

Research Notes

31

Alexia Meyermann • Jennifer Elsner • Jürgen Schupp •
Stefan Liebig

**Pilotstudie einer surveybasierten
Verknüpfung von Personen- und
Betriebsdaten**

Berlin, März 2009

IMPRESSUM

DIW Berlin, 2009
Deutsches Institut für Wirtschaftsforschung
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1860-2185

Alle Rechte vorbehalten.
Abdruck oder vergleichbare
Verwendung von Arbeiten
des DIW Berlin ist auch in
Auszügen nur mit vorheriger
schriftlicher Genehmigung
gestattet.

Research Note 31

Pilotstudie einer surveybasierten Verknüpfung von Personen- und Betriebsdaten

**Durchführung sowie Generierung einer Betriebsstudie als nachgelagerte
Organisationserhebung zur SOEP-Innovationsstichprobe 2007**

Alexia Meyermann*, Jennifer Elsner*, Jürgen Schupp und Stefan Liebig*****

* **Institut für Soziologie, Universität Duisburg-Essen**

** **SOEP am DIW Berlin sowie Institut für Soziologie der Freien Universität Berlin**

*** **Fakultät für Soziologie, Universität Bielefeld sowie DIW Berlin**

Zusammenfassung

Der Beitrag diskutiert hinsichtlich der Erstellung verknüpfter Datensätze von Arbeitnehmern und Betrieben unterschiedliche Vorgehensweisen und die damit verbundenen spezifischen Vor- und Nachteilen. Ausführlich wird dabei die employee-first-method vorgestellt, zu dem bislang in der Surveyforschung lediglich entsprechende Vorbildstudien aus den USA vorliegen. Es kann für Deutschland gezeigt werden, dass mit Hilfe dieser Methodik erstens die Angaben der Personen hinreichend vollständig und fehlerfrei sind. Die Analysen basieren auf den Ergebnissen einer nachgelagerten Erhebung von Betrieben, die im SOEP-Pretest im Jahr 2007 erstmals bei mehr als 300 Beschäftigten erprobt wurde. Hierzu weiterhin zeigen die Ergebnisse, dass mittels der Methode der Datenedition relativ einfach eine beachtliche Verbesserung der Güte der durch offene Abfrage erhaltenen Betriebskontaktadressen erreicht werden. Die Güte und Anzahl der Nennungen der Betriebskontaktdaten ebenso wie die Anzahl der teilnehmenden Betriebe kann zudem durch eine Reihe an Verbesserungen in den unterschiedlichen Phasen des Generierungsprozesses gesteigert werden.

Abstract

***Pilot Study for Survey-Based Linkage of Individual and Firm Data:
Conducting and generating a firm study as a subsequent organisational survey of firms in
the 2007 SOEP innovation sample***

In this paper, we compare different approaches to the generation of datasets linking employees and firms, and present the specific advantages and disadvantages of each method. We discuss in detail the employee-first method, which has only been investigated so far in the survey research in initial US studies. We show that for Germany, this method provides reasonably comprehensive and error-free data on individuals. Our analyses are based on the results of a follow-up study of firms that was first tested in a SOEP Pretest in the year 2007 on more than 300 employees. Our results also show that the data edition method provides a relatively easy means of considerably improving the quality of contact addresses obtained through an open-ended survey question. The quality and number of the firm contact data as well as the number of participating firms can be increased as well through improvements in the different phases of the generation process.

Keywords: *German Socio-Economic Panel Study, SOEP, Linked Employer-Employee-Data*

JEL Classification: A12, C81, C83, J2,

Inhaltsverzeichnis

| | |
|---|----|
| 1. Einleitung | 4 |
| 2. Verknüpfte Individual- und Organisationsdatensätze | 4 |
| 2.1. Methodik der Datensatzerstellung..... | 6 |
| 2.2. Aspekte des Datenschutzes | 8 |
| 2.3. Besonderheiten der Analyseeinheit..... | 9 |
| 3. Vorgehen und Fragestellung | 10 |
| 3.1. Fragestellungen | 10 |
| 3.2. Vorgehen | 12 |
| 4. Ergebnisse | 14 |
| 4.1. Selektivität der Stichproben: Ausschöpfungsquoten..... | 14 |
| 4.2. Nonresponse-Determinanten auf Personenebene..... | 14 |
| 4.3. Generierung der Betriebsstichprobe..... | 17 |
| 4.3.1 Ausschöpfungsquoten | 17 |
| 4.3.2. Edition der Adressdatei | 18 |
| 4.3.3. Durchführbarkeit und Befragungsbereitschaft | 21 |
| 4.4. Erhebung der Betriebsinformationen | 22 |
| 4.4.1. Telefonischer Vorkontakt..... | 22 |
| 4.4.2. Probleme der Fallzahlenrealisierung..... | 23 |
| 4.4.2.1. Erreichbarkeit | 23 |
| 4.4.2.2. Identifikation | 24 |
| 4.4.2.3. Zustimmung | 26 |
| 4.4.3. Online- und postalische Befragung | 28 |
| 4.4.3.1 Ergebnisse | 28 |
| 4.4.3.2 Mögliche Ursachen für Unit-Nonresponse | 28 |
| 4.5. Stichprobenvergleich..... | 30 |
| 5. Zusammenfassung und Ausblick | 32 |
| Literaturverzeichnis..... | 33 |
| Anhang | 35 |

1 Einleitung

Das für die empirische Sozial- und Wirtschaftsforschung zugängliche Angebot an Massendatensätzen in Deutschland zeichnet sich durch ein konzeptionelles Defizit aus: Es fehlen sogenannte Linked-Employer-Employee-Datensätze (LEE-Datensätze), die sowohl ausführliche Informationen über die ökonomischen, personalpolitischen oder auch arbeitsorganisatorischen Strukturen und Prozesse von Betrieben als auch Informationen über die soziale und ökonomische Situation ihrer Beschäftigten und deren Haushalte enthalten. Derartige verknüpfte Daten sind für die Wissenschaft zunehmend wichtiger, weil sich durch Prozesse der Deregulierung und Verbetrieblichung zentrale Entscheidungs- und Gestaltungskompetenzen und Einflüsse in Bezug auf die Gratifikation und die Ausgestaltung von Arbeitsverhältnissen auf die einzelbetriebliche Ebene verlagern (Struck 2005). Neben dem Bedarf an Auf- und Ausbau des Bestands an verknüpften Datensätzen selbst, fehlen der Umfrageforschung in Deutschland auch Informationen über die methodische Vorgehensweise der Generierung derartiger Datensätze, welche ganz besondere Herausforderungen an die Umfrageforschung stellen. Mit dem vorliegenden Beitrag soll daher der bestehende Wissensstand durch aktuelle Befunde einer deutschen Pilotstudie erweitert werden. Dazu wurden in der SOEP-Pretest Studie 2006 die Betriebskontaktdaten der befragten Personen erhoben. Auf Basis dieser kann ein verknüpfter Datensatz erstellt werden, indem die SOEP-Individualdaten mit im Anschluss erhobenen Betriebsinformationen ergänzt werden. Die Erfahrungen sollen zudem dazu genutzt werden, ein solches Design für die SOEP-Haupterhebung vorzubereiten.

Ziel ist es, die Güte der personengenerierten Betriebskontaktadressen zu evaluieren. Insbesondere interessieren folgende Fragestellungen: Erstens, wie ist das Antwortverhalten der befragten Personen im Hinblick auf die Offenlegung ihrer Betriebszugehörigkeit (Namen und Adresse des Arbeitgebers), zweitens wie groß ist der Aufwand, den die Edition der Daten verlangt, um von den offen abgefragten Kontaktdaten der Betriebe zu einer kontaktierbaren Betriebsdatei zu gelangen, und drittens sind Hinweise auf die Qualität einer über Personenangaben generierten Betriebsstichprobe im Hinblick auf die Aspekte der Generalisierbarkeit und Selektivität zu untersuchen.

In Abschnitt 2 werden unterschiedliche LEE-Datensätze vorgestellt und die jeweiligen Vor- und Nachteile verschiedener Arten der Erstellung diskutiert. Nach einer Einführung in die vorliegenden Fragestellungen sowie das zugrunde liegende Forschungsdesign in Abschnitt 3, folgt in Abschnitt 4 die ausführliche Darstellung der Ergebnisse der einzelnen Phasen des Forschungsprozesses.

2 Verknüpfte Individual- und Organisationsdatensätze

In der internationalen Arbeitsmarktforschung existieren mittlerweile eine Reihe verknüpfter, sogenannter „matched-“ oder „linked-employer-employee“ Datensätze. (vgl. z.B. Abowd/Kramarz 1999, Bender et al 2008, Hamermesh 2008) Linked Employer-Employee-Datensätze zeichnen sich durch eine hierarchische Mehrebenenstruktur aus, bei der die Beschäftigten eines Betriebs die unterste, die Betriebe und/oder Unternehmen die darüber liegende Ebene bilden. Das entscheidende Kennzeichen dieser LEE-Daten besteht darin, dass sie Information von mehreren oder – im Idealfall – allen im Betrieb beschäftigten Personen enthalten. Grundlage bilden in den meisten Fällen prozessgenerierte Daten entweder auf der Individual- und/oder auf der Betriebsebene (vgl. für einen Überblick Abowd/Kramarz 1999 sowie Kaiser/Wagner 2008 für die aktuelle Situation in Deutschland).

Im Vergleich zur Entwicklung in anderen europäischen und außereuropäischen Ländern wurde das Analysepotential von LEE-Daten in Deutschland erst sehr spät erkannt. Martin Falk

und Viktor Steiner kommen deshalb in ihrer Expertise für die „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI)“ noch 2001 zu dem Schluss: „Vernetzungsmöglichkeiten zwischen Betriebs- und Personendaten wurden in anderen Ländern viel früher genutzt. In Teilbereichen, wie beispielsweise der betrieblichen Beschäftigungs- und Einkommensentwicklung, ist die Forschung in Deutschland nicht mehr wettbewerbsfähig. Forschung auf diesem Gebiet findet in Deutschland daher kaum statt“ (S. 8). Ein Grund für diese Situation bestand darin, dass derartige Datensätze schlichtweg für die Wissenschaft nicht verfügbar waren.

Mittlerweile sind die Gehalts- und Lohnstrukturerhebung der amtlichen Statistik (Stephan 2001) sowie der Linked-Employer-Employee-Datensatz des Instituts für Arbeitsmarkt- und Berufsforschung (LIAB) (Alda et al. 2005) in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder sowie der Bundesagentur für Arbeit für wissenschaftliche Zwecke zugänglich (vgl. Bender et al. 2008). In beiden Fällen handelt es sich um „echte“ Linked-Employer-Employee-Datensätze, in denen für jeden beteiligten Betrieb alle bzw. eine hinreichend große Auswahl der Beschäftigten enthalten sind. Das Analysepotential beider Datensätze ist sehr unterschiedlich. Der zentrale Unterschied besteht im Ausmaß der verfügbaren betrieblichen Informationen. Die Gehalts- und Lohnstrukturerhebung enthält lediglich grundlegende Parameter der Beschäftigtenstruktur, der Branchenzugehörigkeit oder der Tarifbindung und stellt eine Querschnitterhebung dar. Damit eignet sie sich (insbesondere nach der Ausweitung der erfassten Betriebe und Branchen durch die Änderung des entsprechenden Gesetzes zum 1. Januar 2007) primär für die Analyse von Einkommensstrukturen im Querschnitt. Demgegenüber weist der LIAB einen thematisch breiten Informationsgehalt auf, der sich bezüglich der Firmen von einer detaillierten Beschäftigungsstruktur über die ökonomische Situation des Betriebs, die Weiterbildungsmaßnahmen bis hin zu Arbeitszeitregelungen, Art der Entlohnungssysteme oder Angebote zur Vereinbarkeit von Familie und Beruf erstreckt. Auch wenn diese Themenschwerpunkte ergänzungsbedürftig sind – beispielsweise im Hinblick auf die bestehenden Mobilitätsregimes oder den in der Organisationsforschung als wichtig angesehenen Aspekten der Unternehmens- oder Betriebskultur – so eröffnet der LIAB auf betrieblicher Seite durchaus ein Analysepotential, das über die klassischen arbeitsmarktökonomischen und –soziologischen Fragestellungen hinausweist – zumal er als Längsschnittstudie sowohl auf der Ebene der Betriebe als auch der Beschäftigten zeitliche Verläufe abbildet. Dieser zentrale Vorteil wird jedoch durch den geringen Informationsbestand auf der Seite der Beschäftigten geschmälert, womit der LIAB ein grundsätzliches Manko der Gehalts- und Lohnstrukturerhebung teilt. Beide in Deutschland für die Wissenschaft verfügbaren LEE-Datensätze zeichnen sich somit durch eine geringe Verfügbarkeit von Informationen über Personen und Haushalte aus. Dies gilt sowohl für zentrale Merkmale des aktuellen Beschäftigungsverhältnisses (im LIAB lassen sich z.B. befristete Arbeitsverhältnisse oder Arbeitnehmerüberlassungen nicht identifizieren), als auch für Informationen, die sich auf die über das aktuelle Beschäftigungsverhältnis hinausweisende ökonomische Situation der Person beziehen und erst recht für Angaben über den Haushalt, die familiäre Situation, die soziale Herkunft, die sozialen Präferenzen und Persönlichkeitsmerkmale, Normen- und Wertevorstellungen oder auch politische Einstellungen und Mitgliedschaften in Verbänden oder Parteien.

