

**SHARE-RV: Analysen zu Alterssicherung,
Gesundheit und Familie auf Basis von Daten der
Deutschen Rentenversicherung und des Survey of
Health, Ageing and Retirement in Europe**

Endbericht

**Im Namen des Antragstellers
Professor Axel Börsch-Supan, Ph.D.
28.09.2016**

Inhaltsverzeichnis

Zusammenfassung.....	3
1. Einleitung.....	5
2. Datenerhebung.....	8
2.1 Feldphasen des Survey of Health, Ageing and Retirement in Europe.....	8
2.2 Dokumentation der Einwilligungsbögen	9
3. Datenziehung und Verknüpfung	9
3.1 Ermittlung und Prüfung der Sozialversicherungsnummern.....	9
3.2 Ziehung und Lieferung der RV-Daten.....	10
3.3 „Kombidatensatz“	11
3.4 Datenaufbereitung.....	12
3.5 Verknüpfung der Daten am MEA	12
3.6 Dokumentation.....	13
4. Daten-Veröffentlichung.....	13
4.1 SHARE-RV Release 2-0-0.....	14
4.2 SHARE-RV Release 3-0-0.....	15
4.3 SHARE-RV Release 5-0-0.....	16
4.3.1 Geschlechter und Altersverteilung in Release 5-0-0.....	17
4.3.2 Erwerbsstatus in Release 5-0-0	18
5. Datennutzung.....	19
5.1 Maßnahmen zur Erhöhung der Nutzerfreundlichkeit.....	19
5.2 Nutzerstatistik	20
6. Ausblick.....	21
7. Literaturverzeichnis.....	23
8. Anhang.....	24

Zusammenfassung

SHARE-RV ist seit 2009 ein Kooperationsprojekt zwischen dem Munich Center for the Economics of Ageing (MEA) des Max-Planck-Instituts für Sozialrecht und Sozialpolitik und dem Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). Das Ziel von SHARE-RV ist zum einen die Verknüpfung der Befragungsdaten des Survey of Health, Ageing and Retirement in Europe (SHARE) mit ausgewählten administrativen Daten des FDZ-RVs derselben Person und zum anderen den daraus resultierenden verknüpften Datensatz der gesamten Forschungsgemeinschaft zur Verfügung zu stellen. Dieser überaus umfangreiche Datensatz ermöglicht die Erforschung komplexer Zusammenhänge zwischen verschiedenen Lebensaspekten der Befragten. Der Datensatz beinhaltet u. A. Informationen zur Erwerbshistorie und zum sozio-ökonomischen Hintergrund der Befragten, zu deren physischer und psychischer Gesundheit sowie zu den Einkünften und dem Vermögen aller Haushaltsmitglieder.

Die zweite Förderphase des Projektes SHARE-RV durch das Forschungsnetzwerk Alterssicherung begann am 01. September 2013 und war auf drei Jahre angelegt. Sie umfasste die Weiterführung des Projektes SHARE-RV in der fünften und sechsten Erhebungswelle des SHARE. Die fünfte Welle des SHARE war für SHARE-RV aus zwei Gründen von besonderer Bedeutung: Erstens hat die große Auffrischungstichprobe die Anzahl der verknüpfbaren Fälle und somit das Potential von SHARE-RV deutlich erhöht; zweitens konnten durch die erneute Frage nach Zustimmung derjenigen Panelfälle, die in vorangegangenen Wellen nicht zugestimmt hatten, auch die Verknüpfungsraten der vorherigen Wellen erhöht werden. Da die Förderphase zeitlich nicht synchron mit dem Erhebungszyklus der Befragung verlief, bildet dieser Bericht nicht die Erhebungsphase der fünften Welle ab. Alle notwendigen Vorbereitungen sowie die Feldphase waren im September 2013 bereits abgeschlossen. Die Verknüpfung, deren Validierung und Dokumentation sowie die Veröffentlichung der Daten wurden im Berichtszeitraum durchgeführt.

Da die Wiederbefragung der Panelstichprobe sehr erfolgreich war, wurde das Modul auch in der sechsten Welle des SHARE implementiert. Die Abläufe wurden im Pretest im Sommer 2014 erfolgreich getestet, sodass in der Haupterhebung (Februar-November 2015) wieder alle Panelfälle, für die noch keine Zustimmung vorlag, erneut nach ihrer Zustimmung gefragt wurden.

Die administrativen Daten der Rentenversicherung werden mittels der Sozialversicherungsnummer jährlich aus den Stammdatensätzen gezogen. Im Berichtszeitraum wurden die Daten im Format der Versichertenkontenstichprobe um die Berichtsjahre 2013 und 2014 erweitert, die Daten im Format des Rentenbestands wurden um das Berichtsjahr 2014 erweitert.

Zusätzlich zum regulären Ablauf des SHARE-RV Projekts konnten in der Förderphase drei Aktualisierungen von SHARE-RV veröffentlicht werden. Dazu zählt der Release 2-0-0 (März 2014), der sich auf die SHARE-Erhebungswellen 3 und 4 bezieht. Die Verknüpfungsraten betragen für beide Wellen jeweils knapp 50%. Im Release 3-0-0 (März 2015), welcher erstmals die neuen Zustimmungen aus Welle 5 enthielt, betragen die Verknüpfungsraten für die Wellen 3-5 zwischen 60% und 66%. Das bedeutet für Welle 3 eine Steigerung um 13 Prozentpunkte. Die Verknüpfungsraten des aktuellen Release 5-0-0 (Mai 2016) sind mit denen des letzten Release vergleichbar.

Wir danken dem Forschungsnetzwerk Alterssicherung für die Finanzierung des Projektes in der zurückliegenden Förderphase und freuen uns sehr über die Fortführung der Förderung. Für die kommende Förderphase streben wir weitere Verbesserungen des Datenangebotes – unter anderem durch Erhöhung der Fallzahlen sowie Maßnahmen im Bereich Nutzerfreundlichkeit – an.

1. Einleitung

Der demographische Wandel ist eine der Herausforderungen des 21. Jahrhunderts. Doch was bedeutet dies eigentlich und warum ist der demographische Wandel eine Herausforderung? Unter dem Schlagwort „demographischer Wandel“ wird im Allgemeinen die Alterung der Gesellschaft verstanden. Diese Alterung der Gesellschaft ist zum einen durch den Rückgang der Geburtenzahlen und zum anderen durch die höhere Lebenserwartung der Menschen bedingt (Schmidt 2013), welche wiederum auf verbesserte medizinische Versorgung zurückzuführen ist. Die Herausforderung dieser Entwicklung besteht vor allem darin, dass dies eine neue Entwicklung ist und Erfahrungswerte im Umgang mit dieser Situation fehlen. Die Menschen sind noch nie zuvor so alt geworden wie heute (Weiland et al. 2006) und in diesem Zusammenhang stellen sich viele Fragen: Wie geht es den Menschen im Alter? Wie unterscheiden sich die Menschen in ihrem Umgang mit dem Alter und Altern? Doch nicht nur ältere Menschen selbst, sondern die ganze Gesellschaft wird direkt oder indirekt von der sich verändernden Altersstruktur beeinflusst. So besteht durch das umlagefinanzierte Alterssicherungssystem ein starker Zusammenhang zwischen der Entwicklung des Arbeitsmarktes einerseits und der Rentenentwicklung andererseits. Dadurch werden Ungleichheiten auf dem Arbeitsmarkt zu Ungleichheiten im Alter (Trischler 2014), denn wer auf dem Arbeitsmarkt schlecht Fuß fassen kann, steht häufig auch im Alter vor finanziellen Problemen.

Das Alterssicherungssystem steht somit hauptsächlich durch zwei parallel verlaufende Tendenzen zunehmend vor Herausforderungen. Auf der einen Seite bedeutet ein Rückgang der Geburtenzahlen eine sinkende Anzahl derjenigen, die Beiträge zur Rentenversicherung zahlen, also geringere Einnahmen für die Rentenkassen. Auf der anderen Seite aber nimmt die Zahl der Rentenempfänger durch die höhere Lebenserwartung zu und auch die Rentenbezugsdauer verlängert sich. Um dennoch eine ausreichende Versorgung im Alter gewährleisten zu können, wurden verschiedene Reformen verabschiedet, die den negativen Effekt der beiden Tendenzen auf die Rentenhöhe abfedern sollen. In diesem Zusammenhang stellt sich immer wieder die Frage, wie erfolgreich diese Reformen sind und ob sie tatsächlich denjenigen Menschen zugutekommen, für die sie auch gedacht waren. Ein Beispiel dafür ist die „Rente mit 63“: Diese Altersrente soll besonders langjährig Versicherten, also Versicherten mit einer Wartezeit von 45 Jahren, einen abschlagsfreien Rentenzugang mit bereits 63 Jahren ermöglichen (Deutscher Bundestag, 2014). Untersuchungen hierzu konnten zeigen, dass die Anspruchsberechtigten zum einen am Ende ihres Erwerbslebens nicht häufiger krank sind und dass Anspruchsberechtigte zum anderen deutlich kürzere sozialversicherungspflichtige Erwerbszeiten aufweisen und außerdem besser verdienen (Börsch-Supan et al., 2014; Börsch-Supan et al., 2015).

Um etwa solche Zusammenhänge besser beleuchten zu können, gibt es in den Sozial- und Wirtschaftswissenschaften eine große Nachfrage nach Daten, die es ermöglichen, Fragestellungen zur Alterssicherung nachzugehen. Um die verschiedenen Facetten des Alterns untersuchen zu können, werden jedoch vielschichtige und präzise Informationen benötigt. Zur Verfügung stehende Datenarten sind einerseits administrative Daten und andererseits Umfragedaten. Beide Datenarten weisen sowohl Vor- als auch Nachteile auf. Die in administrativen Daten enthaltenen Informationen werden von Ämtern, Behörden oder Firmen als Nebenprodukt erzeugt – also nicht primär zu Forschungszwecken. Sie umfassen zum Beispiel Informationen zu Heirat, zu der Geburt von Kindern oder zum Renteneintritt, aber auch genaue Erwerbszeiten, die genaue Lohnhöhe, die Arbeitsstundenzahl oder die Entgeltpunkte. Die Vorteile von administrativen Daten liegen auf der Hand: Sie decken meist die gesamte Zielpopulation ab, sind sehr viel akkurater als Umfragedaten und können mit sehr geringem finanziellen und zeitlichen Aufwand bereitgestellt werden. Die Angaben in den Daten sind exakt und weisen meist nur einen geringen Anteil an fehlenden Angaben beziehungsweise Informationslücken auf und sind nicht durch Erinnerungsfehler beeinflusst (Caldewood und Lessof, 2009, Kröger et al. 2011). Ein entscheidender Nachteil dieser Datensätze ist jedoch, dass sie zumeist keine oder nur wenige Kontextvariablen enthalten, die zur Beantwortung bestimmter Forschungsfragen benötigt werden. Konkret bedeutet dies, dass zum Beispiel Angaben zur Bildung, zum Familienstand oder zur Haushaltszusammensetzung der Personen fehlen. Diese Angaben wiederum gehören zum Standardrepertoire einer sozialwissenschaftlichen Umfrage, bei der die Themen und Inhalte von Wissenschaftlern beeinflusst werden können. Jedoch ist man bei Umfragedaten erstens auf die Ehrlichkeit und Auskunftsbereitschaft der Befragten angewiesen und zweitens im Falle von länger zurückliegenden Begebenheiten auch auf ihr Erinnerungsvermögen. Die Kombination von Umfrage- mit administrativen Daten ist daher besonders reizvoll, da sie die jeweiligen Nachteile der beiden Datenquellen auszugleichen vermag.

Das Projekt SHARE-RV setzt hier an und stellt durch eben die Kombination von Umfrage- und administrativen Daten einen Datensatz bereit, der ein reichhaltiges Analysepotential im Bereich der Altersforschung bietet. Die Basis des Datensatzes bildet die deutsche Stichprobe des Survey of Health, Ageing and Retirement in Europe (SHARE), welche mit administrativen Daten der Deutschen Rentenversicherung verknüpft wird.

