Forschungsdatenzentrum
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt-
und Berufsforschung

FDZ

FDZ-Methodenreport

09/2011
DE

Methodische Aspekte zu Arbeitsmarktdaten

Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten

Record Linkage mit Daten des Panels „Arbeitsmarkt und soziale Sicherung“ und administrativen Daten der Bundesagentur für Arbeit

Jonas Beste

Bundesagentur für Arbeit

Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten

Record Linkage mit Daten des Panels „Arbeitsmarkt und soziale Sicherung“ und administrativen Daten der Bundesagentur für Arbeit

Jonas Beste (IAB)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

Inhaltsverzeichnis

Zusammenfassung		4
Abstract		4
1.	Einleitung	5
2.	Datenverknüpfung	6
2.1	Datenquellen	6
2.2	Zweistufiger Ausfallprozess	7
3.	Empirische Analysen	10
3.1	Methode und Stichprobe	10
3.2	Forschungsfragen und unabhängige Variablen	11
4.	Ergebnisse	16
4.1	Resultate des Zustimmungsmodells	19
4.2	Resultate des Verknüpfungsmodells	21
4.3	Auswirkungen auf Forschungsergebnisse	22
5.	Konsequenzen und Aussichten	25
Literatur		27

Zusammenfassung

Die Verknüpfung von Befragungs- mit Prozessdaten kann die Vorteile beider Datentypen vereinen und die Analysemöglichkeiten gegenüber der Verwendung der Einzeldatensätze vergrößern. Die exakte Verknüpfung von Informationen desselben Individuums kann jedoch nur mit der Zustimmung der befragten Person erfolgen. Zudem können nicht allen zustimmenden Personen tatsächlich Prozessdaten zugespielt werden. Da es nicht auszuschließen ist, dass bestimmte Merkmale des Befragten einen Einfluss auf das Zustimmungsverhalten haben und dass die Wahrscheinlichkeit einer tatsächlich zu realisierenden Verknüpfung ebenfalls von personenspezifischen Merkmalen abhängt, ist zu untersuchen, ob es zu einer selektiven Stichprobe und letztlich zu verzerrten Ergebnissen bei der Verwendung von verknüpften Daten kommt. In diesem Bericht wird die Selektivität bei der Verknüpfung des Panels „Arbeitsmarkt und Soziale Sicherung“ mit administrativen Daten der Bundesagentur für Arbeit (BA) untersucht. Die Ergebnisse zeigen eine geringe Selektivität der Stichprobe durch das Antwortverhalten auf die Frage nach der Bereitschaft zur Zuspielung der BA-Daten. Die Selektivität durch die tatsächlich zu realisierenden Verknüpfungen beschränkt sich vorrangig auf Personengruppen ohne gespeicherte Informationen in den BA-Daten. Um zu prüfen, ob die Forschungsergebnisse beeinflusst werden, wenn nur Personen die einer Verknüpfung zustimmen und anschließend auch tatsächlich verknüpft werden können in die Analyse eingehen, wird eine Regression mit dem logarithmieren Brutto-Stundenlohn als zu erklärende Variable durchgeführt. Dabei können keine signifikanten Unterschiede zwischen den Regressionskoeffizienten eines vollständigen und eines reduzierten Modells festgestellt werden, die für eine Verzerrung der Daten sprechen würden.

Abstract

Linking survey data with register data can combine the advantages of both types of data and increase the possibilities for analysis compared to the use of the single data sets. The exact linkage of records from the same individual is only legal with the content of the respondent. Moreover, not all consenting persons can actually be merged with the process data. It is possible that certain characteristics of the respondents have an impact on the consent behavior. Also the likelihood of actually realizable linkage in a technical sense may depend on person-specific characteristics. Therefore it must be examined, whether this leads to a selective sample, and eventually to biased results when using linked data. In this report selectivity bias due to the linkage of the panel 'Labour Market and Social Security' with administrative data from the Federal Employment Agency (BA) is investigated. The results show a low selectivity of the sample due to the response to the question of the content to link BA register data. The selectivity due to the realized merges is limited primarily to groups of people without any information stored in the BA register data. To verify whether the research results will be affected if solely individuals are included in the analysis that agree to the linkage, and subsequently can actually be linked, a regression with the logarithm of gross hourly wage as the dependent variable is performed. In this case, no significant differences could be found between the regression coefficients of a full and a reduced model, which would indicate a bias of the data.

Keywords: **Record Linkage, PASS, data linkage, survey data, register data**

1. Einleitung

Die Verknüpfung zweier unterschiedlicher Datenquellen kann die Vorteile beider Datentypen miteinander vereinen und erfreut sich daher zunehmender Beliebtheit. Bei der Zusammenführung von Datenquellen kann zwischen einer Verknüpfung von sich ähnlichen Fällen in beiden Datenquellen, sogenannten statistischen Zwillingen, und einer exakten Verknüpfung von zwei Datenquellen desselben Untersuchungssubjektes, dem sogenannten Record Linkage, unterschieden werden. Dabei ist zu berücksichtigen, dass die Fälle für die keine Verknüpfung hergestellt werden kann, bei multivariaten Analysen die auf Informationen aus beiden Datenquellen zurückgreifen wie Unit-Nonresponse zu verstehen sind und somit für die Analysen ausfallen.

Einen mit spezifischen Problemen behafteten Sonderfall stellt die im Folgenden behandelte exakte Verknüpfung von Befragungsdaten mit Personen als Untersuchungseinheit mit anderen, in diesem Fall administrativen Daten[1] auf individueller Ebene dar. Anders als beim Matching statistischer Zwillinge ist hier aus Datenschutzgründen eine Zustimmung der betroffenen Personen erforderlich. Ist die Zustimmung erteilt, müssen den Befragungsdaten noch die entsprechenden Informationen aus den Prozessdaten zugespielt werden. Wenn für die Personen keine eindeutigen Schlüssel wie z.B. die Sozialversicherungsnummer enthalten sind, müssen zur Verknüpfung geeignete Identifikatoren wie z.B. Name, Geburtstag und Adresse in beiden Datenquellen vorliegen. Jedoch garantiert das Vorliegen dieser Identifikatoren noch keine einwandfreie Verknüpfung (vgl. Jenkins et al. 2008).

Im vorliegenden Bericht wird eine Modellierung des Ausfallprozesses bei der exakten Verknüpfung der Daten des Panels „Arbeitsmarkt und Soziale Sicherung“ mit den bei der Bundesagentur für Arbeit (BA) gespeicherten Prozessdaten erfolgen, um anhand der Analysen eventuelle Selektivitäten, die zu verzerrten Schätzungen führen können, zu erkennen. Anschließend werden die Folgen der eingeschränkten Stichprobe auf die Ergebnisse einer konkreten Forschungsfrage untersucht. Das Vorgehen orientieren sich dabei an der Arbeit von Hartmann und Krug (2009), in der ebenfalls die Selektivität durch fehlende Zustimmung der Befragten bei der Verknüpfung von Befragungsdaten mit Prozessdaten der BA behandelt wird. Ziel des Berichtes ist es, die Nutzer des PASS auf mögliche Nachteile bei der Verwendung von verknüpften Daten aufmerksam zu machen und eine Einschätzung der Auswirkungen dieser Nachteile zu erleichtern.

[1] Diese werden im Weiteren auch kurz Prozessdaten genannt.

2. Datenverknüpfung

In diesem Kapitel werden die Vorteile einer Ergänzung der Befragungs- mit Prozessdaten sowie die beiden in dieser Untersuchung eingesetzten Datenquellen kurz beschrieben. Anschließend wird der durch fehlende Zustimmung und nicht zu realisierende Verknüpfungen bedingte zweistufige Ausfallprozess dargestellt.

2.1 Datenquellen

Für die Beantwortung von Fragen im Bereich der Arbeitsmarktforschung stehen unter anderen Befragungsdaten und administrative Daten zur Verfügung. Eine Datenverknüpfung bietet Forschern die Möglichkeit die Stärken beider Datentypen zu kombinieren. Bei der Ergänzung der Befragungsdaten durch Prozessdaten ergeben sich zwei entscheidende Vorteile.

Zum einen kann die Datenqualität verbessert werden, indem ungenügende Informationen in den Befragungsdaten ersetzt bzw. ergänzt werden. Bei Untersuchungen auf Grundlage von Befragungsdaten muss immer auch mit dem Problem des Item-Non-Response umgegangen werden. Und besonders die Antworten auf retrospektive Fragen sind für Erinnerungslücken anfällig. Zudem können für die Forschungsfrage relevante Details von den Befragungspersonen vergessen werden. Diese Probleme tauchen in dieser Form bei administrativ erhobenen Daten nicht auf. Die Informationen in den administrativen Daten können daher genutzt werden, um fehlende Werte in den Befragungsdaten zu füllen und Erinnerungslücken bzw. -fehler zu korrigieren. Auch sind andere Probleme wie sozial erwünschtes Antwortverhalten oder Akquieszenz, wie sie bei Befragungsdaten auftreten, hier nicht relevant. So kann bei bestimmten Forschungsfragen die in den Prozessdaten erhobene Information die validere darstellen. Der zweite Vorteil der Datenverknüpfung besteht darin, die Menge der Informationen zu jeder Personen zu erhöhen, indem zusätzliche in den Befragungsdaten nicht enthaltene Informationen genutzt werden. Da es sich bei den Prozessdaten um bereits vorhandene Daten handelt, entstehen für deren Erhebung keine weiteren Kosten[2]. Die administrativen Daten können Informationen über einen sehr langen Zeitraum, auch weit vor dem Befragungszeitraum und darüber hinaus, beinhalten. Durch die Verknüpfung können somit sowohl umfangreichere als auch aussagekräftigere Informationen für die Forschung genutzt werden (vgl. Schmucker und Huber 2009).