Vor diesem Hintergrund erscheint eine Erweiterung des bestehenden Angebots an verknüpften Personen-Betriebsdaten dringend geboten.¹ Dies gilt vor allem deshalb, weil der für Deutschland geltende eingeschränkte Informationsbestand im internationalen Vergleich eine Ausnahme darstellt. Die Erweiterung des bisherigen Datenangebots muss sich insbesondere

¹ Eine vergleichbare Lücke im bisherigen Datenangebot verknüpfter Datenbestände ist Gegenstand eines derzeit laufenden Projekts an der Universität Lüneburg, in dem es um die Verknüpfbarkeit von Betriebsdaten aus unterschiedlichen Datenbeständen geht (Bender/Wagner/Zwick 2007)

auf Informationen beziehen, die über die Eckdaten der aktuellen Erwerbstätigkeit hinausweisen und beispielsweise die soziale Herkunft, die familiäre und häusliche Situation, die Integration in soziale Netzwerke sowie Wertvorstellungen und politische Einstellungen der Personen erfassen. Auf betrieblicher Seite sind genauso Ergänzungen denkbar, die sich etwa in Erweiterung des Fragenkatalogs des IAB-Betriebspanels auf Lohn- und Gehaltsbildungsprozesse, Elemente der Unternehmens- und Betriebskultur, der industriellen Beziehungen oder der nationalen und internationalen Wettbewerbssituation der Betriebe beziehen. Ein derartiger Katalog an Informationen lässt sich nur auf der Grundlage von miteinander verknüpften Personen- und Betriebsstättenbefragungen erheben.

2.1 Methodik der Datensatzerstellung

Zur Generierung verknüpfter Datensätze können zwei unterschiedliche Wege beschritten werden, die sich in ihrem Ausgangspunkt – den Arbeitgebern oder -nehmern – und den damit einhergehenden Implikationen unterscheiden.

(1) Establishment-first Methode

Bei der Establishment-first Methode – dem auch in der Gehalts- und Lohnstrukturerhebung der amtlichen Statistik verfolgten Vorgehen – werden in einem ersten Auswahlschritt Betriebe ausgewählt, in denen dann von allen oder einer Teilstichprobe der Beschäftigten entsprechende Individualinformationen gesammelt werden (vgl. etwa der 2000 National Employer Survey, Capelli 2001). Vorteilhaft ist hier, dass die in den üblichen LEE-Datensätzen bestehende hierarchische Mehrebenenstruktur der Daten bestehen bleibt, die es ermöglicht Unterschiede zwischen einzelnen Betrieben sowie innerhalb der Betriebe analytisch voneinander zu unterscheiden. Eine derartige Erhebung stößt jedoch relativ schnell an ihre Realisierungsgrenzen, wenn die Informationen über die Beschäftigten nicht vom Betrieb selbst bereitgestellt werden können, sondern dazu eine gesonderte Mitarbeiterbefragung durchgeführt werden muss. Die Koordinierung und Umsetzung derartiger Mitarbeiterbefragungen in mehr als 100 oder 200 Betrieben erscheint auch bei Einschaltung eines kommerziellen Instituts im Rahmen der üblichen Forschungsprojekte schwierig zu realisieren. Entsprechend orientiert sich etwa ein aktuell diesem Design verpflichtetes Projekt in Deutschland auf eine einstellige Anzahl an Betrieben (Brose et al 2006).

(2) Employee-first Methode

Bei der Employee-first Methode stellen nicht Betriebe, sondern Personen, d.h. abhängig Beschäftigte, den Ausgangspunkt dar. Die über Befragungen gewonnenen Individualdaten werden mit Betriebsdaten ex post ergänzt. Dies kann wiederum über drei unterschiedliche, in der bisherigen Forschung beschrittene Wege geschehen (vgl. Kmec 2003):

- (1) Es werden Informationen über die in der Personenbefragung genannten Unternehmen, wie sie in den gängigen Wirtschaftsdatenbanken (für Deutschland: Creditreform oder Hoppenstedt) verfügbar sind, verwendet und den Individualinformationen zugespielt. Beispiele dafür sind etwa die „New Worker Establishment Characteristics Database“ (Troske et al 2000) oder der „Decennial Employer-Employee Dataset“ (Bayard et al 2002).
- (2) Auf der Personenebene werden Informationen erhoben, die eine Identifikation des beschäftigenden Betriebs in bestehenden Betriebsdatensätzen erlauben. So wurden im Rahmen einer Studie des Max-Planck Instituts für Bildungsforschung in Berlin die Befragten nach ihrer Sozialversicherungsnummer gefragt, um daran anschließend eine Verknüpfung mit den Daten des IABs vornehmen zu können (vgl. Reimer/Kuenster 2004). Falls der beschäftigende Betrieb einer Person im IAB-Betriebspanel enthalten

ist, können die entsprechenden Betriebsinformationen den Personendaten hinzugefügt werden.

- (3) Schließlich kann man die gewünschten Betriebsinformationen über eine gesonderte Betriebsstättenbefragung der genannten Betriebe erheben und den Individualdaten hinzufügen. Beispiele dafür sind der National Organization Survey (NOS) oder die Multi-City Study of Urban Inequality (vgl. Kmec 2003).

Der Rückgriff auf Informationen von Wirtschaftsdatenbanken anstatt der Verknüpfung von Daten auf eigenständigen Erhebungen beruhen stellt die kostengünstigere Alternative dar, doch sind damit eine Reihe grundlegender Probleme sowie analytische Beschränkungen verbunden. Denn die verfügbaren Informationen (z.B. Anzahl der Beschäftigten, Gründungsjahr, Geschäftsvolumen) sind zwar zur Beurteilung der Liquidität oder der Bonität eines Unternehmens geeignet, für sozialwissenschaftliche Fragestellungen jedoch nur sehr bedingt brauchbar. Zudem werden in den Wirtschaftsdatenbanken nur Betriebe aufgenommen, deren Geschäftsvolumen eine bestimmte Grenze überschreitet und die eine bestimmte Mindestanzahl an Beschäftigten haben (Kmec 2003: 87). Mittelständische und Großbetriebe sind deshalb gut vertreten, Kleinbetriebe fehlen jedoch. Da im Jahr 2006 rund ein Viertel der Arbeiter und Angestellten in Betrieben mit weniger als 20 Mitarbeitern tätig waren (SOEP 2006), können somit immerhin für die verbleibenden drei Viertel der abhängig Beschäftigten in Deutschland Betriebsinformationen genutzt werden.

Beim zweiten Vorgehen besteht das Problem, dass der Anteil an Beschäftigten aus einer Bevölkerungsbefragung, die in einem Betrieb des IAB-Betriebspanels beschäftigt sind, erwartungsgemäß gering sein wird. Denn das IAB-Betriebspanel ist keine Vollerhebung, sondern selbst wiederum eine Stichprobenerhebung, deren Stichprobenumfänge zudem regional sehr unterschiedlich sind.

Vor diesem Hintergrund ist das dritte Vorgehen, die Ergänzung von umfragebasierten Individualdaten durch eine Betriebsstättenbefragung, weitaus vielversprechender. Als beispielhaft gilt hier der National Organization Survey (NOS) von 1991 und 2002. Im Rahmen des General Social Survey (GSS) 1991 und 2002 wurden alle (1991) bzw. ein Teil (2002) der aktuell erwerbstätigen Personen nach dem Namen und der Adresse ihrer Arbeitsstelle befragt.² Als Zieleinheiten galten die lokalen Betriebsstätten, in denen die Personen ihrer Erwerbstätigkeit nachgingen. Auf der Grundlage dieser Nennungen wurde eine eigenständige telefonisch-schriftliche Betriebsstättenbefragung durchgeführt. Diese Informationen wurden dann den Individualdaten des GSS wieder zugespielt. Das Ergebnis war ein verknüpfter Personen-Betriebsdatensatz (vgl. Kalleberg et al. 1996; Smith et al. 2004). Insgesamt konnten 1991 von 51 Prozent und 2002 von 48 Prozent der Erwerbstätigen eine Verknüpfung von Individual- und Betriebsdaten vorgenommen werden³.

Im Unterschied zu den „klassischen“ Linked-Employer-Employee-Datensätzen weist dieser Datensatz jedoch keine hierarchische Mehrebenenstruktur auf. Denn es sind für einen Betrieb in der Regel nur die Individualdaten eines Beschäftigten verfügbar. Gleichwohl lässt sich auf dieser Grundlage untersuchen, welche Strukturen und Prozesse in Betrieben für unterschiedliche, sozial- und wirtschaftswissenschaftliche Phänomene wichtig sind. Denn erstens können über kombinierte Personen-Betriebsbefragungen weitaus mehr Betriebsinformationen gesammelt werden als in einer Personenbefragung allein und zweitens werden die Betriebsdaten

² In der Studie aus dem Jahr 1991 bezog sich dies auch auf den beschäftigenden Betrieb des erwerbstätigen Partners.

³ Lengfeld (2008) nutze jüngst diese verknüpften Daten für organisationssoziologisch motivierte Analysen zur sozialen Ungleichheit.

in einer kombinierten Befragung von den Einschätzungen und Wahrnehmungen der Befragten der Personenbefragung unabhängig erhoben (vgl. Gupta et al. 2000).

Tabelle 1 veranschaulicht zusammenfassend die unterschiedlichen Generierungsmethoden und ihre Vor- und Nachteile.

Tab.1: Unterschiedliche Arten der Erstellung verknüpfter Datensätze im Vergleich

| | 1. Establishment-first-method | 2. Employee-first-method a) Datenbanken | 2. Employee-first-method b) eig. Erhebung |
|----------------------------|--|--|---|
| <i>1. Schritt</i> | 1) Betriebsstättenerhebung | 1) Personenerhebung | |
| <i>2. Schritt</i> | 2) Personenerhebung | 2) Betriebserhebung | |
| | | a) Betriebsdaten aus Datenbanken oder bestehenden Betriebsdatensätzen | b) Betriebsstättenbefragung |
| <i>Vorteile</i> | - Hierarchische Mehrebenenstruktur | - Praktikabel innerhalb von ALLBUS oder SOEP - Kostengünstiger | - Adäquate Abbildung der beschäftigenden Unternehmen - Unterschiedliche Forschungsinteressen verfolgbar |
| <i>Nachteile</i> | - Schnell erreichte Realisierungsgrenzen der hier i.d.R. notwendigen Mitarbeiterbefragungen - Länderspezifisch unterschiedliche Verfügbarkeit von Adresslisten - Notwendigkeit zweier Erhebungen | - Fehlende Bereitschaft oder Fähigkeit der Befragten den Namen ihres AG zu nennen - Selektionsproblem (bzgl. kleiner Betriebe) - Fehlende Mehrebenenstruktur Im Falle kommerzieller Datenbanken: - Für sozialwiss. Fragestellungen wenig brauchbar - Aktualitätsproblem | - Fehlende Bereitschaft oder Fähigkeit der Befragten den Namen ihres AG zu nennen - i.d.R. keine hierarchische Mehrebenenstruktur (aber Marsden!) - Notwendigkeit zweier Erhebungen |
| <i>Anwendungsbeispiele</i> | Niedersächsische Gehalts- und Lohnstrukturhebung (Stephan 2001), National Employer Survey 2000 (NES III) (Capelli 2001) | New Worker Establishment Characteristics Database (NWECD) (Troske et al 2000), Decennial Employer-Employee Dataset (DEED) (Bayard et al 2002) | NOS 1991, 2002 (vgl. Kalleberg 1996, Smith et al 2004, Lengfeld 2008), Multi-City Study of Urban Inequality (vgl. Kmec 2003), ISMIE (Jäckle et al 2004) |

Quelle: eigene Zusammenstellung.