Der Survey of Health, Ageing and Retirement in Europe ist eine multidisziplinäre Längsschnittstudie zur Erforschung sozialer, wirtschaftlicher und gesundheitlicher Lagen alternder Menschen in Europa. Er wurde erstmals 2004 als repräsentative Befragung der Bevölkerung in der Altersgruppe 50plus in 11 europäischen Ländern erhoben. Mittlerweile stehen Daten aus 20 europäischen Ländern und Israel zur Verfügung. Dazu wurden rund

123.000 Menschen in mehr als 293.000 persönlichen Interviews befragt. Die aufbereiteten Daten und deren Dokumentation stehen weltweit allen Wissenschaftlerinnen und Wissenschaftlern kostenfrei zur Verfügung. SHARE ist in den letzten Jahren zu einem wichtigen Pfeiler des Europäischen Forschungsraums geworden. Im März 2011 wurde SHARE zum ersten European Research Infrastructure Consortium (SHARE-ERIC). Die Leitung von SHARE ist beim Munich Center for the Economics of Aging (MEA) einer der beiden Abteilungen des Max-Planck-Instituts für Sozialrecht und Sozialpolitik in München angesiedelt.

Vonseiten der Deutschen Rentenversicherung werden zwei Datensätze (Versicherungskontenstichprobe und Versichertenrentenbestand) für die Verknüpfung bereitgestellt, welche anhand der im Interview erfragten Sozialversicherungsnummern der Befragten eindeutig den Befragungsdaten zugeordnet werden können. Die Daten im Format der Versicherungskontenstichprobe geben über die erreichten Anwartschaften der Versicherten Auskunft. Zudem stehen Informationen zur gesamten Erwerbsbiografie und den aus den Beschäftigungen resultierenden Entgeltpunkten zur Verfügung. Die Daten aus dem Versichertenrentenbestand liefern wertvolle Zusatzinformationen, für Personen die bereits verrentet sind. Dazu gehören beispielsweise Informationen zu der Rentenart, dem Rentenbeginn oder der tatsächlich gezahlten Rentenhöhe. Der aus dieser Verknüpfung resultierende Datensatz bietet somit einerseits Informationen zu etwa dem sozialen Umfeld, dem gesundheitlichen Befinden und den demografischen Merkmalen der Befragten als auch andererseits präzise Informationen aus den Versicherungskonten bei der Deutschen Rentenversicherung, wie etwa zu Erwerbsbiographie, Kindererziehungszeiten und daraus resultierenden Entgeltpunkten. Damit stellt SHARE-RV eine wichtige Verbesserung der Dateninfrastruktur zur Erforschung von Alter und Alterung in Deutschland dar.

Mit den kombinierten SHARE-RV Daten ist es nunmehr möglich, Alterseinkommen auch im Haushaltskontext zu betrachten, was mit administrativen Daten allein nicht möglich wäre. So konnte auf Grundlage von SHARE-RV für Frauen in Westdeutschland ein deutlicher Unterschied in der Höhe der eigenständigen Alterssicherung mit steigender Kinderzahl beobachtet werden (Mika et al., 2016).

Das Kooperationsprojekt zwischen dem Munich Center for the Economics of Aging (MEA) und dem Forschungsdatenzentrum der Deutschen Rentenversicherung (FDZ-RV) Projekt begann im Jahr 2009 mit einer durch die Volkswagenstiftung finanzierten Pilotstudie in der dritten Erhebungswelle des SHARE. Durch die Fortführung des Projektes in zwei Phasen von 2012 bis 2016 durch das Forschungsnetzwerk Alterssicherung (FNA) wird der Datensatz kontinuierlich erweitert und verbessert.

Der vorliegende Bericht fasst den Verlauf des Projektes SHARE-RV für die zweite Förderphase durch das FNA vom 1. September 2013 bis zum 30. September 2016 zusammen. In Kapitel 2 wird zum einen die Implementierung des Projektes in den Wellen 5 und 6 des SHARE beschrieben und zum anderen die Dokumentation der Zustimmung der Befragten zur Datenverknüpfung dargestellt. In Kapitel 3 werden die Arbeitsschritte zur Datenverknüpfung näher erläutert und in Kapitel 4 werden die verschiedenen Versionen des Datensatzes vorgestellt und die Entwicklung der Fallzahlen thematisiert. Kapitel 5 widmet sich der Nutzerfreundlichkeit von SHARE-RV und neben einer Zusammenfassung zur Nutzerstatistik werden verschiedene Maßnahmen zur Erhöhung der Nutzerzahlen dargelegt. Im Schlusskapitel wird schließlich auf zukünftige Vorhaben im Projekt SHARE-RV eingegangen und eine Zukunftsperspektive entwickelt.

2. Datenerhebung

2.1 Feldphasen des Survey of Health, Ageing and Retirement in Europe

Die zurückliegende Förderphase des Projektes bezieht sich auf zwei Erhebungswellen des Survey of Health, Ageing and Retirement in Europe – Welle 5 und Welle 6. Die Feldphase von Welle 5 erfolgte von Februar 2013 bis September 2013. Die Zahl der Interviews wurde durch eine große Auffrischungstichprobe auf ca. 5.800 erhöht¹. Alle Befragten dieser Auffrischungstichprobe wurden nach der Zustimmung zur Datenverknüpfung gefragt. Zusätzlich wurden auch die Befragten der Panelstichprobe, die bisher noch nicht zugestimmt hatten oder deren Daten trotz Zustimmung nicht verknüpft werden konnten, erneut nach der Zustimmung zur Datenverknüpfung gefragt.

In Welle 5 sind insgesamt 3.131 gültige Zustimmungen eingegangen, die eindeutig einer/einem Befragten zugeordnet werden konnten. Für 3.006 Fälle konnten Daten gezogen und verknüpft werden. Der größte Anteil (89%) geht auf die Auffrischungstichprobe zurück. Die restlichen 11% (321 Fälle) stammen aus der Panelstichprobe, dies sind also alles Befragte, die in vorherigen Wellen die Zustimmung verweigert hatten.

Die Feldphase für Welle 6 startete zwei Jahre später im Februar 2015 und dauerte bis November 2015 an. Die Anzahl der Befragten, die um Zustimmung gefragt wurden, war im Vergleich zu Welle 5 deutlich geringer, da es keine Auffrischungstichprobe in Deutschland gab. Daher wurden in Welle 6 nur die bisherigen Verweigerer sowie Personen, deren Daten noch nicht verknüpft werden konnten, nach der Zustimmung gefragt. Es sind insgesamt 535 neue Einwilligungen aus der sechsten Welle eingegangen. Sobald die Datenaufbereitung der Befragungsdaten der sechsten Welle abgeschlossen ist, kann mit der Verknüpfung begonnen werden.

¹ SHARE Welle 5 Release 5.0.0

2.2 Dokumentation der Einwilligungsbögen

Voraussetzung für die Verknüpfung der Daten war die schriftliche Einwilligung der Befragten, das heißt, ein unterschriebener Einwilligungsbogen (siehe Anhang I) musste bei der Rentenversicherung vorliegen. Auf dem Einwilligungsbogen wurden neben der Sozialversicherungsnummer und der Unterschrift des Befragten auch alle personenbezogenen Daten abgefragt, die benötigt werden, um die Sozialversicherungsnummer zu prüfen und gegebenenfalls zu ermitteln. Die Zustimmung wurde während des Interviews vom Interviewer erfragt und zusätzlich im Interviewinstrument dokumentiert. Das Interviewinstrument wurde dafür so programmiert, dass die Fragen zur Zustimmung (siehe Anhang II) nur den ausgewählten Befragten gestellt wurden.

Die Einwilligungsbögen zur Datenverknüpfung wurden aus datenschutzrechtlichen Gründen vom Interviewer direkt an das Forschungsdatenzentrum der Deutschen Rentenversicherung gesandt. Um eine bessere Kontrolle des Ablaufs zu haben, wurden die Interviewer instruiert auch die Bögen der Verweigerer zurückzuschicken.

Bei der Deutschen Rentenversicherung wurden alle Einwilligungsbögen in der Reihenfolge ihres Eingangs mit einer laufenden Nummer versehen und die Angaben der Befragten in eine Datenbank eingegeben. Dabei wurde geprüft, ob für die jeweilige Person bereits ein Einwilligungsbogen vorlag, sodass der neue Einwilligungsbogen derselben Person zugeordnet werden konnte.

3. Datenziehung und Verknüpfung

3.1 Ermittlung und Prüfung der Sozialversicherungsnummern

Die Sozialversicherungsnummer bildet den Kern der Datenverknüpfung. Nur anhand einer gültigen Sozialversicherungsnummer können die Datensätze der SHARE-Befragten in den Daten der Rentenversicherung identifiziert und die korrekten Datensätze verknüpft werden. Daher war die Prüfung und Ermittlung der Sozialversicherungsnummer ein besonders wichtiger Arbeitsschritt. Die Befragten wurden gebeten ihre Sozialversicherungsnummer auf dem Einwilligungsbogen einzutragen. Daher wurden die Interviewer instruiert, die Angabe der persönlichen Daten zu priorisieren, falls die Befragten ihre Sozialversicherungsnummer nicht wissen. Damit sollte verhindert werden, dass Befragte ihre Einwilligung zurückhalten, weil sie ihre Sozialversicherungsnummer nicht wissen. Dementsprechend wage ist auch die Bitte nach der Angabe der Sozialversicherungsnummer formuliert. Als Konsequenz daraus gingen viele Einwilligungsbögen ein, bei denen die Sozialversicherungsnummer nicht angegeben wurde und ermittelt werden musste (siehe Tabelle 1). Die Erfahrungen der vergangenen Wellen haben gezeigt, dass die Ermittlung der Sozialversicherungsnummer sehr erfolgreich ist.

Tabelle 1: Angabe und Ermittlung der Sozialversicherungsnummern

Welle	Insgesamt eingegangenen Einwilligungsbögen	VSNR angegeben			VSNR nicht angegeben		
		N	Davon richtig	Davon falsch und ermittelt	N	Davon ermittelt	Davon nicht ermittelt
W5	3.251	1.960 (60%)	1.928 (98%)	32 (2%)	1.291 (40%)	1.195 (93%)	96 (7%)
W6	535	236 (44%)	209 (89%)	27 (11%)	299 (56%)	243 (81%)	56 (19%)

Alle angegebenen Sozialversicherungsnummern wurden stets auf Plausibilität hin überprüft. Dies geschah - falls auf dem Einwilligungsbogen eingetragen - anhand des Geschlechts, des vollständigen Geburtsdatums sowie des Geburtsnamens. Diese Angaben sind Bestandteil der Sozialversicherungsnummer. Stimmt die Angaben überein, waren keine weiteren Schritte nötig. In den Fällen, in denen die Sozialversicherungsnummer nicht auf dem Einwilligungsbogen angegeben wurde oder die Prüfung gezeigt hatte, dass sie fehlerhaft war, wurde versucht, sie zu ermitteln. So kam es zum Beispiel nicht selten vor, dass Personen versehentlich die Nummer des Partners/der Partnerin mitteilten, was mit Hilfe des Prüfverfahrens umgehend erkannt wurde. Die Ermittlung der Sozialversicherungsnummern erfolgte anhand der persönlichen Daten, die auf dem Einwilligungsbogen erfasst wurden. Tabelle 1 zeigt die Anzahl der Befragten, die in den Wellen 5 oder 6 zum ersten Mal der Datenverknüpfung zugestimmt haben. In Welle 5 haben 60% der Befragten ihre Sozialversicherungsnummer angegeben, 98% der Nummern bestanden das Prüfverfahren. Auch die Ermittlung der fehlenden Nummern war sehr erfolgreich, nur knapp 100 Nummern konnten nicht ermittelt werden, was nur 3% aller Bögen ausmacht. Die Ergebnisse in Welle 6 sind im Vergleich zu Welle 5 etwas schlechter. Das könnte daran liegen, dass hier nur ehemalige Verweigerer gefragt wurden. Der häufigste Grund für nicht ermittelbare Nummern ist, dass die Befragungsperson in einem Beamtenverhältnis beschäftigt war und daher kein Versicherungskonto bei der DRV (mehr) hatte.

3.2 Ziehung und Lieferung der RV-Daten

Für Befragte mit gültiger Sozialversicherungsnummer wurden durch das FDZ-RV in Würzburg Daten aus den Versicherungskonten gezogen. Diese Datenziehung wird zu festen Zeitpunkten eines jeden Jahres für die Daten der Versichertenrentenbestands (RTBN) und der Versicherungskontenstichprobe (VSKT) getrennt vorgenommen. In Würzburg wurden neben der Ziehung der Daten Anonymisierungsmaßnahmen durchgeführt, indem die

Sozialversicherungsnummer der Befragten gelöscht und durch die Laufnummer der Einwilligungsbögen ersetzt wurde. Zudem wurden die Daten in das Format des Scientific Use Files umgewandelt und gelten damit als faktisch anonymisiert.