Für die folgenden Untersuchungen wird das Panel "Arbeitsmarkt und Soziale Sicherung" (PASS), ein vom Institut für Arbeitsmarkt- und Berufsforschung (IAB) im Rahmen der SGB-II-Forschung erhobenes Haushaltspanel, verwendet. Der inhaltliche Aufbau der Befragung eignet sich besonders für Analysen zu Determinanten des Zugangs bzw. Abgangs aus dem Leistungsbezug und erlaubt detaillierte Beschreibungen und Analysen der Lebenslagen von Leistungsempfängern und Personen in einkommensschwachen Haushalten. Die Studie umfasst zwei Teilstichproben. Für die erste Teilstichprobe wurden Bedarfsgemeinschaften aus Prozessdaten der BA gezogen. Bei der zweiten Teilstichprobe handelt es sich um eine Be-

[2] Allerdings können bis zur wissenschaftlichen Nutzbarkeit der administrativ erhobenen Daten Kosten (z.B. durch die Datenaufbereitung) anfallen.

völkerungsstichprobe mit überproportionaler Berücksichtigung einkommensschwächerer Haushalte. Im PASS können die Befragten aufgrund des gemischten Erhebungsmodus zwischen der Befragung als persönliches Interview (CAPI) oder als Telefoninterview (CATI) wählen. Die folgenden Analysen werden mit der ersten Welle des PASS durchgeführt, deren Erhebungszeitraum von Dezember 2006 bis Juli 2007 reicht.

Bei Auswertungen mit dem PASS können Auszüge aus Daten einbezogen werden, die beim IAB in Nürnberg vorliegen. Während die in Befragungen erhobenen Merkmale sich aus einem bestimmten Forschungsinteresse ableiten, erfolgt die Erhebung der Prozessdaten zu administrativen Zwecken und erfasst Informationen über bestimmte Personen, welche in diesem Fall die BA im Rahmen ihrer Aufgaben benötigt. Dabei handelt es sich zum Beispiel um zusätzliche Informationen zu vorausgegangenen Zeiten der Beschäftigung, der Arbeitslosigkeit und der Teilnahme an Maßnahmen während der Arbeitslosigkeit. Die Aufbereitung der hier verwendeten administrativen Daten geschieht durch das IAB, die Bereitstellung für die Forschung durch das Forschungsdatenzentrum (FDZ) der BA im IAB.

2.2 Zweistufiger Ausfallprozess

Durch die besonderen Anforderungen bei der Verknüpfung von Befragungsdaten mit Prozessdaten auf individueller Ebene ergeben sich zwei Prozesse, an denen es zu Ausfällen von Personen und somit einer Reduzierung der eigentlichen Stichprobe kommen kann. Erstens durch die Bereitschaft der Befragten zur Verknüpfung der beiden Datenquellen und zweitens, gegeben der Zustimmung, durch die tatsächliche Realisierbarkeit einer Verknüpfung für den individuellen Fall. Man kann die fehlende Verknüpfung als einen Spezialfall eines Missing Data Problems betrachten. Falls es sich bei beiden Prozessen um zufällige Ausfälle handelt (MCAR), muss nur mit dem Problem einer geringeren Fallzahl umgegangen werden. Wenn sich die Gruppen der Personen, die ihre Zustimmung erteilen bzw. bei gegebener Zustimmung tatsächlich verknüpft werden können und der Personen, die diese verweigern bzw. nicht verknüpft werden können, systematisch voneinander unterscheiden, kann es jedoch zu einer verzerrten Stichprobe kommen.

Beim ersten Prozess, der Zuspielfrage, kann dies zu einem Zustimmungs-Bias führen. Aus Datenschutzgründen ist zum Zweck der Zuspielung dieser Daten an die Befragungsdaten das Einverständnis der Zielpersonen notwendig. Den Befragungsteilnehmern wird die Frage, ob sie einer Zuspielung ihrer bei der BA gespeicherten Daten mit den Befragungsdaten zustimmen, im Rahmen des Erwerbsmoduls bestellt und lautet wie folgt:

> *„Um das Interview möglichst kurz zu halten, könnte das Institut für Arbeitsmarkt- und Berufsforschung in Nürnberg Daten über Zeiten der Beschäftigung, der Arbeitslosigkeit oder der Beteiligung an Maßnahmen des Arbeitsamtes einbeziehen. Für die Ergebnisse dieser Untersuchung wäre dies von großem Vorteil. Aus Gründen des Datenschutzes kann das nicht ohne Ihr Einverständnis geschehen, um das ich Sie herzlich bitte. Dies ist natürlich ebenso freiwillig, wie das Interview, das Sie freundlicherweise geben. Sie können Ihr Einverständnis selbstverständlich auch jederzeit wieder zurückziehen. Alle Vorschriften des Datenschutzes und der anonymen Ergebnisdarstellung gelten selbstverständlich für diese zusätzlichen Daten in gleicher Weise. Ich darf also notieren: Sind Sie mit der Hinzunahme dieser Daten einverstanden?“*

Das Thema der Stichprobenverzerrung durch Selektivität bei der Zustimmung zur Anspielung von zusätzlichen individuellen Daten an Befragungsdaten hat in den letzten Jahren an Bedeutung gewonnen. Mehrere Studien haben verschiedene Einflussmerkmale auf die Bereitschaft zur Zuspielung von zusätzlichen Datenquellen an Befragungsdaten feststellen können, dazu gehören u.a. Alter, Migrationshintergrund, Geschlecht, Bildung, Familienstand, Gesundheitszustand und Erwerbsstatus (vgl. Jenkins et al. 2006; Hartmann und Krug 2009; Kho et al. 2009). Jedoch zeigen sich bezüglich der Wirkungsrichtung der Einflussfaktoren zwischen den Studien gegenläufige Resultate und so ist es schwer eindeutige Einflussfaktoren auszumachen. Bei den Untersuchungen unterscheiden sich die Raten der Zuspielerlaubnis von Befragung zu Befragung mitunter stark voneinander. Einen Einfluss darauf haben vor allem das Thema der jeweiligen Studie bzw. die Art der zu verknüpfen beabsichtigten Daten. Die bisherigen Forschungsarbeiten zum Thema der Stichprobenverzerrung durch Selektivität in der Zuspielbereitschaft sind vor allem im Bereich der Medizin zu finden. Hierbei handelt es sich in der Regel um die Verknüpfung von Befragungsdaten mit Gesundheitsdaten (vgl. Dunn et al. 2004; Woolf et al. 2000; Young et al. 2001; Olson 1999).

Beim zweiten Prozess, der tatsächlichen Realisierung einer Verknüpfung unter der Bedingung einer Zustimmung, kann es zu einem Verknüpfungs-Bias kommen. Für den Verknüpfungsprozess ergeben sich für die beiden Stichproben des PASS, die Leistungsempfängerstichprobe und die Bevölkerungsstichprobe, unterschiedliche Voraussetzungen. Die Haushalte der Personen in der Leistungsempfängerstichprobe des PASS wurden aus den Prozessdaten der BA gezogen. Daher erfolgt hier die Zusammenführung der Datenquellen über einen eindeutigen Schlüssel auf Haushaltsebene. Innerhalb der Haushalte werden die einzelnen Personen dann über Vorname und Geburtsdatum verknüpft. Die exakten Übereinstimmungen werden automatisch zusammengeführt. Der Rest innerhalb der Haushalte wird anhand einer Sichtkontrolle manuell zusammengeführt. Hierbei wird geprüft, ob eine eindeutige Übereinstimmung auf Grund der Datenaufbereitung (z.B. Schreibfehler) durch das automatische Verfahren nicht erkannt wurde. Die Personen die daraufhin noch nicht verknüpft werden konnten, werden in das Verfahren für die Bevölkerungsstichprobe weitergeführt. Hierbei wird ein probabilistisches Record Linkage anhand der Merkmale Nachname, Vorname, Geburtsdatum, Geschlecht, Straße, Hausnummer, Postleitzahl und Ort verwendet[3]. Zur Bestimmung, ob ein Fall verknüpft wird oder nicht, werden zwei Schwellen festgelegt und die Fälle darin eingeordnet. Für Fälle oberhalb der höheren Schwelle erfolgt eine automatische Zusammenführung. Fälle unterhalb der unteren Schwelle werden nicht verknüpft. Fälle die sich zwischen den beiden Schwellen befinden werden per Sichtkontrolle händisch klassifiziert (vgl. Bachteler 2008).

Als Datenbasis auf Seite der Prozessdaten dienen die Integrierten Erwerbsbiographien des IABs (IEB). Es gilt bei den hier verwendeten Prozessdaten der BA zu beachten, dass bestimmte Gruppen von Personen eine höhere Wahrscheinlichkeit aufweisen, über Informatio-

[3] Für weiterführende Informationen zu der hier verwendeten Merge Toolbox (MTB) siehe Schnell et al. (2005).

nen in den Prozessdaten zu verfügen. Die IEB enthalt Daten zu sozialversicherungspflichtig Beschäftigten, Leistungsempfänger/innen, Arbeitssuchenden, Arbeitslosen und Maßnahmeteilnehmer/innen. Nicht berücksichtigt sind nicht sozialversicherungspflichtige Beamte, Selbständige, unbezahlt mithelfende Familienangehörige und Personen, die ausschließlich in so genannten Mini-Jobs tätig sind. (vgl. Jacobebbinghaus und Seth 2007). Personen, deren Informationen gar nicht in beiden Datenquellen vorhanden sind, können dementsprechend nicht verknüpft werden. In Tabelle 1 ist dargestellt, in wie weit sich die Zusammensetzung hinsichtlich der Merkmale „Erwerbsstatus" und „Bezug von Arbeitslosengeld II" (ALG II) zwischen den Personen die einer Verknüpfung zugestimmt haben und denen die bei gegebener Zustimmung nicht verknüpft werden konnten unterscheidet. Dies gibt einen Hinweis darauf, welche Personengruppen wahrscheinlicher über sensible Informationen in den hier verwendeten BA-Prozessdaten verfügen. Es wird deutlich, dass vor allem Beamte und Selbständige, aber auch Rentner und Hausfrauen/-männer sowie Schüler und Studenten bei den im Verknüpfungsprozess nichtgefunden Personen überproportional vertreten sind. Unterproportional vertreten sind hingegen Arbeiter, Angestellte, Arbeitslose und Personen in Arbeitsbeschaffungsmaßnahmen bzw. Ein-Eurojobs. Auffällig ist auch, dass der Anteil der Personen mit ALG-II-Bezug bei den nichtgefundenen Personen deutlich geringer ist als bei der Gruppe der Zustimmenden.