2.2 Aspekte des Datenschutzes

Aufgrund der Spezifika der Generierung verknüpfter Datensätze und ihres erweiterten, sich auf zwei Analyseeinheiten beziehenden Informationsbestands, erhalten datenschutzrechtliche sowie forschungsethische Aspekte besondere Bedeutung (vgl. Calderwood/Lessof 2009). Um den datenschutzrechtlichen Bestimmungen (BDSG 2001) zu genügen, aber auch negativen Auswirkungen auf Antwortbereitschaft von befragten Personen wie Betrieben vorzubeugen und gleichzeitig die Analysepotentiale nicht zu stark einzuschränken, sind besondere Datenschutzkonzepte zu entwickeln, die potentiell auf zwei Stufen ansetzen können: (1) Im Rahmen der Datenerhebung und (2) bei der Bereitstellung der kompletten Daten.

(1) Datenschutz im Rahmen der Erhebung

Was den Datenschutz im Rahmen der Datenerhebung eines verknüpften Datensatzes angeht, so sollten die klarschriftlichen Kontaktdaten – bei einem employee-first Datensatz die Betriebsadresse, bei einem establishment-first Datensatz die Personenadresse - versehen mit einer Betriebs- bzw. Personen-ID direkt nach der Erhebung vom Ausgangsdatsatz und dessen ID getrennt werden. Eine Zuordnung von Betrieb und Beschäftigtem kann von da an ausschließlich über eine systemfreie Identifikationsnummer, eine dritte ID, geschehen. Diese ist

bei einem Treuhänder (bspw. Datenschutzbeauftragten oder Notar) zu verwahren. Zur Durchführung der Befragung der zweiten Analyseeinheit übermittelt das die Ausgangserhebung ausführende Umfrageinstitut den Forschern die Kontaktdaten ohne Angaben zu den Adressdaten der ersten Analyseeinheit. Nach erfolgter (zweiter) Befragung werden die Daten der die systemfreie ID aufbewahrenden Stelle zurückgesandt. kann anhand des Schlüssels die Betriebs- mit den Individualdaten *ohne* Name und Adresse des Betriebs und der Personen verknüpfen. Die verknüpfte Datei enthält also Betriebs- und Personen-ID sowie Betriebs- und Personeninformationen ohne Adress- oder Namensdaten und erfüllt deshalb grundsätzlich die Kriterien faktisch anonymisierte Datensätze (siehe Punkt 2). Für die eigentlichen inhaltlichen Analysen wird die Betriebs- und Personen-ID's schließlich vom verknüpften Datensatz wieder getrennt und mit einer neu zu erstellenden systemfreien ID versehen. Die Informierung oder Einwilligung der befragten Personen über die nachträgliche Kontaktierung ihrer Arbeitgeber auf Basis ihrer Auskunft ist deshalb nicht notwendig, da auf betrieblicher Ebene eine mögliche Verknüpfung sowie Rückverfolgung der personenbezogenen Daten der befragten Personen durch die getrennte Adressenpflege ausgeschlossen werden kann.

(2) Datenschutz im Rahmen der Bereitstellung

Die zweite Stufe des Datenschutzes greift auf der Ebene der Datenanalyse und der Freigabe der verknüpften Daten. Es gilt, eine Identifikation von Personen und/oder Betrieben zu vermeiden. Die üblichen Strategien der Anonymisierung durch Zusammenlegung von Kategorien und Ausblendung bestimmter Variablen sind im Fall eines verknüpften Datensatzes möglicherweise nicht ausreichend. Sobald ein Betrieb identifiziert werden kann, erhöht sich das Reidentifikationsrisiko der in diesem Betrieb beschäftigten Personen stark, da diese nur noch in der relativ kleinen Einheit Betrieb zu suchen sind.(vgl. Wirth 2003) Datenschutzrechtlich ist die Identifizierung eines Betriebs unproblematisch, da der Datenschutz ausschließlich den Schutz personenbezogener Daten umfasst. Sobald jedoch Organisationen identifiziert werden können und diese auch Namen von Personen enthalten oder über Merkmale wie z.B. Funktionsbeschreibungen bestimmten Personen zugeordnet werden können, gelten auch diese Daten als personenbezogen. Einer Identifizierung des Betriebs ist weiter deshalb vorzubeugen, da die Zusicherung und nachträgliche Sicherstellung der Anonymität von Betrieben eine wichtige Voraussetzung für deren Teilnahme darstellt.

Sollen die Betriebs- und Personendatensätze einzeln zur Analyse bereitgestellt werden, wird es notwendig, sämtliche Überschneidungsmerkmale analog zu anonymisieren. Da diese Anonymisierungsvorgänge möglicherweise die Analysepotentiale der Einzeldatensätze einschränken, sind Generierungsvarianten verknüpfter Datensätze, die auf bereits bestehenden und somit einzeln verfügbaren Betriebs- oder Personendatensätze aufbauen, mit besonderen datenschutzrechtlichen Problemen behaftet. Zudem sollten Überschneidungsmerkmale mit bestehenden, öffentlich zugänglichen Datenbanken (z.B. LexisNexis, Hoppenstedt, Creditreform, AMADEUS, MARKUS⁴) geprüft werden, da diese ebenfalls zu einer Reidentifikation von Betrieben herangezogen werden könnten.⁵

2.3 Besonderheiten der Analyseeinheit

Betriebsstättenbefragungen im allgemeinen und verknüpfte Datensätze im besonderen haben mit dem methodischen und theoretischen Problem der Uneindeutigkeit ihrer Analyseeinheit zu kämpfen. „Organizations are themselves multilevel entities“(Wharton 1994), die zudem häufig keine klar abgrenzbaren Einheiten darstellen und regelmäßigen Schwankungen ihrer

⁴ MARKUS basiert auf Creditreform. LexisNexis beinhaltet Hoppenstedt und Creditreform-Informationen.

⁵ Für eine ausführliche Betrachtung der verschiedenen Methoden der Anonymisierung sowie ihren Folgen für Analysepotentiale von Datensätzen, vgl. z.B. Rosemann 2006.

Außengrenzen durch Reorganisationsmaßnahmen ausgesetzt sind. Ob der Konzern, das Unternehmen als rechtliche Einheit, der Betrieb als lokale Einheit (als Niederlassung, Zweigstelle oder Filiale) oder die Abteilung und Arbeitsgruppe erfasst werden soll, ist zunächst einmal eine theoretisch zu beantwortende Frage. Ein methodisches Problem wird diese Frage, wenn es darum geht, wie bzw. ob die gewählten Einheiten überhaupt zu erfassen sind. Hierzu wäre eine entsprechende Adressdatei der Betriebe, Unternehmen oder Konzerne notwendig. Weiter ist es denkbar, dass die interessierende Analyseeinheit und die Einheit, welche die Kompetenz und/oder Befugnis besitzt, Fragen zu beantworten auseinander fallen. Auch die theoretische Auswahl der „richtigen“ Analyseeinheit ist nicht unproblematisch. Gegebenenfalls sind zur Beantwortung einer einzigen Fragestellung unterschiedliche Organisationsebenen relevant⁶ und eventuell unterscheiden sich die relevanten Entscheidungs- und Strukturierungsebenen, die zur Erklärung eines bestimmten Phänomens gemessen werden sollen, von Organisation zu Organisation. Bspw. könnten die Karrierechancen der Beschäftigten in einem Betrieb durch innerbetriebliche Aushandlungsprozesse determiniert sein, in einem anderen jedoch von Regelungen, die überbetrieblich gestaltet werden.

In der sozialwissenschaftlichen Forschung wird üblicherweise so vorgegangen, Betriebe zu befragen, welche als lokale Einheiten bzw. Arbeitsstätten definiert werden. Zusätzlich kann durch Einschluss entsprechender Fragen erhoben werden, ob es sich bei der befragten Einheit um ein Mutter- oder Tochterunternehmen bzw. eine Zentrale oder eine Zweigstelle handelt, um derartige Informationen ex post statistisch berücksichtigen zu können. Die Erfassung der Einheit Betrieb erscheint auch für zu verknüpfende Organisationserhebungen angemessen, da die örtliche und soziale Eingebundenheit der Individuen sich zunächst einmal innerhalb von Betrieben und nicht Unternehmen vollzieht. So wird implizit angenommen, dass diese Ebene theoretisch relevanter und von größerer Strukturierungsbedeutung ist. Eine theoretisch umfangreiche Betrachtung dieses Problems steht bislang jedoch aus.

3 Vorgehen und Fragestellung

Hinsichtlich der Nutzung der employee-first-Methode zur Generierung eines verknüpften Personen- und Betriebsdatensatzes sind kaum Informationen verfügbar, da es nach Kenntnisstand der Autoren bislang keine Studie in Deutschland gibt, die entsprechend vorgeht. Das empirische Design eines solchen Vorgehens sähe wie folgt aus:

- (1) Einschaltung der Frage nach der Betriebskontaktadresse in ein allgemeines Bevölkerungsumfrage-Instrument
- (2) Durchführung einer Betriebsstättenbefragung auf der Grundlage der bereitgestellten Kontaktdaten
- (3) Verknüpfung der Individualdaten der Bevölkerungsumfrage mit den Betriebsdaten aus der Betriebsstättenbefragung

3.1 Fragestellungen

Die diesem Forschungsdesign immanente, zentrale Ungewissheit bezieht sich auf den Aspekt der Selektivität des entstehenden, verknüpften Datensatzes und damit der Generalisierbarkeit der Ergebnisse. Auf Basis vorhandenen Wissens ist nicht einzuschätzen, wie selektiv die realisierten vollständigen Personen-Betriebsinformationen sind, und ob bzw. wie Korrekturen der Stichprobenverzerrung (und auf welcher Ebene des Forschungsprozesses) möglich sind. Durch die spezielle Art der Generierung der Betriebsinformationen – über Personenangaben –

⁶ In der Schulforschung wird dies seit längerem berücksichtigt, indem school effects von class effects getrennt beobachtet werden.(z.B. Cohen 2000, Lee 2000). Für einen Versuch der theoretischen Integration der unterschiedlichen Ebenen, siehe z.B. Gamoran et al 2000) Auch im Rahmen der Arbeitsmarktforschung sind Effekte sowohl der Betriebsebene als auch der Unternehmensebene relevant.

liegt für einen verknüpften Datensatz keine Zufallsauswahl auf Betriebsebene vor, und die Vorteile der Zufallsauswahl auf Individualebene bleiben nur dann erhalten, wenn sämtliche Individuen den Betriebskontakt angeben, die Angaben vollständig für eine Kontaktierung brauchbar sind und sämtliche der kontaktierten Betriebe an der folgenden Betriebsbefragung teilnehmen. Es ergibt sich ein zweistufiges Ausfallproblem aufgrund von Ausfällen bei der Nennung der Betriebskontakte durch den Beschäftigten und Ausfällen bei der Teilnahme der Betriebe bei der Betriebsstättenbefragung. Unsere erste Fragestellung bezieht sich somit auf die Beobachtung und Beurteilung der Ausfallquoten auf den unterschiedlichen Schritten des Generierungsprozesses.