Die Liste mit den Sozialversicherungsnummern der zu ziehenden Fälle der VSKT für das Berichtsjahr 2013 wurde im August 2014 nach Würzburg gesandt und enthielt 4.338 gültige Sozialversicherungsnummern. Die Ziehung der Daten war sehr erfolgreich: Für 4.304 Nummern konnten auch Daten gezogen werden, das entspricht etwa 99%. Im August 2015 standen für die Ziehung der VSKT 2014 4.921 Sozialversicherungsnummern zur Verfügung und für 4.577 Befragte (93%) konnten tatsächlich Daten gezogen werden.

Die jährliche Ziehung des Versichertenrentenbestands erfolgte immer im Dezember. Für die Ziehung des RTBN für das Berichtsjahr 2014 wurden 4.649 gültige Sozialversicherungsnummern nach Würzburg geschickt und für 2.390 Nummern konnten Daten gezogen werden. Der Rücklauf war hier deutlich geringer, da der Rentenbestand nur für Personen verfügbar ist, die schon in Rente sind, aber immer alle Nummern nach Würzburg geschickt werden. Für den RTBN 2015 waren es 5.076 Sozialversicherungsnummern. Die Lieferung der Daten wird im Herbst 2016 erwartet. Nach Erhalt der Daten ist im FDZ-RV Berlin erneut zu prüfen, ob die gezogenen Daten mit den Angaben auf den Einwilligungsbögen übereinstimmen. Diese Prüfung wird mit Hilfe der persönlichen Angaben zu Geschlecht und Geburtsdatum durchgeführt. Im Falle des Versichertenrentenbestands ist darüber hinaus zu prüfen, ob aus den Versicherungskonten tatsächlich Altersrenten bezogen werden, oder eventuell Hinterbliebenenleistungen (z.B. Witwen-/Witwerrenten oder Waisenrenten). In diesem Fall wird dokumentiert, dass der Inhaber des Versicherungskontos verstorben ist.

3.3 „Kombidatensatz“

Ziel der Datenverknüpfung war es, für jeden Befragten die aktuellsten Rentendaten zu ermitteln und bereitzustellen. Es kam jedoch vor, dass die Rentendaten eines Befragten nicht für jedes Berichtsjahr verfügbar waren, da sich das Konto zum Beispiel zum Ziehungszeitpunkt in Bearbeitung befand, aber im Jahr davor und im Jahr darauf wieder Daten gezogen werden konnten. Für verstorbene Befragte konnten auch keine Daten der VSKT mehr gezogen werden. In diesem Fall wurden die zuletzt gezogenen Daten für die Verknüpfung verwendet. Der Zweck des sogenannten „Kombidatensatzes“ war es, für alle Befragten das aktuellste Berichtsjahr zu ermitteln und die Daten anschließend zusammenzufügen. Dazu wurde ein Datensatz mit allen Personen erstellt, für welche Daten der Rentenversicherung gezogen werden konnten, unerheblich aus welchem Berichtsjahr. Der so entstandene Datensatz bildete die Grundlage für die Verknüpfung mit den Befragungsdaten.

3.4 Datenaufbereitung

Bevor die Daten verknüpft und veröffentlicht werden konnten, war es notwendig sie nutzerfreundlich aufzubereiten. Dazu wurden die Rohdaten der administrativen Daten zunächst um diejenigen Variablen gekürzt, welche lediglich für die Rentenberechnung und weniger für die wissenschaftliche Analyse von Bedeutung sind. Des Weiteren galt es zu überprüfen, ob alle Ausprägungen der einzelnen Variablen eine ausreichend große Fallzahl aufwiesen. Da die Befragungspersonen des SHARE 50 Jahre und älter sind, ist es bei einigen Variablen möglich, dass bestimmte Ausprägungen nicht besetzt sind. Bei keiner oder zu geringer Besetzung der Ausprägungen wurden diese umcodiert und zusammengefasst. Diese Umcodierungsmaßnahmen wurden sowohl für die VSKT als auch den RTBN jährlich überprüft und gegebenenfalls angepasst. Das bedeutete wiederum auch, dass die Variablenausprägungen in den Datensätzen mit denjenigen im Codeplan angegebenen, abgeglichen und gegebenenfalls angepasst werden mussten. Zudem wurde die Codierung der fehlenden Werte überprüft. Zum einen galt es sich zu vergewissern, dass die fehlenden Werte plausibel waren, und zum anderen, dass sie für alle Variablen einheitlich waren. Abschließend wurden noch einige Korrekturen durchgeführt. Dabei wurde insbesondere auf Einheitlichkeit und eine gute Handhabung der Daten geachtet. Das bedeutete beispielsweise, dass Labels verändert oder Dezimalzahlen gerundet wurden.

3.5 Verknüpfung der Daten am MEA

Die endgültige Verknüpfung der SHARE-Befragungsdaten mit den administrativen Daten der Rentenversicherung erfolgte am MEA. Hierzu wurden die Dokumentation der Zustimmungen sowie die administrativen Daten von der Rentenversicherung an das MEA übermittelt. Alle sensiblen persönlichen Informationen wie die Sozialversicherungsnummer, der Name und das Geburtsdatum wurden aus Datenschutzgründen vorher gelöscht.

Die Verknüpfung der Daten erfolgte anhand einer eindeutigen Identifikationsnummer, die sowohl auf dem Einwilligungsbogen stand als auch vom Interviewer in das Umfrageinstrument eingegeben wurde. Für die Fälle, die schon in früheren Ziehungen enthalten waren, basierte die Verknüpfung auf bereits vorhandenen Routinen und lief größtenteils automatisiert ab. Die Verknüpfung neuer Fälle war jedoch deutlich aufwendiger, da die Routinen nur bei eindeutigen Fällen griffen. Fälle, die nicht eindeutig verknüpft werden konnten, wurden somit manuell zugewiesen. Hierzu bedurfte es einer engen Kooperation zwischen den Mitarbeiter/innen in Berlin und München. Konnte hinreichend sichergestellt werden, dass Befragungsdaten und administrative Daten einer Person zusammenpassen, wurde eine eindeutige Identifizierungsnummer vergeben, die dann in die Routine für die folgenden Verknüpfungen einging. Nach erfolgter Verknüpfung wurde der sogenannte „Overview“-Datensatz erstellt. Die aktuelle Version enthält alle deutschen SHARE-Befragten der Wellen 3 bis 5 und liefert Informationen darüber, in welchen SHARE Wellen jeder

Befragte teilgenommen hat, ob der Datenverknüpfung zugestimmt wurde und welche Rentendaten für die Personen vorliegen. Dieser Datensatz wurde zusammen mit den SHARE Daten veröffentlicht. Er ist für die Datennutzer sehr hilfreich, um einen ersten Überblick über die Verfügbarkeit der Daten zu bekommen. Abschließend wurde nochmals überprüft, ob das Zusammenfügen der administrativen Daten mit den Befragungsdaten – was letztlich vom Nutzer selbst durchgeführt wird – problemlos abläuft.

3.6 Dokumentation

Eine detaillierte Dokumentation ist für die Datennutzer unerlässlich. Dazu gehört neben den Codebooks auch ein User Guide (siehe Anhang VI), der alle wichtigen Informationen zu dem Projekt und der Nutzung der Daten zusammenfasst. Im Zuge des neuen Release der SHARE und SHARE-RV Daten im Mai 2016 (siehe Abschnitt 4.3) wurden alle Dokumente überarbeitet und online bereitgestellt. Die Codebooks wurden auch komplett überarbeitet und können nun auch in englischer Sprache angeboten werden, um die Nutzung des Datensatzes auch für internationale Wissenschaftler zu vereinfachen. In diesem Zuge wurden auch die Labels der Variablen übersetzt, sodass jeder Datennutzer selbst entscheiden kann, ob er oder sie mit englischen oder deutschen Labels arbeiten möchte. Alle Dokumente sind online verfügbar².

4. Daten-Veröffentlichung

Die Veröffentlichung der SHARE-RV Daten erfolgt in regelmäßigen Abständen. Im Allgemeinen findet die Veröffentlichung parallel zur Veröffentlichung der SHARE Daten statt. Sobald ein neuer Release der SHARE Daten verfügbar ist, wird auch ein neuer Release der aktuellsten administrativen Daten erstellt, um zu gewährleisten, dass die beiden Datensätze auf dem gleichen Stand sind. Die Befragungsdaten werden immer separat nach Welle veröffentlicht und können dann von den Datennutzern individuell zusammengefügt werden. In diesem Punkt weichen wir im Projekt SHARE-RV von den SHARE Richtlinien ab und veröffentlichen die Daten wellenübergreifend. Konkret bedeutet dies, dass die einzelnen Module immer die administrativen Daten derjenigen Personen enthalten, die jemals zugestimmt haben. Diese können dann von den Datennutzern individuell mit allen SHARE Wellen verknüpft werden, in denen die Befragten teilgenommen haben.

In der zurückliegenden Förderphase gab es drei Daten-Veröffentlichungen. Den Release 2-0-0 (März 2014), den Release 3-0-0 (März 2015) und den Release 5-0-0 (Mai 2016). Diese werden im Folgenden kurz beschrieben.

² <http://www.share-project.org/data-access-documentation/record-linkage-share-rv.html>

4.1 SHARE-RV Release 2-0-0

Der Release 2-0-0 bezieht sich auf die SHARE Daten der Wellen 3 und 4. Aufgrund der unterschiedlichen Ziehungsdaten der administrativen Daten beziehen diese sich auf verschiedene Berichtsjahre: der Versichertenrentenbestand geht bis zum Berichtsjahr 2012 und die Versicherungskontenstichprobe bis 2011. Da zu diesem Zeitpunkt die Verweigerer noch nicht erneut um ihre Zustimmung gefragt wurden, sind die Verknüpfungsraten im Vergleich zu späteren Release Versionen noch relativ gering. Insgesamt sind in Release 2-0-0 894 Fälle enthalten, die mit den Wellen 3 und 4 verknüpft werden können. Tabelle 2 zeigt die Verknüpfungsraten und Fallzahlen bezogen auf die Daten der deutschen SHARE Stichprobe für die damals aktuellen Release Versionen dieser Wellen. Die Verknüpfungsraten werden gebildet, indem die Anzahl aller Befragten, die der Datenverknüpfung zugestimmt haben und deren Daten verknüpft werden konnten, durch die Anzahl aller Interviews des jeweiligen Releases geteilt wird. Für Release 1-0-0 der Welle 3 bedeutet das, dass für 47% aller Befragten verknüpfbare Daten vorhanden sind.

Tabelle 2 : Verknüpfungsraten nach Welle (Rel. 2-0-0)

SHARE-RV (Rel. 2-0-0)	SHARE Deutschland	
	Welle 3 (Rel. 1-0-0)	Welle 4 (Rel. 1-1-1)
Verknüpfungsraten	47%	49%
Anzahl verknüpfbarer Fälle	863	762
Fälle in SHARE	1.852	1.572

Die Unterscheidung der verknüpften Fälle nach Familienstand ergibt, dass circa 60% der Befragten, die in Welle 3 und/oder Welle 4 teilgenommen haben, und deren Daten verknüpft wurden, mit einem Partner zusammen leben, dessen Daten ebenfalls verknüpft wurden³. Der Anteil der Befragten, deren Partner nicht verknüpft wurden, liegt in beiden Wellen bei 21%; Singles machen ebenfalls etwa 20% aus (vgl. Tabelle 3).

³ Das ist unabhängig davon, in welcher Welle die Zustimmung zur Verknüpfung erteilt wurde.