Tabelle 1: Erwerbsstatus und ALG-II-Bezug bei Zustimmenden[4] und Nichtgefundenen[5] (Angaben in %)

Variable	Kategorie	Zustimmende	Nichtgefundene[6]
Erwerbsstatus	Arbeiter	8,2	2,6
	Angestellte	17,5	8,1
	Beamte	2,2	20,3
	Selbständige	3,8	15,0
	Arbeitslos	28,7	4,2
	ABM/ Ein-Eurojob	5,1	1,1
	Rentner	4,3	7,3
	Hausfrau/ -mann	8,9	12,7
	Elternzeit	3,0	1,3
	Schüler	7,2	15,4
	Studenten	1,8	4,3
	Ausbildung	4,5	2,0
	Sonstige	4,9	5,7
Aktueller ALG-II-Bezug	Nicht im Bezug	55,5	92,7
	Im Bezug	44,5	7,4

[4] Personen, die einer Zuspielung zugestimmt haben.
[5] Nicht verknüpfbare Personen bei gegebener Zustimmung.
[6] Zwischen den beiden Gruppen der Zustimmenden und der Nichtgefundenen unterscheiden sich, bis auf die Kategorie „Sonstige" bei der Variable „Erwerbsstatus", alle Mittelwerte auf einen Signifikanzniveau von $p < 0.001$.

Den beiden Ausfallprozessen, dem Zustimmungs-Bias und dem Verknüpfungs-Bias, liegen unterschiedliche Mechanismen zugrunde. Während der Zustimmungs-Bias durch eine Entscheidung des Befragten erzeugt werden kann, eventuell begründet durch Misstrauen, Unwissenheit oder den Wunsch auf Privatsphäre, entsteht der Verknüpfungs-Bias nicht durch den bewussten Einfluss des Befragten. Eine entscheidende Rolle auf die Zuspielbereitschaft dürfte die wahrgenommene Legitimität und Seriosität der Erhebung haben. Es ist davon auszugehen, dass die Befragten eher ihre Zustimmung geben, wenn sie das Gefühl haben, dass die Befragung im Auftrag eines seriösen Instituts stattfindet und sorgfältig mit den Daten umgegangen wird. Bezogen auf die zu realisierenden Verknüpfungen können zwei Gründe eine Verknüpfung trotz gegebener Zustimmung verhindern. Zum einen, wie oben erwähnt, nichtvorhandene Informationen in den eigentlichen Prozessdaten. Zum anderen ist für Personen, bei denen nicht alle Identifikatoren zur Verfügung stehen, diese in beiden Datenquellen fälschlicherweise voneinander abweichen oder diese nicht eindeutig diskriminieren, eine geringere Wahrscheinlichkeit für eine positive Verknüpfung zu erwarten. Ob und für welche Personengruppen dies gilt, bleibt zu untersuchen.

3. Empirische Analysen

Im Folgenden soll eine Identifikation von Einflussfaktoren auf die Selektivität im zweistufigen Verknüpfungsprozess erfolgen. Es werden daher die interessierenden Forschungsfragen formuliert und die in den Modellen verwendeten unabhängigen Variablen und deren vermuteten Zusammenhänge mit der jeweiligen Ausfallwahrscheinlichkeit vorgestellt. Zunächst wird jedoch kurz auf die verwendete Methode und die Stichprobe eingegangen.

3.1 Methode und Stichprobe

In der ersten Welle des PASS wurden ca. 18.000 Personen in 12.000 Haushalten befragt. Alle befragten Personen unter 65 Jahren wurden um die Erlaubnis gebeten, ihre Befragungsdaten mit den Prozessdaten verknüpfen zu dürfen. Alle Personen ab 65 Jahren haben den Seniorenfragebogen erhalten, in dem die Frage zur Zuspielung nicht enthalten ist. 17.249 Personen wurden von 551 Interviewern nach der Erlaubnis zur Zuspielung ihrer Prozessdaten gefragt. Insgesamt stimmten 13.766 Befragte (79,8%) diesem Anliegen zu. Den Befragungsdaten der Personen die einer Verknüpfung zugestimmt haben, konnten wiederum für 12.659 Personen (92,0%) tatsächlich Prozessdaten zugespielt werden. Demnach können im PASS für 73,4 Prozent der Stichprobe der unter 65-jährigen eine Verknüpfung hergestellt werden. Durch den Ausfall von 27 Prozent der ursprünglichen Stichprobe ist eine Verzerrung nicht auszuschließen. Dabei ist zu berücksichtigen, dass es bei den realisierten Verknüpfungen auch zu „falschen“ Verknüpfungen nicht identischer Personen gekommen sein kann[7].

[7] Das heißt, dass diese Zusammenführungen von Informationen nicht identischer Individuen als erfolgreiche Verknüpfungen in die Analysen mit eingehen.

Zur empirischen Überprüfung der theoretisch vermuteten Einflussfaktoren auf die Zustimmungswahrscheinlichkeit bzw. die Verknüpfungswahrscheinlichkeit wird ein logistisches Regressionsmodell verwendet. Aufgrund der zwei Analyseebenen (Interviewer, die jeweils mehrere Personen befragen) wird dieses als Mehrebenen-Modelle geschätzt. Hiermit soll berücksichtigt werden, dass neben den Merkmalen der Befragten auch Merkmale der Interviewer mit der Zustimmungsbereitschaft der Befragten in Zusammenhang stehen können. Die Analyse der Stichprobenverzerrung durch die Verknüpfung von Prozess- und Befragungsdaten erfolgt, entsprechend den zwei Selektivitätsprozessen, in zwei Schritten.

Tabelle 2: Zustimmungs- und Verknüpfungsquoten nach PASS-Stichproben (Angaben in %)

	Leitungsempfängerstichprobe	Bevölkerungsstichprobe	**Gesamt**
Zustimmung	79,9	79,7	79,8
Verknüpfbarkeit	98,6	84,3	92,0
Gesamt	78,8	67,2	73,4

Für die Analyse des Zustimmungs-Bias, welcher bei der Zuspielerlaubnis auftreten kann, wird als abhängige Variable eine Dummy-Variable gebildet, die den Wert eins annimmt, wenn der Befragte der Zuspielung zustimmt und null wenn der Befragte verweigert. Dabei wurden ungültige Antworten[8] als Verweigerung bewertet, da für diese Personen keine Verknüpfungen durchgeführt werden dürfen. In die Analyse werden alle Personen einbezogen, denen die Frage zur Zuspielung gestellt wurde. Die zweite Selektivität, der Verknüpfungsbias, kann durch das Verknüpfungsverfahren entstehen. Wie in Tabelle 2 zu sehen ist, können 98,6 der zustimmenden Personen aus der Leistungsempfängerstichprobe erfolgreich verknüpft werden. Die Selektivität durch die realisierte Verknüpfung beschränkt sich damit fast ausschließlich auf die Stichprobe der Wohnbevölkerung. Daher wird die Analyse auf die dichotome zu erklärende Variable „Verknüpfung erfolgreich“ oder „Verknüpfung nicht erfolgreich“ auf die einer Zuspielung zustimmenden Personen in der Bevölkerungsstichprobe eingeschränkt. Um zu überprüfen, ob es bei einer konkreten Forschungsfrage durch die gegebenenfalls selektive Zusammensetzung der verbleibenden Stichprobe zu Verzerrungen in den Ergebnissen kommt, werden beispielhaft für den logarithmierten Brutto-Stundenlohn zwei Modelle geschätzt. Das eine nur mit den Personen, bei denen eine Verknüpfung hergestellt werden konnte und das andere mit allen Personen, denen die Zuspielfrage gestellt wurde. Anschließend werden die Ergebnisse miteinander verglichen, um Abweichungen zu identifizieren.

3.2 Forschungsfragen und unabhängige Variablen

In diesem Abschnitt werden die in den Analysen verwendeten unabhängigen Variablen und deren vermuteten Zusammenhänge mit der Verweigerung der Zuspielerlaubnis bzw. der

[8] Zu den ungültigen Antworten zählen Fälle, die die Frage nicht verstanden haben sowie die Antwortkategorien „weiß nicht“ und „keine Angabe“.

realisierten Verknüpfung beschrieben. Für jede der fünf relevanten Forschungsfragen lässt sich jeweils (mindestens) ein Variablenblock herausarbeiten. Tabelle 3 sind die sechs Gruppen von Einflussfaktoren[9], die in den Analysen berücksichtigt werden, zu entnehmen. Die für die beiden Selektivitätsprozesse zu erwartenden Wirkungsrichtungen der Einflussfaktoren werden getrennt voneinander aufgeführt.

