

# Statistikdaten der gesetzlichen Rentenversicherung – ein grober Überblick

*Hilmar Luckert (VDR Würzburg)*

## 1. Einleitung

Die folgenden Ausführungen konzentrieren sich weitgehend auf regelmäßig durchgeführte Routineerhebungen des statistischen Berichtswesens beim VDR. Dabei wird lediglich auf jene Statistiken näher eingegangen, die auf Individualdaten basieren und die somit im Rahmen des Forschungsdatenzentrums der Rentenversicherung von Interesse sein könnten. Solche Statistiken, zu denen auch beim VDR die Daten lediglich in aggregierter Form vorliegen, werden zur inhaltlichen Abrundung erwähnt.<sup>1</sup> Dieser grobe Überblick über die Daten der Rentenversicherung beschreibt die momentane Situation und lässt offen, welche Daten mit welchen Ausprägungen der Wissenschaft durch das Forschungsdatenzentrum der Rentenversicherung künftig zugänglich gemacht werden können.

Vorweg zu schicken ist, dass es sich beim statistischen Berichtswesen der Rentenversicherung nicht um eine Statistik aus einem Guss handelt, sondern vielmehr um eine Ansammlung historisch gewachsener Statistiken. Diese werden aus einem bestehenden, jedoch Veränderungsprozessen unterliegenden Verwaltungsverfahren gespeist. Infolge dessen, aber auch wegen der verschiedenartigen Anforderungen, liegen den Statistiken durchaus unterschiedliche Datenkonzepte zugrunde. Während vor allem die Leistungsstatistiken überwiegend auf der Basis von Mikrodaten durchgeführt werden, gibt es im Bereich der Verwaltungsstatistiken, bspw. über die Antragserledigung, auch beim VDR lediglich aggregierte Informationen in Form von Summendaten. Viele Statistiken haben eine feste, für alle Fälle identischen Datenstruktur, dabei können allerdings je nach Ausgestaltung des Falles durchaus unterschiedliche Merkmalsbereiche von Interesse sein. Bei Statistiken mit Verlaufscharakter, bei denen je nach Anzahl und Dauer der Episoden mehr oder weniger viele chronologisch einordenbare Daten abzubilden sind, werden auch variable Datenstrukturen benutzt. In der Regel werden Statistiken in der Rentenversicherung als Vollerhebung durchgeführt und beinhalten vorwiegend prozessproduzierte Daten. Nur an wenigen Stellen wird mit Stichproben gearbeitet, weil sich eine Vollerhebung aus arbeitsökonomischen Aspekten verbietet. Nicht übersehen werden darf, dass manche Erhebungen stichtagsbezogen sind, d.h. sie beschreiben die Verhältnisse zu einem bestimmten Zeitpunkt, während andere Statistiken Zeitraumerhebungen darstellen, also über Sachverhalte und Ereignisse berichten, die in einem fest definierten Zeitraum - meist ein Kalendermonat oder ein Kalenderjahr - liegen. Schließlich gibt es auch Statistiken, die sowohl zeitraumbezogene als auch stichtagsbezogene Informationen vereinen, wobei je nach Fragestellung die besser geeigneten Variablen zu verwenden sind.

Generell gilt jedoch, dass die Datensätze verschiedener statistischer Erhebungen in der Rentenversicherung schon so weit anonymisiert sind, dass Individualdaten auf der Personenebene nicht verbunden werden können. D.h., wenn über eine Person in einer bestimmten Statistik in verschiedenen Zeiträumen oder zu verschiedenen Stichtagen berichtet wird, lassen sich diese

<sup>1</sup> Für einen umfassenden Überblick des Informationswesens des VDR siehe Rehfeld (2001, 160 ff.) bzw. im Downloadbereich in [www.fdz-rv.de](http://www.fdz-rv.de).

Informationen nicht miteinander verknüpfen. Solche Informationen zu einer Person, die in verschiedenen Statistiken enthalten sind, selbst wenn sie den gleichen Stichtag oder Berichtszeitraum betreffen, lassen sich ebenfalls nicht verknüpfen.

## 2. Welche Statistiken gibt es?

Wie eingangs erwähnt, beschränkt sich dieser Überblick auf die Routinestatistiken und lässt bereichsspezifische Erhebungen, Umfragen etc. außen vor. Generell können die Statistiken nach ihrem Inhalt in folgende sachlichen Bereiche untergliedert werden:

- Versichertenstatistiken,
- Rehabilitationsstatistiken,
- Rentenstatistiken,
- Finanzstatistiken und
- sonstige Statistiken.

Hinzu kommen noch die Sondererhebungen, die unter Umständen mehrere dieser Bereiche gleichzeitig betreffen.

### 2.1 Versichertenstatistiken

Insgesamt werden beim VDR sieben verschiedene Versichertenstatistiken geführt: drei Summendatensätze<sup>2</sup> und vier Datensätze, die auf Individualdaten basieren. Auf letztere wird hier näher eingegangen.

Die Stammsatzbestandsstatistik (**vgl. Abbildung 1**) wird zweimal jährlich, zum 1. Januar und zum 1. Juli eines jeden Jahres, erhoben. Gegenstand dieser Erhebung sind die vergebenen Versicherungsnummern, welche den bei den Rentenversicherungsträgern geführten Versicherungskonten entsprechen. Für die Zählung in dieser Statistik ist es völlig unerheblich, ob und in welchem Umfang Versicherungszeiten vorliegen. Neben dem Überblick über die Gesamtheit der von der Rentenversicherung zu verwaltenden Versicherungskonten ist diese Statistik insbesondere im Hinblick auf die Kontoführungswechsel und auf die Neuvergabe von Versicherungsnummern von Interesse. Da sie aus dem Stammsatzbestand der Datenstelle der Rentenversicherung gewonnen wird, beschränken sich ihre Inhalte auf eine kleine Auswahl der wenigen dort verfügbaren Merkmale.

---

<sup>2</sup> Die Statistik der latent Versicherten, die Statistik über unmittelbare Beitragserstattungen und die neue Beitragsrückstandstatistik werden an den VDR in summarischer Form gemeldet.

**Abbildung 1:** Stammsatzbestandsstatistik

Termine	Jährlich zum 1.1. und 1.7.
Datenform	Einzeldatensätze
Gegenstand	Vergebene Versicherungsnummern = Anzahl der Versicherungskonten
Quelle	Stammsatzbestand der Datenstelle der Rentenversicherung
Wichtige Merkmale	Geburtsjahr und -monat, Geschlecht, Staatsangehörigkeit, Wohnort (Kreis), Vergabeanstalt, aktueller Kontoführer (Versicherungsträger), Vergabedatum, Kontoführungswechsel, Stilllegungs-, Renten- und Lebensmerkmal
Wozu?	Struktur der betreuten Versicherungskonten, Kontoführungswechsel, Neuvergaben von Versicherungsnummern
Probleme	Keine Informationen über Umfang und Zeitpunkt der zurückgelegten Versicherungszeiten bzw. ob überhaupt versicherungsrelevante Zeiten vorhanden „Karteileichen“

Quelle: Eigene Zusammenstellung.