Tabelle 3: Verknüpfte Fälle nach Welle und Beziehungsform (Rel. 2-0-0)

SHARE-RV (Rel. 2-0-0)	SHARE Deutschland (Rel. 5-0-0): Teilnahme in Welle...		
	...3	...4	...3 & 4
Paare	514 (60%)	444 (58%)	428 (61%)
einzelne Partner	185 (21%)	156 (21%)	147 (21%)
Singles	161 (19%)	156 (21%)	128 (18%)
Summe	860	756	703

4.2 SHARE-RV Release 3-0-0

Der Release 3-0-0 wurde parallel mit dem ersten Release der Welle 5 Daten veröffentlicht und enthält demnach auch die Fälle, die in Welle 5 zugestimmt haben. Die administrativen Daten beziehen sich auf die Berichtsjahre 2013 (RTBN) und 2012 (VSKT). Da auch die Verweigerer aus früheren Wellen erneut gefragt wurden, wirkte sich diese Welle auch positiv auf die Verknüpfungsraten der Wellen 3 und 4 aus (siehe Tabelle 4). Die Anzahl der verknüpften Fälle erhöhte sich bei dieser Daten-Veröffentlichung deutlich.

Tabelle 4: Verknüpfungsraten nach Welle (Rel. 3-0-0)

SHARE-RV (Rel. 3-0-0)	SHARE Deutschland		
	Welle 3 (Rel. 1-0-0)	Welle 4 (Rel. 1-1-1)	Welle 5 (Rel. 1-0-0)
Verknüpfungsraten	60%	66%	61%
Anzahl verknüpfbarer Fälle	1.104	1.035	3.485
Fälle in SHARE	1.852	1.572	5.690

Auch im Hinblick auf die Beziehungsformen war der Anstieg der Fälle deutlich. Der Anteil der „verknüpften“ Paare stieg in jeder der drei Wellen auf etwa 70% an (vgl. Tabelle 5). Für Welle 5 bedeutet dies, dass für 1.224 Personen und ihre Partner Informationen aus administrativen Daten und SHARE Daten zur Verfügung stehen.

Tabelle 5: Verknüpfte Fälle nach Welle und Beziehungsform (Rel. 3-0-0)

SHARE-RV (Rel. 3-0-0)	SHARE Deutschland (Rel. 5-0-0): Teilnahme in Welle...			
	...3	...4	...5	... 3 & 4 & 5
Paare	767 (70%)	691 (69%)	2.448 (70%)	459 (69%)
einzelne Partner	103 (10%)	90 (9%)	343 (10%)	57 (9%)
Singles	219 (20%)	223 (22%)	693 (20%)	152 (22%)
Summe	1.089	1.004	3.484	668

4.3 SHARE-RV Release 5-0-0

Der aktuelle Release 5-0-0 wurde wieder analog zu dem SHARE Release 5-0-0 veröffentlicht. Es wurden die Daten aller fünf Wellen überarbeitet und bisher unveröffentlichte Fälle integriert. Der Versichertenrentenbestand und die Versicherungskontenstichprobe wurden jeweils um ein Berichtsjahr erweitert (RTBN 2014; VSKT 2013). Befragte, deren Einwilligungen erst nach der letzten Datenziehung angekommen waren, sind nun auch in den administrativen Daten enthalten, sodass sich die Zahl der verknüpfbaren Fälle noch einmal leicht erhöhte.

Tabelle 6: Verknüpfungsrates nach Welle (Rel. 5-0-0)

SHARE-RV (Rel. 5-0-0)	SHARE Deutschland		
	Welle 3 (Rel. 5-0-0)	Welle 4 (Rel. 5-0-0)	Welle 5 (Rel. 5-0-0)
Verknüpfungsrates	59%	66%	63%
Anzahl verknüpfbarer Fälle	1.138	1.065	3.596
Fälle in SHARE	1.921	1.621	5.758

Im aktuellen Release sind bereits fast 700 Personen vorhanden, die sowohl in Welle 3 als auch in Welle 4 und 5 teilgenommen haben und deren Daten auch verknüpft werden konnten (vgl. Tabelle 7).

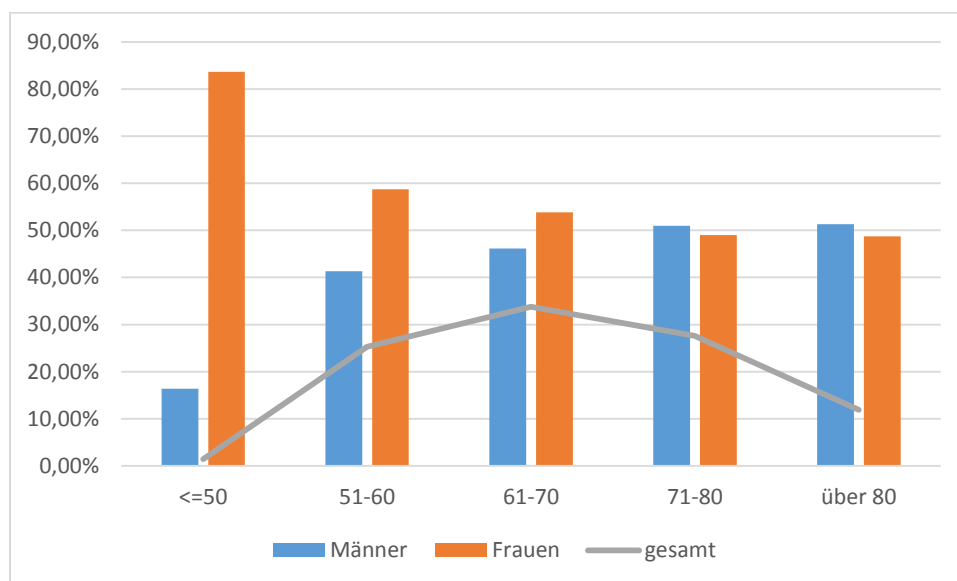
Tabelle 7: Verknüpfte Fälle nach Welle und Beziehungsform (Rel. 5-0-0)

SHARE-RV (Rel. 5-0-0)	SHARE Deutschland (Rel.5-0-0): Teilnahme in Welle...			
	...3	...4	...5	...3 & 4 & 5
Paare	799 (70%)	710 (67%)	2.514 (70%)	468 (69%)
einzelne Partner	113 (10%)	124 (12%)	368 (10%)	51 (8%)
Singles	226 (20%)	231 (21%)	714 (20%)	157 (23%)
Summe	1.138	1.065	3.596	676

4.3.1 Geschlechter und Altersverteilung in Release 5-0-0

Die Zielgruppe des SHARE sind Personen ab 50 Jahren. Es werden jedoch auch die im gleichen Haushalt lebenden Partner befragt, unabhängig von ihrem Alter. Bei den verknüpften Fällen der SHARE-RV Stichprobe reicht die Spannweite des Alters der Befragten von 33 bis 100 Jahre. Grafik 1 stellt die jeweiligen Anteile von Männern und Frauen in den verschiedenen Altersgruppen dar. Zusätzlich bildet die graue Linie ab, wie groß die Anteile der Altersgruppen an der Gesamtstichprobe der verknüpften Fälle sind. Die Altersverteilung zeigt, dass auch Daten von Personen jünger als 50 Jahre verknüpft wurden, die größte Gruppe bilden aber die 61-70 Jährigen. Die Geschlechterverteilung ist in den Altersgruppen nahezu ausgeglichen, nur bei den unter 50-Jährigen finden sich deutlich mehr Frauen.

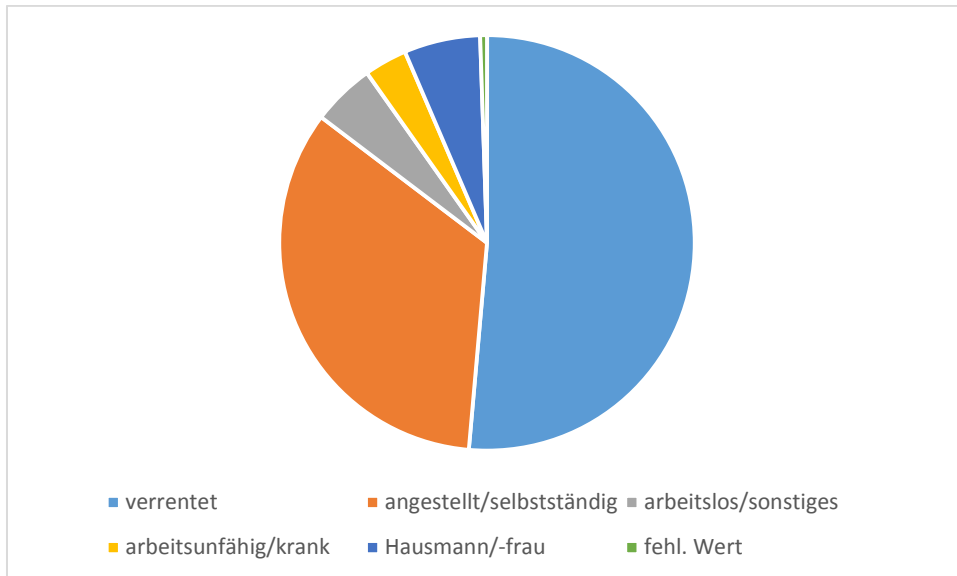
Grafik 1: Geschlechter- und Altersverteilung



4.3.2 Erwerbsstatus in Release 5-0-0

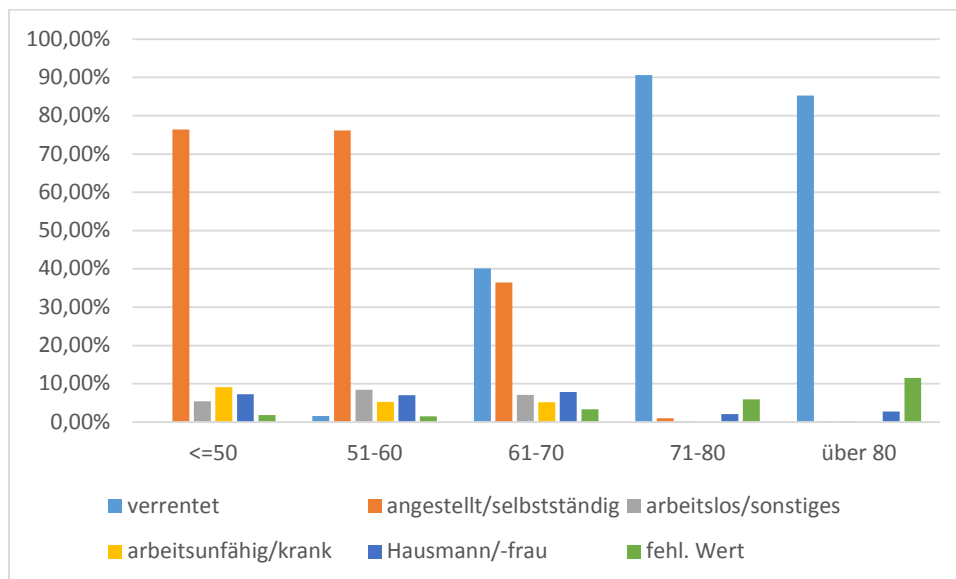
Bei der Betrachtung des selbstberichteten Erwerbsstatus zeigt sich, dass gut die Hälfte der befragten Personen bereits verrentet ist, was mit der oben erwähnten Altersstruktur übereinstimmt. Rund ein Drittel der Befragten gibt im Befragungsjahr 2013 an, noch berufstätig zu sein. Die restlichen circa 15% entfallen etwa zu gleichen Teilen auf Personen, die arbeitslos, arbeitsunfähig oder als Hausmann bzw. Hausfrau tätig sind (vgl. Grafik 2).

Grafik 2: Verteilung des Erwerbsstatus (Angaben in %)



Betrachtet man den Erwerbsstatus nach Altersgruppen getrennt, ist in der Gruppe der 61-70 Jährigen am meisten Heterogenität zu erkennen (vgl. Grafik 3). Mit jeweils circa 40% sind verrentete und erwerbstätige Personen gleichermaßen vertreten. Sowohl die 71-80 Jährigen als auch die über 80 Jährigen sind zu 80%-90% verrentet. Die unter 50-Jährigen sind zu knapp 80% erwerbstätig. In dieser Gruppe ist noch kein Befragter verrentet.

Grafik 3: Erwerbsstatus nach Alter



5. Datennutzung

5.1 Maßnahmen zur Erhöhung der Nutzerfreundlichkeit

SHARE-RV soll der wissenschaftlichen Gemeinschaft nicht nur in Deutschland, sondern weltweit für Forschungsprojekte zur Verfügung stehen. Daher arbeiteten wir kontinuierlich daran, den Bekanntheitsgrad und die Nutzerfreundlichkeit von SHARE-RV zu erhöhen. Zu diesem Zweck wurden einige Verbesserungsmaßnahmen umgesetzt: erstens Maßnahmen die öffentlichkeitswirksame Präsentation betreffend, zweitens die Neugestaltung der Homepage und drittens eine Erweiterung des Angebots für internationale Nutzer.

