

Abschließender Sachbericht

SOEP-Record-Linkage: Befragung von Zuwanderern der Sozialversicherungs- statistik im Längsschnitt (SOEP-REC-LINK) – Prozessdatenbasierte Ziehung sowie Verknüpfung mit Befragungsdaten zur Migration

Leibniz-Einrichtung: Sozio-oekonomisches Panel (SOEP) am DIW Berlin
Aktenzeichen: SAW-2013-SOEP-2
Projektlaufzeit: 01.06.2013 bis 28.02.2017
Ansprechpartner: Prof. Dr. Martin Kroh (Projektverantwortlicher)

Inhalt

Executive Summary	0
1. Ausgangsfrage und Zielsetzung des Vorhabens	1
2. Entwicklung der durchgeführten Arbeiten	1
3. Ergebnisse	13
4. Wirtschaftliche Verwertbarkeit der Ergebnisse	15
5. Kooperationspartner	15
6. Qualifikationsarbeiten	15
7. Liste der Publikationen aus dem Vorhaben	16
8. Verfügbarmachung Forschungsdaten.....	17
9. Liste möglicher Pressemitteilungen und Medienberichte	17

Executive Summary

Die Themen Migration und Integration haben in den letzten Jahren eine hohe Bedeutung in sozialpolitischen und wissenschaftlichen Debatten erlangt. Bereits vor der seit 2014 gestiegenen Fluchtzuwanderung lag der jährliche Zuzug nach Deutschland bei etwa einer Million Personen pro Jahr, wobei die Zuwanderung aus den neu beigetretenen EU-Staaten Mittel-, Ost- und Südosteuropas eine zunehmend wichtige Rolle einnahm. Das SOEP-Record-Linkage Projekt (SOEP-REC-LINK) liefert einen Beitrag zu dieser Debatte, indem es der Forschung hochwertige Daten zur Integration von Neuzuwanderern als Dateninfrastruktur zur Verfügung stellt.

Zu diesem Zweck wurden in einer Kooperation zwischen dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA) und dem Sozio-oekonomischen Panel (SOEP) am DIW Berlin in den Jahren 2013 sowie 2015 repräsentative Stichproben von Zuwanderer-Haushalten gezogen, die seither längsschnittlich befragt werden. Diese Daten stehen in integrierter Form mit den Befragungsdaten der anderen Stichproben des Sozio-oekonomischen Panel (SOEP) Forscherinnen und Forschern weltweit zur Analyse zur Verfügung. Außerdem wurden die Befragungsdaten – soweit die Einwilligung der StudienteilnehmerInnen vorlag – mit administrativen Daten der Bundesagentur für Arbeit zu Erwerbsverläufen der Befragten verknüpft. Diese Verknüpfung erlaubt die gleichzeitige Analyse sehr detaillierter registerbasierter Erwerbsinformationen und -biografien sowie differenzierter Angaben der StudienteilnehmerInnen aus der Befragung. Die erste IAB-SOEP Migrationsbefragung wurde 2013 erstmalig durchgeführt und ist derzeit in ihrem fünften Erhebungsjahr und die zweite IAB-SOEP Migrationsbefragung aus dem Jahr 2015 ist in ihrer dritten Welle. Insgesamt konnten in den jeweils ersten Erhebungsjahren 7.675 erwachsene Personen aus 3.819 Haushalten in Face-to-Face Befragungen interviewt werden. Diese sind entweder selbst zwischen 1995 und 2013 nach Deutschland eingewandert oder aber sie sind Kinder von Einwanderern und ab 1995 als junge Erwachsene in den Arbeitsmarkt eingetreten. Die erste Gruppe besteht vielfach aus EU-Migranten Süd- sowie Mittel- und Osteuropas und die zweite Gruppe vielfach aus Personen, deren Eltern aus den ehemaligen „Gastarbeiterländern“ stammen.

Das REC-LINK Projekt bestand aus vier Arbeitspaketen. In einem ersten Schritt des Projekts wurde erstmals für Deutschland eine Stichprobe von Zuwanderern bzw. deren Nachkommen aus den Registerdaten der Bundesagentur für Arbeit gezogen. Dafür wurde ein Ziehungsdesign entwickelt, das eine regionale Klumpung der vorhandenen Adressen für den Feldeinsatz erlaubte sowie die Zielpopulation mit den vorhandenen Registerdaten möglichst exakt eingrenzte. Das zweite Arbeitspaket ergab sich aus der notwendigen Einwilligung der Befragten zur Verknüpfung von Befragungs- und administrativen Daten. Dies wurde durch ein experimentelles Design zur Erhebung der informierten Einwilligung zur Verknüpfung begleitet, um Folgen der Erhebung der Einwilligung für das spätere Antwortverhalten zu kontrollieren. Im dritten Arbeitspaket wurden die Befragungsdaten mit den administrativen Daten direkt über die Sozialversicherungsnummer bzw. über Vor- und Nachname, Geschlecht, Adresse sowie das Geburtsdatum verlinkt. Schließlich wurden seit 2014 die Befragungsdaten durch das FDZ des SOEP sowie das FDZ des IAB als Infrastrukturleistung an Forschungsgruppen weltweit zur Verfügung gestellt. Nahezu 1.000 individuelle Forschende erhalten jedes Jahr alleine im Rahmen der regulären SOEP Datenweitergabe die Befragungsdaten der IAB-SOEP Migrationsbefragungen. Seit Ende 2016 stehen auch die verknüpften Befragungs- und Registerdaten externen NutzerInnen zur Verfügung. Das Projekt eröffnet neue Analysepotentiale sowohl für die sozial- und wirtschaftswissenschaftliche Forschung wie auch Politikberatung und hat für die – auch internationale – Stellung der Forschungsinfrastruktur in Deutschland hohe Bedeutung.

1. Ausgangsfrage und Zielsetzung des Vorhabens

Schon vor der seit 2014 stark angestiegenen Fluchtmigration zogen jährlich etwa eine Million Personen nach Deutschland. Geprägt wurde diese Zuwanderung zunehmend durch Personen aus den neu der EU beigetretenen Staaten Mittel-, Ost- und Südosteuropas. Auch vor dem Hintergrund der diesbezüglichen sozialpolitischen und wissenschaftlichen Debatten besteht ein Bedarf an belastbaren Befunden zur Arbeitsmarktintegration der aktuellen Neuzuwanderer. Diese unterscheiden sich von den klassischen Arbeitsmigranten in Deutschland u.a. durch ein hohes Humankapital, aber auch durch eine höhere regionale Mobilität innerhalb der EU. Vielfach arbeiten sie nur phasenweise in Deutschland. Administrative Daten erlauben zwar die Identifikation entsprechender Migrantengruppen und differenzierte Analysen von Erwerbsverläufen, liefern aber, aufgrund fehlender Kontextinformationen, kaum Daten zur substantiellen Erklärung von Migrations- und Integrationsprozessen. Paneldaten zu neuen Zuwanderern, die dies im Grundsatz erlauben würden, liegen nicht in ausreichendem Maße vor und beinhalten Lücken in der Detailliertheit der tatsächlichen Erwerbsverläufe. Die Gewinnung von Zufallsstichproben von Zuwanderern und die Erhebung von Befragungsdaten zu diesen Personengruppen werden u.a. durch das Fehlen einer geeigneten Auswahlgrundlage in Deutschland beschränkt. Die Einwohnermeldeämter erlauben nur indirekt und mit sehr hohem logistischen und Kostenaufwand, die interessierenden Personen in der Grundgesamtheit näherungsweise zu identifizieren. Das beantragte Projekt lieferte einen innovativen Vorschlag zur Gewinnung einer repräsentativen Stichprobe von längsschnittlichen Befragungsdaten von Neuzuwanderern bei gleichzeitiger hoher Qualität von Informationen zu den individuellen Erwerbsverläufen.

Zu diesem Zweck wurden in einer Kooperation des Sozio-oekonomischen Panels (SOEP) und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) erstmalig in Deutschland zwei Stichproben von Zuwanderer-Haushalten auf Basis der administrativen Daten der Bundesagentur für Arbeit (BA) gezogen. Diese Stichproben – die erste IAB-SOEP Migrationsbefragung wurde 2013 erstmals befragt und die zweite IAB-SOEP Migrationsbefragung im Jahr 2015 – sind als jährliche Wiederholungsbefragungen angelegt und sind derzeit (Sommer 2017) in ihrer fünften bzw. dritten Erhebungswelle. Außerdem können die Befragungsdaten zu einem Großteil der interviewten Personen mit den Erwerbsverläufen der administrativen Daten des IAB / der BA verknüpft werden. Die für die Verknüpfung notwendige informierte Einwilligung der Befragten wurde experimentell-kontrolliert erhoben, um Selektivitäten der realisierten Stichprobe prüfen zu können. Das Projekt eröffnet neue Analysepotentiale sowohl für die sozial- und wirtschaftswissenschaftliche Forschung wie auch Politikberatung und hat für die – auch internationale – Stellung der Forschungsinfrastruktur in Deutschland hohe Bedeutung.