Wird die Selektivität, die ein solches Sample vorweist, untersucht, sind Aussagen generierbar über die Forschungsrelevanz bzw. die Brauchbarkeit eines derart erstellten LEE-Datensatzes. Hinweise auf systematische Verzerrungen können den Vergleichen der Verteilungen verschiedener, relevanter Individualmerkmale der jeweiligen Personenstichproben der unterschiedlichen Selektionsstufen entnommen werden (Population der Befragten insgesamt, Population der nach ihrer Betriebszugehörigkeit befragten Personen, Population der die Betriebszugehörigkeitsabfrage beantwortenden Personen und Population der Personen, für deren Betriebe Daten aus der Betriebsbefragung vorliegen). Um die Qualität eines aus Personenangaben generierten Betriebsdatensatzes zu bestimmen, für den Fall, dass die Analyseeinheiten die Betriebe darstellen sollen, ist es sinnvoll, die potentiell realisierbare Stichprobe sowie den sich nach der Betriebsbefragung ergebenden Datensatz (verbleibende Stichprobe nach Non-response) mit einem (externen) Kriteriumsdatensatz zu vergleichen. Die Funktion eines Kriteriumsdatensatzes kann beispielsweise das IAB-Betriebspanel übernehmen. Anhand einer Gegenüberstellung der Verteilungen relevanter Betriebsmerkmale der unterschiedlichen Datensätze können Aussagen über Repräsentativität oder systematische Verzerrungen gemacht werden. Sind systematische Verzerrungen zu beobachten, können wiederum Rückschlüsse über Art und Richtung der Verzerrung der gefundenen Effekte in der Stichprobe gezogen werden. Diese Information kann in die Ergebnisinterpretationen der durchgeführten Analysen einfließen und so die Brauchbarkeit des Datensatzes trotz evtl. Verzerrungen erhalten.

Neben der Berücksichtigung vorhandener Selektivität gilt es, den Prozess der Entstehung von Selektivität, d.h. die hier wirksamen Mechanismen, näher zu betrachten. Hierzu werden zunächst die Ursachen für Ausfälle auf Personenebene durch eine Analyse des Nonresponseverhaltens der befragten Individuen untersucht. Die Analyse von Determinanten der Nennung der Betriebszugehörigkeit bzw. Non-Response kann sowohl methodisch als auch inhaltlich interessant sein. Ersteres unterstützt Verbesserungsmöglichkeiten zukünftiger Datenerstellung (z.B. Intervieweranweisungen, Platzierung im Fragebogen). Inhaltlich interessant ist es zudem, nach systematischen Zusammenhängen von Non-Response-Determinanten mit Betriebsmerkmalen zu suchen. Potentiellen Einfluss auf die Auskunftsverweigerung können sowohl organisationale Merkmale (z.B. Organisationsgröße, öffentlicher Dienst) als auch individuelle Merkmale (Soziodemographie, Merkmale bzgl. Erwerbstätigkeit wie z.B. Arbeitszufriedenheit), Merkmale des spezifischen Antwortverhaltens (z.B. Verweigerung der Frage nach dem Einkommen) oder Merkmale, die in Zusammenhang mit der speziellen Erhebungsmethode stehen, aufweisen.

Weitere Hinweise auf Selektivität und Reduktionsmöglichkeiten dieser ergeben sich durch die genauere Betrachtung des Prozesses der Kontrolle und Korrektur der Betriebskontaktangaben im Rahmen der Datenedition. Hierdurch lassen sich Rückschlüsse auf die Auskunftsbereitschaft und die -fähigkeiten der Befragten ebenso wie über den (personellen und finanziellen) Aufwand der Forscher ziehen. Die Dokumentation des Vorgehens der Edition, sowie des Aufwands der Edition an sich gibt Hinweise auf zukünftige Verbesserungsmöglichkeiten.

Diese sind zum Beispiel zu erwarten hinsichtlich veränderter Interviewerinstruktionen, geänderter Erläuterungstexte oder Frageformulierungen.

Das zentrale Problem der Selektivität des fertigen Datensatzes führt somit zu drei näher zu betrachtenden Fragen:

- (1) Wie ist die Selektivität des Datensatzes bzw. wie hoch sind die Ausschöpfungsquoten auf den unterschiedlichen Schritten des Datenerhebungsprozesses?
- (2) Welches sind die hauptsächlichen Determinanten der Nennung der Betriebszugehörigkeit auf Individualebene?
- (3) Wie ist die Güte der Betriebskontaktangaben gemessen am Aufwand der Edition des Datensatzes der Betriebskontakte einzuschätzen?

Die Beantwortung soll Rückschlüsse auf die Brauchbarkeit des Datensatzes und zukünftige Vorgehensweisen der Erstellung ermöglichen.

3.2 Vorgehen

Im Rahmen der SOEP-Innovationsbefragung „Persönlichkeit und Gesundheit“ (vgl. Siegel/Stocker 2007) wurden im ersten Quartal 2007 erwerbstätige Personen nach ihrer Betriebszugehörigkeit gefragt und so der erste Teil des oben vorgestellten Forschungsdesigns realisiert. Auf Basis dieser Betriebsadressen wurden im 1. Quartal 2008 Betriebe kontaktiert, um sie zu einer Online-Befragung einzuladen, die auf Wunsch auch postalisch durchgeführt werden konnte.

Zunächst wurden die Personen um den Namen des Betriebs, in dem die Befragungsperson aktuell oder zuletzt beschäftigt war, gebeten. „*Wie heißt die Firma, die Behörde, bei der Sie derzeit beschäftigt sind (zuletzt beschäftigt waren)*“. Diese Frage wurde offen vercodet. Als Nächstes wurde die Adresse der Betriebsstätte mit Angabe der Straße, Postleitzahl und Ort offen abgefragt: „*Wo befindet sich die Arbeitsstelle?*“. Beide Fragen richteten sich an aktuell oder bis vor kurzem beschäftigte Berufstätige ohne Selbständige. Im SOEP-Pretest 2007 ergab sich so eine Stichprobengröße von 370 Betriebskontaktdaten bei 419 Befragten, die nach dem Namen des Betriebes gefragt wurden aus 1023 Befragten der Stichprobe insgesamt. $N=370$ bildet somit die Zielgröße bzw. das Ausgangsbrutto für die nachgelagerte Betriebsbefragung, die es zu realisieren gilt.

Möglicherweise wird die Bereitschaft zur Nennung der Arbeitgeberinformationen dadurch geschwächt, dass bei den Befragten Zweifel bestehen hinsichtlich der zugesicherten Wahrung der Anonymität und daher auch negative Konsequenzen am Arbeitsplatz durch eine mögliche Rückkoppelung ihrer Daten befürchten. Um diesem Problem bereits während des Interviews entgegenzusteuern, wurde bei Verweigerung vom Interviewer ein entsprechender Text (siehe Anhang 1) vorgelesen, der das Ziel der Abfrage erläutert und Anonymität zusichert. Daran anschließend wurde der Name noch einmal erfragt. Zusätzlich wurde dieses Vorgehen gesplittet und bei einem Teil der Befragten der Erläuterungstext bereits vor Fragestellung vorgelesen, so dass der Einfluss der Texteinblendung auf das Antwortverhalten kontrollierbar ist.

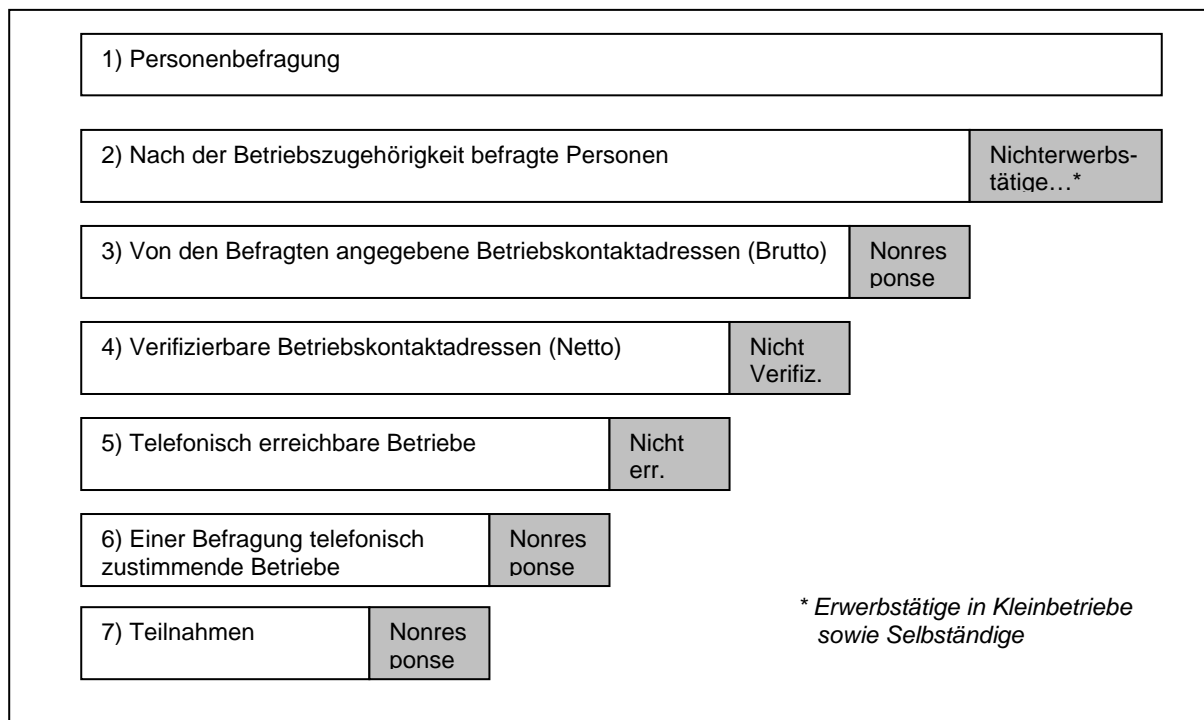
Der Abfrage der Betriebskontakte vorangestellt wurden Fragen nach öffentlicher/nicht öffentlicher Dienst, Branchenzugehörigkeit und Mitarbeiterzahl. Diese Informationen können dazu genutzt werden, Gründe für die Antwortbereitschaft der Befragten ebenso wie für die Teilnahmebereitschaft der Betriebe in der späteren Betriebserhebung zu identifizieren.

Nach Abfrage der Betriebskontaktadressen wurden diese überprüft und aufbereitet. Der Prozess der Datenedition wurde ausführlich dokumentiert und wird im folgenden dargestellt.

Für die Erhebung der Betriebsinformationen zur späteren Verknüpfung wurde ein zweistufiges Vorgehen gewählt. Dieses bestand aus einem telefonischen Vorgespräch und einer Online- (auf Wunsch postalischen) Befragung. Zunächst wurde telefonisch eine Kontaktperson des Betriebs ausfindig gemacht, die um Zustimmung zur Teilnahme und Herausgabe der E-mailadresse gebeten wurde. Eine Ansprechperson im Betrieb telefonisch ausfindig zu machen, birgt den Vorteil, dass die Kontrolle darüber, wer der Informant für den Betrieb ist, so (eher) beim Forscher bzw. Interviewer verbleibt. Die eröffnete Option der Zusendung eines schriftlichen Fragebogens sollte dem Erhebungsmodus geschuldete Fallzahlenverluste reduzieren.

Auf der zweiten Stufe der Betriebsstättenbefragung wurden die Betriebe per Mail (bzw. postalisch) zur Teilnahme eingeladen. Die Einladung erfolgte zeitnah nach dem telefonischen Vorgespräch in der Regel am gleichen Tag. Nach Ablauf von ca. 14 Tagen wurde eine erste Erinnerung versandt; nach Ablauf weiterer 14 Tage folgte die zweite. Der postalischen Kontaktierung wurde bei der zweiten Erinnerung ein weiterer Fragebogen mit adressiertem und frankiertem Rückumschlag beigelegt. Der Fragebogen beinhaltete 80 Fragen⁷ (20 Seiten), die in ca. einer halben Stunde beantwortet werden konnten und befasste sich mit einem weit gefassten Themenspektrum unter dem Titel „Betriebliches Personalmanagement“. Die Fragen bezogen sich insb. auf die Basismerkmale des Betriebs (Gründungszeitpunkt, Rechtsform, Größe u.a.) und dessen Personalstruktur und -politiken.