Wichtige Merkmale sind: das Geburtsjahr, das Geschlecht und die Staatsangehörigkeit des Versicherten, welcher Versicherungsträger aktueller Kontoführer ist, das Vergabedatum und welcher Versicherungsträger die Versicherungsnummer vergeben hat, die Tatsache, ob im vergangenen Jahr Kontoführungswechsel stattgefunden haben, sowie Informationen darüber, ob der Datensatz technisch stillgelegt ist, ob ein Rentenbezug vorliegt oder ob der betreffende Versicherte verstorben ist. Bei der Stammsatzbestandsstatistik handelt es sich primär um eine Verwaltungsstatistik mit Mengengerüsten zu Versicherten nach deren Alter, Geschlecht und Staatsangehörigkeit sowie zur Durchführung von Kontoführungswechseln und zu den im Berichtsjahr neu vergebenen Versicherungsnummern. Diese Statistik beinhaltet keine Angaben darüber, ob und wann Beiträge zur Rentenversicherung entrichtet wurden und welchen zeitlichen Umfang oder welche Höhe diese ggf. aufweisen. Bei latent Versicherten ohne Leistungsbezug besteht ein so genanntes „Karteileichenproblem“.

Wesentlich informativer ist die Statistik der aktiv Versicherten (**vgl. Abbildung 2**), die jährlich zum 1. Oktober des Folgejahres durchgeführt wird und zu der erfahrungsgemäß bis Ende des Folgejahres erste Ergebnisse vorliegen.

Der betrachtete Personenkreis ist deutlich enger gefasst als bei der Stammsatzbestandsstatistik: Im Wesentlichen beinhaltet diese Statistik Versicherte, die aktuell Rentenanwartschaften aufbauen, sei es durch Beitragszahlung oder dadurch, dass sie Anrechnungszeiten erwerben. Die dieser Statistik zugrunde liegenden Daten werden weitgehend bestimmt vom Meldefluss im integrierten Verfahren zur Datenmeldung an die Sozialversicherung (DEÜV). Daten anderer Versichertengruppen, wie zum Beispiel für freiwillig Versicherte oder Handwerker, finden auf anderen Wegen unmittelbar Eingang in die Versicherungskonten. Es gilt jedoch trotz des Time-Lags von zehn Monaten zu beachten, dass die Meldungen noch nicht vollständig sind: Erfahrungsgemäß fehlen bei den Beschäftigten nach zehn Monaten noch zwei bis drei Prozent der Daten zum Berichtsjahr. Bei anderen Zeiten, wie z. B. Wehr-/Zivildienst, ist der Erfassungsgrad nach zehn Monaten sogar noch deutlich niedriger und bei Kindererziehungszeiten – mangels

**Abbildung 2:** Statistik der aktiv Versicherten

Termine	Jährlich zum 1.10. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	Versicherte, die durch Beiträge oder Anrechnungszeiten aktuell Rentenanwartschaften aufbauen
Quelle	DEÜV-Meldungen ⇒ Versicherungskonten
Wichtige Merkmale	Geburtsjahr, Geschlecht, Staatsangehörigkeit, Tätigkeitsschlüssel, Versicherungsbeginn (Jahr), Wohnort (Kreis), Versicherungsart am Stichtag, Arten der Versicherung (Dauern/ Beiträge/ versicherte Entgelte) im Berichtsjahr und im Vorjahr, ggf. Rentenleistungsart und Rentenbeginn
Wozu?	Aktiv Versicherte nach Art der Versicherung, Umfang und Höhe der Beiträge bzw. versicherten Entgelte, Stichtags- und Zeitraumgrößen (2-Jahres-Fenster)
Probleme	Störungen im Meldeverfahren, nie zu 100% vollständig keine Branche

Quelle: Eigene Zusammenstellung, vgl. hierzu auch VDR Statistik Versicherte: Vorwort zu Band 147.

zeitnaher Erfassung im Versicherungskonto – ist er so gering, dass sinnvolle Auswertungen nicht möglich sind.

Die Statistik der aktiv Versicherten bietet ein breites Spektrum von Merkmalen: Es reicht vom Geburtsjahr, dem Geschlecht und der Staatsangehörigkeit der Versicherten über den Tätigkeitsschlüssel<sup>3</sup>, den Versicherungsbeginn und den Wohnort (Kreis) bis hin zu einer Differenzierung nach den verschiedenen Versicherungsarten, zu denen jeweils die Dauer sowie die Beiträge bzw. Höhe der versicherten Entgelte dargestellt sind. Abgebildet wird jeweils ein Zweijahresfenster, das aus dem Berichtsjahr und aus dem Vorjahr besteht. Da in Teilbereichen erfahrungsgemäß immer wieder Störungen beim Meldeverfahren auftreten, kann die Interpretation der Ergebnisse problematisch sein. Innerhalb der Statistik der aktiv Versicherten – wie auch in allen anderen Statistiken der Rentenversicherung – ist eine Differenzierung nach Branchen nicht möglich.

Die Versicherungskontenstichprobe (**vgl. Abbildung 3**) beschäftigt sich mit den Rentenanwartschaften von Versicherten. Im Unterschied zur Statistik der aktiv Versicherten werden in dieser Statistik sämtliche in der Chronologie gespeicherten Versicherungstatbestände dargestellt. Vorhandene Merkmale sind das Geburtsjahr, das Geschlecht und die Staatsangehörigkeit des Versicherten; ein regionales Zuordnungskriterium ist nicht vorhanden. Die übrigen Merkmale beziehen sich hauptsächlich auf die Höhe und die Zusammensetzung der Rentenanwartschaften, wobei verschiedene Arten von versicherungsrechtlich relevanten Zeiten und ihr „Wert“ in Form von Entgeltpunkten bei der Rentenberechnung dargestellt sind.

<sup>3</sup> Der Tätigkeitsschlüssel beruht auf einer Meldung im DEÜV-Verfahren und indiziert die aktuellste Tätigkeit der Versicherten: Das Merkmal enthält derzeit in den Stellen 1 bis 5 die Angaben zur Tätigkeit (vgl. Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Versicherungsnachweisen der Bundesanstalt für Arbeit): Ausgeübte Tätigkeit (Stellen 1 - 3), Stellung im Beruf (Stelle 4), Ausbildung (Stelle 5). Ist kein Tätigkeitsschlüssel zum Jahr des Leistungsfalles oder davor gespeichert, ist das Merkmal mit Nullen zu belegen. Insbesondere gilt dies für Zeiträume vor dem Jahr 2000.