2. Entwicklung der durchgeführten Arbeiten

Arbeitspaket 1: Stichprobenziehung

Die Ziehung von Stichproben von Zuwanderern und von in Deutschland geborenen Personen mit Migrationshintergrund stellt die Forschungsgemeinschaft in Deutschland seit jeher vor große Herausforderungen. Die größte Schwierigkeit besteht dabei im Fehlen einer geeigneten Auswahlgrundlage („Sampling-Frame“) der Zielpopulation von Menschen mit Migrationshintergrund. Zur Ziehung einer Stichprobe von AusländerInnen könnte prinzipiell das

Ausländerzentralregister (AZR) verwendet werden, jedoch ist der Datenzugang für ForscherInnen außerhalb des Bundesamts für Migration und Flüchtlinge restriktiv und praktisch schwer realisierbar. Auch enthält das AZR keine Daten zu Personen mit Migrationshintergrund, die eingebürgert sind beziehungsweise die seit Geburt die deutsche Staatsangehörigkeit besitzen. Dies waren 2015 laut Statistischem Bundesamt immerhin 55 Prozent der Personen mit Migrationshintergrund.¹ Die Migrations- und Integrationsforschung greift daher oftmals auf Einwohnermeldeamtsstichproben zurück, die, neben erheblichen Kosten, auch keine direkte Möglichkeit bieten, deutsche Personen mit Migrationshintergrund zu identifizieren.

Die Integrierten Erwerbsbiographien (IEB) als Auswahlgrundlage

Die Stichprobenziehungen der IAB-SOEP Migrationsbefragungen 2013 (M1) und 2015 (M2) basieren auf einem innovativen und mehrstufigen Stichprobendesign. Erstmals wurden im Kontext der Migrations- und Integrationsforschung dazu die sogenannten „Integrierten Erwerbsbiografien“ (IEB) der Bundesagentur für Arbeit als Auswahlgrundlage verwendet.

Die IEB werden vom Institut für Arbeitsmarkt- und Berufsforschung (IAB) erstellt und sind das Ergebnis einer Verknüpfung mehrerer prozessproduzierter Datenquellen. Ausgangspunkt der IEB sind Meldungen, die die Rentenversicherungsträger zu Beschäftigtenzeiten erfassen und an die Bundesagentur für Arbeit (BA) weiterleiten, sowie Informationen, die im Rahmen der Arbeitsmarktförderung und der Arbeitslosenversicherung durch die BA erfasst werden. Alle Meldungen liegen in Kontenform vor und enthalten tagesgenaue Angaben zu Erwerbszeiten. Ergänzt werden diese Informationen mit Personenmerkmalen und mit Leistungs- und Lohnangaben, die für den Zeitraum der jeweiligen Erwerbsmeldung gültig sind.

Die IEB bestehen aus der Gesamtheit der Personen, die im Beobachtungszeitraum mindestens einmal einen der folgenden Erwerbsstatus aufweisen:

- Sozialversicherungspflichtige Beschäftigung (erfasst ab 1975) und geringfügige Beschäftigung (erfasst ab 1999)
- Bezug von Leistungen nach dem Rechtskreis SGB III (erfasst ab 1975) oder SGB II (erfasst ab 2005)
- Bei der Bundesagentur für Arbeit (BA) als arbeitsuchend gemeldet (erfasst ab 2000)
- (Geplante) Teilnahme an einer arbeitsmarktpolitischen Maßnahme (erfasst ab 2000)

Diese aus unterschiedlichen Datenquellen stammenden Informationen werden in den IEB zusammengeführt und der Erwerbsstatus jeweils tagesgenau abgebildet. Sie bestehen aus ca. 87,9 Millionen Personen.

Informationen über aktuelle und vorherige ausländische Staatsangehörigkeiten der Personen sowie Informationen über die Teilnahme an für MigrantInnen ausgerichtete Maßnahmen der Arbeitsagentur ermöglichen eine weitgehend trennscharfe Identifikation von Menschen mit Migrationshintergrund in Deutschland. Die ebenfalls gespeicherten individuellen Adressinformationen schaffen darüber hinaus die Grundlage für einen zeitnahen und effizienten Feldzugang. Durch die Auswahl der IEB als Auswahlgrundlage werden zwei Personengruppen von der initialen Ziehung ausgeschlossen. Dazu gehören Beamte und Selbständige, die nie einer sozialversicherungspflichtigen Beschäftigung nachgegangen sind, nie

¹ Statistisches Bundesamt. (2016). Bevölkerung und Erwerbstätigkeit Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2015. *Fachserie 1 Reihe* 2.2, Tabelle 1.1.

Arbeitslosenleistungen erhalten haben sowie nie an arbeitsmarktpolitischen Maßnahmen teilgenommen haben. Auf Basis von Analysen mit dem Sozio-oekonomischen Panel (SOEP) wurde die Größe dieser Gruppen auf unter 10% der Zielpopulation geschätzt. Aufgrund der Konzeption der Befragung als Haushaltsbefragung, in der neben aus der IEB gezogenen Ankerpersonen auch alle anderen Haushaltsmitglieder befragt werden, gelangen die genannten Personengruppen jedoch als weitere Haushaltsmitglieder in die Stichprobe.

IAB-SOEP-Migrationsbefragung 2013 (M1)

Im Fokus der IAB-SOEP Migrationsbefragung 2013 (M1) standen erwachsene MigrantInnen der ersten Generation, die nach 1995 nach Deutschland zugewandert sind sowie erwachsene MigrantInnen der zweiten Generation, also Kinder von nach Deutschland zugewanderten Personen. In einem ersten Schritt wurden die deutschlandweiten Adressen der Zielpersonen in den IEB in 6.725 Regionaleinheiten zusammengefasst, aus denen wiederum 250 Regionaleinheiten für den Feldeinsatz zufällig gezogen wurden. Zur Identifikation von Menschen mit Migrationshintergrund wurden bei der Ziehung neben Informationen aus den IEB wie der Nationalität auch Informationen über die regionale Herkunft von Vor- und Nachnamen (Onomastik) herangezogen. In den 250 Regionaleinheiten des Feldeinsatzes wurden in einem nächsten Schritt je 80 Adressen zufällig gezogen. Die Ziehung basierte dabei auf einem Algorithmus, der anhand der geographischen Lage der Adressen einen sogenannten „Random Walk“ simuliert und damit eine Zufallsstichprobe darstellt, jedoch eine für den Feldeinsatz geeignete Stichprobe mit kurzem Anfahrtswegen für InterviewerInnen garantiert.

Tabelle 1: Haushalte der IAB-SOEP Migrationsbefragung 2013 (M1) nach Herkunftsgruppen

Herkunft	Haushalte der Nettostichprobe		Relative Ziehungswahrscheinlichkeit
	n	%	
Italien	202	7.4	++
Spanien & Griechenland	229	8.4	++
Türkei	381	14.0	–
Ehemaliges Jugoslawien	262	9.6	+
Spätaussiedler	347	12.7	–
Polen	269	9.9	o
Rumänien	189	6.9	++
Frühere GUS	367	13.5	o
Arabische und andere musl. Staaten	167	6.1	–
Restliche Welt	310	11.4	–
Insgesamt	2723	100	

Quelle: IAB-SOEP-Migrationsbefragungen 2013 (M1). Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1.

Im Zuge der Ziehung erhielten bestimmte Herkunftsgruppen außerdem eine höhere Ziehungswahrscheinlichkeiten, um hinreichend große Fallzahlen für spezifische Gruppen zu

gewährleisten. Dazu gehören insbesondere EU-BürgerInnen aus Südeuropa, aber auch Personen aus den neuen EU-Mitgliedsländern (siehe Tabelle 1). Im Zuge der IAB-SOEP Migrationsbefragung 2013 (M1) wurden im ersten Befragungsjahr insgesamt 4.964 persönliche Interviews in 2.723 Haushalten durch professionelle InterviewerInnen durchgeführt.