Die Maßnahmen zur öffentlichkeitswirksamen Präsentation verfolgten das Ziel, den Bekanntheitsgrad von SHARE-RV zu steigern. Hierzu wird im November dieses Jahres der erste SHARE-RV Nutzerworkshop in München stattfinden. Der Workshop ist für zwei Tage angesetzt. Zum einen werden die Datensätze SHARE und SHARE-RV vorgestellt, zum anderen werden praktische Hinweise für den Umgang mit den Datensätzen gegeben. Darüber hinaus sollen Nutzer, die bereits mit SHARE und/oder SHARE-RV gearbeitet haben, die Gelegenheit bekommen, ihre Arbeiten vorzustellen (siehe Anhang III). Somit bietet der Workshop nicht nur die Möglichkeit, Informationen zu den Datensätzen an Nutzer weiterzugeben, sondern soll auch zum Austausch der Nutzer untereinander sowie zwischen den Nutzern und den Datenproduzenten anregen. Dieser Austausch kann weiterhin genutzt werden, um Anregungen der Nutzer zur Kenntnis zu nehmen und bei der Weiterentwicklung des Projektes zu berücksichtigen.

Zudem haben die Mitarbeiter/innen des Projekts regelmäßig auf Konferenzen präsentiert, um das Datenangebot und das reichhaltige Analysepotenzial von SHARE-RV vorzustellen (siehe

Anhang V). Auf Konferenzen bietet sich ebenfalls Gelegenheit, Flyer auszulegen. Dazu wurden die Inhalte des SHARE-RV Flyers aktualisiert. Dieser soll zukünftig Interessenten eine Zusammenfassung der Inhalte und des Potentials von SHARE-RV sowie Informationen zum Datenzugang und Kontaktadressen liefern. Der Flyer wird derzeit von einem Grafikdesigner gestaltet und wird voraussichtlich im November gedruckt. In diesem Zuge wurde ebenfalls ein neues SHARE-RV Logo entworfen, welches an das neue Farbschema von SHARE angepasst wurde. Somit erhalten alle Dokumente ein einheitliches Layout und Design.

Durch die Neugestaltung sowohl der SHARE Homepage (www.share-eric.eu) als auch der Homepage des Forschungsdatenzentrums (www.fdz-rv.de) konnte auch hier die neue einheitliche Linie umgesetzt werden. Auf der Homepage des FDZ-RV sind die Unterseiten zu SHARE-RV, nun direkt über einen eigenen Menüpunkt von der Startseite aus zu erreichen. Den Nutzern werden dadurch übersichtlich Informationen sowohl in deutscher als auch in englischer Sprache zur Verfügung gestellt. Neben einem allgemeinen Überblick zu dem Datenangebot SHARE-RV und den Inhalten der Datensätze finden sich auch Informationen zum Zugang zu den Daten und ein Querverweis zur SHARE Homepage.

5.2 Nutzerstatistik

Die Nutzung des Datensatzes SHARE-RV wurde von verschiedensten Forschungseinrichtungen und Hochschulen beantragt. Im Förderzeitraum gingen insgesamt 23 neue Datennutzungsanträge beim Forschungsdatenzentrum der Rentenversicherung ein (siehe Tabelle 8). Darunter waren zwei Nutzungsanträge von nicht-deutschen Universitäten, was zeigt, dass das Datenangebot auch über die Grenzen hinweg Bekanntheit erlangt hat.

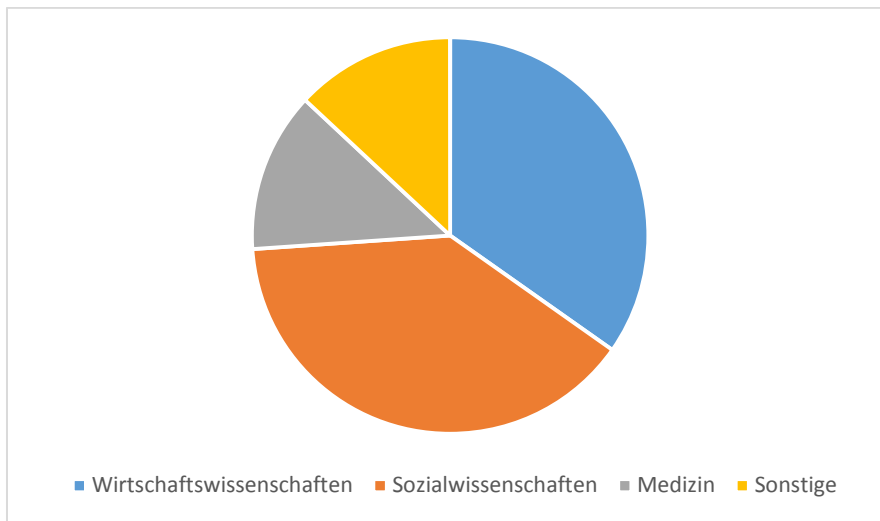
Tabelle 8: Anzahl der Datennutzungsanträge im Förderzeitraum nach Jahren

Jahr	Anzahl
2013	2
2014	5
2015	12
2016	4
Summe	23

Die Forschungsprojekte für welche SHARE-RV beantragt wurde sind thematisch breit gefächert. Den größten Anteil der Datennutzer stellten Sozial- und Wirtschaftswissenschaftler dar. Sie machen zusammen etwa 75% aus. Die Projektvorhaben reichen von Lebenslaufanalysen über Untersuchungen zu Gesundheit und/oder

Erwerbsarbeit bis hin zu methodischer Forschung. Eine Auflistung der Publikationen basierend auf SHARE-RV findet sich im Anhang (siehe Anhang IV).

Grafik 4: Datennutzung nach Fachbereich (Angaben in %)



6. Ausblick

Das Projekt SHARE-RV entwickelte sich in den letzten Jahren kontinuierlich weiter und doch gibt es weiterhin eine Fülle an Ideen und Möglichkeiten wie sowohl Abläufe als auch Ergebnisse optimiert werden können. Dazu gehört neben der Erhöhung der Fallzahlen auch die Einbettung des Datensatzes in die Forschungslandschaft.

Eine Möglichkeit, sozial- und wirtschaftswissenschaftliche Datenquellen vorzustellen, bietet die Rubrik „Data Observer“ der „Jahrbücher für Nationalökonomie und Statistik“ (ehemals „European Data Watch“ in Schmollers Jahrbuch). Die Beiträge zu dieser Rubrik geben detailliert Auskunft über interessante Datensätze für die empirische Forschung, informieren über Möglichkeiten des Datenzugangs und stellen beispielhaft Studien vor, die mit diesen Datensätzen durchgeführt wurden. An einem Artikel zu SHARE-RV in dieser Reihe wird bereits gearbeitet, dieser soll bis Anfang 2017 eingereicht werden.

Des Weiteren gibt es Überlegungen dazu, wie den Nutzern der Umgang mit den Rentendaten noch weiter erleichtert werden könnte. Da die Rentendaten viele Variablen enthalten, die aus der Rentenberechnung stammen und nicht für die wissenschaftliche Forschung generiert wurden, ist ihre Aufbereitung und Handhabung meist aufwendig. Daher wäre es denkbar, den Nutzern Aufbereitungsarbeiten abzunehmen, beispielsweise im Hinblick auf das Konzept der Entgeltpunkte. Dieses bietet auf der einen Seite sehr detaillierte Informationen, ist auf der anderen Seite vielen Nutzern aber nur wenig bekannt und es wäre somit vorteilhaft, den Nutzern Codes zur weiteren Verarbeitung der Variablen zur Verfügung zu stellen.

Ein weiterer Punkt der SHARE-RV Planung betrifft die internationalen Nutzer. Die Verwendung der Daten wurde durch die Übersetzung der Codepläne und die Bereitstellung

der Label-Syntax bereits erleichtert. Schwierigkeiten bereitet jedoch noch der Zugang zu den Daten, da die Homepage des Forschungsdatenzentrums und somit die Bestellformulare für die Rentendaten bisher nur in deutscher Sprache vorhanden sind. Hierzu wird die Möglichkeit geprüft, eine englische Version des Datennutzungsvertrages anzubieten.

7. Literaturverzeichnis

Börsch-Supan, Axel, Alt, Benedikt, Bucher-Koenen, Tabea (2015): *Early retirement for the underprivileged? Using the record-linked SHARE-RV data to evaluate the most recent German pension reform*. In: Börsch-Supan, Axel, Kneip, Thorsten, Litwin, Howard, Myck, Michal, Weber, Guglielmo (Hrsg.): *Ageing in Europe - Supporting Policies for an Inclusive Society*, Berlin 2015, S. 267–278.

Börsch-Supan, Axel, Coppola, Michaela, Rausch, Johannes (2014): Die „Rente mit 63“: Wer sind die Begünstigten? Was sind die Auswirkungen auf die Gesetzliche Rentenversicherung? MEA Discussion Papers 17-2014. München: Munich Center for the Economics of Aging (MEA).

Caldewood, Lisa; Lessof, Carli (2009): *Enhancing Longitudinal Surveys by Linking to Administrative Data*. In: Lynn, Peter (Hrsg.): *Methodology of Longitudinal Surveys*. Chichester: John Wiley & Sons, Ltd. 55-72.

Deutscher Bundestag (2014): *Gesetz über Leistungsverbesserungen in der gesetzlichen Rentenversicherung (RV-Leistungsverbesserungsgesetz)*, Bundestagsdrucksache 18/909, Berlin.

Kröger, Katharina; Fachinger, Uwe; Himmelreicher, Ralf K. (2011): *Empirische Forschungsvorhaben zur Alterssicherung. Einige kritische Anmerkungen zur aktuellen Datenlage*. *RatSWD Working Paper Series Nr.170*. Berlin: Rat für Sozial- und Wirtschaftsdaten.

Mika, Tatjana, Czaplicki, Christin, Post, Julia C. (2016): *Alterseinkommen von Müttern und kinderlosen Frauen im Haushaltskontext*. *RVaktuell* 2016, S. 117-123.

Schmidt, Christoph M. (2013): *Die demographische Herausforderung – ein aktueller Überblick*. In: Hüther, Michael; Naegele, Gerhard (Hrsg.): *Demografiepolitik. Herausforderungen und Handlungsfelder*. Wiesbaden: Springer Fachmedien, 71-95.

Trischler, Falko (2014): *Erwerbsverlauf, Altersübergang, Alterssicherung. Zunehmende soziale Ungleichheit im Alter*. Wiesbaden: Springer Fachmedien.

Weiland, S. K., Rapp, K., Klenk, J., & Keil, U. (2006). *Zunahme der Lebenserwartung. Größenordnung, Determinanten und Perspektiven*. *Dtsch Arztebl*, 103(16). Köln: Deutscher Ärzteverlag. A 1072-1077.

8. Anhang

Anhang I: Einwilligungsbogen zur Verknüpfung von Befragungs- und Kontendaten



Sehr geehrte(r) Interviewteilnehmer(in),

vielen Dank für Ihre Teilnahme an unserer Studie „50+ in Europa“!

Um den gesellschaftlichen Wandel in der Bundesrepublik genauer beschreiben zu können, benötigt die Wissenschaft möglichst exakte Informationen. Daher möchten wir mit Ihrer Einwilligung auch Daten der gesetzlichen Rentenversicherung nutzen. Dabei handelt es sich um Ihren Versicherungsverlauf, der Informationen zu Beschäftigungszeiten, Zeiten der Kindererziehung sowie der Arbeitslosigkeit oder Arbeitsunfähigkeit und zur Höhe Ihrer Rente nach heutigem Stand enthält. Für uns sind hierbei alle Interviewteilnehmer gleichermaßen von Bedeutung, auch wenn Sie keine oder nur vorübergehend Beiträge zur gesetzlichen Rentenversicherung geleistet haben.

Das Datenschutzgesetz setzt für die Verbindung der Rentenversicherungsdaten mit Ihren Befragungsdaten Ihre Einwilligung voraus, welche selbstverständlich freiwillig ist.