Tabelle 3: Unabhängige Variablen und Variablenblöcke

Variablenblock		Variable
1.	Soziodemographische Merkmale des Befragten	Geschlecht
		Alter
		Höchster Schulabschluss
		Migrationshintergrund
		Wohnregion
		Bruttoerwerbseinkommen
		Erwerbsstatus
2.	Fehlende Angaben bei sensitiven Fragen	Nennung der Kategorie „Weiß nicht“ bzw. „Keine Angabe“ bei soziodemographischen Variablen
3.	Charakteristika der Befragung	Arbeitslosengeld-II-Bezug
		Sanktionierung des Haushalts seit Bezug
		Panelbereitschaft
		Zustimmungsbereitschaft des HBV
4.	Merkmale der Interviewsituation	Befragungsmodus
		Interviewsprache
5.	Merkmale des Interviewers	Alter des Interviewers
		Höchster Schulabschluss des Interviewers
		Interaktion zw. gleichen Bildungsniveau des Befragten und des Interviewers
		Geschlecht des Interviewers
		Anzahl der geführten Interviews zum Interviewzeitpunkt
6.	Verknüpfungsrelevante Merkmale	Häufigkeit des Vor- und Nachnamens
		Fehlende Werte in den Schlüsselvariablen

[9] Der Variablenblock zu den fehlenden Angaben bei sensitiven Fragen wird in der Ergebnisdarstellung (Tabelle 4) in einem Block mit den soziodemographischen Merkmalen des Befragten aufgeführt.

Die im Weiteren interessierenden Forschungsfragen lauten:

1. Unterscheiden sich die Personen in ihrer Zuspielbereitschaft selektiv nach soziodemographischen Merkmalen?

Man kann davon ausgehen, dass die soziodemographischen Merkmale des Befragten eine Rolle bei der Beantwortung der Zuspielfrage spielen, auch wenn die Wirkungsrichtungen nicht immer eindeutig vorhersagbar sind. Für das Alter ist einerseits vorstellbar, dass ältere Personen die Zuspielung eher verweigern, was sich durch die geringere Erfahrung im Umgang mit Technologie und ein daraus resultierendes Misstrauen begründen lässt. Andererseits kann es sein, dass gerade jüngere Personen die Seriosität bestimmter Institutionen nicht richtig einschätzen können und dies daher eher zu einer Verweigerung der Zustimmung führt. Für das Alter werden fünf Dummy-Variablen in die Schätzungen mit aufgenommen, um auch nichtlineare Zusammenhänge erfassen zu können. Für Personen mit Migrationshintergrund wird ein negativer Zusammenhang mit der Zustimmungsbereitschaft vermutet. Dies kann durch Sprachschwierigkeiten und daraus resultierenden Problemen beim inhaltlichen Verständnis der Frage begründet sein. Weiterhin kann auch hier fehlendes Wissen über administrative Vorgänge dazu führen, dass die Seriosität des auftragsgebenden Institutes und die Vorkehrungen der Datenschutzsicherung nicht richtig einschätzt werden können und daher ein Missbrauch ihrer Daten befürchtet wird. Gleiches kann für Personen mit niedriger Bildung gelten. Als Proxyvariable für die Bildung einer Person wird der höchste Schulabschluss durch die Aufnahme von insgesamt vier Dummy-Variablen in den Schätzungen berücksichtigt. Neben dem Geschlecht einer Person wird als weitere Kontrollvariable berücksichtig, ob der Haushalt einer Person in Ost- oder Westdeutschland liegt. Zusätzlich zu den bereits genannten Merkmalen kann auch der Erwerbsstatus des Befragten Einfluss auch das Antwortverhalten nehmen. Da besonders zu sozialversicherungspflichtig Beschäftigten und ALG-II-Empfängern sensible Daten zur Verknüpfung bei der BA bereit stehen, kann vermutet werden, dass diese gegenüber anderen Personengruppen wie Schülern, Beamten, Selbstständigen und Hausfrauen bzw. -männern eine geringe Zuspielbereitschaft aufweisen.

2. Ist die Verweigerung der Zuspielerlaubnis als ein ähnliches Phänomen wie Item-Non-Response bei sensitiven Fragen zu verstehen?

Es kann angenommen werden, dass als sensibel empfundene Fragen nur ungern beantwortet werden. Fehlende Angaben[10] bei solchen Fragen können deshalb als Proxy für das Misstrauen bezüglich des vorsichtigen Umgangs mit den Informationen gewertet werden. Neben der Zuspielerlaubnis werden im PASS weitere sensitive Fragen gestellt, deren Beantwortung selektiv sein kann. Dazu zählen unter anderem die Fragen zu den soziodemographischen Merkmalen. Besonders aussagekräftig sollte hierbei die Frage nach dem persönlichen Bruttoerwerbseinkommen sein. Aus diesem Grund werden bei den soziodemographischen Merkmalen des Befragten zusätzliche Dummy-Variablen in die Schätzungen aufgenommen, die angeben, ob der Befragte auf diese sensitiven Fragen nicht geantwortet hat. Als Refe-

[10] Dazu gehören in diesem Fall die Antwortkategorien „weiß nicht“ und „keine Angabe“.

renzgruppe dient dabei jeweils eine gültige Antwortkategorie der Variable. Weisen Personen, die hier zu keiner Antwort bereit waren, ebenfalls eine geringere Wahrscheinlichkeit der Zustimmung zur Zuspielung auf, kann dies als ein Hinweis darauf verstanden werden, dass es sich bei der Verweigerung der Zuspielerlaubnis um eine ähnliche Dimension wie bei dem Item-Non-Response bei sensitiven Fragen handelt.

3. Beeinflusst die Nähe des Befragten zum Inhalt der Befragung oder zum auftragsgebenden Institut die Wahrscheinlichkeit einer Zustimmung zur Zuspielung?

Diese Frage zielt auf eine Art Content-Effekt bzw. Sponsorship-Effekt ab. So ist zu vermuten, dass das Thema der Befragung, die Art der zu verknüpfen beabsichtigten Daten, ihr Verwendungszweck sowie das auftragsgebende Institution, in diesem Fall das IAB in der BA, das Zustimmungsverhalten des Befragten beeinflussen können. Bei Personen mit ALG-II-Bezug ist eine Nähe des Befragten zum auftragsgebenden Institut gegeben. Dadurch dürfte eine realistische Einschätzung der Seriosität leichter fallen. Des Weiteren können sich Personen mit aktuellen ALG-II-Bezug eventuell dem Thema verbunden oder der Institution verpflichtet fühlen und einer Zuspielung ihrer Daten eher zustimmen. Andererseits liegen gerade zu ALG-II-Empfängern sensible Daten in den Prozessdaten vor, und so kann es sein, dass diese Personen der Zustimmung negativer gegenüber stehen als Personen, zu denen keine sensiblen Informationen vorhanden sind. Der Referenzgruppe der Personen in Haushalten die noch nie ALG II bezogen haben, stehen Personen in Haushalten die aktuell ALG II beziehen und Personen in Haushalten die aktuell nicht beziehen, aber bereits schon mindestens einmal ALG II bezogen haben gegenüber. Eine weitere Kategorie enthält Personen in Haushalten, die den Bezug von ALG II in der Befragung nicht mitgeteilt haben, obwohl sie zur Leistungsempfänger-Stichprobe gehören. Für diese Gruppe wird erwartet, dass sie einer Zuspielung ihrer Prozessdaten weniger wahrscheinlich zustimmen. Zudem können die mit dem Bezug von ALG-II verbundenen Erfahrungen einen Einfluss auf das Antwortverhalten haben. Eine Proxyvariable für negative Erfahrungen stellt die Sanktionierung des Haushaltes und die damit verbundene Kürzung des zu erhaltenden ALG-II-Betrags dar. Bei vorhandenen negativen Erfahrungen ist eine geringere Bereitschaft für die Zuspielung zu erwarten. Die Frage nach der Panelbereitschaft kann als Proxyvariable für die Einstellung der Zielperson zur Befragung gewertet werden. Sie könnte ein Hinweis darauf sein, dass Personen, die zwar bei der Befragung teilnehmen, dann aber die Panelbereitschaft verweigern, ein geringes Interesse an der Befragung haben und eine Zuspielung daher auch eher ablehnen (vgl. Prost und von Briel 1995). Ebenfalls in diesen Variablenblock enthalten ist die Information, ob der Hausbevollmächtigte[11] (HBV) seine Zustimmung zur Zuspielung erteilt hat oder nicht. Es ist zu erwarten, dass die weiteren Haushaltsmitglieder sich an der Entscheidung dieser Person orientieren und diese daher ihr Antwortverhalten beeinflusst.

[11] Der HBV wird definiert durch die Beantwortung des Haushaltsfragebogens.