Abb.1: Stufen des Generierungsprozesses eines verknüpften Personen-Betriebsdatensatzes auf Basis einer personengenerierten Betriebsstichprobe



Das komplette Forschungsdesign aufgegliedert in die einzelnen Schritte des Forschungsprozesses, ist in Abbildung 1 exemplarisch dargestellt. Wie deutlich wird, ist auf jeder Stufe des

⁷ Vgl. Anhang 5.

Forschungsprozesses mit potenziellen Ausfällen von zunächst unbestimmter Größenordnung (grau unterlegt) zu rechnen, denen mit entsprechenden Schritten forschenseitig begegnet werden muss.

4 Ergebnisse

4.1 Selektivität der Stichproben: Ausschöpfungsquoten

Tabelle 2 veranschaulicht die auf jeder Stufe realisierten Fallzahlen und die jeweiligen Ausfälle entsprechend der einzelnen Phasen des Generierungsprozesses innerhalb dieser Studie. Ausgehend von der maximal erreichbaren Fallzahl der Betriebsstichprobe, die sich ergibt durch die Anzahl befragter Personen der Personenbefragung, folgen fünf weitere Schritte bis zur Generierung des Betriebsdatensatzes. Wie zu sehen ist, fand in der vorliegenden Studie auf jedem dieser Schritte ein Fallzahlenverlust statt. Stichprobe B ergibt sich nach Abzug des individuellen Nonresponse, d.h. dem Ausfall auf Personenebene. 12% der insgesamt möglichen Fallzahl ist der Antwortverweigerung auf der Individualebene geschuldet. Ob es sich hierbei um neutrale Ausfälle handelt oder systematische, wird im folgenden (in Punkt 4.2.) geprüft werden. Stichprobe C ergibt sich nach Abzug der in externen Datenbanken nicht auffindbaren Betriebsadressen in Höhe von weiteren acht Prozentpunkten (von 88 auf 80%). Die Anzahl telefonisch erreichbarer Betriebe ist durch Stichprobe D erfasst und beläuft sich auf 67%. Zu den Ausfällen dieses Schrittes zählen Betriebe, deren Telefonnummer nicht auffindbar oder fehlerhaft war, oder die telefonisch bis zum Ende der Feldphase (nach mehrmaligen Kontaktversuchen) nicht erreicht werden konnten. Diejenigen Betriebe, die nach telefoni-schem Kontakt einer Befragung zugestimmt haben, bilden Stichprobe E. Stichprobe F schließlich ergibt sich nach Abzug des betrieblichen Nonresponse, der nach Zusendung des Fragebogens (online bzw. postalisch) entsteht.

Tabelle 2: Ausschöpfungsquoten unterschiedlicher Stichproben

| | Stichprobe A (Befragte Personen) | Stichprobe B (Angabe der Personen) | Stichprobe C (auffindbar in Datenbanken) | Stichprobe D (tel. erreichbar) | Stichprobe E (Zustimmung zur Teilnahme) | Stichprobe F (Rücksendung Fragebogen) |
|-------|--|--|--|-----------------------------------|---|---|
| BASIS | 370 | 325 | 298 | 247 | 161 (125 online, 36 postalisch) | 29 (9 postalisch, 20 online) |
| 1023: | 36% | | | | | |
| 370: | 100% | 88% | 81% | 67% | 44% | 8% |
| 325: | | 100% | 92% | | | |
| 298: | | | 100% (=286)* | 86% | 56% | 10% |
| 247: | | | | 100% | 65% | 12% |
| 161: | | | | | 100% | 18% |

*298 abzüglich 12 Fälle, die im Datensatz mehrfach vertreten waren.

Quelle: SOEP Innovationsbefragung 2007.

Die jeweiligen Phasen der Stichprobengenerierung (A-F) unterscheiden sich hinsichtlich ihrer Fallzahlenrisiken und -potentiale. Die einzelnen Schritte werden im Folgenden näher erläutert und im Hinblick auf die Verbesserungsmöglichkeiten der jeweils realisierten Ausschöpfung diskutiert werden.

4.2 Nonresponse-Determinanten auf Personenebene

Die Verweigerungen der Angabe der Betriebszugehörigkeit beziehen sich auf die Ausfallproblematik zwischen Stichprobe A und B. Hier kam es zu Ausfällen in Höhe von 12%. Um die erklärenden Faktoren dieser Ausfälle auf Personenebene zu identifizieren, wurden zunächst bivariante und multivariate Analysen berechnet. Die zu erklärende Variable bildet die

binär-codierte Nicht-Angabe des Namens des beschäftigenden Betriebs. 327 Antwortenden (=0) stehen 43 Antwortverweigerer (=1) gegenüber.

In den bivariaten Analysen zeigten sich signifikante Zusammenhänge der Antwortverweigerung mit Merkmalen der Interviewsituation, des Antwortverhaltens, arbeitsbezogenen Einstellungen und betriebspezifischen Merkmalen; keine hingegen mit soziodemographischen Merkmalen. Als Merkmal der Interviewsituation wurde die Variable Texteinblendung spezifiziert, die den Einfluss des vorab vorgelegten Erläuterungstexts kontrolliert (siehe Anhang 1). Als Merkmal des Antwortverhaltens gelten die Verweigerung des Einkommens sowie die Bereitschaft in Zukunft an weiteren Befragungen teilzunehmen. Personen, denen ein Erläuterungstext vorab vorgelesen wurde ebenso wie solche, die bereits die Angabe des Einkommens verweigerten, verweigerten signifikant häufiger die Angabe ihrer Betriebszugehörigkeit.⁸ Die zukünftige Teilnahmebereitschaft von Antwortverweigerern ist niedriger ausgeprägt als bei Antwortenden. In den bivariaten Analysen mit den betriebspezifischen Merkmalen Betriebsgröße sowie öff. Dienst zeigten sich ebenfalls signifikante Zusammenhänge: Beschäftigte in größeren Betrieben und Beschäftigte im Öffentlichen Dienst verweigern weniger als ihre Kollegen in kleineren Betrieben anderer Branchen.

Tab.3: Merkmale der Interviewsituation und des Antwortverhaltens

| Variablen | Maß | P-Wert | Anmerkungen |
|----------------------------------|------|--------|--|
| Texteinblendung*** (=1) | Chi2 | 0.000 | Verweigerer befinden sich häufiger in der Gruppe, der ein Erläuterungstext vorgelesen wurde. |
| Verweigerer Einkommensangabe*** | Chi2 | 0.000 | Verweigerer, verweigern eher die Angabe des Einkommens. |
| Erneute Teilnahmebereitschaft*** | Chi2 | 0.000 | Unter den Verweigerern sind mehr Befragte mit zukünftiger negativer Teilnahmebereitschaft. |

*p<.05; **p<.01; ***p<.001

Tab.4: Individualmerkmale (soziodemographisch und arbeitsplatzbezogen)

| Variablen | Maß | p-Wert /Sign. | Anmerkungen |
|---|--------------|---------------|---|
| Bildung (Casmin) | t-test, chi2 | n.s. | Keine Richtung erkennbar |
| Berufliche Stellung | | n.s. | Keine Richtung erkennbar |
| Geschlecht | Chi2 | n.s. | Keine Richtung erkennbar |
| Alter (metrisch) | t-test | n.s. | Keine Richtung erkennbar |
| Wohnort vor 89 (Ost/West/Ausl.) | Chi2 | n.s. | Verweigerer häufiger mit Wohnort vor 89 in Westdeutschland. |
| Wohnort aktuell (Ost/West) | Chi2 | n.s. | Keine Richtung erkennbar |
| Einkommen (metrisch und kategorial) | t-test, chi2 | n.s. | Das Einkommen der Verweigerer ist tendenziell niedriger. |
| Arbeitszufriedenheit (metrisch und kategorial) | t-test, chi2 | n.s. | Verweigerer sind eher unzufrieden. |
| Einkommenszufriedenheit (metrisch, kategorial) | t-test, chi2 | n.s. | Verweigerer sind eher unzufrieden. |
| Sorgen um den Arbeitsplatz | Chi2 | n.s. | Keine Richtung erkennbar |
| Betriebszugehörigkeitsdauer (metrisch und kategorial) | t-test, chi2 | n.s. | Verweigerer haben eher eine kürzere Betriebszugehörigkeit. |

*p<.05; **p<.01; ***p<.001

Auf Basis dieser Ergebnisse wurde ein multivariates logistisches Regressionsmodell spezifiziert. Wie in Tabelle 6 zu erkennen, wird die Verweigerung der Angabe des Betriebskontakts bereits zu einem hohen Grad durch die Variablen Texteinblendung, Verweigerung des Ein-

⁸ Es gab auch einen (pos.) Zusammenhang zwischen der Verweigerung der Angabe der Schulzugehörigkeit der eigenen Kinder und der Angabe der Betriebszugehörigkeit, die aber aufgrund sehr geringer Fallzahl hier nicht explizit interpretiert werden.

kommens und des betriebsspezifischen Merkmals Betriebsgröße erklärt.⁹ Der zunächst signifikante Effekt öffentlicher Dienst ($p=0.018$) wird insignifikant sobald die Betriebsgröße kontrolliert wird. Das Hinzufügen soziodemographischer Merkmale (Bildung, Erwerbsstatus, Alter, Einkommenshöhe) führte zu keiner Modellverbesserung, was den Ergebnissen der bivariaten Analysen entspricht. Die Richtung des Zusammenhangs zwischen Einkommenshöhe und Verweigerung ist erwartungsgemäß in der Art, dass mit sinkendem Einkommen die Verweigerung zunimmt.

Tabelle 5: Betriebliche Merkmale

| Variablen | Maß | P-Wert | Anmerkungen |
|--|--------------|--------|--|
| Betriebsgröße (Anzahl Beschäftigte) ^{***} (metrisch und kategorial) | t-test, Chi2 | 0.000 | <i>Verweigerer sind eher in kleineren Betrieben beschäftigt.</i> |
| Öff. Dienst (=1) [*] | Chi2 | 0.037 | <i>Verweigerer sind eher nicht im öff. Dienst beschäftigt.</i> |

* $p<.05$; ** $p<.01$; *** $p<.001$

Hinsichtlich weiterer Merkmalszusammenhänge sind keine eindeutigen Richtungen zu entdecken. Die Berücksichtigung der arbeitsspezifischen Variablen Arbeitszufriedenheit und Sorgen um den Arbeitsplatz, ergab bi- wie multivariat keine signifikanten Zusammenhänge, wengleich mit sinkender Arbeitszufriedenheit die Antwortverweigerung steigt. Der Zusammenhang zwischen Sorgen um den Arbeitsplatz und der Antwortverweigerung weist hingegen in keine eindeutige Richtung.

Tabelle 6: Logistische Regression

| | Verweigerung Betriebskontakt (=1) | |
|------------------------------------|-----------------------------------|--------------------------|
| | Odds Ratios | Coef |
| Betriebsgröße | .6783544 ^{***} | -.3880854 ^{***} |
| Verweigerung Einkommensangabe (=1) | 3.66059 ^{***} | 1.297597 ^{***} |
| Texteinblendung (=1) | 16.40478 ^{***} | 2.797572 ^{***} |
| Konstante | | -3.367794 ^{***} |
| Pseudo R ² | | .2248 |

Anmerkungen: N=370; Signifikanzniveau: * $p<0.05$, ** $p<0.01$, *** $p<0.001$
Quelle: SOEP Innovationsbefragung 2007.