IAB-SOEP-Migrationsbefragung 2015 (M2)

Im Fokus der zweiten Stichprobe, der IAB-SOEP Migrationsbefragung 2015 (M2), standen Zugewanderte der letzten Jahre ab 2011. Auf Basis der IEB wurden zunächst alle Personen identifiziert, die seit dem 1. Januar 2011 zum ersten Mal in den IEB gelistet und seitdem mindestens einmal mit einer ausländischen Staatsangehörigkeit im Register geführt wurden. In einem weiteren Schritt wurden die Adressen der so ausgewählten Personen zu Regionaleinheiten zusammengefasst, die sich aus der Kombination von Gemeinden und Postleitzahlen ergeben. Aus diesen wurden 125 Regionaleinheiten zufällig gezogen. In jeder Regionaleinheit wurden dann jeweils 80 Adressen zufällig ausgewählt wobei bestimmte Herkunftsgruppen überproportional in die Stichprobe gezogen wurden. Dies gilt insbesondere für Zugewanderte aus Bulgarien, Rumänien sowie Personen aus Südeuropa (siehe Tabelle 2). Die IAB-SOEP Migrationsbefragung 2015 (M2) umfasste im ersten Befragungsjahr 1.689 persönliche Interviews in 1.096 Haushalten.

Tabelle 2: Haushalte der IAB-SOEP Migrationsbefragung 2015 (M2) nach Herkunftsgruppen

Herkunft	Haushalte in Nettostichprobe		Relative Ziehungswahrscheinlichkeit
	n	%	
Polen	132	12.0	++
Rumänien	109	10.0	++
Italien	44	4.0	+
Bulgarien	69	6.3	+
Ungarn	37	3.4	o
Portugal & Spanien	45	4.1	+
Griechenland	23	2.1	+
Türkei	9	0.8	–
Arabische und andere musl. Staaten	147	13.4	o
Westeuropa (Rest)	100	9.1	o
Osteuropa (Rest)	220	20.1	o
Restliche Welt	152	13.9	–
Insgesamt	1096	100	

Quelle: IAB-SOEP-Migrationsbefragungen 2015 (M2). Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1.

Um die überdurchschnittliche Berücksichtigung bestimmter Gruppen im Ziehungsdesign sowie zusätzlich die unterschiedliche Teilnahmebereitschaft bei den Analysen berücksichtigen zu können, werden für beide Migrationsbefragungen Gewichtungsfaktoren zur Verfügung gestellt.

In einem ersten Schritt wurden sogenannte Designgewichte erstellt. Designgewichte gleichen die im Ziehungsdesign vergebenen unterschiedlichen Ziehungswahrscheinlichkeiten bestimmter Migrationsgruppen („disproportionales Ziehungsdesign“) aus. Da nicht alle Personen in der Stichprobe auch bereit waren an der Befragung teilzunehmen, wurden in einem zweiten Schritt zusätzlich Ausfallgewichte berechnet. Basierend auf Informationen über die Personen der gesamten Bruttostichprobe, d.h. teilnehmende und nicht teilnehmende Haushalte, werden auf Basis multivariater Analyseverfahren diejenigen Personengruppen ermittelt, die eine vergleichsweise geringe Wahrscheinlichkeit der Teilnahme an der Befragung aufweisen. Dabei werden sowohl Informationen aus den IEB, Beobachtungen der InterviewerInnen zum Wohnumfeld einer Adresse wie auch die regionalen Datenbanken des Statistischen Bundesamtes herangezogen.² Personen in der Nettostichprobe mit einer niedrigen Teilnahmewahrscheinlichkeit werden dementsprechend mit einem höheren Gewicht versehen. Die Kombination von Designgewichten und Ausfallgewichten wird in einem dritten Schritt anhand der Daten des Mikrozensus auf die zugrundeliegende Zielpopulation hochgerechnet. Dies erlaubt den Datennutzern eine unverzerrte Schätzung von Populationsparametern. Im Hinblick auf die Nutzerfreundlichkeit wurden zwei Versionen von Gewichtungsfaktoren erstellt. Die erste Version erlaubt die isolierte Analyse der IAB-SOEP Migrationsbefragungen 2013 (M1) und 2015 (M2). In der zweiten Version wurde eine Integration beider Stichproben in die bestehenden Stichproben des Sozio-oekonomischen Panels (SOEP) vorgenommen. Die Befragungsdaten können somit auch komfortabel innerhalb eines erweiterten Kontexts – etwa in gemeinsamen Analysen mit früheren Migrationskohorten oder in Deutschland geborenen Personen - analysiert werden.

Arbeitspaket 2: Informed Consent und Teilnahmebereitschaft

Die Datennutzung und -verarbeitung von so genannten Sozialdaten – wie sie die Integrierten Erwerbsbiographien (IEB) des Institut für Arbeitsmarkt- und Berufsforschung (IAB) darstellen – erfordert die ausdrückliche, schriftliche und informierte Zustimmung der betreffenden Person (siehe SGB X). Aus diesem Grund wurden die Befragten der IAB-SOEP Migrationsbefragungen 2013 und 2015 (M1 und M2) am Ende ihres persönlichen Interviews um die Zustimmung zur Verknüpfung ihrer Befragungsdaten mit den Sozialversicherungsdaten des IAB gebeten. Dabei wurde einerseits die Wichtigkeit der Datenverknüpfung für die Forschung betont, andererseits jedoch auch die Freiwilligkeit der Zustimmung unterstrichen. Ein durch die InterviewerInnen überreichte Informationsbroschüre klärte die Befragten weiter über den Zweck der Datenverknüpfung und die Inhalte der zu verknüpfenden Registerdaten auf. Darüber hinaus wurde die Einhaltung der Datenschutzrichtlinien versichert. Zustimmungswillige Befragte wurden um Angabe ihres vollen Namens gebeten, welcher später als Schlüssel für die Datenverknüpfung genutzt wurde³ (siehe nächster Abschnitt). Anschließend wurde das datierte und unterschriebene Formblatt an den / die InterviewerIn zurückgegeben (für eine

² Für eine detaillierte Darstellung von Ziehungsdesign, Teilnahmeraten und Gewichtungsstrategie der IAB-SOEP-Migrationsbefragungen 2013 (M1) und 2015 (M2) siehe Kroh et al. (2015) sowie Kroh und Kühne (2017).

³ Die zunächst vorgesehene Verwendung der Sozialversicherungsnummer als Schlüssel wurde verworfen, da zu befürchten war, dass diese von den Befragten häufig nicht direkt und korrekt angegeben werden kann.

ausführlichere Beschreibung des Anfrageablaufes, siehe Eisnecker et al. 2017). Der Ablauf der Anfrage war für beide Stichproben und in allen Wellen identisch.

Trotz der zunehmenden Nutzung von verknüpften Befragungsdaten in bedeutenden Wiederholungsbefragungen war bisher ungeklärt, ob die Zustimmungsraten durch eine Verzögerung der Erstanfrage für Datenverknüpfung auf die zweite oder eine spätere Welle der Befragung erhöht werden kann. Vermutet werden könnte, dass Befragte zunehmendes Vertrauen in die Befragung entwickeln und somit zustimmungsbereiter werden. Ebenfalls nicht untersucht wurde bisher, ob die eventuell als sensitiv empfundene Anfrage auf Datenverknüpfung die Teilnahmebereitschaft in späteren Panelwellen absenkt. Aus diesem Grund wurde für die IAB-SOEP Migrationsbefragung 2013 (M1) ein experimentelles Forschungsdesign implementiert. Dieses teilt die StudienteilnehmerInnen in Gruppen auf, in denen die Frage der Einwilligung zur Datenverknüpfung entweder im ersten Jahr, im zweiten Jahr, im dritten Jahr der Befragung oder gar nicht implementiert wurde⁴ (Tabelle 3). Die Ergebnisse zeigen jedoch weder einen signifikant erhöhten Panelausfall in Folge der Anfrage 2013 noch eine verbesserte Zustimmungsraten durch die Verzögerung der Erstanfrage auf die zweite Welle 2014 (die Ergebnisse des Experiments wurden bereits in Eisnecker und Kroh (2016) veröffentlicht).

Tabelle 3: Einsatzplan Anfrage auf Datenverknüpfung nach Befragung und Befragungswelle

		Welle				N Realisierte Haushalte erste Befragungswelle (2013/2015)
		2013	2014	2015	2016	
IAB-SOEP-Migrationsbefragung 2013 (M1)	Kontrollgruppe	-	-	-	-	377
	Experimentalgruppe I	X	-	X*	X*#	525
	Experimentalgruppe II	-	X	X*	X*#	362
	Experimentalgruppe III	X	X*	X*#	X*#	1277
	Experimentalgruppe IV	-	-	X	X*	182
IAB-SOEP-Migrationsbefragung 2015 (M2)				X	X*	1096

Quelle: IAB-SOEP-Migrationsbefragungen 2013 (M1) und 2015 (M2), Wellen 2013 und 2015. Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1.

* Anfrage nur wenn Befragter nicht bereits in Vorwelle zugestimmt hat.

Anfrage nur wenn Befragter nicht bereits in Vorwellen zweimal oder häufiger abgelehnt hat.

⁴ Die ursprünglich angedachte experimentell variierte monetäre Incentivierung der Befragten, der Datenverknüpfung zuzustimmen, wurde nicht durchgeführt. Begründet lag dies darin, dass ExpertInnen und das Feldinstitut eine solche Incentivierung als nicht zielführend für die Erhöhung der Zustimmungsraten einschätzten.