4. Weisen die Interviewer untereinander unterschiedliche Zustimmungsquoten auf?

Ein besonderer Stellenwert kommt dem Interviewer zu. Dabei orientieren sich die theoretischen Annahmen an Forschungsergebnissen zum Einfluss von Interviewern auf die Teilnahme an Panel-Studien (vgl. Campanelli und O'Muircheartaigh 1999; Haunberger 2009). Es wird angenommen, dass der Interviewer mit bestimmten Eigenschaften die Zuspielbereitschaft des Befragten beeinflusst. Eine entscheidende Rolle können dabei die soziodemographischen Charakteristika des Interviewers haben. Diese dienen dem Befragten als Hinweis auf die Seriosität des Institutes und den Schutz seiner Daten. Falls ein Interviewer vom Befragten als vertrauenswürdig eingestuft wird, fällt es dem Befragten vermutlich leichter einer Zuspielung seiner Prozessdaten zuzustimmen. Hierbei kann besonders ein hoher Bildungsgrad vertrauensstiftend wirken. Außerdem wird mit der Aufnahme des Alters und des Geschlechts des Interviewers für weitere soziodemographische Merkmale kontrolliert. Neben den soziodemographischen Merkmalen spielt vor allem die Erfahrung des Interviewers eine wichtige Rolle. Durch die Vielzahl der durchgeführten Interviews erfolgt ein Lernprozess, wie Fragen zu stellen sind, um die Wahrscheinlichkeit einer positiven Antwort zu erhöhen. Für die Interviewer stehen nur eingeschränkt Informationen zur Verfügung und daher wird die Erfahrung eines Interviewers anhand der Anzahl der bereits in der aktuellen Welle durchgeführten Interviews angegeben. Da davon auszugehen ist, dass erfahrene Interviewer über bessere Strategien verfügen, einen vertrauensvollen Eindruck zu erwecken und positive Antworten zu generieren, ist hier mit einem positiven Zusammenhang mit der Zustimmungswahrscheinlichkeit zu rechnen. Ansonsten sind die Interviewermerkmale vor allem in der Interaktion mit den Merkmalen der Befragten interessant, da diese in Kombination unterschiedlich wirken können. Daher wird eine Interaktion des gleichen Bildungsniveaus zwischen Befragten und Interviewer in die Analysen mit aufgenommen. Es ist davon auszugehen, dass bei gleichem Bildungsniveau ein höheres Vertrauen des Befragten zum Interviewer vorliegt und somit die Zuspielfrage eher positiv beantwortet wird. Auch ein Einfluss der Interviewsituation ist nicht auszuschließen. Daher werden die Sprache in der das Interview durchgeführt wird sowie der Erhebungsmodus in das Modell mit aufgenommen. Diese beiden Merkmale sind für die Interviewer konstant.

5. Kommt es, unter der Bedingung, dass der Verknüpfung zugestimmt wurde, durch die tatsächlich realisierte Datenverknüpfung zu weiteren Selektivitäten?

Ein weiterer Mechanismus, der bei der Verknüpfung von Befragungs- mit administrativen Daten zu einer Selektivität der Stichprobe und so eventuell auch zu verzerrten Ergebnissen führen kann, hängt mit den tatsächlich realisierten Verknüpfungen zusammen. Ein Grund für nichtrealisierte Verknüpfungen können fehlende administrative Daten sein. Wie bereits in Abschnitt 2.2 erwähnt, ist es plausibel anzunehmen, dass Personen die einer sozialversicherungspflichtigen abhängigen Erwerbstätigkeit nachgehen oder Arbeitslosengeld erhalten, eher Informationen in den Prozessdaten der BA aufweisen als Schüler, Selbstständige oder Beamte und so auch wahrscheinlicher tatsächlich verknüpft werden können. Ein weiterer Grund können fehlende bzw. unpräzise Schlüsselvariablen sein. Für die Analyse des Ver-

knüpfungs-Bias werden zusätzliche Informationen, zu den beim Verknüpfungsvorgang relevanten Schlüsselvariablen, berücksichtigt. Da die Daten auch über den Vor- und Nachnamen einer Person miteinander verknüpft werden, kann ein vom Interviewer falsch aufgenommener Name die Wahrscheinlichkeit einer Verknüpfung verringern. Ein Hinweis darauf dürfte ein nur einmal in den Befragungsdaten vorkommender Name sein. Und auch ein in den Befragungsdaten richtig aufgenommener, jedoch sehr ungewöhnlicher und eventuell schwer zu schreibenden Name, könnte besonders fehleranfälliger sein und dazu führen, dass dieser in den Prozessdaten nicht identisch vorliegt. Bezogen auf Merkmale der Personen kann bei der Verknüpfbarkeit ein negativer Zusammenhang mit dem Migrationshintergrund vermutet werden. Es ist anzunehmen, dass ausländische Vor- bzw. Nachnamen dem Interviewer nicht so präsent sind und damit eine höhere Fehleranfälligkeit aufweisen. Vor allem aber dürfte das Fehlen einer oder mehrere verwendeter Schlüsselvariablen die Wahrscheinlichkeit einer positiven Verknüpfung spürbar senken.

4. Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der beiden Selektivitätsanalysen und des Anwendungsbeispiels kurz vorgestellt. In Tabelle 4 sind die Ergebnisse der Schätzungen für die beiden verwendeten Selektivitätsmodelle abgebildet.

Tabelle 4: Ergebnisse des Zustimmungs- und des Verknüpfungsmodells

	Zustimmung			Verknüpfung		
	OR		Z	OR		Z
Soziodemographische Merkmale des Befragten						
Geschlecht						
weiblich	0,955		-0,98	1,166		1,75
männlich	Ref.			Ref.		
Alter						
bis unter 25 Jahre	Ref.			Ref.		
25 bis unter 35 Jahre	0,968		-0,37	1,197		0,85
35 bis unter 45 Jahre	0,906		-1,10	0,998		-0,01
45 bis unter 55 Jahre	0,837		-1,93	0,832		-0,84
über 55 Jahre	0,964		-0,37	0,576	*	-2,49
keine Information	0,268	**	-2,61	0,077	**	-2,59
Höchster Schulabschluss						
höchstens Volks-/Hauptschulabschluss bzw: POS	Ref.			Ref.		
Mittlere Reife/Realschulabschluss	1,081		1,37	0,899		-0,86
Fachhochschulreife/ Hochschulreife	0,953		-0,73	0,473	***	-6,41
Anderer Schulabschluss	0,453		-1,50	-		-
noch Schüler	0,948		-0,36	0,198	***	-7,87
keine Information	0,299	**	-2,71	0,670		-0,45
Region						
neue Bundesländer	1,100		1,64	1,231		1,79
alte Bundesländer	Ref.			Ref.		
Migrationshintergrund						
Kein Migrationshintergrund	Ref.			Ref.		
Selbst zugezogen	0,830	**	-2,83	1,138		0,86
Mind. 1 Elternteil zugezogen	0,933		-0,78	0,955		-0,27
Mind. 1 Großelt. zugez., Elt. in D geb.	1,458	*	2,17	0,609	*	-2,03
keine Information	0,663	*	-2,36	0,621		-1,48
Bruttoeinkommen						
kein Einkommen	Ref.			Ref.		
unter 1500 Euro	1,313	*	2,18	1,110		0,58
1500 bis 3000 Euro	1,363	*	2,06	0,877		-0,55
über 3000 Euro	1,759	***	3,33	0,679		-1,62
Angabe verweigert	0,402	***	-5,98	0,981		-0,07
weiß nicht	1,029		0,18	1,080		0,30
Erwerbsstatus						
Arbeiter/ Angestellte	Ref.			Ref.		
Beamte/ Selbständige	1,027		0,27	0,038	***	-24,54
arbeitslos/ ABM/ Ein-Eurojob	1,186		1,18	0,658		-1,29
Rentner/ Hausfrau/ Hausmann/ Elternzeit	1,172		1,09	0,165	***	-7,34
Schüler/ Studenten/ Ausbildung	1,309		1,94	0,168	***	-6,72
sonstige/ keine Information	0,962		-0,31	0,218	***	-6,73

	Zustimmung			Verknüpfung		
	OR		Z	OR		Z
Charakteristika der Befragung						
ALG-II-Bezug						
noch nie im ALG-II-Bezug	0,952		-0,72	0,403	***	-3,71
kein aktueller ALG-II-Bezug	0,900		-1,35	0,815		-0,56
aktueller ALG-II-Bezug	Ref.			Ref.		
ALG-II-Bezug nicht angegeben	0,749	**	-2,93	-		-
Sanktion des Haushaltes seit Anfang 2005						
HH noch nie sanktioniert	Ref.			Ref.		
HH bereits sanktioniert	0,959		-0,55	4,973	*	2,14
Panelbereitschaft						
Zustimmung nicht verweigert	Ref.			Ref.		
Zustimmung verweigert	0,271	***	-14,50	0,779		-1,19
HBV						
Person ist HBV	Ref.			Ref.		
HBV stimmt Zuspielung zu	2,125	***	11,90	0,843		-1,74
HBV verweigert Zuspielung	0,610	***	-5,91	0,804		-1,18
keine Information	1,081		0,52	0,726		-1,47
Merkmale der Interviewsituation						
Befragungsmodus						
CATI	Ref.			Ref.		
CAPI	0,814		-1,11	0,644	**	-2,96
Interviewsprache						
deutsch	Ref.			Ref.		
nicht deutsch	0,528	*	-2,11	0,909		-0,16
Merkmale des Interviewers						
Alter des Interviewers						
bis unter 25 Jahre	0,658		-1,69	0,709		-1,82
25 bis unter 35 Jahre	0,452	***	-3,38	0,772		-1,46
35 bis unter 45 Jahre	0,500	**	-3,22	0,907		-0,61
45 bis unter 55 Jahre	0,828		-0,99	0,863		-1,01
über 55 Jahre	Ref.			Ref.		
Höchster Schulabschluss des Interviewers						
Volks-/Hauptschule	Ref.			Ref.		
mittlere Reife	0,477	***	-3,41	0,931		-0,50
Abitur	0,563	**	-2,62	1,186		1,16
keine Information	0,703		-0,61	0,527		-1,61
Interaktionseffekt gleichen Bildungsniveaus						
kein vergleichbares Bildungsniveau	Ref.			Ref.		
vergleichbares Bildungsniveau	0,964		-0,68	0,936		-0,69
Geschlecht des Interviewers						
weiblich	0,776	*	-2,01	0,987		-0,14
männlich	Ref.			Ref.		