**Abbildung 3:** Versicherungskontenstichprobe

Termine	Jährlich zum 1.10. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	Versicherte im Alter von 15 - 67 Jahren mit rentenrechtlichen Zeiten im Versicherungskonto oder einem Bonus aus Versorgungsausgleich
Quelle	Versicherungskonten
Wichtige Merkmale	<ol style="list-style-type: none"> <li>1. Geburtsjahr, Geschlecht, Staatsangehörigkeit, Tätigkeitsschlüssel, ggf. Rentenleistungsart und Rentenbeginn, Merkmale zur Anwendung des Fremdretenrechts, Werte zur Gesamtleistungsbewertung, Merkmale zum Umfang der vorhandenen Zeiten sowie zur Höhe und Zusammensetzung der momentanen Rentenanwartschaft</li> <li>2. chronologischer Versicherungsverlauf <ul style="list-style-type: none"> <li>• Art der versicherten Zeit</li> <li>• Beitragsart-Besonderheiten</li> <li>• Zeitpunkt und Zeitraumdauer (Monate)</li> <li>• Beitragshöhe / versichertes Entgelt</li> <li>• „Wert“ in Entgeltpunkten</li> </ul> </li> </ol>
Wozu?	Erworbene Rentenanwartschaften nach Höhe und Zustandekommen, Umfang der angerechneten Zeiten, Auswirkungen von rechtlichen Regelungen, Untersuchungen des chronologischen Versicherungsverlaufs
Probleme	<ul style="list-style-type: none"> <li>• unvollständige Kontoklärung ⇒ Unterschätzung der Zeiten und Anwartschaften keine Informationen zu ausländischen Versicherungszeiten</li> <li>• keine Regionalangabe</li> </ul>

Quelle: Eigene Zusammenstellung.

Dabei – und das ist das Wichtigste an dieser Statistik – ist in den Daten der Versicherungskontenstichprobe der chronologische Versicherungsverlauf dokumentiert, in dem Art, Zeitpunkt und Zeitdauer, Beitragshöhe und versichertes Entgelt sowie schließlich der Wert dieser Zeit in Form von Entgeltpunkten ausgewiesen sind. Mit dieser Statistik lassen sich daher die erworbenen Anwartschaften nach Höhe und Zustandekommen analysieren und der Umfang der angerechneten Zeiten sowie die Auswirkungen von rechtlichen Regelungen damit in Verbindung bringen. Losgelöst von rentenrechtlichen Aspekten kann der chronologische Verlauf der (versicherten) Entgelte aber ebenso von Interesse sein wie das Auftreten und die Abfolge bestimmter Arten der Versicherung, z. B. versicherungspflichtige oder geringfügige Beschäftigung, Lohnersatzleistungen, versicherungspflichtige Selbständigkeit usw. Problematisch an dieser Statistik ist, dass zuverlässige Aussagen oftmals nur zu gewinnen sind, wenn die Versicherungskonten vollständig geklärt sind. Dazu bedarf es aber der Mitwirkung der Versicherten, die in der Praxis nicht immer gegeben ist. Hinsichtlich der faktischen Anonymisierung stellen Längsschnittinformationen bekanntlich besondere Anforderungen.

Die Versorgungsausgleichsstatistik (**vgl. Abbildung 4**), die jährlich zum 1. April des Folgejahres durchgeführt wird, befasst sich mit Versicherten, für die ein Versorgungsausgleich im Versicherungskonto gespeichert ist und beinhaltet dabei jeweils sämtliche Bestandsfälle. Die Untersuchung des Anwartschaftsausgleichs in Versorgungsfällen, insbesondere die übertragenen und erworbenen Anwartschaften nach Ausgleichsformen, dürften auch die hauptsächliche Nutzungsmöglichkeit dieser Statistik bedeuten.

**Abbildung 4:** Versorgungsausgleichsstatistik

Termine	Jährlich zum 1.04. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	Versicherte mit einem Versorgungsausgleich (alle Bestandsfälle)
Quelle	Versorgungsausgleichsverfahren ⇒ Versicherungskonten
Wichtige Merkmale	Geburtsjahr / Geschlecht der Versicherten und des geschiedenen Ehepartners, Ehedauer, Art und Höhe des Versorgungsausgleichs, Beitragszahlungen zur Wiederauffüllung
Wozu?	Übertragene / erworbene Anwartschaften nach Ausgleichsformen
Probleme	Keine Informationen zu Anwartschaften insgesamt und zum Versicherungsverlauf

Quelle: Eigene Zusammenstellung.

Wichtige Merkmale, die hierbei zur Verfügung stehen, sind das Geburtsjahr und das Geschlecht – jeweils vom Versicherten und Ehepartner –, die Ehedauer, die Art und die Höhe des Versorgungsausgleichs sowie ggf. Beitragszahlungen zur Wiederauffüllung der Anwartschaften. Da aber datenmäßig kein Zusammenhang zwischen den geschiedenen Partnern besteht und jegliche Informationen über den Versicherungsverlauf und die insgesamt vorhandenen Rentenanwartschaften fehlen, sind die Nutzungsmöglichkeiten stark eingeschränkt.

## 2.2 Rehabilitationsstatistiken

Neben der summarisch abgewickelten Reha-Antrags- und Erledigungsstatistik sowie der Reha-Antragslaufzeiten-Statistik ist hier insbesondere die so genannte Reha-Statistik-Datenbasis (RSD) zu nennen. Diese wird als Jahrerhebung zum 15. April des Folgejahres und als Verlaufserhebung zum 15. November des Folgejahres gemeldet. Etwa 3 Monate später stehen üblicherweise Ergebnisse zur Verfügung.

Bei der Jahrerhebung der Reha-Statistik-Datenbasis (**vgl. Abbildung 5**) liegen dem VDR Individualdatensätze mit einem eingeschränkten Merkmalspektrum von sämtlichen Versicherungsträgern vor. Diese beschreiben die Struktur der abgeschlossenen Leistungen zur Teilhabe am Arbeitsleben (medizinische und berufliche Rehabilitation) im Berichtsjahr.

Im Rahmen des so genannten Reha-Kurzberichtes, zu dem von der BfA keine Individualdaten vorliegen, werden auch Anträge und Bewilligungen einbezogen. Wichtige Merkmale der Jahrerhebung sind das Geburtsjahr, das Geschlecht, der Wohn- und der Maßnahmeort (jeweils nur

**Abbildung 5:** Reha-Statistik-Datenbasis (RSD) Jahreserhebung

Termine	Jährlich zum 15.04. des Folgejahres
Datenform	Einzeldatensätze, von BfA zum Teil nur Summensätze
Gegenstand	1. Reha-Kurzbericht mit groben Strukturangaben zu Anträgen, Bewilligungen (und durchgeführten Leistungen im Berichtsjahr) 2. Abgeschlossene Leistungen zur Teilhabe (medizinische und berufliche Rehabilitation) im Berichtsjahr
Quelle	Reha-Antragsverfahren / Reha-Entlassungsberichte
Wichtige Merkmale	Geburtsjahr, Geschlecht, Familienstand, Wohn- und Maßnahmeort (Kreis), Antragsart, Maßnahmeart, Verfahrensart, Art der Durchführung, Bewilligte Tage und Dauer, Diagnosen und Ursache der Erkrankung, Beruf und Stellung im Beruf, Arbeit vor Antragstellung, Entlassungsform und Leistungsfähigkeit
Wozu?	Ursachen und Struktur der Leistungen zur medizinischen Rehabilitation und sonstigen Leistungen zur Teilhabe sowie der betroffenen Versicherten
Probleme	Keine einheitliche Datenstruktur (zwei verschiedene Einzeldatensatzformate und ein Summendatensatz)

Quelle: Eigene Zusammenstellung, vgl. hierzu auch VDR Statistik Rehabilitation: Vorwort zu Band 146.

der Kreis), die Antrags-, Maßnahme- und die Verfahrensart, die Entlassungsdiagnosen, der zuletzt ausgeübte Beruf sowie Angaben über Maßnahmebeginn und Maßnahmeende. Mit der Jahreserhebung der Reha-Statistik-Datenbasis lässt sich daher die Struktur der im jeweiligen Berichtsjahr durchgeführten Reha-Leistungen beschreiben.