Die kumulierten Zustimmungsraten bis 2016 fielen mit 62% aller Angefragten für M1 und 64% für M2 eher gering aus⁵ (Tabelle 4), was jedoch teilweise durch die insgesamt niedrigere Zustimmungsbereitschaft von MigrantInnen⁶ erklärt werden kann.

Tabelle 4: **Zustimmung Datenverknüpfung nach Befragung und Befragungswelle**

		Welle				Kumuliert
		2013	2014	2015	2016	
IAB-SOEP- Migrationsbefragung 2013 (M1)	N Zustimmung	1664	608	429	13	2713
	N Ablehnung	1673	751	447	31	1634
	% Zustimmung	49.9	44.7	49.0	29.6	62.4
IAB-SOEP- Migrationsbefragung 2015 (M2)	N Zustimmung	.	.	875	254	1129
	N Ablehnung	.	.	814	272	639
	% Zustimmung	.	.	51.8	48.3	63.9

Quelle: IAB-SOEP-Migrationsbefragungen 2013 (M1) und 2015 (M2), Wellen 2013 bis 2016. Personen ohne P-Fragebogen im Abfragejahr wurden ausgeschlossen. Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1 sowie Vorabdaten der Daten für das Jahr 2016, zukünftig veröffentlicht in Version 33.

Neben der Reduktion der Effizienz des Samples besteht durch die Ablehnung der Verknüpfung die Gefahr, dass eine selektive Zustimmung bestimmter Gruppen zur Verzerrung von Parameterschätzungen, die nur auf der zustimmenden Untergruppe beruhen, führen kann. Um diese Befürchtung zu überprüfen, wurden multivariate Regressionsmodelle zur Vorhersage der Verknüpfungszustimmung mit einer großen Anzahl von Prädiktorvariablen der Befragten, des Haushalts und der Wohnumgebung geschätzt (siehe Eisnecker et al. (2016)). Hierbei erwies sich jedoch nur ein sehr kleiner Teil der Prädiktoren als statistisch bedeutsam (Tabelle 5), was auf eine positiv zu wertende niedrige Selektivität der Zustimmung verweist. Um trotzdem die Gefahr von selektivitätsbedingter Verzerrung zu minimieren, wurden auf Grundlage der multivariaten Regressionsmodelle Gewichte zur Anpassung an die Grundgesamtheit erstellt. Für diese Gewichte kann gezeigt werden, dass ihre Anwendung die Selektivität eher oder weniger zustimmender Gruppen korrigiert ohne dabei nicht-selektive Gruppen signifikant zu beeinflussen. Die Gewichte werden den NutzerInnen neben den verknüpften Datensätzen zur Verfügung gestellt.

⁵ Für einen Überblick der Zustimmungsraten in anderen Befragungen siehe Sakshaug, Joseph W., Mick P. Couper, Mary Beth Ofstedal and David R. Weir. (2012). Linking Survey and Administrative Records: Mechanisms of Consent. *Sociological Methods & Research* 41: 535–69.

⁶ Siehe Beste, Jonas. (2011). Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten. Nürnberg: Institut für Arbeitsmarkt und Berufsforschung sowie Hartmann, Josef und Gerhard Krug. (2009). Verknüpfung von personenbezogenen Prozess- und Befragungsdaten—Selektivität durch fehlende Zustimmung der Befragten? *Zeitschrift für Arbeitsmarkt Forschung* 42: 121–39.

Tabelle 5: **Signifikante Effekte bei der Vorhersage der Verknüpfungszustimmung**

	Durchschnittlicher Marginaler Effekt	Sig.
Tertiäre Bildung	0.06	**
Einkommen aus Erwerbsarbeit		
<i>Ref: Einkommen angegeben</i>		
Kein Einkommen	-0.04	**
Einkommen verweigert	-0.11	**
Arabisches/Muslimisches Herkunftsland	-0.11	**
Spricht Deutsch schlecht/überhaupt nicht	-0.09	**
Haushaltseinkommen verweigert	-0.15	**
N Befragte	6,009	
N Haushalte	3,361	

Quelle: IAB-SOEP-Migrationsbefragungen 2013 (M1) und 2015 (M2), Wellen 2013 bis 2015. Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1.

** $p < 0.01$

Multivariates Logistisches Modell mit abhängiger Variable Verknüpfungszustimmung (0/1), Clusterung auf Haushaltsebene wurde berücksichtigt, Kumulierte Zustimmung bis Welle 2015.

Arbeitspaket 3: Record Linkage

Die Befragungsdaten der TeilnehmerInnen der IAB-SOEP Migrationsbefragungen, die der Verknüpfung ihrer Daten zugestimmt haben, wurden mit Daten aus den Integrierten Erwerbsbiographien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) Nürnberg verknüpft. Die IEB, die auch als Auswahlgrundlage beider Stichproben verwendet wurden, umfassen individuelle und tagesgenaue Informationen zu Erwerbsarbeit, Leistungsbezug und Arbeitsmarktförderungsmaßnahmen (zur IEB siehe auch Abschnitt *Arbeitspaket 1: Stichprobenziehung*). Im Folgenden wird das Verfahren zur Verknüpfung dargestellt sowie über Ergebnisse der Qualitätskontrolle des entstandenen Datensatzes berichtet.

Verknüpfung der Ankerpersonen sowie sonstiger Haushaltsmitglieder

Die Ankerpersonen, die aus den administrativen Daten gezogen wurden, können mit Hilfe eines auch in den IEB vorhandenen Indikators direkt wieder zu den administrativen Daten gespielt werden. Die weiteren Familienmitglieder, die somit nur indirekt über die Ankerperson aus der IEB gezogen wurden, müssen hingegen über ihre persönlichen Merkmale zu den Personen in den IEB gespielt werden. Dazu wurde mehrstufig vorgegangen: Als erster Schritt wurden die Adressdaten aus der Befragung und die Adressdaten aus den IEB einer Vereinheitlichung unterzogen. Im zweiten Schritt wurden die beiden Datensätze direkt verglichen und gleiche Datenzeilen zugeordnet. Im dritten Schritt wurden die Zuordnungen mit einem probabilistischen Merging für die verbleibenden Personen wiederholt. Hier wurden Treffer mit hoher Übereinstimmungswahrscheinlichkeit gefunden, die wegen Schreibfehlern oder leicht anderer Schreibweise sonst nicht gefunden worden wären. Dieses Verfahren ist speziell bei den hier vorliegenden ausländischen Namen geeignet, eventuelle alternative Schreibweisen zuordnen zu können.

Die oben genannten personenbezogenen Merkmale wurden bereinigt, um eine möglichst hohe Übereinstimmung zwischen beiden Datenquellen zu erreichen. Konkret wurden Umlaute und Sonderzeichen durch ihre jeweiligen ASCII-konformen Äquivalente ersetzt, alle Buchstaben in Großbuchstaben umgewandelt sowie Leerzeichen und Piktationszeichen entfernt. Mehrere Vor- bzw. Nachnamen wurden in einem Feld ohne Leerzeichen zusammengefasst. Neben diesen grundlegenden Bereinigungen wurden spezifische Routinen zur Standardisierung einzelner Variablen durchgeführt. Die Hausnummer wurde – falls notwendig – aus dem Straßenfeld in das Hausnummernfeld verschoben. Alle übrigen Adresszusätze (z.B. Stockwerk, Appartement-Nr.) wurden gelöscht. Bei Straßennamen wurden häufig vorkommende Muster und Abkürzungen ausgeschrieben (z.B. Straße statt Str.), bei Ortsnamen wurden Anhänge standardisiert (z. B. „am Rhein“ / a. Rhein / a. R.) bzw. entfernt (z.B. Stadtteile wie bei Berlin-Kreuzberg).

Nach der Standardisierung wurde ein exakter Abgleich der Felder zwischen den Einträgen der beiden Datenquellen für die Familienmitglieder durchgeführt. Im ersten exakten Linkage-Schritt war die vollständige Übereinstimmung in allen zur Verfügung stehenden Feldern ausschlaggebend, also Vor- und Nachname, Geburtsdatum, Geschlecht, Straße, Hausnummer, Postleitzahl und Ort. In den nächsten Schritten wurde jeweils bei noch nicht verknüpften Personen eine Abweichung in bestimmten Feldern erlaubt, während die jeweils übrigen Felder weiterhin exakt übereinstimmen mussten. Es wurden nur Abweichungen in Feldern zugelassen, bei denen das Risiko, zwei unterschiedlichen Personen zu verknüpfen (sog. falsch-positive Matches), sehr gering war. Insgesamt wurde die Übereinstimmung zwischen den beiden Datenquellen im Rahmen des exakten Record Linkage mit den in Tabelle 6 dargestellten Merkmalskombinationen geprüft.

