entgeltpunkte als Indikator für das Lebensarbeitseinkommen von Männern der betrachteten Kohorten grundsätzlich in Frage stellen würde. Andererseits bildet das Merkmal „Entgeltpunkte aus Beitragszeiten“ das tatsächlich erzielte Lebensarbeitseinkommen in den unteren Einkommensklassen (weniger als 35 Entgeltpunkte) offensichtlich besser ab.

Trotz dieser weitgehend ähnlichen Befunde, die die Analysen mit den Datensätzen SUF Demografie 2003 und Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ gezeigt haben, wurde insbesondere zur Erhöhung der statistischen Verlässlichkeit der Ergebnisse eine weitere Verbesserung der Datenbasis angestrebt. Zum einen lag dies daran, dass der Stichtag der Rentenbestandstatistik (31.12.) nicht exakt zu dem Berichtszeitraum der Rentenwegfallstatistik (Dezember bis November) passt. Damit müssen einige restriktive Annahmen darüber getroffen werden, dass man tatsächlich die Todesfälle auf die Population bezieht, die dem Mortalitätsrisiko ausgesetzt war. Aufgrund der Logik der Stichprobenziehung können diese mit dem Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ nachträglich nicht korrigiert werden. Schließlich zeigte sich, dass die Stichprobengröße vor allem für weitere Untergliederungen der Population, wie z.B. der Vergleich von Sterbewahrscheinlichkeiten in den alten und neuen Bundesländern, noch nicht hinreichend groß war. Aufgrund dieser Überlegungen wurde ein Datensatz zum Fernrechnen entwickelt.

3.3 Kontrolliertes Fernrechnen zur ‚Differentiellen Sterblichkeit‘

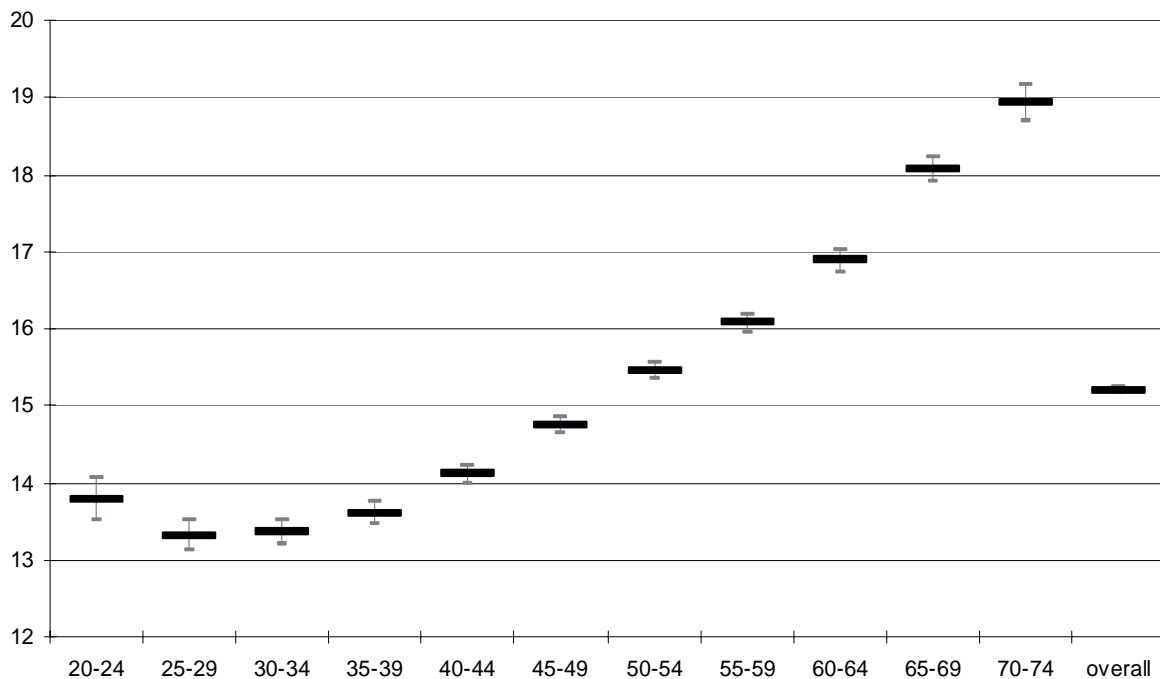
Aus der Perspektive des Datenschutzes bietet das Verfahren des kontrollierten Fernrechnens den Vorteil, dass die mit dem FDZ-RV kooperierenden und vertraglich gebundenen Wissenschaftler nicht mit den Datensätzen in Kontakt kommen, die den Analysen zu Grunde liegen. Entsprechend aufbereitete Datensätze verbleiben im FDZ-RV und die Wissenschaftler senden ihre in STATA, SPSS bzw. SAS geschriebenen Syntaxen an einen Mitarbeiter des FDZ-RV. Ein Mitarbeiter führt die Berechnungen mit der jeweiligen Software durch. Im Anschluss daran werden die empirischen Befunde dahingehend überprüft, ob die datenschutzrechtlichen Vorschriften eingehalten wurden und falls ja, werden die Ergebnisse den Wissenschaftlern zugesendet. Im Falle von in Abbildung 4 zu sehenden Schätzungen zur ferneren Lebenserwartung in einer Differenzierung nach Entgeltpunktklassen ist eine Deanonymisierung einer bestimmten Person ausgeschlossen, insofern gab es keine datenschutzrechtlichen Bedenken.