Wesentlich mächtiger, aber auch dementsprechend schwieriger zu interpretieren, ist die Verlaufserhebung der Reha-Statistik-Datenbasis (**vgl. Abbildung 6**). Allerdings liegen dazu nicht von allen Versicherungsträgern Einzeldatensätze vor. Betrachtungsgegenstand der Verlaufserhebung sind die gestellten und die erledigten Anträge auf Leistungen zur Rehabilitation sowie vor allem die abgeschlossenen Reha-Leistungen innerhalb eines 8-Jahreszeitraumes.

Damit lassen sich sowohl die zeitliche Entwicklung auf diesem Sektor gut beobachten als auch die mehrfache Inanspruchnahme von Reha-Leistungen durch Versicherte analysieren bzw. die aus mehreren Teilleistungen bestehenden Reha-Verfahren auswerten. Darüber hinaus sind weitere Verlaufsdaten enthalten, nämlich Informationen darüber, ob und mit welcher Leistungsart ein Rentenantrag gestellt bzw. bewilligt wurde, ob und in welcher Höhe eine Rente gezahlt wird usw. Ergänzt werden die Daten mit Informationen über den Versicherten für einen elf Jahre betragenden Zeitraum. Schließlich gibt es für jeden Rehabilitanden Merkmale zu Geburtsjahr, Geschlecht, Staatsangehörigkeit und Wohnort (Kreis). Die Verlaufsdaten helfen insbesondere die Inanspruchnahme und den „Erfolg“ von Leistungen zur Rehabilitation und das Rehabilitationsgeschehen vor einer Rentengewährung zu analysieren. Die Fülle der Information bringt es jedoch mit sich, dass die Daten komplex strukturiert sind und – da sie einen 8- bzw. 11-Jahreszeitraum beschreiben – häufig von Änderungen in der Verschlüsselungsweise der zugrunde liegenden Verwaltungsdaten betroffen sind. Soweit dem VDR Individualdaten mit Verlaufscharakter vorliegen ist zu beachten, dass diese einer strengen Zweckbindung unterlie-

**Abbildung 6:** Reha-Statistik-Datenbasis (RSD) Verlaufserhebung

Termine	Jährlich zum 15.11. des Folgejahres
Datenform	Einzeldatensätze, von BfA zum Teil nur Summensätze
Gegenstand	<ol style="list-style-type: none"> <li>1. Gestellte und erledigte Reha-Anträge sowie abgeschlossene Leistungen zur Teilhabe (medizinische und berufliche Rehabilitation) im 8-Jahreszeitraum (Berichtsjahr und 7 Jahre davor)</li> <li>2. Reha-Verlauf einschl. Beiträge und Rente im 8-Jahreszeitraum (Berichtsjahr und 7 Jahre davor)</li> </ol>
Quelle	Versicherungskonto + Reha-/Rentenverfahren
Wichtige Merkmale	<ol style="list-style-type: none"> <li>1. Demographische Merkmale: Geburtsjahr, Geschlecht, Familienstand, Wohnort (Kreis)</li> <li>2. Reha-Daten im 8-Jahresfenster: Antragsart, Maßnahmeart, Verfahrensart, Art der Durchführung, Maßnahmeort (Kreis), Antrags- und Bewilligungsdatum, ggf. Ablehnungsgründe, bewilligte Tage und Dauer, Diagnosen und Ursache der Erkrankung, Beruf und Stellung im Beruf, Arbeit vor Antragstellung, Entlassungsform, Leistungsfähigkeit, positives und negatives Leistungsbild</li> <li>3. Rentendaten im 8-Jahresfenster: Antrags und Bescheidsdatum, Rentenleistungsart, Beginn- und Ende der Leistung, Rentenbetrag</li> <li>4. Versicherungsdaten im 11-Jahresfenster: Beitragsart, Dauer und Höhe des/der versicherten Entgelts / Lohnersatzleistung</li> </ol>
Wozu?	Entwicklung der Rehabilitation im 8-Jahreszeitraum, Inanspruchnahme und „Erfolg“ von Rehabilitation
Probleme	Keine einheitliche Datenstruktur (zwei verschiedene Einzeldatensatzformate und ein Summendatensatz), komplexe Datenstruktur

Quelle: Eigene Zusammenstellung.

gen und die Frage der Weitergabemöglichkeit im Rahmen des Forschungsdatenzentrums daher noch offen ist. Der Verlaufsscharakter des Datenmaterials stellt ein schwieriges Problem für eine faktische Anonymisierung dar.

### 2.3 Rentenstatistiken

Die als interne Arbeitsstatistik anzusehende Rentenantrags- und Erledigungsstatistik wird lediglich in summarischer Form durchgeführt. Gleiches gilt für die wenige Merkmale umfassende Rentennachbehandlungsstatistik. Drei weitere Statistiken im Rentenbereich dürften jedoch, wie die Umfrage unter potenziellen Datennutzern des FDZ-RV<sup>4</sup> gezeigt hat, von besonderem Interesse sein, da sie die Hauptaufgabe der gesetzlichen Rentenversicherung beschreiben.

Die Rentenzugangs-, Rentenwegfalls- und Rentenänderungsstatistik (**vgl. Abbildung 7**) wird zum 10. Februar des auf das Berichtsjahr folgenden Jahres erstellt. Ergebnisse liegen in der Re-

<sup>4</sup> Vgl. hierzu den Beitrag von Ralf Himmelreicher im vorliegenden Band.



**Abbildung 7:** Rentenzugangs-, Rentenwegfalls- und Rentenänderungsstatistik

Termine	Jährlich zum 10.02. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	1. Neu zugewandene Renten 2. Weggefallene Renten 3. Änderungen von Rentenleistungsart oder Teilrentenanteil
Quelle	Rentantrags-/ Rentenzahlverfahren
Wichtige Merkmale	Geburtsjahr, Geschlecht, Familienstand, Wohnort (Kreis), Diagnosen, Beruf, Rentenleistungsart, Rentenbeginn, Rentenbetrag und Zusammensetzung, Auffüllbetrag/Rentenzuschlag, Kranken-/ Pflegeversicherung (Art und Beitrag), Einkommensanrechnung, Zusammentreffen mit Einkommen, Relevanz bestimmter rentenrechtlicher Vorschriften, Anwendung von Fremdrechten und Vertragsrentenrecht, Relevanz der Arbeitsmarktlage, Versicherungsstatus und Jahresentgelt vor Leistungsfall, Werte aus der Gesamtleistungsbewertung, Werte aus der Rentenberechnung (Zeitdauer und Entgeltpunkte) Bezugsdauer/Wegfallsgrund beim Rentenwegfall, alte/neue Leistungsart bzw. Teilrentenanteil bei Rentenänderung
Wozu?	Struktur (Anzahlen, Alter, Beträge, Zeiten, Berechnungswerte, usw.) von „neuen“ Renten, Rentenwegfällen, Rentenänderungen
Probleme	Keine Rentenkumulation, kein Haushaltsbezug, keine Informationen zu anderen Alterssicherungseinkünften

Quelle: Eigene Zusammenstellung, vgl. hierzu VDR Statistik Rentenzugang: Vorwort zu Band 149.

gel etwa 2 Monate später vor. Grundsätzlich gliedert sich diese Statistik in drei Teilbereiche, nämlich die

- Rentenzugangsstatistik, in der alle im Laufe eines Berichtsjahres neu bewilligten (zugewandenen) Renten erfasst sind,
- Rentenwegfallsstatistik, in der alle im Laufe eines Berichtsjahres endgültig weggefallenen Renten enthalten sind, sowie
- Rentenänderungsstatistik, in der alle Fälle dokumentiert werden, bei denen sich im Berichtsjahr die Rentenleistungsart oder der Teilrentenanteil geändert hat oder die von einer Teilrente in eine Vollrente oder umgekehrt umgestellt wurden.