Da mit soziodemographischen Merkmalen keine Zusammenhänge gefunden wurden, besteht Anlass anzunehmen, dass die Ausfälle dieser Stufe zufällig und nicht systematisch sind. Problematisch bleibt der Zusammenhang mit der Betriebsgröße. Da bekannt ist, dass Betriebsbefragungen regelmäßig insbesondere kleinere Betriebe untererfassen, wird hier ein bereits existierendes Ausfallproblem gedoppelt. Der besonders große Einfluss, den ein vorab eingeblendeter Erläuterungstext auf das Antwortverhalten hat, wird dann problematisch, wenn dieser eine datenschutzrechtliche Verpflichtung darstellen sollte, was im Einzelfall zu prüfen ist.

Eine Möglichkeit die Ausschöpfungsquote an dieser Stelle zu erhöhen, besteht darin, Zusatzinformationen der Verweigerer aus der Individualbefragung zu nutzen. In der NOS 2002 wurde so vorgegangen, dass die fehlenden Werte der Betriebskontaktadressen imputiert wurden. Die Imputation erfolgte anhand der Branchen- und geographischen Angaben, die aus der Individualbefragung (GSS 2002) bekannt waren.

⁹ Betriebsgröße wurde 7-kategorial erfasst mit den Kategorien 5-10 Beschäftigte, 11-unter 20 Besch., 20-unter 100 Besch., 100-unter 200 Besch., 200-unter 2000 Besch., 2000 Besch. und mehr.

4.3 Generierung der Betriebsstichprobe

4.3.1 Ausschöpfungsquoten

In der SOEP-Innovationsbefragung wurden von den 1023 Teilnehmenden bzw. den 419 Befragten, die nach dem Betrieb gefragt wurden insgesamt bei 370 nach Namen und Adresse des sie beschäftigenden Betriebes gefragt, da Selbständige, Nicht-Erwerbstätige und Beschäftigte in Betrieben mit weniger als 5 Beschäftigten aus der Stichprobe herausfielen.¹⁰ Bereinigt um vier nicht sinnvolle Angaben¹¹ ergab die Frage nach dem Namen des Arbeitgebers 325 Nennungen. Dies entspricht einer Ausschöpfungsquote von 88%. 253 der Befragten nannten in der vorliegenden Erhebung Name, PLZ, Ort und Straße (=68%).(siehe Tabelle 7) Ähnliche Ergebnisse erzielte die NOS von 2002, in der 68,7% der 888 Befragten die Adresse und 81,8% der Befragten den Namen ihres Arbeitgebers nannten.¹²

Tabelle 7: Ausschöpfungsquoten

| Personen-Stichprobe | Stichprobe A | Stichprobe B | | |
|---------------------|---|--------------|--------------------|-------------------------|
| Teilnehmende | BZG-Befragte | Angabe | Angabe | Angabe |
| | excl. Nichterwerbstätige, Selbständige und Beschäftigte in Betrieben <5Besch. | Name | Name und mind. Ort | Name, PLZ, Ort, Strasse |
| 1023 | 370 | 325 | 312 | 253 |
| | 100% | 88% | 84% | 68% |

Tabelle 8: Ausschöpfungsquoten einzelner Adresskomponenten

| Angabe: | Anzahl getätigter Angaben von maximal 370 |
|------------|---|
| Name | 325 = 88% |
| PLZ | 293 = 79% |
| Ort | 335 = 91% |
| Strasse | 278 = 75% |
| Hausnummer | 139 = 38% |

Wie Tabelle 8 zu entnehmen ist, variieren die Ausschöpfungsquoten hinsichtlich der einzelnen Adresskomponenten. Der Ort, als eine für die Identifizierung wichtige Adresskomponente, wurde in 91% aller Fälle angegeben. Die Adresskomponenten PLZ und Straße wurden von rund drei Viertel der Befragten genannt. Die Hausnummer wird in vielen Fällen nicht angegeben. Ob einzelne und wenn ja welche Adresskomponenten ausreichen zur eindeutigen Identifizierung eines Betriebs, ist für die deutsche Adressstruktur nicht einheitlich zu beantworten. Je nach Betrieb (und insb. Betriebsgröße) und Standort (Klein- vs. Großstadt) ist bzw. sind entweder der Name, Name und Ort, Name, Ort und Straße oder Name, Ort, Straße und Haus-

¹⁰ Inklusive Selbständiger und Beschäftigter in Betrieben mit weniger als 5 Beschäftigten ergäbe sich eine Stichprobenzahl von N=419. In der US-Amerikanischen National Organizations Study wurden die Selbständigen in die Betriebsstichprobe integriert und im Rahmen der Betriebsbefragung erneut befragt.

¹¹ Hierzu zählten die Angaben der Betriebsnamen „Gaststätte“, „eine Baufirma“, „Hotel“, „Gebäudereinigung“.

¹² Zu einem deutlich schlechteren Ergebnis kommt eine Studie auf der Grundlage des British Household Panels (BHPS-ECHP) aus dem Jahr 2001 (Improving survey measurement of income and employment [ISMIE] Survey) (Jaeckle et al. 2005; Jenkins et al. 2004). Lediglich 58,5 Prozent der Erwerbstätigen nannten den Namen, die Adresse und die Telefonnummer ihres Arbeitgebers. Ein Grund für die relativ geringe Auskunftsbereitschaft liegt aus unserer Sicht jedoch darin, dass als Grund für die Kontaktierung des Arbeitgebers eine Überprüfung (Kontrolle) der von den Befragten gemachten Einkommensangaben angeführt wurde. Es ist deshalb auch nicht verwunderlich, dass die Auskunftsverweigerung bei denjenigen, die keine Antworten zum Einkommen geben wollten, am höchsten ist.

nummer hinreichend bzw. notwendig, um diesen Betrieb eindeutig identifizieren zu können und die Kontaktierbarkeit (per Telefon oder Post) zu gewährleisten.¹³

Um die Ausfälle auf dieser Stufe in Höhe von 12% zu reduzieren, kann an mehreren Stellen angesetzt werden: Erstens besteht Spielraum, die Gruppe der befragten Personen auszuweiten. Im National Organizations Survey (NOS) 2002 beispielsweise wurden die Selbständigen nicht aus der Stichprobe der Betriebe für die spätere Betriebsbefragung ausgeschlossen. Man könnte sich ein Design vorstellen, bei welchem die Befragung der Selbständigen direkt im Rahmen der Personenbefragung durchgeführt wird. Zweitens kann bei der Quantität der Nennungen bzw. der Verweigerungen dieser angesetzt werden und drittens, bei der Qualität der Nennungen gemessen an der Kontaktierbarkeit der Betriebe.

4.3.2 *Edition der Adressdatei*

Vor Kontaktierung der Betriebe wurden zunächst die individuellen Betriebskontaktangaben editiert, um Aussagen über Quantität und Qualität der gemachten Angaben zu treffen und die Möglichkeiten nachträglicher Erhöhung beider Kriterien zu evaluieren. Der Prozess der Datenedition unterscheidet sich hier definitionsgemäß von der üblichen Datenbereinigung. Die hier vorgenommene Editierung bezeichnet die Überprüfung der Betriebskontaktangaben der Befragten anhand externer Quellen, ihre eventuelle Korrektur und die Vervollständigung fehlender Adresskomponenten. Die 325 Beobachtungen in Stichprobe B bilden den Ausgangspunkt der vorzunehmenden Editionen. Die angegebenen Namen und Adressen wurden anhand von Informationen aus externen Datenbanken (LexisNexis¹⁴, Internetsuchmaschine sowie Telefonbuch) korrigiert, falls sie fehlerhaft waren, und sie wurden ergänzt, falls einzelne Adresskomponenten nicht vorhanden waren. Ziel der Edition und gleichzeitiges Kriterium der Güte der Betriebskontaktangaben bildete die eindeutige Identifizierbarkeit des beschäftigten Betriebs in den externen Quellen. Die Anzahl derjenigen Kontaktangaben, die zu korrigieren waren, wurde festgehalten.¹⁵ Ebenso wurde die Anzahl derjenigen Angaben, die in den externen Datenquellen neu gefunden und ergänzt wurden, dokumentiert. Ergänzungen ebenso wie Korrekturen wurden nur vorgenommen, wenn es der Logik der sonstigen gemachten Angaben entsprach.¹⁶ Trotz angestrebter Objektivierung des Prozesses der Datenedition durch Offenlegung des Vorgehens und Spezifizierung von Kriterien bleiben die Vorgänge der Korrektur fehlerhafter und der Ergänzung unvollständiger Adressangaben zu einem gewissen Ausmaß subjektiv, d.h. von der die Edition ausführenden Person abhängig. Einzelfallentscheidungen sind (teilweise) notwendig.

Die Benutzung externer Datenquellen (LexisNexis, Internet und online-Telefonbuch) zur Editierung impliziert die Annahme, der Wahrheitsgehalt (und die Aktualität) dieser Quellen läge über denen der Respondentenangaben. Diese Annahme ist womöglich nicht in jedem Fall haltbar. Ein ideales Vorgehen wäre es, sämtliche der vorhandenen externen Quellen parallel

¹³ Im Fall eines fiktiven Metzger Cochinskis reicht möglicherweise in einer Kleinstadt der Ortsname aus, in einer Großstadt hingegen evtl. nicht. Die Nennung der Deutschen Bank AG als Arbeitgeber ohne Angabe der Straße (und evtl. sogar der Hausnummer) wird jedoch selten (v.a. nicht in größeren Städten) ein hinreichendes Kriterium sein, da diese vor Ort i.d.R. mehrfach vertreten ist.

¹⁴ LexisNexis ist ein kommerzieller Datenanbieter, der Informationen über Unternehmen bereitstellt. U.a. greift LexisNexis auf Hoppenstedt und Creditreform zu.

¹⁵ Korrekturen, die über rein orthographische Verbesserungen nicht hinausgingen, wurden dabei nicht berücksichtigt, da diese dem üblichen Prozess der Datenbereinigung und nicht der Spezifität personengenerierter Betriebsadressen zuzurechnen sind.

¹⁶ In einem Fall bspw. waren Ort und PLZ angegeben aber keine Straße. In diesem Ort/PLZ war der Betrieb nicht zu finden, allerdings aber in einem anderen Ort mit einem Straßennamen, der dem ursprünglich angegebenen Ortsnamen glich. Da hier von bewusster Falschangabe auszugehen ist, wurde von der Korrektur des Ortes und der Ergänzung des Straßennamens Abstand genommen.

zur Datenedition zu nutzen. Dieses erhöht den Aufwand der Edition in nicht unerheblichem Maße, verbessert aber die Qualität des Editionsprozesses. Dann ist allerdings eine Regelung zu treffen, wie bei divergierenden Informationen aus unterschiedlichen Quellen zu verfahren ist.¹⁷

Von den 325 sinnvollen Betriebskontaktangaben (Name und/oder Adresse) konnten 298 Beobachtungen anhand externer Quellen verifiziert werden (siehe Tabelle 9). 27 bzw. 8% der 325 Angaben waren weder in LexisNexis, Internet noch Telefonbuch vorfindbar, so dass in diesen Fällen anzunehmen ist, dass entweder fehlerhafte Angaben gemacht wurden oder der entsprechende Betrieb nicht mehr existiert. Da diese Angaben nichtsdestotrotz gemacht wurden und davon auszugehen ist, dass (zumindest in einigen Fällen) nicht die Bereitschaft, sondern ausschließlich die Fähigkeit mangelte, korrekte und vollständige Angaben zu machen, besteht hier Potential die Ausschöpfungsquote zu erhöhen.¹⁸ Durch (individuell) exaktes Nachfragen und Hilfestellungen (z.B. Nachschlagen in externen Unterlagen ermöglichen) des Interviewers in der Interviewsituation könnten die durch fehlende Fähigkeit verursachten Ausfälle reduziert werden. Zudem wäre zu versuchen, diese Betriebe dennoch anzuschreiben, wie es die Forscher des National Organizations Survey 2002 gemacht haben, mit dem Argument, „postal workers can sometimes deliver packages despite errors identified by smart Mailer or interviewing clerks“ (Smith et al 2004). Liegt die Ursache in der Zwischenzeit vom Markt verschwundenen Betrieben, kann dieses Problem forschenseitig durch eine zeitnah an die Personenbefragung folgende konzipierte Betriebserhebung minimiert werden.