	Zustimmung		Verknüpfung		
	OR	Z	OR		Z
Anzahl der geführten Interviews zum Interviewzeitpunkt					
1 bis 10 Interviews	0,969	-0,38	0,919		-0,61
11 bis 30 Interviews	0,958	-0,55	0,943		-0,45
31 bis 60 Interviews	0,991	-0,11	1,108		0,75
mehr als 60 Interviews	Ref.		Ref.		
Verknüpfungsrelevante Merkmale					
Häufigkeit des Vor- und Nachnamens					
Vor- und/ oder Nachname nur einmal vorhanden	-	-	0,822	*	-2,15
Vor- und Nachname mehrfach vorhanden	-	-	Ref.		
mindestens eine der beiden Informationen fehlt	-	-	1,644		1,50
Missings (Vorname, Nachname, Geburtsdatum)					
alle Informationen vorhanden	-	-	Ref.		
mind. eine Information fehlt	-	-	0,409	***	-4,06
Stichprobengröße	**17247**		**6392**		
Pseudo R-Quadrat	**0,15**		**0,25**		
sigma_u	**1,12**		**0,20**		
rho	**0,28**		**0,01**		

*Signifikanzniveau: * P<0.05; ** P<0.01; *** P<0.001*

4.1 Resultate des Zustimmungsmodells

Die Ergebnisse der Analyse zum Zustimmungs-Bias zeigt die Spalte „Zustimmung“ in der Tabelle 4. Im Hinblick auf die soziodemographischen Einflussfaktoren ergibt sich für das Geschlecht, das Alter, den höchsten Bildungsabschluss und die Wohnregion keine signifikante Selektivität der Zustimmung. Zielpersonen ohne Migrationshintergrund stimmen einer Zuspielung eher zu als Personen die selbst nach Deutschland zugezogenen sind. Der vermutete Zusammenhang einer geringen Zuspielbereitschaft bedingt durch Sprachschwierigkeiten und fehlendes Wissen über administrative Vorgänge, lässt sich aber nur für Personen erster Generation ausmachen. Für Personen mit Migrationshintergrund der dritten Generation ergibt sich hingegen eine höhere Zuspielbereitschaft im Vergleich zu Personen ohne Migrationshintergrund. Auch für das monatliche Bruttoerwerbseinkommen sind signifikante Effekte feststellbar. Für Personen die ein Erwerbseinkommen erzielen ergibt sich gegenüber Personen ohne Erwerbseinkommen eine höhere Bereitschaft der Datenverknüpfung zuzustimmen. Zudem ist mit steigenden Erwerbseinkommen eine höhere Zustimmungsbereitschaft feststellbar. Der Erwerbsstatus einer Person führt hingegen zu keiner signifikanten Selektivität.

Bezogen auf das Misstrauen gegenüber dem sicheren Umgang mit den sensiblen Informationen, bestätigen die Ergebnisse die genannten Vermutungen. Personen, die bei den Fragen zum Alter, dem höchsten Schulabschluss, dem Migrationshintergrund oder bei anderen sen-

sitiven Fragen wie nach dem Bruttoerwerbseinkommen[12] keine gültigen Angaben gemacht haben, weisen eine signifikant geringere Wahrscheinlichkeit zur Zuspielbereitschaft als die Referenzgruppen auf.

Bezüglich der untersuchungsspezifischen Merkmale und der Frage, ob und wie die Beziehung des Befragten zur Befragung das Zustimmungsverhalten beeinflusst, sind nur wenige signifikante Einflüsse feststellbar. Für die Verweigerung zur Panelbreitschaft, welche als Indikator für ein geringes Interesse an der Befragung geführt wurde, zeigt sich wie erwartet ein positiver Zusammenhang mit der Ausfallswahrscheinlichkeit. Für die vermuteten Zusammenhänge zwischen der Nähe der Befragungspersonen zum Inhalt der Befragung bzw. dem auftragsgebenden Institut und ihrer Zuspielbereitschaft können keine Hinweise festgestellt werden. Lediglich für die Gruppe der Personen in Haushalte, die einen ALG-II-Bezug in der Befragung verheimlicht haben, ergibt sich ein signifikant negativer Effekt. Die voraussichtlich negative Erfahrung einer Sanktionierung während des Bezugs hat keinen signifikanten Einfluss. Das Antwortverhalten des Haushaltsbevollmächtigten hingegen wirkt sich signifikant auf die Zuspielbereitschaft der weiteren Haushaltsmitglieder aus. Falls der HBV einer Zuspielung zustimmt, stimmen auch die weiteren Haushaltsmitglieder eher eine Zuspielung zu. Verweigert der HBV diese aber, so sinkt auch die Wahrscheinlichkeit einer positiven Antwort der anderen Personen im Haushalt.

Auch die Merkmale der Interviewer haben einen signifikanten Einfluss auf das Zustimmungsverhalten der Befragten. Zum einen spielt das Geschlecht des Interviewers eine Rolle. Für männliche Interviewer ist im Gegensatz zu ihren weiblichen Kolleginnen eine niedrigere Ausfallwahrscheinlichkeit feststellbar. Und auch das Alter des Interviewers hat einen Einfluss. So weisen ältere Interviewer (ab 55 Jahren) eine höhere Wahrscheinlichkeit auf eine Zustimmung des Befragten auf die Frage nach der Zuspielung zu erhalten als ihre Kollegen im Alter von 25 bis 44 Jahren. Und auch Interviewer mit einem Hauptschulabschluss als höchstem Schulabschluss erhalten eher eine positive Antwort als Interviewer mit höheren Schulabschlüssen. Ein Interaktionseffekt, der sich aus einem gleichen höchsten Schulabschluss des Befragten und des Interviewers ergeben könnte, lässt sich nicht feststellen. Beide Merkmale, das Alter und der höchste Schulabschluss, können als Indikatoren für die Berufserfahrung des Interviewers verstanden werden. Dies ergibt sich aus der Überlegung, dass Personen mit höherem Schulabschluss[13] die Tätigkeit als Interviewer in den meisten Fällen vermutlich nur einen kurzen Zeitraum ausüben, wohingegen Personen mit niedrigem Schulabschluss öfters auch einen langen Zeitraum als Interviewer tätig sind. Dies könnte eine Erklärung sein, warum der zuvor vermutete Zusammenhang von steigender Bildung des Interviewers und zunehmender Bereitschaft des Befragten, der Zuspielung zuzustimmen, empirisch nicht belegt wird. Die Erfahrung des Interviewers mit der aktuellen Befragung, gemessen an der Anzahl der bereits durchgeführten Interviews innerhalb der aktuellen Welle, zeigt kein signifikantes Ergebnis. Bei den Merkmalen die zur Überprüfung des Einflusses der Interviewsitua-

[12] Beim Erwerbseinkommen werden die Antwortkategorien „weiß nicht“ und „keine Angabe“ getrennt ausgegeben.

[13] Darunter dürfte sich vermutlich ein hoher Anteil an Studenten befinden.

tion in das Modell mit aufgenommen wurden, lassen sich für die Interviewsprache, nicht aber für den Erhebungsmodus signifikante Effekte ausmachen. In Interviews, die nicht in deutscher Sprache geführt werden, ist die Wahrscheinlichkeit einer Zustimmung geringer als in deutschsprachigen Interviews, was wohl weniger auf die Fähigkeit der fremdsprachigen Interviewer als auf ein höheres Misstrauen der nicht deutschsprechenden Befragten zurückzuführen ist.

Die Ergebnisse lassen den Rückschluss zu, dass das Antwortverhalten auf die Frage nach der Zustimmungsbreitschaft nur zu einer geringen Selektivität der Stichprobe führt. Das Pseudo-R-Quadrat von 0,15 zeigt, dass die im Modell berücksichtigten Merkmale die Varianz im Zustimmungsverhalten nur eingeschränkt erklären. Das Rho beschreibt die Intra-Klassen-Korrelation und ein Wert von 0,28 lässt sich dahingehend verstehen, dass die Zwei-Ebenen-Struktur von Interviewer und Befragten für die Erklärung des Zustimmungsverhaltens relevant ist.

4.2 Resultate des Verknüpfungsmodells

In der Spalte „Verknüpfung“ in Tabelle 4 sind die Ergebnisse der Analyse des Verknüpfungs-Bias dargestellt. Da der Verknüpfungsprozess für die Personen in den beiden Stichproben unterschiedlich verläuft und nahezu alle Personen der BA-Stichprobe erfolgreich verknüpft werden können, erfolgt die Analyse des Ausfallprozesses nur für die Teilstichprobe der Gesamtbevölkerung. Betrachtet man die Selektivität, die durch die tatsächlich zu realisierenden Verknüpfungen bei gegebener Zustimmung entsteht, bestätigen sich die zuvor vermuteten Zusammenhänge größtenteils.

Es wurde vermutet, dass zu bestimmten Personengruppen unwahrscheinlicher Informationen in den Daten der BA vorhanden sind und diese daher eine höhere Ausfallwahrscheinlichkeit haben. Für die soziodemographischen Merkmale ergeben sich nur wenige signifikante Effekte. Beim Alter weist die Gruppe der Personen ab 55 Jahren im Vergleich zu jüngeren Altersgruppen eine niedrigere Verknüpfungswahrscheinlichkeit auf. Für Befragte die noch zur Schule gehen sowie solche mit (Fach-)Hochschulreife ergeben sich im Vergleich zu Personen mit höchstens Hauptschulabschluss geringere Verknüpfungswahrscheinlichkeiten. Schüler dürften aus bereits genannten Gründen weniger wahrscheinlich administrativen Daten der BA aufweisen. Für die geringere Verknüpfungswahrscheinlichkeit der höher Gebildeten lässt sich hier keine eindeutige theoretische Erklärung ausmachen. Für Arbeitslose[14] zeigt sich keine signifikant geringere Verknüpfungswahrscheinlichkeit als bei der Referenzgruppe der Arbeiter und Angestellten. Beamte, Selbständige sowie Nichterwerbspersonen[15] hingegen haben eine signifikant niedrigere Wahrscheinlichkeit als die Referenzgruppe erfolgreich verknüpft zu werden. Dies stimmt mit den theoretischen Überlegungen überein. Ebenfalls entsprechend den Erwartungen wirkt sich der aktuelle Bezug des ALG II positiv auf die Wahr-

[14] Diese Kategorie beinhaltet auch Personen in Arbeitsbeschaffungsmaßnahmen (ABM) und Ein-Euro-Jobs.
[15] Hierzu zählen Rentner, Hausfrauen/ -männer und Personen in Elternzeit.

scheinlichkeit einer Verknüpfung aus. Wie die Personen in sozialversicherungspflichtiger Beschäftigung, weisen auch ALG-II-Leistungsempfänger wahrscheinlicher Informationen in den BA-Prozessdaten auf.