Der Arbeitsaufwand, den ein solches Verfahren mit sich bringt, ist für beide Seiten, jedoch insbesondere für die Beschäftigten im FDZ-RV hoch. Bereits kleine Programmierfehler können eine lange Fehlersuche zur Folge haben. Insofern ist es von großer Bedeutung, dass die Syntaxen vorher ausgiebig getestet wurden. Im hier vorgestellten Projekt war dies im Rahmen der mehrere Monate andauernden Analysen mit dem SUF Demographie sowie mit dem Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ möglich. Ein großer Vorteil ist, dass die zu Grunde liegende Datensätze der Rentenversicherung als einheitliche Statistikdatensätze als Rentendatensatz SK 90 konzipiert sind (vgl. Vorworte in VDR 2004 a/b). Diese Standardisierung mit ihren konsistent aufgebauten Datensätzen erleichtert ein solches Verfahren. Allerdings ist von den Mitarbeitern im FDZ-RV zunächst der zu analysierende Datensatz aufzubereiten. Da es sich hierbei um Vollerhebungen handelt, sind erhebliche Datenmengen zu bewältigen. Die Fallzahl GKV-pflichtversicherter männlicher Versichertenrentner (ohne Witwerrenten, mit mindestens 20 persönlichen Entgeltpunkten und mindestens 25 Jahren rentenrechtlicher Zeiten) beläuft sich auf 3.802.282 Personen.

Die Ergebnisse aus den Berechnungen mit dem Fernrechenfile ‚Differentielle Sterblichkeit‘ finden sich in Abbildung 4. Beim Vergleich mit Abbildung 3 fällt das ähnliche Muster der Punktschätzungen auf. Bezüglich der Höhe der ferneren Lebenserwartung der Männer in Abhängigkeit von der Entgeltpunktklasse lassen sich kaum Unterschiede feststellen. Das wesentliche Unterscheidungsmerkmal ist die statistische Trennschärfe: Bei allen Einkommensklassen über 35 persönlichen Entgeltpunkten kann man nun mit 99-prozentiger Sicherheit sagen, dass die Lebenserwartung von Klasse zu Klasse ansteigt. Eine solche statistisch signifikante Aussage zu formulieren war mit den anderen beiden Datensätzen

nicht möglich. Weitere auf Datenbasis des Fernrechenfile ‚Differentielle Sterblichkeit‘ durchgeführte Analysen sind in Gaudecker und Scholz (2006) dokumentiert.

Abbildung 4: Fernere Lebenserwartung von Männern im Alter 65 nach Zahl der persönlichen Entgeltpunkte, nur GKV-Pflichtversicherte mit mindestens 25 Jahren rentenrechtlicher Zeiten und mindestens 20 persönlichen Entgeltpunkten, Jahr 2002.



Quelle: FDZ-RV, Fernrechenfile ‚Differentielle Sterblichkeit 2002‘, eigene Berechnungen. Die dünnen vertikalen Linien repräsentieren 99-prozentige Konfidenzintervalle.