Die Daten entstammen dem Rentenbewilligungs- bzw. den Rentenzahlverfahren. Neben dem Geburtsjahr, dem Geschlecht, dem Wohnort (Kreis), der Rentenleistungsart, den Diagnosen (bei Renten wegen verminderter Erwerbsfähigkeit), dem Beruf sowie dem Rentenbetrag und dessen Zusammensetzung enthält diese Statistik eine Reihe von Merkmalen, mit denen die Auswirkungen einzelner rentenrechtlicher Vorschriften bzw. der Umfang bestimmter rentenrechtlicher Zeiten statistisch dokumentiert werden. Beim Rentenwegfall kommen etwa noch die Rentenbezugsdauer und der Wegfallsgrund hinzu. Im Falle von Rentenänderungen ist insbesondere ausgewiesen von welcher Leistungsart in welche neue Leistungsart die Umstellung erfolgt ist, bzw. welcher Teilrentenanteil vor bzw. nach der Änderung vorgelegen hat. Bei der Interpretation dieser Statistik kann problematisch sein, dass weder Informationen über eine Kumulation meh-

rerer Renten – insbesondere Witwenrente mit einer eigenen Altersrente – noch Angaben über andere Alterseinkünfte vorliegen. Auch lassen sich beispielsweise Ehepartner im Sinne eines Haushaltsbezuges nicht zusammenführen.

Die Rentenbestandsstatistik aus Konten der Rentenversicherungsträger (**vgl. Abbildung 8**) ist im Prinzip ein Spiegelbild der Rentenzugangstatistik, sie bezieht sich jedoch auf sämtliche laufenden Rentenzahlungen am 31. Dezember eines Berichtsjahres und wird ebenfalls zum 10. Februar des Folgejahres erstellt. Die Merkmale lehnen sich grundsätzlich an die Rentenzugangstatistik an – rein formal liegt sogar die gleiche Datenstruktur vor -, allerdings sind bestimmte Informationen, wie etwa die Diagnose oder der Beruf, bei Rentenbestandsfällen nicht mehr verfügbar.

**Abbildung 8:** Rentenbestandsstatistik aus den Konten der Rentenversicherungsträger

Termine	Jährlich zum 10.02. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	Laufende Rentenzahlungen am 31.12. des Berichtsjahres
Quelle	Rentenbestand / Versicherungskonten
Wichtige Merkmale	Geburtsjahr, Geschlecht, Familienstand, Wohnort (Kreis), Rentenleistungsart, Rentenbeginn, Rentenbetrag und Zusammensetzung, Auffüllbetrag/Rentenzuschlag, Kranken-/ Pflegeversicherung (Art und Beitrag), Einkommensanrechnung, Zusammentreffen mit Einkommen, Relevanz bestimmter rentenrechtlicher Vorschriften, Anwendung von Fremdrentenrecht und Vertragsrentenrecht, Relevanz der Arbeitsmarktlage, Werte aus der Gesamtleistungsbewertung, Werte aus der Rentenberechnung (Zeitdauern und Entgeltpunkte) bei Renten nach SGB VI
Wozu?	Struktur (Anzahlen, Alter, Beträge, Zeiten, Berechnungswerte, usw.) von laufenden Rentenzahlungen
Probleme	Keine Rentenkumulation, kein Haushaltsbezug, keine Informationen zu anderen Alterssicherungseinkünften, unterschiedliche Berechnungsarten (umgewertete RVO/AVG/RKG-Renten, umgewertete Renten aus dem Beitrittsgebiet)

Quelle: Eigene Zusammenstellung, vgl. hierzu VDR Statistik Rentenbestand: Vorwort zu Band 148 und Anhang 2.

Mit der Rentenbestandsstatistik aus den Konten der Rentenversicherungsträger lässt sich die Struktur der laufenden Rentenzahlungen beschreiben, also vor allem Mengengerüste, Rentenzahlungsbeträge einschließlich Teilkomponenten, Auftreten und Umfang verschiedener rentenrechtlicher Zeiten und die wertmäßigen Auswirkungen bestimmter Anrechnungsvorschriften. Als Mangel dieser Rentenbestandsstatistik ist festzuhalten, dass nur einzelne Renten ausgewiesen sind, nicht aber das Zusammentreffen mehrerer Rentenleistungen bei einer Rentenbezieherin bzw. einem Rentner. Wenn schon die Rentenkumulation bei einer Person nicht nachvollziehbar ist, so gilt dies natürlich auch für Ehepaare, d.h. die laufenden Rentenzahlungen können nicht auf Haushaltsebene aggregiert werden. Ob und in welcher Höhe Einkünfte aus anderen Alterssicherungssystemen oder u.U. aus eigener Erwerbstätigkeit vorliegen, lässt sich aus dieser Statistik nicht erkennen.

In welchem Umfang Renten aus der gesetzlichen Rentenversicherung kumulativ ausgezahlt werden, zeigt die Rentenzahlbestandsstatistik (**vgl. Abbildung 9**), die aus den laufenden Rentenzahlungen vom Renten Service der Deutschen Post und der Bundesknappschaft jeweils zum 1. Juli eines jeden Kalenderjahres gewonnen wird. D.h., in dieser Statistik können nicht nur einzelne Renten isoliert betrachtet werden, sondern auch Rentnerinnen und Rentner mit der Gesamtheit ihrer laufenden Rentenleistungen aus der gesetzlichen Rentenversicherung.

**Abbildung 9:** Rentenzahlbestandsstatistik (Post/Bundesknappschaft)

Termine	Jährlich zum 1.07. des Folgejahres
Datenform	Einzeldatensätze
Gegenstand	Laufende Rentenzahlungen am 01.07. des Berichtsjahres
Quelle	Postzahlbestand / Knappschaftszahlbestand
Wichtige Merkmale	Geburtsjahr, Geschlecht, Wohnort (Kreis), Rentenleistungsart, Rentenbeginn, Rentenbetrag und Zusammensetzung, Auffüllbetrag/Rentenzuschlag, Kranken-/ Pflegeversicherung (Art und Beitrag), Erstattungsbeträge
Wozu?	Struktur (Anzahlen, Alter, Beträge) von laufenden Rentenzahlungen, insbesondere Mehrfachrentenbezug
Probleme	Keine Versicherungszeiten, kein Haushaltsbezug, keine Information zu anderen Alterssicherungseinkünften

Quelle: Eigene Zusammenstellung, vgl. hierzu VDR Statistik Rentenbestand: Vorwort zu Band 148, insbesondere den Abschnitt zu „Rentenkumulationen“.