Für das Zustandekommen einer erfolgreichen Zuspielung sind, neben der Wahrscheinlichkeit, dass für eine Person Informationen in den Prozessdaten gespeichert sind, vor allem die Schlüsselvariablen ausschlaggebend. Falls Informationen in den relevanten Schlüsselvariablen Vorname, Nachname oder Geburtsdatum in den Befragungsdaten fehlen, hat dies gravierenden Einfluss auf die Realisierung einer Verknüpfung. Falls die entsprechenden Schlüsselvariablen vorhanden sind, kommt deren Qualität eine entscheidende Rolle zu. Ist der Vorname des Befragten nur einmal in den Befragungsdaten enthalten, lässt dies zwei Vermutungen zu, die beide dazu führen, dass dieser Fall weniger wahrscheinlich verknüpft werden kann. Zum einen kann ein selten vorkommender Name anfälliger für Schreibfehler sein, so dass dieser den Prozessdaten wahrscheinlicher falsch vorliegt als häufiger vorkommende Namen. Zum anderen dürften falsch geschriebene Vornamen auch nur selten - in der Regel einmal - vorkommen. Daher kann das einmalige auftauchen eines Namens ein Hinweis darauf sein, dass dieser in den Befragungsdaten fehlerhaft aufgenommen wurde. Die Ergebnisse bestätigen auch diese Vermutungen, dass in den Befragungsdaten einmalig vorkommende Namen mit einer geringeren Verknüpfungswahrscheinlichkeit einhergehen. Der signifikante Effekt bei Personen mit Migrationshintergrund der dritten Generation kann nur bedingt als Hinweis darauf dienen, dass eventuell fehleranfälligere ausländische Namen mit einer höheren Ausfallwahrscheinlichkeit einhergehen. Auffällig ist der signifikante Einfluss des Befragungsmodus auf die Wahrscheinlichkeit, die beiden Datenquellen erfolgreich miteinander verknüpfen zu können. Die höhere Verknüpfungswahrscheinlichkeit von Telefoninterviews (CATI) gegenüber persönlichen Interviews (CAPI) lässt sich durch den Sachverhalt erklären, dass Personen zu denen vor der Befragung keine Telefonnummern vorlagen ins CAPI-Feld übernommen wurden. Dieses kann daher möglicherweise als Proxy für eine schlechtere Adressqualität dienen und somit zu einer geringeren Verknüpfungswahrscheinlichkeit führen.

Der Verknüpfungs-Bias kann durch die im Modell enthaltenen Variablen recht gut erklärt werden, was durch ein Pseudo-R-Quadrat von 0,25 ausgedrückt wird. Die Varianz beim Verknüpfungserfolg ist also zu einem guten Teil auf die untersuchten Merkmale zurückzuführen.

4.3 Auswirkungen auf Forschungsergebnisse

In diesem Abschnitt sollen anhand eines Beispiels, wie bei Krug und Hartmann (2009) empfohlen, die möglichen Unterschiede in den Forschungsergebnissen untersucht werden, die durch den Ausfall von Fällen und einer selektiven Reduzierung der Stichprobe entstehen können. Dazu werden die Ergebnisse einer Analyse unter Einschluss aller Personen mit den Ergebnissen einer Analyse unter Ausschluss der nicht verknüpfbaren Personen miteinander verglichen und auf Differenzen überprüft. Um die Verknüpfung zu simulieren, wird das erste Modell mit allen verfügbaren Fällen durchgeführt, während im zweiten Modell nur solche Be-

fragten berücksichtigt werden, die der Datenzuspielung zustimmen und auch tatsächlich verknüpft werden können. Von Interesse ist dabei, ob die Modelle zu unterschiedlichen Ergebnissen bei der Identifikation signifikanter Einflussfaktoren kommen. Unterschiedliche Regressionskoeffizienten zwischen den beiden Modellen können als Hinweis dienen, dass die Selektivität durch den Ausfallprozess groß genug ist, um die Ergebnisse zu verzerren.

Als Forschungsbeispiel wird eine Regression mit dem logarithmieren Brutto-Stundenlohn als zu erklärende Variable für Personen zwischen 25 und 64 Jahren durchgeführt. Die Analysen werden nur mit den Befragungsdaten durchgeführt, da nicht für alle Personen im ersten Modell eine Zuspielerlaubnis vorliegt. Diese Anwendung kann nur Aussagen über den Selektivitätsprozess bezüglich der im Modell verwendeten Variablen machen. Eine Übertragung auf andere Analysen oder eine Verallgemeinerung ist damit nicht möglich. Da die Modelle auf sich überschneidenden Stichproben beruhen, werden für die Berechnung der Koeffizienten „Seemingly Unrelated Estimations“ verwendet. Die Regressionskoeffizienten der beiden Modelle werden dann mit Hilfe von Wald-Tests auf Unterschiedlichkeit getestet.

Tabelle 5: Lineare Regression des logarithmierten Bruttostundenlohns

	Vollständige Stichprobe			Eingeschränkte Stichprobe		
	Koef.		Std. Fehler	Koef.		Std. Fehler
Geschlecht						
weiblich	Ref.			Ref.		
männlich	0,171	***	0,020	0,184	***	0,023
Alter						
metrisch	0,028	***	0,008	0,027	**	0,010
Alter quadriert/100						
metrisch	-0,024	*	0,010	-0,024	*	0,011
Bildung						
niedrig	Ref.			Ref.		
mittel mit Berufsausbildung	0,164	***	0,036	0,191	***	0,041
mittel ohne Berufsausbildung	0,094		0,059	0,100		0,073
Fachhochschulreife/ Hochschulreife	0,221	***	0,040	0,223	***	0,046
keine Information	0,064		0,112	0,110*		0,045
Region						
neue Bundesländer	Ref.			Ref.		
alte Bundesländer	0,305	***	0,019	0,314	***	0,021
Migrationshintergrund						
Kein Migrationshintergrund	Ref.			Ref.		
Selbst zugezogen	-0,126	***	0,028	-0,148	***	0,032
Mind. 1 Elternteil zugezogen	-0,070	*	0,032	-0,065		0,037
Mind. 1 Großelt. zugez., Elt. in D geb.	0,047		0,061	0,070		0,080
keine Information	-0,036		0,070	-0,002		0,103

	Vollständige Stichprobe			Eingeschränkte Stichprobe		
	Koef.		**Std. Fehler**	**Koef.**		**Std. Fehler**
ISCO 88						
0 "Soldaten"	-0,098		0,085	0,028		0,194
1 "Angehörige gesetzgebender Körperschaften, leitende Verwaltungsbedienstete und Führungskräfte in der Privatwirtschaft"	0,229	***	0,041	0,229	***	0,051
2 "Akademiker"	0,197	***	0,027	0,202	***	0,033
3 "Techniker und gleichrangige nichttechnische Berufe"	Ref.			Ref.		
4 "Bürokräfte"	-0,159	***	0,025	-0,180	***	0,028
5 "Dienstleistungsberufe"	-0,38	***	0,028	-0,393	***	0,032
6 "Fachkräfte in der Landwirtschaft und Fischerei"	-0,357	***	0,084	-0,456	***	0,096
7 "Handwerks- und verwandte Berufe"	-0,234	***	0,032	-0,258	***	0,037
8 "Anlagen- und Maschinenbediener sowie Montierer"	-0,298	***	0,034	-0,346	***	0,039
9 "Hilfsarbeitskräfte"	-0,539	***	0,034	-0,542	***	0,036
Stellung im Beruf						
Arbeiter	-0,004		0,023	0,009		0,026
Angestellter	Ref.			Ref.		
Beamter	0,052	*	0,021	-0,020		0,051
Selbstständiger	-0,251	***	0,042	-0,369	***	0,058
keine Information	-0,621	***	0,085	-0,748	***	0,107
Teilzeit						
nein	Ref.			Ref.		
ja	0,078	***	0,022	0,085	***	0,025
Familienstand						
verheiratet	Ref.			Ref.		
nicht verheiratet	-0,060	**	0,021	-0,062	*	0,024
keine Information	0,013		0,033	0,008		0,039
Haushaltstyp						
Single	-0,100	***	0,029	-0,090	**	0,033
Paar ohne Kind	0,017		0,019	0,019		0,022
Alleinerziehend	-0,113	***	0,031	-0,096	**	0,034
Paar mit Kind	Ref.			Ref.		
andere Haushaltstypen	-0,072		0,051	-0,042		0,062
keine Generierung möglich	-0,087		0,055	-0,051		0,062
Konstante						
	1,642	***	0,172	1,652	***	0,201
Stichprobengröße	4824			3471		
R-Quadrat	0,34			0,36		
Wald-Test auf Gleichheit aller Koeffizienten in beiden Modellen				chi2(33) = 42,9		
				Prob > chi2 = 0,116		

*Signifikanzniveau: * P<0.05; ** P<0.01; *** P<0.001*

Auf eine inhaltliche Interpretation der Analyse wird verzichtet, da das Forschungsbeispiel lediglich einem Vergleich der in beiden Modellen erzielten Regressionskoeffizienten dienen soll. Wie in Tabelle 5 dargestellt, lassen sich bei dem gewählten Forschungsbeispiel keine entscheidenden Unterschiede zwischen den beiden Modellen, dem vollständigen Modell mit allen Befragten und den eingeschränkten Modell nur mit den zustimmenden und tatsächlich verknüpften Personen, erkennen. Die nur geringen Unterschiede in den Signifikanzniveaus der einzelnen Regressionskoeffizienten können auf die geringere Fallzahl im zweiten Modell zurückgeführt werden. Es zeigt sich, dass es bei keinem signifikanten Regressionskoeffizient zu einem Wechsel der Wirkungsrichtung kommt. Zwar unterscheiden sich einzelne Regressionskoeffizienten signifikant zwischen den beiden Modellen[16], es kann aber kein signifikanter Unterschied aller Koeffizienten in den beiden Modellen festgestellt werden. Es soll hier nochmal betont werden, dass Abweichungen in den Ergebnissen durch Selektivitäten der Stichprobe immer von den verwendeten Variablen abhängen und sich das hier erzielte Ergebnis nicht auf andere Untersuchungen mit den PASS-Daten übertragen lässt.