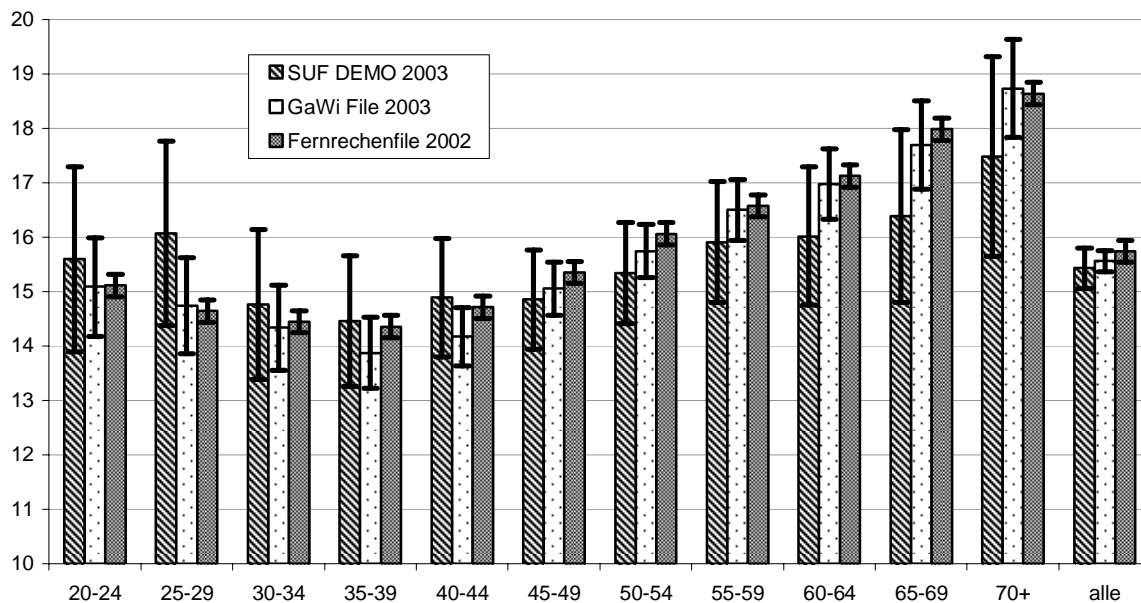
Abschließend lässt sich feststellen, dass der Datenzugang über die Methode des kontrollierten Fernrechnen nach kleineren Anlaufschwierigkeiten hervorragend funktionierte. Dennoch ist es für alle Beteiligten sehr zeitaufwändig. Das Gros der wissenschaftlichen Fragestellungen kann mit einem wesentlich geringeren Stichprobenumfang analysiert werden. Im Fall der Analysen im Projekt ‚Differentielle Sterblichkeit‘ sind deshalb sehr hohe Fallzahlen erforderlich, weil das Verfahren der Berechnung von Sterbetafeln in der Demographie entwickelt wurde und Demographen üblicherweise sehr große Fallzahlen von Bevölkerungen zur Verfügung stehen. Bei im Bereich der Wirtschafts- und Sozialwissenschaften angesiedelten Forschungsvorhaben dürften in der Regel mit den Stichprobenumfängen der SUFs mit rund 100.000 Fällen hinreichende Fallzahlen zur Verfügung stehen. Diese sind im Hinblick auf ihre Fallzahl erheblich größer als die in den Wirtschafts- und Sozialwissenschaften üblichen Datensätze.

6. Zusammenfassung

Analysen zum differentiellen Mortalitätsgeschehen mit Sterbetafeln benötigen sehr hohe Fallzahlen. Diese sind im Rahmen von umfragebasierten Erhebungen nicht verfügbar. Auf Basis von prozessproduzierten Daten der gesetzlichen Rentenversicherung wurden durch das am Mannheimer Forschungsinstitut Ökonomie und demographischer Wandel (MEA) angesiedelte Projekt ‚Differentielle Sterblichkeit‘ Analysen zum Mortalitätsgeschehen durchgeführt. In Abbildung 5 sind die Ergebnisse zusammenfassend dargestellt, die auf Basis der drei verschiedenen Datensätze berechnet wurden. Da das Merkmal „rentenrechtliche Zeiten“ im SUF Demographie 2003 nicht enthalten ist, wird auf die Sterblichkeit aller Versicherten mit mehr als 20 persönlichen Entgeltpunkten zurückgegriffen.⁸ Die Fallzahlen betragen damit beim SUF Demographie 81.610, beim Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ 295.894 und beim Fernrechenfile ‚Differentielle Sterblichkeit 2002‘ 4.413.575 Fälle. Aufgrund der unterschiedlichen Bezugszeiträume der drei verschiedenen Datensätze sind die Befunde zur Sterblichkeit lediglich bedingt vergleichbar, zu Illustrationszwecken sind sie jedoch sehr hilfreich.