Daher bitten wir Sie, das Formular auf der Rückseite auszufüllen und unterschrieben dem Interviewer/der Interviewerin zu übergeben. Diese(r) wird es in dem beigefügten Briefumschlag direkt an die Deutsche Rentenversicherung Bund schicken. Das zweite Exemplar der Einwilligung ist für Ihre Unterlagen bestimmt.

Die Angaben zu Ihrer Person werden nur verwendet, um die Rentenversicherungsdaten zu ermitteln. Diese werden dann ohne Ihre persönlichen Angaben und ohne die Versicherungsnummer an das Münchener Zentrum für Ökonomie und Demographischer Wandel (MEA) des Max-Planck-Instituts für Sozialrecht und Sozialpolitik übermittelt und mit den Befragungsdaten verbunden. **Die Verknüpfung und Auswertung der Daten erfolgt anonym und ausschließlich zu wissenschaftlichen Zwecken.** Übrigens: Die Rentensachbearbeiter(innen), die Ihre Anfragen oder Rentenanträge bearbeiten werden, erfahren nichts über Ihre Teilnahme an diesem Projekt. Ihre Teilnahme oder Nichtteilnahme wirkt sich keinesfalls auf die Bearbeitung von Anträgen oder Ihre Rente aus.

Für Rückfragen steht Ihnen Frau Tatjana Mika unter der Rufnummer

030/865 89 541 oder per Mail (tatjana.mika@drv-bund.de) bei der Deutschen Rentenversicherung Bund jederzeit gerne zur Verfügung.

Wir bedanken uns sehr herzlich für Ihre Mithilfe.

Prof. Axel Börsch-Supan, Ph.D.

SHARE-Koordinator und Direktor des
Max-Planck-Instituts für
Sozialrecht und Sozialpolitik
Amalienstraße 33
80799 München

Dr. Axel Reimann

Präsident der Deutschen
Rentenversicherung
Bund
Ruhrstraße 2
10709 Berlin

KENNZIFFER: 500001

**Einwilligung zur Zusammenführung von Befragungs- und Kontendaten
der Rentenversicherungsträger (gem. § 67 b des Sozialgesetzbuches X)**

Ich erkläre mich einverstanden, dass der für mich zuständige Rentenversicherungsträger alle über mich gespeicherten rentenrechtlichen Zeiten und Entgeltzeiten - den sogenannten Versicherungsverlauf - und die sich daraus ergebende Rentenhöhe nach heutigem Stand an das Münchener Zentrum für Ökonomie und Demographischer Wandel (MEA) des Max-Planck-Instituts für Sozialrecht und Sozialpolitik übermittelt. Dabei werden mein Name und meine Versicherungsnummer durch eine Kennziffer ersetzt.

Die Daten werden mit meinen Befragungsdaten verbunden und ausschließlich für das Forschungsprojekt „50+ in Europa“ ausgewertet. Ich wurde darüber informiert, dass alle datenschutzrechtlichen Bestimmungen eingehalten werden. Ich behalte mir vor, jederzeit mein Einverständnis wieder zurückzuziehen.

Familienname	Vorname	Geburtsname (Anfangsbuchstabe ist ausreichend)
	T T . M M . J J J J	
Geburtsort	Geburtsdatum	

Datum **Unterschrift**

Sie helfen uns bei der Ermittlung der Daten sehr, wenn Sie hier noch Ihre **Versicherungsnummer der gesetzlichen Rentenversicherung** eintragen. Sie finden diese Nummer u.a. in Ihrem Sozialversicherungsausweis, bei Arbeitnehmern in der jährlichen Meldung Ihres Arbeitgebers an die Rentenversicherung (eine Kopie geht an Sie) oder neuerdings in den jährlichen Meldungen der Rentenversicherung, in den monatlichen Gehaltsabrechnungen oder den Rentenbescheiden bei Rentenbeziehern.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Rentner/innen PAN Versicherungsnummer BX/ZA

Personen, die noch keine Rente beziehen

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

 Versicherungsnummer

Sozialversicherungsnummer ist nicht bekannt Ich habe keine Sozialversicherungsnummer

Bitte geben Sie den Umschlag mit dem ausgefüllten Formular auch dann an den Interviewer zurück, wenn Sie Ihre Sozialversicherungsnummer nicht zur Hand haben. Vielen Dank!

Anhang II: Modul zur Datenverknüpfung aus SHARE-Fragebogen

LI004_Intro:

INTERVIEWER: „Wir wechseln nun das Thema. Für wissenschaftliche Untersuchungen ist es sehr interessant, mehr über die Erwerbsbiografie von Menschen zu erfahren. Zu diesem Zweck möchten wir gerne die Befragungsdaten aus den Interviews mit den Sozialversicherungsdaten der Deutschen Rentenversicherung verknüpfen. Dies ist aus datenschutzrechtlichen Gründen nur mit Ihrer Einwilligung möglich. Selbstverständlich ist Ihre Einwilligung freiwillig.

Nehmen Sie sich bitte ein paar Minuten Zeit, um das Formular zu lesen.“

Anweisung für Interviewer:

Nehmen Sie die 2 Einwilligungsformulare und übergeben Sie eines davon der Zielperson.

LI001_Number:

Anweisung für Interviewer:

Nehmen Sie jetzt die Kopie des Einwilligungsformulars zur Zuspiegelung der Sozialversicherungsdaten zur Hand.

Geben Sie bitte die 6-stellige Kennziffer oben rechts auf dem Einwilligungsformular ins CAPI-Programm ein.

LI002_Number_Check

Anweisung für Interviewer:

Geben Sie die Kennziffer zur Kontrolle nochmal ein.

LI003_Consent

Anweisung für Interviewer:

Helfen Sie der Zielperson, falls nötig. Streichen Sie das Formular durch, falls die Zielperson verweigert.

Fügen Sie das Formular in den an die DRV adressierten Briefumschlag und werfen Sie den Brief ein.

Die Kopie verbleibt bei der Zielperson.

Hat die Zielperson der Datenverknüpfung zugestimmt?

1. Ja. Zielperson hat zugestimmt, das Formular ausgefüllt und es mir im Umschlag zurückgegeben.
2. Zustimmung unsicher. Formular verbleibt bei Zielperson und wird bei Zustimmung von der Zielperson selbst zurückgesandt.
5. Nein. Zielperson hat der Datenverknüpfung nicht zugestimmt. Formular durchstreichen und auch an die DRV versenden.

Anhang III: SHARE & SHARE-RV User Workshop – Call for Papers



Workshop on the Survey of Health, Ageing and Retirement in Europe (SHARE)

November 3-4, 2016

Munich, Germany

Learn how to get more out of the SHARE data and get your work known to other data users!

This workshop aims at students and researchers interested in getting a deeper insight into the SHARE data SHARE as well as the German Pension Insurance Linkage SHARE-RV.

SHARE - The Survey of Health, Ageing and Retirement in Europe is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status, and social and family networks. It encompasses data about approximately 123,000 individuals (more than 293,000 interviews) from 20 European countries (+Israel) aged 50 or older. For more information see: <http://www.share-eric.eu>.

SHARE-RV - The German Pension Insurance Linkage is a project within the German sub-study SHARE linking respondents' survey data with the administrative records of the German Pension Insurance (Deutsche Rentenversicherung, DRV for short). For more information see: <http://www.share-eric.eu/data-access-documentation/record-linkage-share-rv.html>.

Objective: The aim of this workshop is to provide a general overview of SHARE and SHARE-RV and practical advices for analyzing the data. In addition, we welcome presentations of own work and experiences based on SHARE or SHARE-RV data.

The workshop will be split in three parts:

- The first part focuses on general aspects of SHARE such as the content and structure of the questionnaire, the organization of the data files, and the available documentation. This includes sections on how to generate your own data, and about the retrospectively collected life histories (SHARELIFE).
- The second part concentrates on SHARE-RV. Its high research potential lies in the investigation of connections between various aspects of respondents' lives and their working history or their socio-economic status in later life. Its composition is unique in the German data set world, because the process-produced longitudinal DRV data start at age 14. Besides a detailed introduction about its content, which is determined from the German pension law, the structure will be explained. Furthermore, some data handling advices will be given.
- Finally, the third part of this workshop is for presentations of selected scientific results based on SHARE or SHARE-RV. Researchers are invited to submit short ongoing research on all topics. Submissions by colleagues at earlier stages of their scientific career are particularly welcome.

A detailed program will be made available as soon as possible on our website.

Fees: Participation is free of charge. Light lunch is included. Travel expenses for students can be reimbursed by prior agreement.

Registration: Registration for participation in the workshop is mandatory. Please fill out the registration form and send it to Sabine Friedel (friedel@mea.mpisoc.mpg.de). Note that submission of an abstract is not a prerequisite for participation in the workshop.

Submission of abstracts: Submission of abstracts is voluntary. We invite abstracts on all topics based on SHARE or SHARE-RV. All submissions should include about 300 words providing information about the research question, type of empirical analysis, and if possible, main results. Please send it to Sabine Friedel (friedel@mea.mpisoc.mpg.de).

Deadline for registration and submissions of abstracts: September 30, 2016.

All applicants will be notified about their acceptance by October 7, 2016 latest.

For more information, contact Sabine Friedel (friedel@mea.mpisoc.mpg.de).

We look forward to seeing you in Munich!

Date: November 3-4, 2016

Time: November 3: 12 p.m. - 6 p.m.
November 4: 10 a.m. - 4 p.m.

Location: Munich Center for the Economics of Ageing (MEA)
Max-Planck-Institute for Social Law and Social Policy
Amalienstraße 33
Room 313
80799 Munich, Germany

Coordination: Tim Birkenbach, Sabine Friedel, Stefan Gruber, Imke Herold, Julie Korbmacher, Tatjana Mika, and Annette Scherpenzeel.



MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
MEA Munich Center for the Economics of Ageing



Anhang IV: Publikationen basierend auf SHARE-RV

Börsch-Supan, Axel, Alt, Benedikt, Bucher-Koenen, Tabea (2015): *Early retirement for the underprivileged? Using the record-linked SHARE-RV data to evaluate the most recent German pension reform*. In: Börsch-Supan, Axel, Kneip, Thorsten, Litwin, Howard, Myck, Michal, Weber, Guglielmo (Hrsg.): *Ageing in Europe - Supporting Policies for an Inclusive Society*, Berlin 2015, S. 267–278. Verfügbar unter: <http://www.degruyter.com/view/books/9783110444414/9783110444414-026/9783110444414-026.xml>

Czaplicki, Christin, Stern, Tobias (2016): *Sozialstrukturelle Determinanten von Tabakkonsum im Alter*. *RVaktuell*, 2/2016, Jahrgang 63, S. 39-46.