Als wichtige Merkmale liegen wiederum das Geburtsjahr, das Geschlecht, der Wohnort (Kreis), die Rentenleistungsart, der Rentenbeginn, sowie der Rentenbetrag und seine Zusammensetzung vor. Außerdem finden sich in dieser Statistik diverse Erstattungsbeträge, die der Rentenversicherung von anderen Stellen ersetzt werden. Problematisch für die Interpretation ist allerdings, dass in den Daten keine Versicherungszeiten enthalten sind, also keine Aussagen darüber gemacht werden können, in welchem zeitlichen Umfang die Versicherten Beiträge gezahlt haben. Auch die Wirkungen einzelner Berechnungsvorschriften lassen sich in der Rentenzahlbestandsstatistik nicht mehr nachvollziehen, weil entsprechende Informationen schon in den Ausgangsdaten nicht vorhanden sind. Die Statistik ist somit insbesondere dafür geeignet, die Struktur und den Umfang von Mehrfachrentenbezug auszuweisen. Darüber hinausgehende Fragestellungen, wie solche zum Haushaltskontext oder über Leistungen aus anderen Systemen der Alterssicherung, können mit dieser Statistik nicht analysiert werden.

## 2.4 Finanzstatistiken

Im Bereich der Finanzstatistiken existieren mehrere Berichterstattungen, von denen als wichtigste die monatlichen und jährlichen Rechnungsergebnisse zu nennen sind. Da jedoch keine Individualdaten vorliegen, wird hier auf diese Statistiken nicht näher eingegangen. Im Übrigen wird darauf hingewiesen, dass das Bundesministerium für Arbeit bzw. Gesundheit und Soziale Sicherung in seiner Reihe „Die gesetzliche Rentenversicherung in der Bundesrepublik Deutsch-

land im Jahre .... – statistischer und finanzieller Bericht“ detaillierte Rechnungsergebnisse der Rentenversicherungsträger veröffentlicht (vgl. BMA 1997).

## 2.5 Sonstige Statistiken

Anders als die Rechtsbehelfsstatistik und die Statistik über die Versendung von Versicherungsverläufen, Renteninformationen und Kontenklärungen beruht die jeweils zum 30. Juni eines Berichtsjahres erhobene Personalstatistik prinzipiell auf Einzeldatensätzen. Die Personalstatistik bezieht sich auf die bei der Rentenversicherung beschäftigten Personen. Allerdings gilt auch hier, dass nicht von allen Versicherungsträgern Einzeldatensätze vorliegen, sondern teilweise lediglich Summendaten geliefert werden. Insofern kann auf Basis der Personalstatistik keine Mikroanalyse durchgeführt werden. Ferner verfügt die amtliche Statistik mit der Personalstatistik nach § 7 Finanz- und Personalstatistikgesetz über eine Erhebung, in der vergleichbare Daten enthalten sind. Daher eignen sich weder die sonstigen Statistiken im Allgemeinen noch die Personalstatistik im Besonderen als Datensätze für das FDZ-RV.

## 2.6 Sondererhebungen

Zu nennen sind hier insbesondere die Sondererhebung „Vollendete Versichertenleben“ aus dem Rentenzugang, die zuletzt für das Berichtsjahr 2001 durchgeführt wurde, sowie die Sondererhebung „Altersvorsorge in Deutschland“, die derzeit für das Jahr 2002 aufbereitet wird.

Bei der Sondererhebung „Vollendete Versichertenleben“ werden die Daten aus dem Rentenzugang mit einem Versicherungsverlauf entsprechend der Versicherungskontenstichprobe kombiniert (vgl. **Abbildung 10**). D.h., es stehen zusätzlich zu den konkreten Rentenzugangsdaten die chronologischen Verlaufsdaten der rentenrechtlich relevanten Zeiten zur Verfügung.

**Abbildung 10:** Sondererhebung „Vollendete Versichertenleben“ aus dem Rentenzugang

Termine	Bei Bedarf einmal im Jahr, zuletzt 2001
Datenform	Einzeldatensätze wie beim Rentenzugang und zusätzlich wie bei der Versicherungskontenstichprobe
Gegenstand	Versicherungsverläufe zu aktuellen Versichertenrentenzugängen (als Stichprobe)
Quelle	Versicherungskonten
Wichtige Merkmale	Vgl. Rentenzugang und Versicherungskontenstichprobe
Wozu?	Modellrechnungen bei geplanten Rentenrechtsänderungen und für finanzielle Vorausschätzungen

Quelle: Eigene Zusammenstellung.

Diese Verbindung erlaubt es, die Einflüsse bestimmter Konstellationen im Versicherungsverlauf auf die spätere Rentenhöhe zu beurteilen. Allerdings werden solche Sondererhebungen nicht regelmäßig, sondern nur bei einem speziellen Bedarf der Rentenversicherung durchgeführt.

Bei der Sondererhebung „Altersvorsorge in Deutschland“ werden Daten aus dem chronologischen Versicherungsverlauf mit weiteren Daten aus einer Fragebogenerhebung in Verbindung gebracht, die u. a. auch andere Alterssicherungsformen betreffen (vgl. **Abbildung 11**).<sup>5</sup> Aufgrund der Datenschutzklausel im Befragungsbogen besteht derzeit keine Möglichkeit, die Mikrodaten der AVID der Wissenschaft zur Verfügung zu stellen.

**Abbildung 11:** Sondererhebung „Altersvorsorge in Deutschland“ AVID

Termine	Bei Bedarf, zuletzt 2002
Datenform	Einzeldatensätze wie bei der Versicherungskontenstichprobe plus Befragungsdaten
Gegenstand	Befragungsdaten zu Ausbildung, Berufsleben, beabsichtigtem Rentenzeitpunkt und anderen Alterssicherungsformen sowie Versicherungsverläufe von 40 - 59-Jährigen im Ehepartnerschaftskontext
Quelle	Versicherungskonten + Befragung
Wichtige Merkmale	Vgl. Rentenzugang und Versicherungskontenstichprobe
Wozu?	Modellrechnungen für künftige Rechtsänderungen und Einordnung der Rentenversicherung ins bestehende System der Alterssicherung

Quelle: Eigene Zusammenstellung, vgl. [www-altersvorsorge-in-deutschland.de](http://www-altersvorsorge-in-deutschland.de).

Der Vorteil gerade dieser Sondererhebung ist vor allem, dass sie zum Einen über die Alterssicherung im Bereich der gesetzlichen Rentenversicherung hinausgeht und Anwartschaften und Zugehörigkeiten bei anderen Systemen der Altersvorsorge beleuchten kann. Zum Anderen ist es möglich, die Erwerbsbiographien von Versicherten und ihren Ehepartnern (im Haushaltskontext) unter Berücksichtigung möglicher Lücken im Versicherungsverlauf zu beschreiben. Die Datenfülle und der Längsschnittcharakter der Daten würden aber erhebliche Anonymisierungsprobleme aufwerfen.