⁸ Zu den Interpretationseinschränkungen für die Bezieher niedriger Renten siehe Abschnitt 3.1

Abbildung 5: Fernere Lebenserwartung von Männern im Alter 65 nach Zahl der persönlichen Entgeltpunkte, Versicherte mit mindestens 20 persönlichen Entgeltpunkten – Vergleich der drei Datensätze



Quelle: FDZ-RV – SUFRTBNWF03DEM, Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘, Fernrechenfile ‚Differentielle Sterblichkeit 2002‘, eigene Berechnungen. Die vertikalen Linien repräsentieren 99-prozentige Konfidenzintervalle.

Der Vergleich von Punktschätzungen und Konfidenzintervallen verdeutlicht, wie wichtig die Einbeziehung der statistischen Schwankungen ist. So liegt die Punktschätzung der Lebenserwartung in der höchsten Einkommensklasse für das SUF Demographie fast eineinhalb Jahre unter derjenigen im Fernrechenfile ‚Differentielle Sterblichkeit 2002‘. Das Konfidenzintervall schließt jedoch die Punktschätzung auf Basis des Fernrechenfiles mit ein und bewahrt vor Fehlschlüssen. So bleiben die Ergebnisse aus dem SUF Demographie zwar instruktiv, erlauben jedoch keine signifikanten Aussagen, da sich sämtliche Konfidenzintervalle überlappen. Die Berechnungen zur Sterblichkeit auf Basis des Gastwissenschaftler-Themenfile ‚Differentielle Sterblichkeit 2003‘ ergeben ein wesentlich präziseres Bild: So lässt sich in fast allen Fällen mit 99-prozentiger Sicherheit feststellen, dass Sterblichkeitsunterschiede zwischen solcher Gruppen bestehen, die nicht direkt nebeneinander liegen. Hierbei sind jedoch die Unschärfen im Hinblick auf die Höhe der Lebensarbeitseinkommen in den unteren Entgeltpunktklassen wegen jenen Einkünften zu beachten, die – wie beispielsweise bei Selbstständigen mit kürzerer rentenversicherungspflichtiger Erwerbsbiografie - nicht verbeitragt sind. Mit anderen Worten, in den unteren Entgeltpunktklassen können sich männliche Versichertenrentner befinden, die bezüglich der Höhe ihres Lebensarbeitseinkommens weiter oben angesiedelt sein sollten.

Die Berechnungen zur ferneren Lebenserwartung von Männern an Hand des Fernrechenfile ‚Differentielle Sterblichkeit 2002‘ zeigen - von den Unschärfen im den unteren Einkommensklassen abgesehen - ab der Einkommensklasse von 35 bis 39 Entgeltpunkten eine mit dem Einkommen steigende Lebenserwartung, wobei die ermittelten Differenzen in den Gruppenmittelwerten innerhalb der jeweiligen Vertrauensintervalle liegen; das heißt, die Differenzen sind signifikant.

Die durchschnittliche fernere Lebenserwartung der Männer liegt bei etwa 15,7 Jahren. Im Hinblick auf die Entgeltpunktklassen haben die Männer mit zwischen 35 und bis einschließlich 49 Entgeltpunkten eine um etwa eineinhalb bis etwa ein Jahr niedrigere Lebenserwartung als der Durchschnitt der Männer. Und andersherum liegt die Lebenserwartung in den Klassen von 50 bis einschließlich 64 Entgeltpunkten etwa ein halbes Jahr bis eineinhalb Jahre über der durchschnittlichen Lebenserwartung von allen Männern. Vor dem Hintergrund dessen, dass in den Klassen von 35 bis 49 und von 50 bis 64 Entgeltpunkten je 37% der Männer angesiedelt sind, lässt sich festhalten, dass für drei Viertel der männlichen Rentner die fernere Lebenserwartung um plus minus eineinhalb Jahre schwankt, wenngleich die Lebenserwartung für Männer in den höheren Entgeltpunktklassen signifikant höher ist. In diesem Zusammenhang ist von Bedeutung, dass die hier vorgenommenen Analysen die statistische Assoziation von Lebenseinkommen und Sterblichkeit untersuchten. Ob eine höheres Lebenseinkommen kausal mit einer höheren Lebenserwartung einhergeht, kann nicht geklärt werden. Dies liegt unter anderem daran, dass das Bildungsniveau die Sterblichkeit beeinflusst und mit dem Einkommen korreliert ist. Ein schlechter Gesundheitszustand sowohl das Einkommenserzielungspotential im Lebensverlauf als auch die fernere Lebenserwartung erheblich verringern wird.