5. Konsequenzen und Aussichten

Die Verknüpfung von Befragungs- mit Prozessdaten erweitert die Analysemöglichkeiten für Fragen der Arbeitsmarktforschung. Der dadurch erzielte Informationszugewinn kann jedoch auch mit einer Selektivität der Stichprobe einhergehen. Die Analysen zur Selektivität bei der Verknüpfung von Datenquellen mit Einwilligung der Befragten weisen bei den in unserem Fall verwendeten PASS-Daten auf eine eher geringe Selektivität durch das Zustimmungsverhalten hin. Dies zeigt die geringe Varianzaufklärung im ersten Modell. Die meisten betrachteten Variablen in dem geschätzten Modell bleiben ohne signifikanten Einfluss. Für die in Kapitel 3.2 formulierten Forschungsfragen lassen sich die Ergebnisse wie folgt zusammenfassen: Bei den soziodemographischen Merkmalen kommt es nur zu marginalen Selektivitäten. Hier lässt sich festhalten, dass Personen mit Migrationshintergrund der ersten Generation sowie Personen ohne Erwerbseinkommen in einem verknüpften Datensatz unterrepräsentiert sein dürften. Es gibt Hinweise für die Annahme, dass Personen, die erhöhten Wert auf den Schutz ihrer Daten legen und dementsprechend bei sensitiven Fragen keine gültigen Angaben gemacht haben, eher die Zuspielung ihrer Daten verweigern. Dem Bezug von staatlichen Transferleistungen kommt im Zustimmungsmodell keine relevante Bedeutung zu. Die Annahme, dass die Nähe des Befragten zum Inhalt der Befragung und dem Auftrag gebenden Institut eine positive Auswirkung auf die Zuspielbereitschaft hat, bestätigt sich nicht. Die Ergebnisse bekräftigen jedoch den Stellenwert des Interviewers und seiner Eigenschaften beim Antwortverhalten des Befragten. Allerdings konnte kein Interaktionseffekt zwischen den Merkmalen der Interviewer und der Befragten festgestellt werden. Gerade durch diese Interaktionen könnte es zu einer Selektivität der reduzierten Stichprobe durch Interviewermerkmale kommen. Die Selektivität durch die tatsächliche Verknüpfung wird, wie

[16] Bei der hier berechneten Anzahl von Koeffizienten kann eine gewisse Anzahl von Abweichungen als zufällig verstanden werden.

vermutet, vor allem durch die Wahrscheinlichkeit Informationen in den Prozessdaten aufzuweisen, vorangetrieben. Da für fast alle der Zuspielung zustimmenden Personen in der Leistungsempfängerstichprobe die Verknüpfung realisiert werden konnte, beschränkt sich die Analyse des Verknüpfungs-Bias auf die Stichprobe der Wohnbevölkerung. Vor allem Beamte und Selbständige, aber auch Nichterwerbspersonen weisen eine geringere Verknüpfungswahrscheinlichkeit auf. Deswegen ist in einer verknüpften Stichprobe mit diesbezüglicher Selektivität zu rechnen.

Zusammenfassend ist festzuhalten, dass es im PASS durch einen zweistufigen Ausfallprozess zu einer Reduzierung der Stichprobe kommt. Der Einfluss der dadurch entstehenden bedingt selektiv zusammengesetzten Stichprobe auf Forschungsergebnisse ist abhängig von den in den jeweiligen Analysemodellen verwendeten Merkmalen. Es wird den Nutzern des PASS empfohlen, bei Analysen mit verknüpften Daten besonderes Augenmerk auf die verwendeten Variablen zu legen. Weitere Arbeiten zu diesem Thema erscheinen notwendig, um die Selektivität der Stichprobe bei der Nutzung von verknüpften Datensätzen genauer einschätzen zu können. Zusätzliche Erkenntnis könnte die Berücksichtigung von Längsschnittfaktoren bringen. So könnten Personen die aufgrund einer vorherigen Verweigerung wiederholt nach ihrer Zuspielerlaubnis gefragt wurden genauer untersucht werden. Auch könnten Informationen aus der Kontaktphase in die Modelle mit aufgenommen werden. Vor allem aber die genaue Bedeutung des Interviewers bedarf weiterführender Forschung. Hier könnten weitere Cross-Level-Interaktionseffekte in die Analysen mit eingebaut werden. Im PASS wurde für die fünfte Welle eine detaillierte Interviewerbefragung durchgeführt. Diese zusätzlichen Informationen zum Interviewer dürften hilfreich sein, um die Fragestellung weiter zu bearbeiten.

Literatur

Bachteler, Tobias (2008): Dokumentation Record Linkage IEB-PASS. Unveröffentlichtes Manuskript.

Campanelli, Pamela und O'Muircheartaigh, Colm (1999): Interviewers, Interviewer Continuity, and Panel Survy Nonresponse. Quality and Quantity 33: 59-76.

Dunn, Kate M.; Jordan, Kelvin; Lacey, Rosie J.; Shapley, Mark und Jinks, Clare (2004): Patterns of Consent in Epidemiologic Research: Evidence from Over 25,000 Responders. American Journal of Epidemiology 159(11): 1087-1094.

Hartmann, Josef und Krug, Gerhard (2009): Verknüpfung von personenbezogenen Prozess- und Befragungsdaten – Selektivität durch fehlende Zustimmung der Befragten? Zeitschrift für Arbeitsmarktforschung 42: 121-139.

Haunberger, Sigrid (2009): The effect of interviewers, respondent and area characteristics on cooperation in panel surveys: a multilevel approach. Quality and Quantity 44: 957-969.

Jacobebbinghaus, Peter und Seth, Stefan (2007): The German Integrated Employment Biographies Sample IEBS. Schmollers Jahrbuch 127(2): 335-342.

Jenkins, Stephan P.; Cappelari, Lorenzo; Lynn, Peter; Jäckle, Annette und Sala, Emanuela (2006): Pattern of consent: evidence from a general household survey. Journal of the Royal Statistical Society: Series A (Statistics in Society) 169(4): 701-722.

Jenkins, Stephan P.; Lynn Peter; Jäckle, Annette und Sala, Emanuela (2008): The Feasibility of Linking Household Survey and Administrative Record Data: New Evidence for Britain. International Journal of Social Research Methodology 11(1): 29-43.

Kho, Michelle E.; Duffett, Mark; Willison, Donald; Cook, Deborah J. und Brouwers, Melissa C. (2009): Written informed consent and selection bias in observational studies using medical records: systematic review. British Medical Journal 338: 866-873.

Olsen, (1999): Linkages with Data from Social Security Administrative Records in the Health and Retirement Study. Social Security Bulletin 62(2): 73-85.

Prost, Rolf und von Briel, Christa (1995): Wären Sie vielleicht bereit, sich gegebenenfalls noch einmal befragen zu lassen? Oder: Gründe für die Teilnahme an Panelbefragungen. ZUMA-Arbeitsbericht 95/04.

Schmucker, Alexandra und Huber, Martina (2009): Linking Process generated employment data with survey data – an assessment of linkage and response quality. FDZ Methodenreport 5/2009.

Schnell, Rainer, Bachteler, Tobias; Reiher, Johannes (2005). "MTB: Ein Record-Linkage-Programm für die empirische Sozialforschung." ZA-Information 56: 93-103.

Woolf, Steven H.; Rothemich, Stephen F.; Johnson, Robert E. und Marsland, David W. (2000): Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research. Archives of Family Medicine 9(10): 1111-1118.

Young, Anne F.; Dobson, Annete J. und Byles, Julie E. (2001): Health services research using linked records: who consents and what is the gain. Australien and New Zealand Journal of Public Health 25(5): 417-420.

Impressum

FDZ-Methodenreport 09/2011

Herausgeber
Forschungsdatenzentrum (FDZ)
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Str. 104
90478 Nürnberg

Redaktion
Stefan Bender, Iris Dieterich

Technische Herstellung
Iris Dieterich

Rechte
Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit
http://doku.iab.de/fdz/reporte/2011/MR_09-11.pdf

Internet
http://fdz.iab.de/

Rückfragen zum Inhalt an:
Jonas Beste
Institut für Arbeitsmarkt- und
Berufsforschung (IAB)
Regensburger Str. 104
90478 Nürnberg
Tel.: +49 (911) 179 2279
E-Mail: jonas.beste@iab.de