### 3. Datenmengen und Verfügbarkeit

Die Daten zu fast allen Statistiken werden jeweils mit gewissem Abstand zum Erhebungsstichtag oder Berichtszeitraum bei den Versicherungsträgern erhoben, nach einer maschinellen Plausibilitätsprüfung dem VDR übermittelt und dort nochmals intensiven Plausibilitätsprüfungen unterzogen und ggf. bereinigt. Anschließend nimmt der VDR umfangreiche Auswertungen vor und sichtet diese, um die Ergebnisse abzusichern. Danach erfolgen die vorgesehenen in-

<sup>5</sup> Für eine detaillierte Darstellung der AVID-Spezifika siehe den Beitrag „Erste Daten der Wissenschaft mit den Daten der GRV: Die Studie ‚Altersvorsorge in Deutschland‘ (AVID)“ von Thorsten Heien im vorliegenden Band. Für Interessenten an spezifischen Mikrodatenauswertungen der AVID zu Fragestellungen, die mit den mehreren tausend Tabellen in 7 Tabellenbänden nicht beantwortet werden können, gibt es nur folgenden Weg: Forschende können sich an das BMGS oder den VDR wenden und die Mikrodaten auswerten lassen (vgl. hierzu die Anmerkungen im Beitrag von Ralf Himmelreicher in diesem Band).

ternen und externen Berichtserstattungen und es werden teilweise tabellarisch Ergebnisse in gedruckter Form publiziert oder im Internet bereitgestellt. Grundsätzlich stehen die Datenbestände für einzelne Statistiken auch zu länger zurückliegenden Zeiträumen für Auswertungsvorhaben zur Verfügung. Ob und in welchem Umfang ältere Mikrodatensätze im Rahmen des FDZ-RV von Interesse sind, bleibt zu klären. Die vorhandenen Ressourcen sollen jedoch primär darauf verwandt werden, aktuelle Statistiken für das FDZ-RV aufzubereiten. Darüber hinaus könnten beim Rentenbezug und beim Rentenwegfall eine Aufbereitung der Daten aus einer langen Reihe von Berichtszeiträumen unter dem Gesichtspunkt des Zugangs bzw. Wegfalls von Renten für Jahrgangskohorten von besonderem Interesse sein. Es wird darauf aufmerksam gemacht, dass einzelne Datenbestände sehr umfangreich sind (**vgl. Abbildung 12**). Allerdings wird zur Anonymisierung der Daten durch Stichprobenziehung im Rahmen des FDZ-RV eine erhebliche Reduzierung des Datenumfangs erreicht werden.

**Abbildung 12:** Datenmengen und Verfügbarkeit

Bezeichnung der Statistik	Verfügbar seit	Anzahl der Datensätze*)	Anzahl der Merkmale *)
Stammsatzbestandsstatistik	1978	108 Mio.	18
Statistik der aktiv Versicherten	1982	42 Mio.	ca. 140
Versicherungskontenstichprobe	1992	0,5 Mio.	ca. 170 + 19*X
Versorgungsausgleichsstatistik	1981	4,6 Mio.	ca. 50
RSD Jahreserhebung (abgeschlossene Reha-Leistungen)	1971	1 Mio.	ca. 60
RSD-Verlaufserhebung	-	11 Mio.	ca. 22+140*X
Rentenzugangsstatik	1973	1,3 Mio.	ca. 210
Rentenwegfallsstatistik	1973	1,2 Mio.	ca. 150
Rentenänderungsstatistik (Rentenumwandlung)	1973	0,2 Mio.	ca. 210 + 150
Rentenbestandsstatistik aus den Konten der Rentenversicherungsträger	1992	24 Mio.	ca. 150
Rentenzahlbestandsstatistik Post/Bundesknappschaft	1981	24 Mio.	ca. 75

\*) bei der aktuellen Erhebung.

Quelle: Eigene Zusammenstellung.

Die Mikrodatensätze der Rentenversicherung umfassen im Querschnitt teilweise mehr als hundert Merkmale, im Fall von Längsschnittinformationen wird ein Vielfaches an Merkmalen erreicht. Nähere Einzelheiten dazu sind in den Dokumentation der jeweiligen Statistik abgedruckt. Dazu zählen zunächst die Benutzerhinweise, die es zu den meisten Statistiken gibt. Sie enthalten neben Angaben zu den zugrundeliegenden Datenquellen sowie zum Erhebungsdesign auch weitere mögliche Besonderheiten bei der Datengewinnung und Angaben zum Meldeweg.

Die jeweiligen Merkmale mit ihren Ausprägungen sind in so genannten Datensatzbeschreibungen dokumentiert. Diese sind in ihrer derzeitigen Form für die interne Nutzung vorgesehen, sie erfordern deshalb vielfach Kenntnisse im Rentenrecht sowie über den Verfahrensablauf. Umfangreiche Verschlüsselungen, wie z.B. für Diagnosen oder Berufe ergeben sich mitunter aus separaten Schlüsselverzeichnissen. Merkmalszusammenhänge erschließen sich teilweise auch aus den Fehlerverzeichnissen, die für die maschinelle Plausibilitätsprüfung verwendet werden. Den Interessenten der verschiedenen Statistiken wird empfohlen, sich die Vorworte zu den publizierten Statistikbänden der Reihe VDR-Statistik sowie die Anhänge dieser Publikationen zu vergegenwärtigen. Diese Texte enthalten wichtige Hinweise zum Erhebungsablauf, zu den Besonderheiten in bestimmten Berichtsjahren und den verwendeten Fachbegriffen.

#### **4. Pilotprojekt Versicherungskontenstichprobe 2001**

Der VDR hat im Jahr 2003 eine projektbezogene Anonymisierung von ausgewählten Mikrodaten aus der Versicherungskontenstichprobe des Jahres 2001 vorgenommen. Der dazu entwickelte Mikrodatsatz war ein Pilotprojekt, anhand dessen der Zugang zu anonymisierten Mikrodaten der Rentenversicherung erprobt wurde. Dieser Datensatz wird im Folgenden als „Pilotprojekt Versichertenkontenstichprobe 2001“ bezeichnet.<sup>6</sup> Neben demographischen Merkmalen der Versicherten enthält dieser Datensatz Verlaufsinformationen, speziell zu Beitragsarten und Beitragshöhen mit Zeitpunkten und Zeitdauern, die erreichten Entgeltpositionen sowie ggf. die Rentenleistungsart und das exakte Alter bei Rentenbeginn. Weil die Determinanten des Renteneintritts im Vordergrund der Analyse stehen, konnte die Fallauswahl auf Versicherte im Alter von 55 bis 65 Jahren beschränkt werden. Bei der Knappschaft Versicherte und Ausländer wurden ausgeschlossen. Außerdem wurden nur „geklärte“ Versicherungskonten in das Pilotprojekt einbezogen, weil ungeklärte Versicherungskonten keine zuverlässigen Rückschlüsse erlaubt hätten. Im Sinne einer Anonymisierung sind Versicherte mit selten in Anspruch genommene Leistungsarten aus dem Datenbestand selektiert worden. Nach mehreren Abstimmungsgesprächen mit den Wissenschaftlern wurden ausschließlich konventionelle Anonymisierungsmethoden eingesetzt. Dazu wurden nicht benötigte Merkmale, die zur Deanonymisierung hätten dienen können, wie etwa der Tätigkeitsschlüssel und sensible Merkmale, generell gelöscht. Metrische Merkmale wurden generell vergrößert. Bei nicht metrischen Merkmalen wurden Ausprägungen soweit als möglich zu sinnvollen Gruppen zusammengefasst, wie z.B. Rentenleistungsarten, die zu Rentenleistungsartgruppen verdichtet wurden.