Im Hinblick auf die vier Kernziele des Forschungsdatenzentrums, (1) Bereitstellung von Mikrodaten der Rentenversicherung und (2) Betreuung bei methodischen Fragen sowie der (3) Bereitstellung von Informationen und (4) Anregung zur Diskussion zwischen Wissenschaft und Datenproduzent hat das Projekt ‚Differentielle Sterblichkeit‘ folgendes gebracht:

Erstens wurde im Rahmen des Projektes zunächst der Scientific Use File (SUF) Demografie genutzt. Am Beginn einer wissenschaftlichen Nutzung der FDZ-RV-Daten sollten stets SUFs stehen. Dies gilt auch für den Fall, dass sich kein direkt relevanter Datensatz im Angebot des Forschungsdatenzentrums befindet. Erste Analysen mit den SUFs dienen dazu, sich in die Charakteristika der prozessproduzierten Daten der Rentenversicherung einzuarbeiten. Sollte das Analysepotenzial der SUFs als Basisfile nicht hinreichend sein, kann u.U. ein spezieller Themendatensatz als SUF erstellt werden. Sind insbesondere größere Fallzahlen wichtig, dann können zweitens Analysen auf Gastwissenschaftler-Arbeitsplätzen eine Alternative darstellen. Als dritte Form der Datennutzung wurde hier das kontrollierte Fernrechnen, bei

dem Wissenschaftler keinen Kontakt mit den Daten haben, vorgestellt. Das kontrollierte Fernrechnen ist in der derzeitigen Aufbauphase des FDZ-RV nicht routinisiert und erfordert hohe personelle Ressourcen.

Zweitens konnte an Hand der Kooperation zwischen dem Projekt und den Mitarbeitern des FDZ-RV einige methodische Probleme gelöst werden, z.B. im Hinblick auf eine Synchronisierung der unterschiedlichen Berichtszeiten von Rentenbestand und Rentenwegfall.


Hierzu wurden drittens neben dem Erfahrungswissen der FDZ-RV Mitarbeiterinnen und Mitarbeiter u.a. publizierte Dokumentationen als Handreichung am Gastwissenschaftler-Arbeitsplatz zur Verfügung gestellt.

Im Hinblick auf das vierte Ziel Anregung zur Diskussion zwischen Wissenschaft und Datenproduzent haben der zweimonatige Aufenthalt von Hans-Martin v. Gaudecker in den Räumen des FDZ-RV in Berlin sowie sein Vortrag im Rahmen des Jour Fixe des Geschäftsbereiches Forschung, Entwicklung und Statistik der Deutschen Rentenversicherung Bund am 8. Februar 2006 einen diskursiven Beitrag geleistet; dieser soll durch diese Publikation sowie durch weitere Workshops des Forschungsdatenzentrums fortgesetzt werden (siehe hierzu www.fdz-rv.de).

Im Hinblick auf die Entwicklung des ersten und wichtigsten Kernzieles des Forschungsdatenzentrums, nämlich die Bereitstellung von Mikrodaten der Rentenversicherung, kann darauf hingewiesen werden, dass das Datenangebot demnächst erweitert wird.

Abbildung 6: Prozessproduzierte Mikrodaten der Gesetzlichen Rentenversicherung und deren Verfügbarkeit im FDZ-RV

Renten-zugangs-statistik	Renten-bestands-statistik	Rentenzahl-bestands-statistik	Versicherten-statistik	Reha-Statistik	Versorgungs-ausgleichs-statistik
Querschnittsdatensätze					
Rentenzugang Rentenwegfall und Renten- änderung	Rentenzahl- ungen aus den Trägerkonten	Mehrfachbe- zug von Renten- zahlungen	Statistik der aktiv Versicherten	RSD- Jahreser- hebung abgeschlos- sener Reha- Leistungen	Bestandsfälle aller Versicherten mit Versorgungs- ausgleich
Längsschnittsdatensätze					
Vollendete Versicherten- leben			Versicherungs- konstentstich- probe	RSD-Verlaufs- erhebung	

Quelle: Himmelreicher und Radl (2006);  = verfügbar im FDZ-RV für 2003/4.

Während sich die zur Verfügung gestellten Daten des FDZ-RV bislang auf die Querschnittsdatensätze der Rentenzugangs- und Rentenbestandsstatistik konzentrierte, wird das Angebot auf die Längsschnittdatensätze ‚Vollendete Versichertenleben‘ und ‚Versicherungskontenstichprobe‘ erweitert. An der Aufbereitung der Statistik der aktiv Versicherten als SUF wird ebenfalls gearbeitet. Insofern wird sich das Datenangebot des FDZ-RV in Zukunft vergrößern. Und die Zugangsmöglichkeiten werden, sofern es verfügbare Ressourcen ermöglichen, an die besonderen Anforderungen der wissenschaftlichen Projekte angepasst.