Tabelle 6: **Merkmalskombinationen für das exakte Record Linkage**

	Exaktes Linkage	Exaktes Linkage ohne Hausnummer	Exaktes Linkage ohne Postleitzahl
Vorname	X	X	X
Geburtsdatum	X	X	X
Geschlecht	X	X	X
PLZ	X	X	-
Straße	X	X	X
Ort	X	X	X
Hausnummer	X	-	X

Personen aus der IAB-SOEP Migrationsstichprobe, die mit den obigen Methoden nicht verknüpft werden konnten, wurden mit Hilfe von fehlertoleranten Matching-Verfahren in den administrativen Daten gesucht.⁷ Dazu wurde für jede mögliche Verknüpfung zwischen den Datensätzen mit Hilfe eines probabilistischen Matching-Verfahrens ein Ähnlichkeitsmaß

⁷ zum Verfahren siehe Herzog, Thomas N., Fritz J. Scheuren und William E. Winkler. (2007). Data Quality and Record Linkage Techniques. New York: Springer.

berechnet und zu einem Qualitätsindex aufsummiert. Dazu wurde das Programm Merge ToolBox (MTB, Version 0.742⁸) verwendet.

Dieses Verfahren berücksichtigt, dass die Wahrscheinlichkeit einer Übereinstimmung von Merkmalen von deren Varianz in der Grundgesamtheit abhängig ist. Hierbei werden so genannte m- und u-Parameter berechnet, die angeben, mit welcher Wahrscheinlichkeit die Angaben zu einer Person zwischen zwei verschiedenen Quellen variieren bzw. wie wahrscheinlich das Auftreten eines identischen Merkmals bei unterschiedlichen Personen ist. Beispielsweise deutet die Übereinstimmung der Hausnummer zweier Personen weniger auf einen tatsächlichen Match hin als die Übereinstimmung des Nachnamens. Sie wird deshalb niedriger gewichtet. Die variablenspezifischen m- und u-Parameter, die dem probabilistischen Matching zugrunde liegen, wurden von Bachteler (2008)⁹ übernommen. Es wurden alle oben genannten Namens- und Adressmerkmale zwischen den beiden Datenquellen verglichen und ein additives Ähnlichkeitsmaß berechnet. Je höher der Wert dieses Index, desto größer die Gesamtähnlichkeit der beiden Zeilen und desto wahrscheinlicher liegt eine tatsächliche Übereinstimmung vor. Aus den zehn besten Übereinstimmungen pro Person wurde mit Hilfe eines Sichtvergleichs derjenige Treffer ausgewählt, der eine echte Übereinstimmung annehmen ließ.

Ergebnis und Qualität des Record Linkage

Die Verknüpfung der Daten wurde auf Grundlage der Personen vorgenommen, die sowohl ihr schriftliches Einverständnis zur Verknüpfung gegeben haben als auch mit einem gültigen Interview in den Daten vorhanden waren. Somit konnten über die beiden ersten Wellen der Stichprobe M1 gesamt 2.066 Personen in beiden Datensätzen gefunden werden, was für die erste Welle mit 1.552 Treffern eine Quote von 93 Prozent ergibt, in der zweiten Welle mit 514 Treffern eine Quote von 85 Prozent (Tabelle 7). Der niedrigere Wert in der zweiten Welle liegt an der relativ geringen Verknüpfungswahrscheinlichkeit bei Nicht-Anker Personen (70 %).

Aus dem Verknüpfungsprojekt entstand ein Datenpool aus qualitativ sehr hochwertigen administrativen Daten zur Berufs- und Einkommenslage von Haushalten kombiniert mit multidisziplinären, längsschnittlichen Befragungsdaten. Diese Daten wurden erstmals gezielt bei Personen mit Migrationshintergrund, einer sozial- und wirtschaftswissenschaftlich besonders relevanten Gruppe, erhoben. So entstand ein in der Breite und Tiefe sehr detaillierter Datensatz. Diese in Deutschland einmalige Kombination birgt nachhaltige Analysepotentiale für die Arbeitsmarkt- und Einkommensforschung und schließt erkennbare Datenlücken im Bereich der aktuellen Migrations- und Integrationsforschung.

⁸ Vergleiche Schnell, Rainer, Tobias Bachteler und Jörg Reiher. (2005). MTB: Ein Record-Linkage Programm für die empirische Sozialforschung. *ZA-Information* 56: 93-103 oder Bachteler, Tobias. (2012). Merge ToolBox – MTB Record Linkage Software, Version 0.742 - Getting Started. Nürnberg: German Record Linkage Center.

⁹ Bachteler, Tobias. (2008). Dokumentation Record Linkage IEB-PASS. mimeo.

Tabelle 7: Anzahl der verknüpften Beobachtungen nach Welle und Verknüpfungsmethode (M1)

Verknüpfungsmethode	Welle 2013			Welle 2014		
	Gesamt	Nicht-Anker	Anker	Gesamt	Nicht-Anker	Anker
Verknüpfung mit Indikator	887	-	887	302	-	302
Exakte Übereinstimmung	297	297	-	122	122	-
Exakt ohne Hausnummer	197	197	-	62	62	-
Exakt ohne Postleitzahl	4	4	-	0	0	-
Probabilistische Verknüpfung	167	167	-	28	28	-
Insgesamt verknüpft	1552	665	887	514	212	302
Grundgesamtheit	1664	739	925	608	305	303
Anteil verknüpfter Personen an der Grundgesamtheit	93%	90%	96%	85%	70%	100%

Quelle: IAB-SOEP-Migrationsbefragung 2013 (M1), Wellen 2013 und 2014, Adressdaten von Kantar Public und aus dem Data Ware House der Bundesagentur für Arbeit. Die Verknüpfung gilt nur dann als erfolgreich, wenn zu einer Personenummer in den administrativen Daten auch weitere Angaben vorhanden sind. Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2015, Version 32.1, SOEP, 2017, doi: 10.5684/soep.v32.1.

Bei einem Vergleich der Angaben zu den Personen in den beiden Datenquellen wurde eine sehr hohe Übereinstimmung in den betrachteten Eigenschaften erreicht. Es traten fast keine Abweichungen im Geschlecht der jeweils verlinkten Personen oder im Geburtsjahr auf (jeweils im einstelligen Bereich bei Ankerpersonen); nur unter 2 Prozent der Personen erschienen in den administrativen Daten vor dem Jahr der ersten Zuwanderung, das sie in der Befragung genannt hatten.

Arbeitspaket 4: Datenaufbereitung und -zugang

Zugang zu den Daten der IAB-SOEP Migrationsbefragungen

Die Befragungsdaten der IAB-SOEP Migrationsbefragungen 2013 (M1) und 2015 (M2) können am IAB und am SOEP über die an den Instituten üblichen Datennutzungsverträge bezogen werden. Darüber hinaus sind die Daten der Befragungen seit Welle 2013 Teil der jährlichen SOEP-Datenlieferung an die vertraglichen Daten-NutzerInnen (pro Jahr nahezu 1.000 individuelle Forschende).

Die Zugänglichkeit der verknüpften Befragungs- und Registerdaten unterliegen hohen datenschutzrechtlichen Auflagen des Sozialdatenschutzes. Verschiedene Zugangswege für externe Forscherinnen und Forscher wurden im Rahmen des REC-LINK Projekts geprüft und soweit datenschutzrechtlich möglich auch umgesetzt.

Die verknüpften, mit §75 SGB-X-Antrag zugänglichen Daten (IAB-SOEP-MIG-ADIAB)

Ein Datensatz mit dem vollen Umfang der SOEP und der IEB-Merkmale sowie der Möglichkeit, Daten aus dem Betriebspanel hinzuzuspielen, wurde am FDZ des IAB erstellt und steht über den Nutzungsweg „Gastaufenthalt“ in einem der gesicherten Standorte mit anschließender Fernverarbeitung vom eigenen Arbeitsplatz aus über die Software JoSuA zur Verfügung.¹⁰

Zur Erhöhung der Nutzerfreundlichkeit wurden die IEB-Daten für den verknüpften Datensatz am FDZ der BA am IAB in das bewährte SIAB-Format umgewandelt.¹¹ Die Originaldaten der IEB¹² weisen z.T. Variablen mit Hunderten von Ausprägungen auf und sind für Datenanalysen nur nach umfangreicher Einarbeitung nutzbar.

Die Nutzung des IAB-SOEP-MIG-ADIAB-Datensatzes kann seit Dezember 2016 beantragt werden. Zum Stand April 2017 liegen knapp 10 NutzerInnenverträge vor.