Besonderes Augenmerk galt der Anonymisierung der Daten im chronologischen Teil der Versicherungskontenstichprobe, zumal die chronologische Abfolge erkennbar bleiben sollte. Zunächst wurde das Berichtsfenster, das sich prinzipiell auf das gesamte Versicherungsleben erstreckte, auf ein 10-Jahresfenster verkleinert und somit auf das Berichtsjahr und die neun davor liegenden Kalenderjahre reduziert. Die einzelnen Beitragsarten wurden wiederum zu Beitragsartgruppen zusammengefasst. Für die Untersuchung nicht benötigte Sachverhalte, wie z.B. Sachbezug, Zeiten der Kindererziehung oder Zeiten nach dem Fremdrentengesetz sowie konkrete Entgeltangaben wurden gelöscht. Sämtliche Datumsangaben wurden vergrößert und schließlich wurden alle angegebenen Entgeltpunkte gerundet.

---

<sup>6</sup> Der Datensatz „Pilotprojekt Versichertenkontenstichprobe 2001“ ist die empirische Basis der im vorliegenden Band abgedruckten Studie „Determinanten des Renteneintritts - eine Untersuchung mit prozessproduzierten Daten“ von Stefan Kempf und Martin Kukuk.

In Sachen Pilotprojekt Versicherungskontenstichprobe 2001 kann festgehalten werden, dass der betriebene Aufwand zur Aufbereitung und Anonymisierung des Originaldatensatzes erheblich war. Bedenkt man ferner, dass sich mit dem Datensatz dieses Pilotprojektes nur noch vergleichsweise wenige Fragestellungen analysieren lassen, dann erhält man eine grobe Vorstellung davon, in welchem Umfang in Zukunft vor allem personelle Ressourcen in das FDZ-RV eingebracht werden müssen.

## 5. Fazit

Das Angebot an überwiegend prozessproduzierten Mikrodaten der Rentenversicherung ist sehr groß. Sie sind dafür geeignet, auf zahlreiche wissenschaftliche Fragestellungen verschiedener Disziplinen empirisch gestützte Antworten zu finden.<sup>7</sup> Da das Gros der Daten vereinfacht gesprochen dafür erhoben wurde, die Höhe der jeweiligen Leistung für die Versicherten zu ermitteln, entsprechen sie nicht grundsätzlich wissenschaftlichen Anforderungen. Anstelle tiefgegliederter Differenzierung, wie sie häufig in wissenschaftsbasierten Umfragen möglich sind, haben die Daten der Rentenversicherung in der Regel die Eigenschaft, dass sie im Rahmen von Totalerhebungen meist sehr exakt erhoben wurden. D.h., wichtiges Kennzeichen der Daten der Rentenversicherung sind hohe Fallzahlen und eine hohe Validität.

Anhand des Datensatzes „Pilotprojekt Versichertenkontenstichprobe 2001“ konnte exemplarisch gezeigt werden, dass es möglich ist, selbst einen Längsschnittdatensatz der Rentenversicherung für die Wissenschaft aufzubereiten. Und zwar aufzubereiten in einer Weise, die sowohl den strengen Kriterien des Datenschutzes als auch den Anforderungen der Wissenschaftler genügt. Wenngleich dieses Beispiel lehrt, dass die zukünftige Zusammenarbeit zwischen Wissenschaft und FDZ-RV keinen finalen sondern eher einen prozesshaften Charakter haben wird, weil jede neue Erkenntnis gleichsam neue Fragen aufwerfen wird.

Die Vorgehensweise beim „Pilotprojekt Versichertenkontenstichprobe 2001“ kann sicherlich ein Modell für Gastwissenschaftler-Arbeitsplätze im FDZ-RV sein. Um im Rahmen des FDZ-RV Scientific Use Files und Public Use Files anbieten zu können, sind jedoch solche Lösungen anzustreben, die für die Aufbereitung und die Anonymisierung der Daten eine wesentlich bessere Aufwand-Nutzen Relation aufweisen.

---

<sup>7</sup> An den Daten der Rentenversicherung sind nach den Befunden der FDZ-RV-Umfrage derzeit 40 Forscher(gruppen) aus 11 verschiedenen wissenschaftlichen Disziplinen, vor allem aus Ökonomie, Demographie, Soziologie und Medizin, interessiert (vgl. hierzu den Beitrag von Ralf Himmelreicher im vorliegenden Band).



## Literatur:

- Bundesministerium für Arbeit und Sozialordnung (1997): Die gesetzliche Rentenversicherung in der Bundesrepublik Deutschland im Jahre 1996 - Statistischer und finanzieller Bericht. Bonn.
- Rehfeld, U. (2001): Die Statistiken der gesetzlichen Rentenversicherung. *Deutsche Rentenversicherung*, 3-4, 169-188.
- Verband Deutscher Rentenversicherungsträger (Hg.) (2004): *VDR Statistik Rehabilitation des Jahres 2002*, Band 146. Frankfurt am Main.
- Verband Deutscher Rentenversicherungsträger (Hg.) (2004): *VDR Statistik Rentenbestand am 31. Dezember 2003*, Band 148. Frankfurt am Main.
- Verband Deutscher Rentenversicherungsträger (Hg.) (2004): *VDR Statistik Rentenzugang des Jahre 2003 einschließlich Rentenwegfall, Rentenänderung/Änderung des Teilrentenanteil*, Band 148. Frankfurt am Main.
- Verband Deutscher Rentenversicherungsträger (Hg.) (2004): *VDR Statistik Versicherte 2001/2002*, Band 147. Frankfurt am Main.

**Hilmar Luckert** hat an der Julius Maximilians Universität Würzburg Mathematik studiert und sich dabei schwerpunktmäßig mit mathematischer Statistik befasst. Nach Abschluss des Studiums kam er 1981 zum VDR und war zunächst als Referent für verschiedene Statistiken zuständig. Nachdem 1987 die Abteilung „Volkswirtschaft und Statistik“ in Frankfurt eingerichtet wurde blieb Hilmar Luckert mit seinem Referat in Würzburg bis heute zuständig für die technische Durchführung sämtlicher Routinestatistiken und Berichterstattungen. Inhaltlich betreute er selbst lange Jahre die Versicherungskontenstichprobe und Sondererhebungen, entwickelte und gestaltet aber auch alle anderen Statistiken mit. Als Leiter der Abteilung „Berufliche Bildung, Statistisches Berichtswesen, Datenschutz und Datensicherheit“ sind inzwischen zum ursprünglichen Arbeitsschwerpunkt Statistik weitere hinzugekommen. Der Aufbau des Forschungsdatenzentrums der Rentenversicherung, dabei insbesondere die programm- und datentechnische Konzeption und Realisierung des Vorhabens, sieht er in diesem und im nächsten Jahr als eine seiner Kernaufgaben an.