Literatur

- Berkel, Barbara; Börsch-Supan, Axel (2006): Auswirkungen verschiedener Reformszenarien auf Renteneintrittsentscheidungen. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 213-225.
- Chiang, C. L. (1984): The Life Table and its Applications. Krieger Pub. Co., Malabar, Fla.
- Gaudecker Hans-Martin von (2004): Intragenerationale Umverteilungswirkungen in der gesetzlichen Rentenversicherung. In: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau. Bericht vom ersten Workshop des FDZ-RV am 28. und 29. Juni in Würzburg. DRV-Schriften Band 55: Frankfurt a. M., S. 126-129.
- Gaudecker, Hans-Martin von (2006): Differentielle Sterblichkeit in der GRV: Problemaufriss und erste Berechnungen. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 242-252.
- Gaudecker, Hans-Martin von; Rembrandt D. Scholz (2006): Lifetime Earnings and Life Expectancy. MEA-Discussion Paper Nr. 101-06, verfügbar unter www.mea.uni-mannheim.de.
- Haider, Steven; Solon, Gary (2006): Life-Cycle Variation in the Association between Current and Lifetime Earnings, *American Economic Review* (im Erscheinen).
- Heese, Claudia (2004): Aspekte des Datenschutzes im Forschungsdatenzentrum der gesetzlichen Rentenversicherung. In: Verband Deutscher Rentenversicherungsträger (Hg.): Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau. Bericht vom ersten Workshop des FDZ-RV am 28. und 29. Juni in Würzburg. Frankfurt a.M., S. 41-48.
- Himmelreicher, Ralf K. (2001): Soziodemographie, Erwerbsarbeit, Einkommen und Vermögen von westdeutschen Haushalten. Eine Längsschnitt-Kohortenanalyse auf Datenbasis des SOEP (1984-1997). Logos: Berlin.
- Himmelreicher, Ralf K.; Frommert, Dina (2006): Gibt es Hinweise auf zunehmende Ungleichheit der Alterseinkünfte und zunehmende Altersarmut? Der Einfluss von Erwerbs- und Familienbiografien auf die Rentenhöhe in Deutschland. In: DIW Vierteljahreshefte zur Wirtschaftsforschung 1/2006, Berlin, S. 108-130.
- Himmelreicher, Ralf K.; Radl, Jonas (2006): Zusammenfassung und Ausblick auf die weitere Entwicklung im FDZ-RV. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 281-286.
- Rische, Herbert (2006): Geleitwort. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 5.
- Scholz, Rembrandt D. (2004): Datenbedarf in der Mortalitätsforschung in Deutschland. In: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im

- Aufbau. Bericht vom ersten Workshop des FDZ-RV am 28. und 29. Juni in Würzburg. DRV-Schriften Band 55: Frankfurt a. M., S. 141-146.
- Scholz, Rembrandt D. (2006): Differentielle Sterblichkeitsanalysen mit den Daten der Deutschen Rentenstatistik. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 253-266.
- Stegmann, Michael; Luckert, Hilmar; Mika, Tatjana (2005): Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV): Grundsätze zur Anonymisierung von Mikrodaten und zu Gastwissenschaftler-Arbeitsplätzen, Deutsche Rentenversicherung, 2-3, S. 203–215.
- Unger, Rainer (2006): Das individuelle Zugangsrentenniveau 2003. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 267-280.
- VDR (2004a): Statistik Rentenbestand am 31. Dezember 2003, Band 148: Frankfurt.
- VDR (2004b): Statistik Rentenzugang des Jahres 2003 einschließlich Rentenwegfall, Rentenänderung/Änderung des Teilrentenanteils, Band 149: Frankfurt.
- Wübbeke, Christina (2006): Der Einfluss betrieblicher Rahmenbedingungen auf Zeitpunkt und Form des Ausscheidens älterer Arbeitnehmerinnen und Arbeitnehmer aus dem Erwerbsleben – Eine Analyse für Westdeutschland auf basis der IAB-Beschäftigtenstichprobe 1975 – 1995 mit Ergänzungsteil I. In: Forschungsrelevante Daten der Rentenversicherung. Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. DRV-Schriften Band 55: Berlin, S. 157 – 174.