Datenschutzkonzept für einen verknüpften Scientific Use File (SUF)

Um der Wissenschaft eine Datenversion zur Verfügung zu stellen, die niedrigschwelliger zugänglich ist als der IAB-SOEP-MIG-ADIAB, wurde an einem Konzept zur Erstellung eines faktisch anonymisierten Scientific Use File gearbeitet. Dieser sollte ohne Antrag nach §75 SGB X und über beide FDZ von IAB und SOEP bezogen werden können. Zu diesem Zweck wurden Vergrößerungskonzepte, z.B. für eine Spelltyp-abhängige Umwandlung von tages- in monatsgenaue Daten, entwickelt. Parallel dazu gab es mehrere Treffen und Telefonkonferenzen, in denen über den Merkmalskranz eines solchen verknüpften SUF beraten wurde.

Im Rahmen des REC-LINK Projekts fand zudem im November 2015 am DIW Berlin ein 2-tägiger Workshop statt, der dem Austausch wissenschaftlicher Ergebnisse und Standpunkte zu Sozialdatenschutz, faktischer Anonymisierung und Deanonymisierungsrisiken diente. TeilnehmerInnen waren Akteure verschiedener Seiten: neben ExpertInnen für statistische Geheimhaltung und WissenschaftlerInnen mehrerer datenherausgebender Institute waren Datenschutzbeauftragte, das Bundesministerium für Arbeit und Soziales und die beiden Produzenten von Registerdaten, die Deutsche Rentenversicherung und das IAB/Bundesagentur für Arbeit, vertreten. Da Befürchtungen, dass "Big Data"-Bestände aktuell oder in Zukunft für eventuelle Datenangriffe genutzt werden könnten, weiterhin unter Datenschutzbeauftragten bestehen, ist ein Zugang zu den anonymisierten, verknüpften Daten von IEB sowie IAB-SOEP Migrationsbefragung (M1 und M2) weiterhin nicht über das FDZ des SOEP möglich.

Darüber hinaus wurde eingehend geprüft, ob die Ergänzung fehlender Befragungsdaten auf Grundlage der IEB Registerdaten datenschutzrechtlich möglich ist. Dies hätte zum einen Informationslücken betroffen, welche sich aus fehlenden Angaben aufgrund von Unkenntnis

¹⁰ Standorte und Nutzungsbedingungen siehe http://fdz.iab.de/de/FDZ_Data_Access/FDZ_On-Site_Use/standorte.aspx und http://fdz.iab.de/de/FDZ_Data_Access/FDZ_Remote_Data_Access.aspx

¹¹ Dokumentation aller Merkmale des Datensatzes IAB-SOEP-MIG-ADIAB siehe Trübswetter, Parvati und Tanja Fendel: IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB, Version 1 (IAB-SOEP-MIG-ADIAB 7514, Version 1), FDZ Datenreport 11/2016, http://doku.iab.de/fdz/reporte/2016/DR_11-16.pdf

¹² Am FDZ des SOEP sind die IAB-SOEP-Migrationsdaten Teil der SOEP-Datendistribution. Die Nutzerzahlen geben daher keinen Aufschluss über das Interesse speziell an den IAB-SOEP-Migrationsdaten.

oder Antwortverweigerung der Befragten ergeben. Zum anderen sollten Informationen aus den Registerdaten zur Nachbildung von Angaben herangezogen werden, die üblicherweise zum Befragungsprogramm des SOEP gehören, aber von den TeilnehmerInnen der IAB-SOEP Migrationsbefragung wegen umfangreicher Sonderfragen nicht erhoben wurden. Damit waren insbesondere die Monatskalendarien (Aktivitäten im letzten Jahr) und die Biografie seit dem 15. Lebensjahr gemeint. Die um Registerdaten ergänzten Befragungsdaten sollten dann den Nutzerinnen und Nutzern mit dem SOEP-üblichen Verfahren (Datennutzungsvertrag) zur Verfügung gestellt werden.¹³ Leider zeigte sich, dass trotz intensiver Bemühungen um ein Datenschutzkonzept die Bereitstellung solcherart aufgefüllter Befragungsdaten aufgrund der hohen Hürden des Sozialdatenschutzes nicht möglich ist.

3. Ergebnisse

Die IAB-SOEP Migrationsbefragungen (M1 und M2) bilden eine derzeit einmalige Dateninfrastruktur, um die Integration von MigrantInnen nachzuvollziehen, wobei die Arbeitsmarktintegration der in jüngerer Zeit zugewanderten EU-MigrantInnen im Fokus steht. Eine Reihe von Policy-Berichten, die vielfach in den Medien rezipiert wurden wie auch peer-reviewed Journal Publikationen, sind durch die Projektbeteiligten entstanden. Darüber hinaus werden die Daten von einer Vielzahl nationaler und internationaler Forscher-Teams verwendet.

Die survey-methodischen Befunde im Rahmen eines experimentellen Designs ergeben keine Hinweise darauf, dass die Verknüpfungsanfrage 2013 zu einem erhöhten Panelausfall in den Folgejahren führte. Dies ist als positiv für den weiteren Einsatz von Datenverknüpfungen im SOEP und in anderen Panelbefragungen zu werten. Weiter fand sich auch kein signifikanter Zusammenhang zwischen der Platzierung der Erstanfrage in Welle 2013 oder 2014 und der Zustimmungsbereitschaft zum Linkage. Da Panelbefragungen fast immer von Welle zu Welle Befragungspersonen verlieren, erscheint eine möglichst frühe Platzierung der Erstanfrage auf Datenverknüpfung daher empfehlenswert, um den Anteil der Zustimmungen zu maximieren. Unsere Ergebnisse stellen somit wichtige Orientierungspunkte für die Durchführenden von Wiederholungsbefragungen dar. Auch für die Zukunft bietet das experimentelle Design weitere interessante Analysemöglichkeiten; etwa im Hinblick auf die längerfristigen Auswirkungen (wiederholter) Abfragen.

Die Zustimmung zur Datenverknüpfung war bisher durch eine niedrige Selektivität im Hinblick auf verschiedenste untersuchte Befragten-Charakteristika gekennzeichnet. Um das Risiko von Verzerrungen durch Selektivitätseffekte weiter abzusenken, wurde eine Gewichtungvariable für die Anwendung durch die Datennutzer erstellt. Trotz einer relativ niedrigen Zustimmungsrate in den Einzelwellen konnte bis zur Welle 2016 durch wiederholte Anfrage eine zufriedenstellende kumulierte Zustimmungsrate von mehr als 60% der Angefragten erreicht werden.

¹³ Von den 12 Aktivitätstypen der SOEP-Kalendarien sind 6 aus den BA-Registerdaten rekonstruierbar: Vollzeit- und Teilzeiterwerbstätigkeit, geringfügige Beschäftigung, betriebliche Ausbildung, Arbeitslosigkeit, Aus- und Weiterbildung. Da es sich auf der einen Seite um selbst berichtete Aktivitäten handelt und auf der anderen Seite um Verwaltungsdaten, die im Rahmen von sozialversicherungsrelevanten Prozessen anfallen, überrascht es nicht, dass der Bezugsrahmen von Register- und Befragungsdaten nicht deckungsgleich ist. So beziehen sich die Registerdaten auf sozialversicherungspflichtige Beschäftigung, Leistungsbezug und Maßnahmen der Arbeitsförderung, während die Befragten auch Episoden außerhalb dieser Grenzen berichten. Ihre Angaben können jedoch durch Erinnerungsfehler verzerrt sein.

Erste Vergleiche von Register- und Befragungsdaten zeigen Abweichungen, die Basis für weitere survey-methodische Forschung sein können.¹⁴ Diese müssen jedoch nicht notwendigerweise auf Fehler in der Befragung hindeuten. Zwar gelten im Allgemeinen Registerdaten im Vergleich zu Befragungsdaten als eine zuverlässigere Datenquelle, da Befragungsdaten von Erinnerungsfehlern und anderen Verzerrungen getrübt sein können, im Gegensatz zu Registerdaten, die "hard facts" repräsentieren. Zumindest im Fall der IEB, die aus verschiedenen Quellen zusammengestellt werden (Meldungen der Arbeitgeber nach DEÜV, Bundesagentur f. Arbeit, Jobcenter, Kommunale Träger), gibt es innerhalb der Verläufe (in den IEB "Konten" genannt) jedoch durchaus auch widersprüchliche Angaben. Demnach bestätigt sich das Ergebnis von Reimer und Künster (2004) auch in unserem Verknüpfungsjekt: Weder die Befragungsdaten noch die Registerdaten können umstandslos zur Validierung der jeweils anderen Datenquelle herangezogen werden. Eine Integration der nebeneinander stehenden Verläufe der beiden Datenquelle wäre nur auf Basis detaillierter Feinanalysen unter Berücksichtigung der Forschungsziele möglich und sinnvoll.¹⁵