Tabelle 9: Anzahl der Beobachtungen, die in externen Quellen verifiziert werden konnten

| Gesamtbestand | Datenbestand gefunden (db = 1) | Datenbestand nicht gefunden (db = 0) |
|---------------|--------------------------------|--------------------------------------|
| 325 | 298 (92%) | 27 (8%) |

Firmendatenbank

Für die Edition wurde zunächst eine Firmendatenbank genutzt. Von 298 Betrieben wurden für 72 Betriebe (24%) Informationen in LexisNexis gefunden, die sich allerdings regelmäßig auf die rechtliche Einheit Unternehmen und nicht auf die erfragte lokale Einheit Betrieb beziehen. Darüberhinaus sind die zusätzlichen Angaben wie z.B. Finanzkennzahlen, Gründungszeitpunkt, Rechtsform etc., die in Firmendatenbanken bereitgestellt werden, nicht für alle Fälle und in sehr unterschiedlichem Ausmaß vorhanden.¹⁹ Insgesamt ist die Nutzung der Wirtschaftsdatenbank (hier: LexisNexis) als externe Quelle für den Zweck der Erhebung von Betriebsinformationen (über prozessgenerierte Daten) ebenso wie zur reinen Adressvalidierung scheinbar nicht zu empfehlen.

Internet

Im weiteren Verlauf wurden die Recherchen der Betriebskontakte mittels Internet durchgeführt. Die Güte des Internets als Validierungskriterium ist kritisch zu betrachten. Besitzen die angegebenen Betriebe eigene websites, ist eine hohe Validierungsqualität zu vermuten. Zudem können hier weitere relevante – aber höchst unterschiedliche - Informationen neben der Adresse eingeholt werden, die zur späteren Kontaktaufnahme und Interviewdurchführung genutzt werden könnten. Existiert keine Firmenhomepage, wird eine Validierung über das

¹⁷ In einem Fall bspw. kam es vor, dass die Adresse vollständig und sinnvoll angegeben wurde, im Internet aber eine andere Adresse ermittelt worden war.

¹⁸ Beispielsweise wurde mehrmals die Deutsche Post als Arbeitgeber genannt mit Angabe des Ortes. Hier war die Betriebsstätte – da es sich bei diesem Unternehmen um ein sehr diversifiziertes handelt - jedoch ohne Angabe der Straße nicht identifizierbar.

¹⁹ z.B. Rechtsform: 155, Beschäftigtenzahl 2006: 33.

Internet problematisch. Unter Objektivitäts- und Validitätsaspekten betrachtet, sollten andere Internetquellen (z.B. lokale oder regionale Adresssammlungen) als die firmeneigene Homepage unseres Erachtens zur Edition der Adressangaben nur mit Vorsicht genutzt werden. Die Güte dieser Informationen ist generell schwierig hinsichtlich Autorenschaft, Aktualität und Wahrheitsgehalt zu bewerten. Die Internetrecherche erwies sich in anderer Hinsicht als vorteilhaft: Wiesen Namen oder Adresskomponenten Orthographiefehler (bspw. lautsprachlich korrekt, aber orthographisch falsch, Namens- und Zahlendreher) auf, aufgrunddessen der Betrieb bspw. im Telefonbuch nicht identifiziert werden konnte, ermöglichte eine Recherche über das Internet in vielen Fällen eine Korrektur.

Telefonbuch

Als weitere und letzte Quelle wurde in dieser Studie das Online-Telefonbuch der Deutschen Telekom genutzt. Dessen Validität dürfte bei Betrieben größer als bei Privatpersonen einzuschätzen sein, da diese ein ökonomisches Interesse daran haben, verzeichnet und korrekt verzeichnet zu sein. Die Nutzung ist zudem relativ einfach, zeitgünstig und kostenlos. Insgesamt ist aus diesen Erfahrungen für zukünftige Editionen unter dem Aspekt geringstmöglichen Aufwands die Empfehlung abzuleiten, zunächst das Telefonbuch zu nutzen und nur in einzelnen Fällen zusätzlich eine Internetsuchmaschine, bspw. um rein orthographische Fehler aufdecken und korrigieren zu können.

Tabelle 10 zeigt die nach Editionsmöglichkeit und Art der Edition aufgespaltenen Nennungen, sowie den Ausgangs- und Endpunkt der Edition der vorliegenden Angaben. Von 298 Namensnennungen im Ergebnisdatensatz waren acht zu korrigieren (und korrigierbar; siehe Spalte 3). Von 298 Ortsangaben des Ergebnisdatensatz waren zwei zu korrigieren, vier zu ergänzen und drei ergänzbar. Ein anderes Bild ergibt sich bei Betrachtung der Adresskomponenten PLZ, Straße und Hausnummer. Die von den Befragten gemachten Angaben diesbezüglich waren deutlich schlechter in der Qualität (siehe Spalte 3) und wesentlich geringer in der Anzahl (siehe Spalte 2). Die guten Möglichkeiten der Verbesserungen zeigen sich hier. Dadurch, dass in den meisten Fällen auf Name und Ort zurückgegriffen werden konnte, konnten nahezu alle der fehlenden Angaben ergänzt werden (siehe Spalte 6).

Tabelle 10: Häufigkeit der gemachten Angaben Anteile der Korrekturen. Häufigkeit der fehlenden Angaben und Anteile der Ergänzungen. (Datenperspektive)

| | Anzahl (1) | Anzahl (2) | Anzahl (3) | Anzahl (4) | Anzahl (5) | Anzahl Ist (6) | Anzahl Soll (7) |
|---------|---------------------|-------------------------------|-----------------------------|----------------------|---------------------|-------------------------------|-----------------|
| | Angabe vorh. | Bereits Korrekt und vorhanden | Fehlerhaft (und korrigiert) | Nicht ang. (unvoll.) | Neu gefunden | Nach Edition vorhanden | |
| Name | 298 (100%) | 290 | 8 | | | | 298 |
| PLZ | 261 (88%) | 240 | 21 | 38 | 36 | 297 (100%) | 298 |
| Ort | 294 (99%) | 292 | 2 | 4 (?) | 3 | 297 (100%) | 298 |
| Strasse | 263 (88%) | 234 | 29 | 34 | 35 | 298 (100%) | 298 |
| Hnr | 134 (45%) | 109 | 25 | 163 | 155 | 291 (98%) | 298 |

Quelle: SOEP Innovationsbefragung 2007.

Nach der Edition anhand der externen Quellen ergibt sich folgendes Bild (Spalte 6): Es verbleiben 291 komplette Adressen (Name, PLZ, Ort, Straße und Hausnummer) inclusive, 297 exclusive der Hausnummer. Letzteres entspricht einer Ausschöpfungsquote von 80% der befragten Personen (370) (wenn davon ausgegangen wird, dass Postsendungen auch ohne Haus-

nummer zustellbar sind), d.h. der maximal realisierbaren Zahl und 91% der erhaltenen Antworten (325), d.h. der maximal realisierbaren Zahl abzüglich Nonresponse.

Aus der Datenperspektive heraus betrachtet, weisen die Ergebnisse auf einen relativ geringen Aufwand an Korrekturen und Ergänzungen, die vorgenommen werden mussten, hin. Insgesamt kann auf den Schritt der Datenedition nicht verzichtet werden, da einerseits die Anzahl notwendiger Korrekturen und Ergänzungen nicht vernachlässigbar gering ist, und andererseits die Möglichkeiten dieser relativ gut sind. Daher erscheint es aus unserer Sicht sinnvoll und vom Aufwand her vertretbar, stets den gesamten Datensatz zu prüfen und nicht nur Ergänzungen unvollständiger Adressangaben vorzunehmen.

4.3.3 Durchführbarkeit und Befragungsbereitschaft

Die dargelegten Ergebnisse des Editionsprozesses geben nicht nur Hinweise auf den Aufwand und die Möglichkeiten der Edition, sondern darüber hinaus auch auf den Grad vorhandener Fähigkeiten und Bereitschaft der Befragten hinsichtlich der Nennung einzelner Adresskomponenten sowie der Adresse insgesamt. Informationen hierüber können dazu genutzt werden, Veränderungen im Erhebungsprozess einzuleiten, bspw. im Hinblick auf die Art der Fragestellung und die Rolle der Interviewer bei der Abfrage.

In den Tabellen 11 und 12 finden sich die Angaben der Betriebskontakte von 298 Respondenten aus der Respondentenperspektive. 125 Personen gaben den Betriebskontakt vollständig an (Name, Straße, Hausnummer, PLZ und Ort). 226 Personen konnten sämtliche Adresskomponenten korrekt angeben. In der überwiegenden Zahl der Fälle mussten Korrekturen oder Ergänzungen nur bzgl. einer Adresskomponente durchgeführt werden. Diese Ergebnisse weisen darauf hin, dass die Güte der Angaben der Befragten relativ hoch einzuschätzen ist.

Tabelle 8: Anzahl Unvollständiger Angaben, Anteile von 298 editierbaren Angaben berechnet

| | Anzahl Beobachtungen | Prozente | Kumulierte Prozente |
|---|----------------------|----------|---------------------|
| Vollständig (Name, PLZ, Ort, Str., Hnr) | 125 | 42,28 | 42,3 |
| 1 Neu gefundene Angabe | 124 | 41,28 | 83,6 |
| 2 Neu gefundene Angaben | 35 | 11,74 | |
| 3 neu gefundene Angaben | 12 | 4,03 | |
| 4 neu gefundene Angaben | 2 | 0,67 | |
| Insgesamt | 298 | 100 | 100 |

Quelle: SOEP Innovationsbefragung 2007.

Hinweise, die auf die fehlende Fähigkeit (statt fehlender Bereitschaft) der Befragten deuten, finden sich zum Beispiel in der Art bestimmter Angaben der Adresskomponente Straße. Vereinzelt wurden hier korrekte Ortsbeschreibungen gegeben („Am Bahnhof“, „Am Schloß“) statt der Nennung des formal korrekten Straßennamens. Ein weiteres Beispiel für fehlende Fähigkeit trotz vorhandener Bereitschaft sind Angaben zu PLZ oder Hausnummer, die sich nur in einzelnen Ziffern oder der Ziffernreihenfolge von der korrekten Angabe unterscheiden.²⁰ Hilfestellungen durch den Interviewer (z.B. durch Nutzung eines Postleitzahlen-/Telefonbuchs während des Interviews) oder die Bitte darum, die Adresse nachzuschlagen, könnten somit zu einer wesentlichen Verbesserung der Angaben in Bezug auf Qualität und Quantität beitragen und so evtl. die Ausfälle von Stichprobe B nach C verringern.

²⁰ Daher wurde bereits bei Vorliegen drei korrekter Ziffern der PLZ, diese als „korrekte“ Angabe gewertet.

Tabelle 9: Anzahl fehlerhafter Angaben

| Anzahl | Anzahl Beobachtungen | % |
|---------------------|----------------------|-------|
| Vollständig korrekt | 226 | 75,84 |
| 1 Korrektur | 62 | 20,81 |
| 2 Korrekturen | 8 | 2,68 |
| 3 Korrekturen | 2 | 0,67 |
| Insgesamt | 298 | 100 |

Quelle: SOEP Innovationsbefragung 2007.

4.4 Erhebung der Betriebsinformationen

4.4.1 Telefonischer Vorkontakt

Die Betriebe sollten zunächst telefonisch kontaktiert werden, um Sie um ihre Teilnahme an der Betriebsbefragung zu bitten (für welche die Emailadresse notwendig war). Auf Basis der editierten Betriebskontaktadressen wurden die Telefonnummern aus dem Telefonbuch ergänzt. Deren Auffinden erwies sich daher als unproblematisch, da bereits im Zuge der Dateneditierung das Telefonbuch genutzt wurde.