Czaplicki, Christin, Post, Julia Charlotte (2015): *SHARE-RV: Neues Analysepotential für die Untersuchung des Zusammenhangs von Fertilität und Erwerbstätigkeit in den Lebensverläufen von Männern und Frauen*. In: DGD - Deutsche Gesellschaft für Demographie e. V. (Hrsg.): *Das Potenzial amtlicher Daten für die demografische Forschung und die Familienforschung*. DGD-Online-Publikation 01/2015, Bonn. Verfügbar unter: http://www.share-project.org/fileadmin/pdf_documentation/SHARE-RV/Czaplicki_Post_2015_SHARE-RV.pdf

Czaplicki, Christin; Korbmacher, Julie (2010): *SHARE-RV: Verknüpfung von Befragungsdaten des Survey of Health, Ageing and Retirement in Europe mit administrativen Daten der Rentenversicherung*, DRV-Schriften Band 55/2010, S. 28-37. Verfügbar unter: <http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/resDisplay.do?id=321&tabelle=6>

Korbmacher, Julie (2014): *Recall Error in the Year of Retirement*, SHARE Working Paper Series 21-2014, Munich Center for the Economics of Aging (MEA) Verfügbar unter: http://mea.mpisoc.mpg.de/uploads/user_mea_discussionpapers/1577_WP_Series_21_2014_Korbmacher.pdf

Korbmacher, Julie (2014): "New Challenges for Interviewers when Innovating Social Surveys: Linking Survey and Objective Data". Dissertation LMU München. Verfügbar unter: <https://edoc.ub.uni-muenchen.de/17966/>

Korbmacher, Julie; Czaplicki, Christin (2013): *Linking SHARE Survey Data with Administrative Records: First Experiences from SHARE-Germany*, Malter, Frederic und Börsch-Supan, Axel, *SHARE Wave 4: Innovations & Methodology*, München, S. 47-53. Verfügbar unter: http://www.share-project.org/fileadmin/pdf_documentation/Method_FRB_FINAL.pdf

Korbmacher, Julie; Schröder, Mathis (2013): *Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer*, *Survey Research Methods*, Vol. 7, 2 115-131. Verfügbar unter: <https://ojs.ub.uni-konstanz.de/srm/article/view/5067>

Mika, Tatjana, Czaplicki, Christin, Post, Julia C. (2016): Alterseinkommen von Müttern und kinderlosen Frauen im Haushaltskontext. RVaktuell 2016, S. 117-123. Verfügbar unter: http://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/5_Services/03_broschueren_und_mehr/03_zeitschriften/rv_aktuell/2016/heft_5_6.pdf?__blob=publicationFile&v=2

Mika, Tatjana; Czaplicki, Christin (2010): SHARE-RV: Eine Datengrundlage für Analysen zu Alterssicherung, Gesundheit und Familie auf der Basis des Survey of Health, Ageing and Retirement in Europe und der Daten der Deutschen Rentenversicherung, RVaktuell, 12/2010, S. 396-400. Verfügbar unter: http://www.deutsche-rentenversicherung.de/Allgemein/de/Inhalt/5_Services/03_broschueren_und_mehr/03_zeitschriften/rv_aktuell/2010/heft_12_mika_czaplicki.html

Anhang V: Vorträge und Präsentationen zu SHARE-RV

- 09/2016 The Survey of Health, Ageing and Retirement in Europe. Vortrag in der German Data Forum Session "Using microdata to analyse demographic changes" Jahrestagung des Vereins für Sozialpolitik, Augsburg.
- 07/2016 The Survey of Health, Ageing and Retirement in Europe. Statistik Tage Bamberg Fürth, Bamberg.
- 06/2016 SHARE User Workshop (Girona)
- 04/2016 SHARE-RV: Neue Daten zu Einkommen und Gesundheitsverhalten. Präsentation im Rahmen des Brown Bag Seminars der Deutschen Rentenversicherung, zusammen mit Christin Czaplicki und Tatjana Mika.
- 02/2016 SHARE User Workshop (Stockholm)
- 06/2014 Validierung der Bildungsangaben in den Daten des FDZ-RV mit Befragungsdaten aus SHARE. Präsentation im Rahmen des 11. Workshops des Forschungsdatenzentrums der Deutschen Rentenversicherung, Berlin.
- 03/2014 SHARE-RV: Neues Analysepotenzial für die Fertilitäts- und Familienforschung. Präsentation mit Julia Post im Rahmen der DGD-Jahrestagung 2014: Demografiestrategie – Work in Progress? Das Potenzial amtlicher Daten für die demografische Forschung und die Familienforschung, Berlin.
- 02/2014 Gesis Workshop SHARE meet the data
- 09/2013 SHARE-RV: Linking survey data of SHARE with social security data of the German Pension Insurance. Beitrag mit Tatjana Mika im Rahmen des Workshops Developing a data linkage system to enable innovative research, Berlin.
- 07/2013 Pflege und Beruf: Betreuung beginnt nicht erst mit der „Pflegestufe“. Ein Vergleich der Anzahl pflegender Personen in der VSKT, SHARE und SHARE-RV. Präsentation im Rahmen des 12. Graduiertenkolloquiums des Forschungsnetzwerks Alterssicherung (FNA), Berlin.
- 06/2013 Messung von informeller Pflege am Beispiel von SHARE-RV. Präsentation im Rahmen des 10. Workshops des Forschungsdatenzentrums der Deutschen Rentenversicherung, Berlin.

User Guide Release 5-0-0

CONTENT

1.	ABOUT SHARE-RV	1
1.1	WHAT IS SHARE-RV ABOUT?	1
1.2	HOW IS SHARE-RV ORGANIZED?	1
1.3	WHO IS ELIGIBLE FOR RECORD LINKAGE?	2
1.4	FOR WHICH SHARE RESPONDENTS ARE ADMINISTRATIVE DATA AVAILABLE?	2
2.	DATA	4
2.1	SHARE	4
2.1.1	SURVEY DATA	4
2.1.2	SHARE GV-LINKAGE MODULE - OVERVIEW	5
2.2	THE ADMINISTRATIVE RECORDS	6
2.2.1	VERSICHERTENKONTENSTICHPROBE (VSKT)	6
2.2.2	VERSICHERTENRENTENBESTAND (RTBN)	6
3.	DATA ACCESS AND DOWNLOAD	7
4.	DATA HANDLING	7
4.1	DATA FORMATS AND LANGUAGES	7
4.2	HOW TO LINK THE DATA?	7
4.3	REPORTING YEAR	8
4.4	How can I find out in which year and month an event happened based on VSKT?	8
4.5	WHAT ARE THE EARNING POINTS EXACTLY?	9
	APPENDIX A	11

1. ABOUT SHARE-RV

1.1 WHAT IS SHARE-RV ABOUT?

SHARE-RV is a cooperation project between the Munich Center for the Economics of Aging (MEA), the Max Planck Institute for Social Law and Social Policy, and the Research Data Center of the German Pension Insurance ([FDZ-RV](http://www.fdz-rv.de)¹) in Berlin. It stands for the direct linkage of survey data of the *Survey of Health, Ageing and Retirement in Europe* ([SHARE](http://www.share-eric.eu)²) with administrative records of the "Deutsche Rentenversicherung", German Pension Insurance ([DRV](http://www.deutsche-rentenversicherung.de)³). More precisely, it is the linkage of the German subsample of SHARE survey data and selected administrative records provided by the FDZ-RV. The project started as a pilot study in 2009 as part of the 3rd SHARE wave and was primarily funded by the Volkswagenstiftung. Thanks to the follow up funding of the Forschungsnetzwerk Alterssicherung (FNA), the project continues at least until SHARE's sixth wave of data collection.

The goals of the project SHARE-RV are to link survey data from SHARE with data from the DRV and to make them available to the entire research community. Hence, SHARE-RV allows the investigation of connections between various aspects of respondents' lives and their working history or their socio-economic status in later life.

1.2 HOW IS SHARE-RV ORGANIZED?

The project is based on a direct linkage, meaning that the data of exactly the same person (here German SHARE respondents) were linked using the respondents' Social Security Number (SSN) as a unique identifier. When linking survey data with the records of the DRV, respondents' written consent is required. Respondents are asked for consent during the regular SHARE interview by the interviewer, which is documented on a separate consent form. The consent form collects respondents' SSN as well as some basic demographics which are needed to check the SSN. Filled-in consent forms without the respondents' signature cannot be considered for the linkage. If respondents are not able or not willing to provide their SSN, it is possible to find it out based on basic demographics provided by

¹ www.fdz-rv.de

² www.share-eric.eu

³ www.deutsche-rentenversicherung.de

the respondents. In order to be absolutely sure that the correct records are linked, a lot of effort is put in checking and correcting the SSN.

1.3 WHO IS ELIGIBLE FOR RECORD LINKAGE?

In SHARE Wave 3, all German respondents were asked to consent to record linkage. From wave 5 on, all respondents who have not consented before or were new in the SHARE sample were asked to consent. The corresponding administrative data of the consenters are drawn yearly by the DRV at fixed dates.

Please note that the survey and the admin data are not necessarily representing and accordingly reporting the same year (see Chapter 4.3 [reporting year](#) for more details).

1.4 FOR WHICH SHARE RESPONDENTS ARE ADMINISTRATIVE DATA AVAILABLE?

Firstly, the DRV is only allowed to provide the records if respondents consented explicitly during the interview. Secondly, all inhabitants with at least one spell of insured activity are included in the administrative records of the DRV. The data contain information on all activities for which social insurance contributions are paid. Not included are, for example, spells as civil servant or self-employee. Some respondents fulfill both conditions, but their administrative data are not available anyway. This is the case if, for example, short-time insured persons ask for reimbursement of their contributions. Another reason for missing records could be that entitlements were accumulated in the former German Democratic Republic and the respondents retired before their records could be transferred to the new system.

The availability of administrative data that could be linked to the corresponding records is summarized in wave-specific linkage rates (see Table 1). This rate is calculated as the number of respondents in a given wave whose data can be linked, divided by all respondents who participated in that wave. The linkage rates considered in Table 1 refer to the most recent release version only. As the number of respondents differs from wave to wave due to attrition and new persons being interviewed, the linkage rate differs from wave to wave, too.

Table 1 Linkage rate by wave

	Wave 3 (Rel. 5-0-0)	Wave 4 (Rel. 5-0-0)	Wave 5 (Rel. 5-0-0)
Number of interviews in SHARE	N=1,921	N=1,621	N=5,758
Record linkage rate	59,2%	65,7%	62,5%
Number of linkable cases	N=1,138	N=1,065	N=3,596

2. DATA

All data sets (SHARE and the administrative records) will be updated regularly, resulting in new release versions. A new release version of the administrative data will be published when either a new release of the SHARE data or a new version of the administrative records becomes available.

Hence, users have to ensure that they use the most recent release of the [data sets](#)⁴. The SHARE survey data and the administrative data sets are provided separately by the corresponding research data center. Figure 1 gives a detailed overview of the two data set providers and the available data sets.

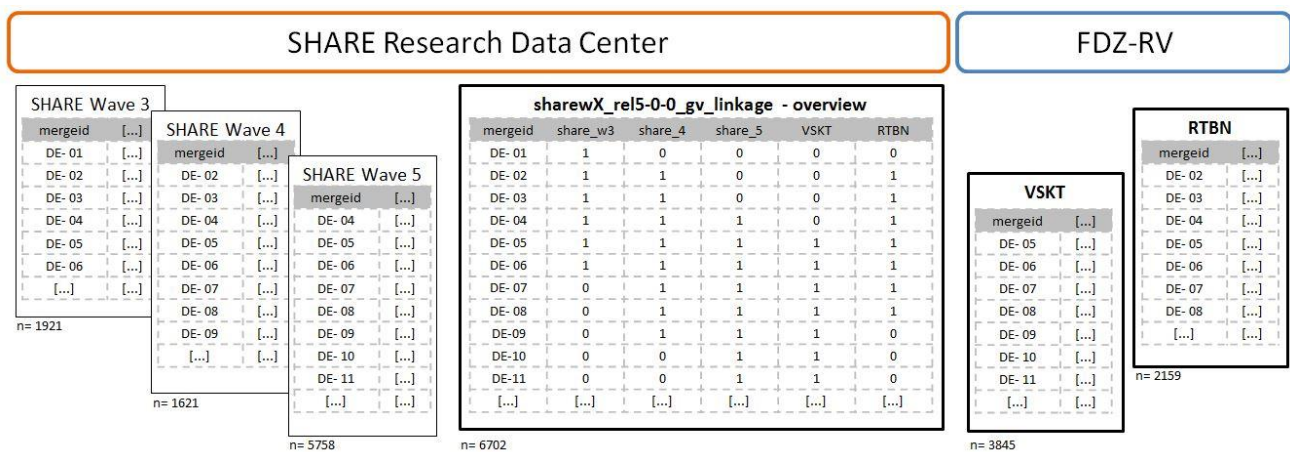


Figure 1 Data sets

2.1 SHARE

2.1.1 SURVEY DATA

SHARE is a multidisciplinary and cross-national panel database. It provides micro data on health, socio-economic status as well as social and family networks of about 123,000 individuals aged 50 or over and their (younger) partners. To date, SHARE has collected more than 293,000 interviews in five panel waves on current living circumstances and one wave on retrospective life histories (SHARELIFE). The first wave was collected in 2004/2005, the second in 2006/2007, SHARELIFE in 2008/2009, the fourth wave mainly in 2011 and the fifth wave in 2013. SHAREs sixth wave has been collected in 2015 and the data will be available in summer 2017. So far, 21 countries have participated in SHARE. However, not

⁴ For an overview of SHARE releases see: www.share-eric.eu/data-access-documentation.html

all countries participated in every wave and the timing of data collection can also slightly differ between countries⁵.

The German SHARE data collection started with the first wave. Until Wave 5 approximately 6,700 individuals have been interviewed at least once.

The SHARE Research Data Center provides the SHARE survey data separately for each wave including all countries. Therefore, the SHARE data should be reduced to German respondents only. A detailed documentation of the SHARE data can be found [here](#)⁶.