Neben diesen Beiträgen zur survey-methodischen Forschung ermöglichten die erhobenen Daten der IAB-SOEP Migrationsbefragungen bereits eine Fülle von inhaltlicher migrationsbezogener Forschung, deren Erkenntnisse hier nur auszugsweise wiedergegeben werden können:

- Traditionelle Muster der Migration, in denen Zuwanderer nach dem Zuzug dauerhaft ihren Lebensmittelpunkt in ein neues Zielland verlagern, werden zunehmend durch wiederholte Migrationsbewegungen über den Lebensverlauf hinweg ersetzt (Brücker et al. 2014b).
- Der Zuzug nach Deutschland verdoppelt im Durchschnitt die Verdienste von erwerbstätigen MigrantInnen, wobei insbesondere deutsche Sprachkenntnisse und die Anerkennung beruflicher Abschlüsse Schlüsselfaktoren für eine erfolgreiche Arbeitsmarktintegration sind (Brücker et al. 2014a).
- Für die Stellenfindung von Geflüchteten spielten in der Vergangenheit sowohl formelle Wege wie Arbeitsagenturen und Stellenanzeigen als auch die sozialen Netzwerke von Familie, Freunden und Bekannten eine große Rolle (Eisnecker und Schacht 2016).
- Die Mehrheit der in der Vergangenheit nach Deutschland Geflüchteten brachte bereits Berufserfahrung aus dem Ausland mit, aber nur ein Teil einen formellen Berufsabschluss (Liebau und Salikutluk 2016).
- Geflüchtete kamen in der Vergangenheit meist ohne Deutschkenntnisse in der Bundesrepublik an, ihr Niveau glich sich jedoch im Laufe der Zeit an jenes anderer MigrantInnengruppen an (Liebau und Schacht 2016).
- MigrantInnen steigerten nach ihrem Zuzug ihre deutsche Sprachkompetenz erheblich. Zusätzlich erwarb ein knappes Drittel der Zugewanderten weitere berufsqualifizierende Abschlüsse in Deutschland (Liebau und Romiti 2014).
- Im Vergleich zu anderen MigrantInnengruppen nahmen die Geflüchteten der vergangenen Jahre später eine Erwerbstätigkeit auf, waren auch Jahre nach der Zuwanderung noch häufiger erwerbslos und verdienten weniger (Salikutluk et al. 2016).

¹⁴ vgl. Reimer, Maike und Ralf Künster. (2004). Linking Job Episodes from Retrospective Surveys and Social Security Data: Specific Challenges, Feasibility and Quality of Outcome. *Arbeitspapiere des Projekts Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*, Nr. 8/2004.

¹⁵ Reimer und Künster (2004: 24f.)

- Mehr als die Hälfte der MigrantInnen in Deutschland gaben an, dass sie schon Benachteiligung aufgrund ihrer Herkunft erfahren haben. Am häufigsten berichteten Personen türkischer und arabisch-muslimischer Herkunft von Diskriminierung, wobei Benachteiligung besonders oft bei der Ausbildungs- und Ausbildungsplatzsuche sowie bei Behörden erfahren wurde (Tucci et al. 2014).

4. Wirtschaftliche Verwertbarkeit der Ergebnisse

Die Daten wurden zu Forschungszwecken erhoben und sind auch Forschenden im In- und Ausland zugänglich. Darüber hinaus haben Ergebnisse, die auf den Daten beruhen, vielfach Eingang in die Politikberatung und die Medien gefunden. Auf eine wirtschaftliche Verwertung war das Projekt nicht ausgerichtet.

5. Kooperationspartner

Institut für Arbeitsmark- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)

Das REC-LINK Projekt wurde in Kooperation mit dem IAB durchgeführt. Das IAB war insbesondere zuständig für die Arbeitspakete 3 sowie 4, also der Verlinkung der Befragungs- mit den Registerdaten und der Bereitstellung der verlinkten Daten an internationale Forscherinnen und Forscher. Die Beiträge von SOEP und IAB sind gemeinsam beschrieben unter Pkt. 2 und 3.

German Record Linkage Center (German RLC)

Das German RLC, siehe: <http://www.record-linkage.de/-Projects.htm> hat das Projekt durch die Bildung von regional-geklumpten Sample Points auf Basis der georeferenzierten Adressen der IEB unterstützt.

Berliner Institut für Empirische Integrations- und Migrationsforschung (BIM)

Eine Kooperation fand insbesondere statt mit dem am BIM-Forschungs-Interventions-Cluster "Solidarität im Wandel?" angesiedelten Teilprojekt "Geschlechtsspezifische Untersuchung zur sozialen Lage von Migrant*innen auf Basis von SOEP und Mikrozensus mit Fokus auf die Situation von Menschen mit Fluchterfahrung.", siehe https://www.bim.hu-berlin.de/de/projekte/2016/04/01/2016/12/31/geschlechtsspezifische-untersuchung-zur-sozialen-lage-von-migrant*innen-auf-basis-von-soep-und-mikrozensus-mit-fokus-auf-die-situation-von-menschen-mit-fluchterfahrung/

6. Qualifikationsarbeiten

Im Rahmen des REC-LINK Projekts wurden zwei Promotionsprojekte entwickelt. Das Projekt von Simon Kühne (DIW Berlin) befasst sich mit Interviewereffekten in persönlichen Befragungen und das Projekt von Philipp Eisnecker (DIW Berlin) mit privaten Kontakten zwischen MigrantInnen und nicht-MigrantInnen. Beide Dissertationsprojekte sind an der Berlin Graduate

School of Social Sciences der Humboldt-Universität zu Berlin angesiedelt und sollen bis Mitte 2018 abgeschlossen sein.

Auf Basis der IAB-SOEP Migrationsbefragungen, die seit 2014 Teil der regulären Datenweitergabe des SOEP an nahezu 1.000 individuelle ForscherInnen weltweit sind, werden neben Dissertationen auch eine hohe Zahl an BA und MA Arbeiten verfasst. Eine konkrete Zahl liegt uns nicht vor.

7. Liste der Publikationen aus dem Vorhaben

Methodische Publikationen

- Brücker, Herbert, **Martin Kroh**, Simone Bartsch, **Jan Goebel**, **Simon Kühne**, **Elisabeth Liebau**, **Parvati Trübswetter**, **Ingrid Tucci** und Jürgen Schupp. (2014). The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Contents. *SOEP Survey Papers* No. 216 - Series C: Data Documentation.
- **Eisnecker, Philipp** und **Martin Kroh**. (2016). The Informed Consent to Record Linkage in Panel Studies: Optimal Starting Wave, Consent Refusals, and Subsequent Panel Attrition. *Public Opinion Quarterly* 80: 1-13.
- **Eisnecker, Philipp**, **Klaudia Erhardt**, **Martin Kroh** und **Parvati Trübswetter**. (2017). The Request for Record Linkage in the IAB-SOEP Migration Survey. *SOEP Survey Papers* No. 291 – Series C: Data Documentation.
- **Kroh, Martin** und **Simon Kühne**. (2017). The 2015 IAB-SOEP Migration Sample (M2): Sampling Design and Weighting Adjustment. *SOEP Survey Papers – Series C: Data Documentation*, im Erscheinen.
- **Kroh, Martin**, **Jan Goebel**, **Simon Kühne**, Frederike Preu. (2015): The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment. *SOEP Survey Papers* No. 271 – Series C: Data Documentation.
- **Trübswetter, Parvati** und Tanja Fendel. (2016). IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB * Version 1 (IAB-SOEP-MIG-ADIAB 7514, Version 1). *FDZ-Datenreport* 11/2016.
- **Tucci, Ingrid** und Nicolas Legewie. (2016). Panel-basierte Mixed-Methods-Studien: Design, Feldzugang, Potentiale und Herausforderungen am Beispiel der Studie "Das Erwachsenwerden der Nachkommen von GastarbeiterInnen in Deutschland". *SOEPpapers* 872.

Inhaltliche Publikationen der Projektbeteiligten auf Basis der gewonnenen Daten

- Brücker, Herbert, Simone Bartsch, **Philipp Eisnecker**, **Martin Kroh**, **Elisabeth Liebau**, Agnese Romiti, Jürgen Schupp, **Parvati Trübswetter**, **Ingrid Tucci** und Ehsan Vallizadeh. (2014). Die IAB-SOEP-Migrationsstichprobe: Leben, lernen, arbeiten - wie es Migranten in Deutschland geht. *IAB-Kurzbericht* 21/2014.
- Brücker, Herbert, **Elisabeth Liebau**, Agnese Romiti und Ehsan Vallizadeh. (2014). Anerkannte Abschlüsse und Deutschkenntnisse lohnen sich. *DIW Wochenbericht* 43: 1144-1151.
- Brücker, Herbert, **Ingrid Tucci**, Simone Bartsch, **Martin Kroh**, **Parvati Trübswetter** und Jürgen Schupp. (2014). Neue Muster der Migration. *DIW Wochenbericht* 43: 1126-1135.

- **Eisnecker, Philipp**, Johannes Giesecke, **Martin Kroh**, **Elisabeth Liebau**, Jan Marcus, Zerrin Salikutluk, Diana Schacht, C. Katharina Spieß und Franz Westermaier. (2016). Die Integration Geflüchteter – Erkenntnisse aus der Vergangenheit: Editorial. *DIW Wochenbericht* 35: 727-730.
- **Eisnecker, Philipp** und Diana Schacht. (2016). Die Hälfte der Geflüchteten in Deutschland fand ihre erste Stelle über soziale Kontakte. *DIW Wochenbericht* 35: 757-764.
- **Kroh, Martin**, Zerrin Salikutluk, Diana Schacht, **Elisabeth Liebau** und Johannes Giesecke. (2017). Die Integration Geflüchteter in der Vergangenheit – Spracherwerb und Arbeitsmarkteinbindung. *Gesellschaft • Wirtschaft • Politik (GWP)* 1/2017: 29-36.
- **Liebau, Elisabeth**. (2016). Psychische Gesundheit und Lebensqualität von Zuwanderern und deren Nachkommen im SOEP in den Jahren 1984-2016. *Psychotherapie, Psychosomatik, medizinische Psychologie* 66: 393-396.
- **Liebau, Elisabeth** und Zerrin Salikutluk. (2016). Viele Geflüchtete brachten Berufserfahrung mit, aber nur ein Teil einen Berufsabschluss. *DIW Wochenbericht* 35: 732-740.
- **Liebau, Elisabeth** und Diana Schacht. (2016). Spracherwerb: Geflüchtete schließen zu anderen MigrantInnen nahezu auf. *DIW Wochenbericht* 35: 741-748.
- **Liebau, Elisabeth** und Agnese Romiti. (2014). Migranten investieren in Sprache und Bildung. *DIW Wochenbericht* 43: 1136-1143.
- Salikutluk, Zerrin, Johannes Giesecke und **Martin Kroh**. (2016). Geflüchtete nahmen in Deutschland später eine Erwerbstätigkeit auf als andere MigrantInnen. *DIW Wochenbericht* 35: 749-756.
- **Tucci, Ingrid**. (2016). Lebenssituation von Migranten und deren Nachkommen. In: *Datenreport 2016*. Bonn: Bundeszentrale für politische Bildung.
- **Tucci, Ingrid, Philipp Eisnecker** und Herbert Brücker. (2014). Wie zufrieden sind Migranten mit ihrem Leben? *DIW Wochenbericht* 43: 1152-1158.

8. Verfügbarmachung Forschungsdaten

Die Befragungsdaten der IAB-SOEP Migrationsstichprobe sind seit dem 12.12.2014 über das FDZ SOEP integriert in die SOEP Haupterhebung verfügbar (doi:10.5684/soep.v30). Bereits zwei Monate zuvor wurden die Befragungsdaten der isolierten Stichprobe über das FDZ IAB der BA bereitgestellt (doi:10.5684/soep.iab-soep-mig.2013). Die mit den Erwerbsverläufen der administrativen Daten des IAB / der Bundesagentur für Arbeit (BA) verknüpften Daten sind seit November im FDZ IAB der BA verfügbar (http://fdz.iab.de/de/FDZ_Individual_Data/iab-soep-mig.aspx).

9. Liste ausgewählter Pressemitteilungen und Medienberichte

Pressemitteilungen und Pressegespräche

- „Solidarität im Wandel? Aktuelle Forschungsergebnisse zur Integration von Geflüchteten.“ Flucht-Forschungs-Symposium des Berliner Instituts für empirische Integrations- und Migrationsforschung (BIM). Berlin, den 23.03.2017.

- „Investitionen in die Integration von Geflüchteten lohnen sich“. Pressemitteilung des Deutschen Instituts für Wirtschaftsforschung, 18.01.2017, https://www.diw.de/de/diw_01.c.100319.de/presse/pressemitteilungen/pressemitteilung_n.html?id=diw_01.c.550520.de
- „Studie zu früherer Fluchtzuwanderung zeigt Ansatzpunkte für Integration aktuell Geflüchteter“. Pressemitteilung des Deutschen Instituts für Wirtschaftsforschung, 01.09.2016, https://www.diw.de/de/diw_01.c.100319.de/presse/pressemitteilungen/pressemitteilung_n.html?id=diw_01.c.541848.de
- „Migration und Integration zahlen sich für Einwanderer aus“. Pressemitteilung des Deutschen Instituts für Wirtschaftsforschung, 17.10.2014, https://www.diw.de/de/diw_01.c.100319.de/presse/pressemitteilungen/pressemitteilung_n.html?id=diw_01.c.485505.de
- „IAB und DIW Berlin starten gemeinsame Befragung von Migranten – Kooperationsvertrag unterzeichnet“. Pressemitteilung des Deutschen Instituts für Wirtschaftsforschung, 03.06.2013, https://www.diw.de/de/diw_01.c.100319.de/presse/pressemitteilungen/pressemitteilung_n.html?id=diw_01.c.422079.de
- „Die Integration Geflüchteter: Erkenntnisse aus der Vergangenheit“. Pressegespräch des DIW Berlin. Berlin, den 01.09.2016.
- „Leben, lernen, arbeiten – wie es Migranten in Deutschland geht“. Gemeinsames Pressegespräch des IAB und des DIW Berlin. Bundespressekonferenz. Berlin, den 17.10.2014.

Presseberichte (Auswahl)

- "Für Flüchtlinge ist der Weg zum Erfolg besonders weit." Frankfurter Allgemeine Zeitung, 02.09.2016, S.19.
- "Die Sprache ist der Schlüssel." Süddeutsche Zeitung, 02.09.2016, S.15.
- "Traditionelle Rollen. Flüchtlinge Frauen arbeiten seltener." taz.die tageszeitung, 02.09.2016, S.6.
- "Die Tücken der Integration." Die Welt, 02.09.2016, S.10.
- "Studie: Schaffen wir das? Ein Blick auf die Flüchtlinge der Neunziger gibt Aufschluss." sueddeutsche.de, 01.09.2016.
- "Integration: Was wurde aus den Flüchtlingen der Neunziger?" ZEIT ONLINE, 01.09.2016.
- "Arbeitsmarkt: Bundeswehr bildet syrische Flüchtlinge aus." FAZ.net, 01.09.2016.
- "Arbeitsmarkt: Flüchtlinge verdienen weniger als andere Migranten." FAZ.net, 01.09.2016.
- "Arbeitsmarktstudie: Flüchtlinge verdienen deutlich weniger als andere Zuwanderer." Spiegel Online, 01.09.2016, S.0.
- "HINTERGRUND: Flüchtlinge finden ohne Hilfestellung nur schwer Arbeit." dpa-AFX Wirtschaftsnachrichten GmbH, 01.09.2016, S.0364.
- "Sprachkenntnisse zahlen sich für Migranten aus." Die Welt, 18.10.2014, S.9.
- "Einwandern lohnt sich." Frankfurter Allgemeine Zeitung, 18.10.2014, S.25.
- "Diskriminierte Migranten." Süddeutsche Zeitung, 18.10.2014, S.27.

- "Kontakte führen zu den meisten Jobs." Süddeutsche Zeitung, 18.10.2014, S.65.
- "Grüne fordern mehr Chancen für Migranten auf dem Arbeitsmarkt." faz.net, 18.10.2014.
- "Könnyen meg lehet duplázni a keresetet - rengeteg embernek bejött már." ma.hu, 18.10.2014.
- "Deutschlands glückliche Einwanderer." zeit.de, 17.10.2014.
- "Göçmenler dilbilgisi ve eğitime yatırım yapıyor." zaman-online.de, 17.10.2014.
- "Migranten leiden unter Diskriminierung auf dem Arbeitsmarkt." stern.de, 17.10.2014.
- "Немецкие социологи выяснили, что иностранцам очень выгодно переезжать в Германию." newsru.com, 17.10.2014.
- "Grüne fordern mehr Chancen für Migranten auf dem Arbeitsmarkt." focus.de, 17.10.2014.
- "Arbeitsmarktsituation von Migranten - Bildung ist der Schlüssel." deutschlandfunk.de, 17.10.2014.
- "Studie: Migranten leiden unter Diskriminierung auf dem Arbeitsmarkt." dpa, 17.10.2014.
- "Oft diskriminiert, aber mit Chancen - Migranten in Deutschland." dpa, 17.10.2017.
- "Grüne fordern mehr Chancen für Migranten auf dem Arbeitsmarkt." dpa-AFX, 17.10.2014, S.0.