2.1.2 SHARE GV-LINKAGE MODULE - OVERVIEW

A generated data file "*sharewX_rel5-0-0_gv_linkage*", the so called 'overview' data set, includes only German respondents and was specifically generated for the project SHARE-RV.

This data set summarizes very useful [information](#)⁷ about the availability of the data, e.g. who consented and whether and which administrative data are available. It can be used in the early stages of analyses to find out for which respondents the administrative data are available.

Different from the standard SHARE modules, which are wave-specific, the 'overview' data set includes all respondents who participated in at least one wave between Waves 3-5. This implies that not all cases included in the 'overview' data set will merge when linking it to a specific wave of the SHARE data. E.g., merging the 'overview' data set with a German SHARE Wave 5 data results in 944 unmatched cases. These are respondents who did not participate in wave 5, such as the ones with the fictitious mergeid "DE-01", "DE-02" and "DE-03" (see [Figure 1](#)).

The variables "share_w3", "share_w4", and "share_w5" indicate in which SHARE waves the respondent participated. As the FDZ-RV provides two different data sources, the "[Versicherungskontenstichprobe \(VSKT\)](#)" and the "[Versichertenrentenbestand \(RTBN\)](#)", the variables "vskt" and "rtbn" summarize which data sets are available via FDZ-RV.

⁵For detailed information see: www.share-eric.eu/home/overview.html

⁶<http://www.share-eric.eu/data-access-documentation/documentation0.html>

⁷For detailed information see www.share-eric.eu/fileadmin/pdf_documentation/SHARE-RV/English_SHARE-RV_5-0-0_Codebook_gv_linkage.pdf

2.2 THE ADMINISTRATIVE RECORDS

2.2.1 VERSICHERTENKONTENSTICHPROBE (VSKT)

The data in the format of the VSKT is longitudinal and contains monthly information on respondents' employment histories from January of the year in which the respondent turned 14 until December of the year the respondent turned 65 (624 months in total). Each longitudinal variable is provided as a separate data set including 624 variables in wide-format which refer to one month each. The first variable (month 1) refers to the January the respondent turned 14, the second variable (month 2) the February the respondent turned 14 and so on. To assign the calendar time of each month, see [Chapter 4.4](#). The suffix of each data set corresponds to the names of the longitudinal variables. The VSKT contains information e.g. on the type of employment, earnings points for each month, and child-care periods. The data set 'fix' includes time-invariant information as demographics as well as accumulated pension entitlements. For detailed information please take a look at the [VSKT codebook](#)⁸.

2.2.2 VERSICHERTENRENTENBESTAND (RTBN)

The RTBN is cross-sectional and available for retirees only. It summarizes respondents' benefits which are accumulated until retirement. In addition, the amount of the pension which is paid is included. For detailed information please take a look at to the [RTBN codebook](#)⁹.

The two administrative data sets (VSKT and RTBN) are also not wave-specific but include all respondents who consented and whose administrative data are available. Again, not all cases will merge when combining the VSKT (the same holds for the RTBN) data set with any wave-specific SHARE data. The administrative data includes only those respondents whose data are available independently from the SHARE wave they participated in.

⁸ www.share-eric.eu/fileadmin/pdf_documentation/SHARE-RV/English_SHARE-RV_5-0-0_Codebook_VSKT.pdf

⁹ www.share-eric.eu/fileadmin/pdf_documentation/SHARE-RV/English_SHARE-RV_5-0-0_Codebook_RTBN.pdf

3. DATA ACCESS AND DOWNLOAD

Researchers must register for both datasets separately. Registered SHARE users will be informed about new releases and other important news automatically.

In order to access the SHARE data, researchers have to [register](#)¹⁰ as SHARE users at the [SHARE Research Data Center](#)¹¹. Terms and conditions of the Research Data Center can be found there as well. Upon registration, users are able to download SHARE data [here](#)¹². Once logged in, you can directly download the '[overview](#)' data set 'Linkage Release 5.0.0'. In order to get access to the administrative records of the FDZ-RV researchers also have to register as users at the FDZ-RV separately. The administrative data of the FDZ-RV which can be linked to the survey data will be provided on Compact Disk. Please follow the step-by-step instruction given in [Appendix A](#) to register as FDZ-RV user and request the administrative data. At the time of writing, the homepage is available in German only.

4. DATA HANDLING

4.1 DATA FORMATS AND LANGUAGES

All data is provided in Stata and SPSS format. For the use with other statistical programs, the data can be transferred by users themselves.

The data sets of the FDZ-RV are published in German. For English speaking users, a syntax for English labeling is provided via the Compact Disk of the FDZ-RV or via the [SHARE-RV website](#)¹³.

Please note that SHARE-Data and administrative RV-data are at least readable by Stata12.

4.2 HOW TO LINK THE DATA?

The administrative data are provided with the same identification number as used in the SHARE data (mergeid), so that it is linkable in the same way as all SHARE modules. [easySHARE](#)¹⁴ can be linked with the administrative data using the same identifier (mergeid) as well.

¹⁰ www.share-project.org/t3/share/fileadmin/pdf_documentation/SHARE_Data_Statement.pdf

¹¹ www.share-eric.eu/data-access-documentation/research-data-center-data-access.html

¹² www.cdata28.uvt.nl/sharedatadissemination/users/login

¹³ <http://www.share-eric.eu/data-access-documentation/record-linkage-share-rv.html>

¹⁴ <http://www.share-eric.eu/data-access-documentation/easyshare.html>

Although the project SHARE-RV started in the third wave of SHARE (SHARELIFE), the linkage is not limited to data of this wave. As all SHARE respondents have a stable identifier, the administrative data can also be linked with data of the same respondents of previous and following waves. As a consequence, the administrative data sets consist of records of respondents who may have participated in different SHARE waves.

SHARE's German target population consists of all persons aged 50 years and over at the time of sampling who have their regular domicile in Germany. In addition, the respondents' current partners living in the same household will also be interviewed, regardless of their age. Therefore, it is possible to do analyses not only on individual level but also on household level if both took part in the SHARE interview. In order to identify persons belonging to the same household please refer to the specific-wave household identifier (hhid*). More information could be found in the [SHARE User Guide](#)¹⁵, Chapter 7 'Merging the data'.

4.3 REPORTING YEAR

The variable 'reporting year' in the administrative data sets gives the user information about the year the data refers to. Given that it was not always possible to update the data for all respondents every year, the data set includes the most recent version available for each respondent. This means, for example, that the administrative data of the longitudinal data set includes information until in December 2013 if the reporting year is 2013.

4.4 How can I find out in which year and month an event happened based on VSKT?

Each longitudinal variable is in the same format (wide), consisting of 624 variables which refer to 624 months of respondents' biographies. These months always correspond to the same period in respondents' live: beginning in January of the year the respondent becomes 14 years old ('var'_1) until December of the year the respondent becomes 65 year old ('var'_624).

¹⁵ http://www.share-eric.eu/fileadmin/pdf_documentation/SHARE_guide_release_2-6-0.pdf

An example: Respondent Anna is born in March 1950. Therefore, the variable SES_1 refers to January 1964 since she turned 14 that year. She started her first job in month 77 (SES_77=13 for the first time → start=77). To find out the 'real' year and month of that event, the following formula should be used:

$$\begin{aligned} \text{Year} &= \text{year of birth} + 14 + \text{int}^{16}[(\text{start}-1)/12] \\ \text{Month} &= \text{start} - (\text{Year} - \text{year of birth} - 14) * 12 \end{aligned}$$

For our example this is:

$$\begin{aligned} \text{Year} &= 1950 + 14 + \text{int} [(77-1)/12] \\ &= 1964 + 6 \\ &= 1970 \\ \text{Month} &= 77 - (1970-1950-14)*12 \\ &= 5 \end{aligned}$$

We now know that the event-month 77 refers to May of 1970.

4.5 WHAT ARE THE EARNING POINTS EXACTLY?

Basically, earning points are based on the level of the income in each individual year. Thereafter, voluntary contributions will be added after they have been converted into remuneration as well as earnings points with a fixed value (e.g. for child-raising periods). For each year, individuals' income will be related to the average income of all insured persons. If the income corresponded to this average income, individuals will receive one point. If the income was less, they will receive less points and if the income was higher, they will receive more points. Income from employment in the new Federal States (East Germany) will be updated to the level of the old Federal States by using a predetermined factor. These points are called earning points (Deutsche Rentenversicherung Bund, 2012). For more information please refer to the [Pension Law](#)¹⁷.

¹⁶In SPSS the command is ,truncate', in STATA ,int' or 'trunc'

¹⁷http://www.deutsche-rentenversicherung.de/cae/servlet/contentblob/233424/publicationFile/50176/leben_und_arbeiten_europa_englisch.pdf

PLEASE NOTE: When using SHARE-RV data for publications, please add the following disclaimer to the obligatory SHARE acknowledgement. *“This paper uses data from SHARE-RV Rel. 5-0-0, released. SHARE-RV is funded by the Volkswagenstiftung and the Forschungsnetzwerk Alterssicherung (FNA).”*

The basic disclaimer for SHARE waves 1-5 can be found [here](#)¹⁸.

In addition, please do not forget to provide us with information about all publications using SHARE data by emailing to info@share-project.org.

Responsible for preparation of the SHARE-RV data and the documentation are:

Imke Herold

Sabine Friedel

Julie Korbmacher

Version: 5-0-0 (14.06.2016)

Contact: info@share-project.org with “SHARE-RV” in the subject

¹⁸<http://www.share-eric.eu/data-access-documentation/research-data-center-data-access.html>

APPENDIX A

1. Go to the [FDZ-RV website](#) and click on “Nutzungsantrag” > “zum Erstantrag”

The screenshot shows the website of the Deutsche Rentenversicherung Forschungszentrum. On the left is a navigation menu with categories like 'AKTUELLES', 'DATENBESCHREIBUNG', and 'NUTZUNGSANTRAG'. The 'NUTZUNGSANTRAG' category is highlighted in red, and its sub-items 'Erstantrag' and 'Ergänzungsantrag' are also highlighted in red. On the right, the 'Nutzungsantrag' section is visible, with sub-sections for 'Zugang zu den Datenprodukten', 'Erstantrag', and 'Ergänzungsantrag'. A dropdown menu under 'Erstantrag' is open, showing the option '... zum Erstantrag' highlighted in red.

2. [Click on](#) “Neuer Antrag” to get the form which has to be filled out.

The screenshot shows the 'Antrag zur wissenschaftlichen Nutzung von Mikrodaten' form. At the top, there is a navigation bar with 'Antragssteller', 'Weitere Personen', 'Projektbeschreibung', 'Datenart', 'Nutzungsform', and 'Übersicht'. Below this, there is a section titled 'Haben Sie Antragsdaten zwischengespeichert?' with a text box containing instructions. Below the text box is an input field for 'Antragsnummer:'. At the bottom, there are two buttons: 'Neuer Antrag' (highlighted in red) and 'Antragsdaten lesen'.

3. [Please provide the requested information on the next two screens and click on](#) “Weiter”.

4. [By choosing “Scientific Use File” and then “Themenfile”, it is possible to select “SHARE-RV”](#)

Deutsche Rentenversicherung
Forschungsdatenzentrum der Rentenversicherung

Antrag zur wissenschaftlichen Nutzung von Mikrodaten

Antragssteller Weitere Personen Projektbeschreibung **Datenart** Nutzungsform Übersicht

Welche Daten werden benötigt

Scientific Use File (SUF)

Basisfile

Themenfile

Datenangebot (Mehrfachauswahl möglich):

- Aktiv Versicherte und Rentengeschehen 2003 (Regionalfile)
- Aktiv Versicherte und Rentengeschehen 2001 (Regionalfile)
- BASID 2007
- SHARE-RV**
- Demographiedatensatz Rentenbestand/-wegfall
- Erwerbsminderung und Diagnosen 2014

weitergehender Datenbedarf

Public Use File (PUF)

Zurück Eingabe löschen Antragsdaten zwischenspeichern Weiter

5. [Click on “Weiter” to get an overview of your entries. Please check your personal data carefully. To finalize your application, click on “Antrag senden”, print, and sign the document before sending it to the FDZ-RV.](#)

Forschungsdatenzentrum der Rentenversicherung (0640)
Ruhrstraße 2
10709 Berlin

6. A Compact Disk with the data of the FDZ-RV will be sent to you free of charge.