Im telefonischen Vorkontakt waren Informationen über die Studie preiszugeben, die den telefonischen Gesprächspartner über Thema, Art des Auftraggebers und den gesuchten Ansprechpartner im Betrieb aufklären sollten. Als besonders sinnvoll erwies es sich hierbei, den nicht-telefonischen Befragungsmodus zu betonen. Dies lässt sich evtl. damit begründen, dass nicht-telefonische, online- oder postalische Befragungen mit der Möglichkeit der freien Zeiteinteilung durch die antwortende Person ebenso einhergehen wie mit der Gelegenheit, sich des (wissenschaftlichen, nicht-kommerziellen) Adressaten der Umfrage zu versichern. Einschränkung ist festzuhalten, dass eine telefonische Zusage möglicherweise als wenig bindend wahrgenommen wird. Es ist mit vorgetäuschten Zusagen – Scheinzusagen – zu rechnen, die getätigt werden, um eine schnelle und reibungslose Beendigung des Gesprächs zu erreichen. Darüber hinaus zeigte es sich insbesondere bei Klein- und Kleinstbetrieben als vorteilhaft, auf die Notwendigkeit der Teilnahme von und das Interesse an Betrieben dieser Subpopulation gesondert hinzuweisen.²¹

Generell scheint es aus Zeitsparmotiven empfehlenswert, die ohne Nachfrage zu gebenden Informationen möglichst gering zu halten. Aufwand und Dauer des Kontakts sind sehr hoch und lang, da bei den häufig vorkommenden telefonischen Weiterleitungen innerhalb einer Organisation (oder auch zwischen Betriebsteilen einer Organisation) die Erstinformationen mehrmals im Rahmen eines Betriebskontaktes zu wiederholen sind. Dennoch ist zu befürworten, dass die Befragungsinformationen stets - das heißt, auch bei Kontakt mit unbeteiligten „Gatekeepern“ (Telefondienst, Zentrale Vermittlungsstellen) - vollständig genannt werden. Erstens können die Informationen, die ein Gatekeeper bei einer telefonischen Gesprächsanfrage an die entsprechenden Entscheidungsträger weiterleitet, für die Zustimmung dieser zum Durchstellen des Telefonanrufs entscheidend sein. Zweitens kann hierdurch manchem strikten Nein vorgebeugt werden, welches auf Missverständnissen beruht (bspw. dem strikten Nein zu kommerziellen oder telefonischen Befragungen).

Im Gespräch mit den direkten Ansprechpartnern waren Rückfragen hinsichtlich der Inhalte des Fragebogens bzw. der Art der Fragen häufig. Rückfragen hinsichtlich des Zwecks der Studie bzw. detailliertes Nachfragen waren jedoch eher selten, berichteten die Interviewer. Nur in Ausnahmen wurde nach dem Auswahlverfahren gefragt, was mit dem Verweis auf

²¹ Von diesen Betrieben kam z.B. der Einwand, „ich glaube nicht, dass wir für Sie von Interesse sind“.

eine statistische Zufallsziehung aus Datenbanken beantwortet wurde, da die Art der Stichprobenziehung (auf einer Personenbefragung basierend) nicht preisgegeben werden darf. Bei sehr wenigen Fällen war es notwendig, explizit und nachdrücklich die Anonymität zu versichern.

Tabelle 13: Ergebnisse des telefonischen Vorkontakts

| Ergebnis / Abschluss des Telefonats | Fallzahlen | Prozente | Prozente |
|--|------------|----------|----------|
| | N=298 | 100=298 | 100=286 |
| Betrieb doppelt im Datensatz | 12 | 4 | - |
| Ja (postalisch: 36 bzw. 22%, online: 125 bzw. 78%) | 157 | 53 | 55 |
| Nein | 84 | 28 | 29 |
| Nicht erreichbar | 45 | 15 | 16 |
| Abbruch wegen erfolgloser Anrufversuche (oder Ende Feldzeit) | 30 | 10 | 10 |
| Telefonnummer falsch und nicht korrigierbar | 11 | 4 | 4 |
| Firma existiert nicht mehr / in Auflösung | 4 | 1 | 1 |

Die Ergebnisse des telefonischen Erstkontakts sind in Tabelle 13 zusammengefasst und belegen, dass die durchschnittliche Ablehnungsquoten am Telefon bei knapp einem Drittel liegt, die Zusagen überwiegen hier deutlich. Die postalische Befragung präferierten 36 der kontaktierten Betriebe. Das zeigt, dass die dem Erhebungsmodus (hier: online) geschuldeten Ausfälle durch eine angebotene Alternative (hier: postalisch) erhalten bleiben können. Zu dem Schluss, dass Betrieben die Auswahl aus unterschiedlichen Erhebungsmodi angeboten werden sollte, kommen auch Tarnai & Paxson (2004), die in ihrer Studie die Präferenzen US-amerikanischer Betriebe für bestimmte Erhebungsmodi untersuchten. Von den Betrieben dieser Erhebung genannte Gründe, die gegen die Teilnahme an der Online-Befragung sprachen, sind die notwendige Herausgabe der Email-Adresse, fehlende Internetaffinität sowie die Tatsache, dass „grundsätzlich nur schriftliche Anfragen“ bearbeitet werden, wie ein Betrieb berichtete.

Die gesamten Ausfälle dieser Stufe (ausgehend von der Stichprobe der in Datenbanken auffindbaren Betriebe, vgl. Tabelle 2, Stichprobe C) in Höhe von 46% sind nicht nur Ergebnis von Zustimmung oder Ablehnung der Betriebe. Bevor es hierzu kommen konnte, musste zunächst einmal der Betrieb erfolgreich kontaktiert und ein kompetenter und entscheidungsbefugter Ansprechpartner im Betrieb identifiziert worden sein. Im Rahmen von telefonischen Kontaktversuchen bei Betriebsbefragungen sind somit (mind.) folgende Probleme der Fallzahlrealisierung zu unterscheiden: das Problem der Erreichbarkeit, das Problem der Identifikation eines angemessenen Informanten und das Problem der Zustimmung.

4.4.2 Probleme der Fallzahlenrealisierung

4.4.2.1 Erreichbarkeit

Das Problem der Erreichbarkeit des Betriebs ergibt sich, wie in Tabelle 13 dargestellt, durch Abbrüche aufgrund erfolgloser Anrufversuche bis zum Ende der Feldzeit (z.B. Anrufbeantworter, Besetztzeichen), falschen Telefonnummern und Firmen, die sich in Auflösung befinden oder bereits nicht mehr existieren. Da es sich hierbei evtl. um systematische Ausfälle handelt von hoher Anzahl (16%), wird dieser Punkt explizit beachtet. Die schlechte telefonische Erreichbarkeit der Betriebe ist sicherlich teilweise begründet durch Besonderheiten dieser Studie: Zwischen Individual- und Betriebsbefragung lag eine relativ große Zeitspanne von über einem Jahr. Um einen systematischen Einfluss aufgelöster oder nicht (mehr) erreichbarer Betriebe messen zu können, wäre es sinnvoll, derartige Betriebe für spätere Analysen nicht aus dem Datensatz auszuschließen. Aufgelöste Betriebe stellen eine inhaltlich relevante Information für spätere Analysen dar. Durch eine zeitnah an die Individualbefragung angelegte telefonische Betriebsbefragung könnte ihre Zahl jedoch verringert werden. Ausfälle aufgrund

erfolgloser Anrufversuche bis zum Ende der Feldzeit sollte mit einem Design begegnet werden, dass Anrufversuche zu unterschiedlichen Wochen- und Tageszeiten vorsieht und systematisch variiert. Diese Empfehlung stützen auch Ergebnisse einer Studie von Willimack & Dalzell (2006), die systematische Zusammenhänge zwischen den Ergebnissen der Anrufversuche und den Tages- und Wochenzeiten, an denen diese stattfanden, entdecken konnten.

Die schwierige Erreichbarkeit von Betrieben ist nicht nur problematisch in Bezug auf die Fallzahlen, sondern beeinflusst insgesamt die Höhe des Aufwands über die Häufigkeit der Anrufe und Anrufversuche, deren Dauer und der Dauer der telefonischen Kontaktphase insgesamt sowie die Kosten. Um 286 Betriebe telefonisch zu erreichen, wurden insgesamt 521 Anrufe in einem Zeitraum von 6 Wochen getätigt. Dies entspricht einer durchschnittlichen Anruhfrequenz von 1,8 Anrufen je Betrieb. 12% der Betriebe wurden dreimal oder häufiger angerufen (siehe Tabelle 14).²²

Tabelle 14: Anzahl Anrufe je Betrieb und insgesamt

| Anrufe je Betrieb | Anzahl Betriebe | % (von 286) |
|----------------------------|-----------------|-------------|
| 1 | 169 | 59 |
| 2 | 55 | 19,0 |
| 3 | 28 | 9,8 |
| 4 | 21 | 7,3 |
| 5 | 6 | 2,1 |
| 6 | 3 | 1,0 |
| 7 | 1 | 0,3 |
| 8 | 2 | 0,7 |
| 9 | 1 | 0,3 |
| Anzahl Anrufe insg. | 521 | |

Lynn & Sala (2004) kommen bei der von Ihnen durchgeführten telefonischen Nachbefragung von Betrieben in Großbritannien, die auf die postalische Anfrage nicht reagiert hatten, zu dem Schluss, dass kein Zusammenhang besteht zwischen der Anzahl der Anrufe und Anrufversuche und der Zustimmungsquote. Sie empfehlen daher die Kontaktbemühungen entsprechend hoch zu halten und keinesfalls zu früh abzubrechen. (Lynn/Sala 2004:12) In dieser Studie, die sich durch eine aufwendige telefonische Kontaktphase auszeichnet, wurde eine Teilnahmequote von 57,1% der telefonisch kontaktierten Betriebe erreicht. Nur 5,5% der Betriebe konnten nicht erreicht werden.

4.4.2.2. Identifikation

Jede Organisationsbefragung sieht sich mit dem Problem konfrontiert, dass zur Erfassung ihres Analyseobjekts eine Person identifiziert werden muss, die stellvertretend für die Organisation, den korporativen Akteur, antwortet. Wer dieser Informant sein sollte, ist eine theoretisch vorab zu treffende Entscheidung (häufig: Vertreter der Personalabteilung, Geschäftsleitung). Gemäß der in dieser Studie gemachten Erfahrungen wird die Identifikation des Ansprechpartners im Betrieb, der Kompetenz und Entscheidungsbefugnis vereinen sollte, im Rahmen eines telefonischen Vorkontakts durch folgende vier Aspekte erschwert:

- 1) die telefonische Erreichbarkeit des Ansprechpartners selbst,
- 2) das Gatekeeper-Verhalten,
- 3) das Auseinanderfallen von Entscheidungsgewalt und Teilnahmebereitschaft,

²² Nicht systematisch festgehalten wurde die Anzahl Weiterleitungen innerhalb eines Gesprächs, die zwischen 0 und 4 schwankten und die Dauer des Telefonats entsprechend beeinflussen. Ebenfalls können auf Basis der vorliegenden Daten keine Aussagen über die durchschnittliche Dauer der Anrufe und die Kosten getätigt werden.

- 4) das Auseinanderfallen der Kenntnisse, die zur Beantwortung des Fragebogens notwendig sind, auf mehr als eine Person.

(1) Die telefonische Erreichbarkeit des Ansprechpartners selbst

Häufig konnte der Ansprechpartner während des ersten Telefonats nicht direkt gesprochen werden („im Urlaub“, „außer Haus“, „in Besprechung“). In diesen Fällen wurde entweder um Rückruf zu anderen (Tages- oder Wochen-)Zeiten gebeten, die E-mailadresse herausgegeben oder zugesagt, die Einladungsmail weiterzuleiten. Der Vorteil, den der telefonische Vorkontakt hinsichtlich des Aufbaus einer persönlichen Beziehung mit den Mechanismen sozialer Kontrolle und Reziprozität bietet, konnte in diesen Fällen nicht ausgeschöpft werden. Insbesondere hier ist die telefonische Zustimmung als wenig valide einzuschätzen, d.h. es ist mit Scheinzusagen zu rechnen.

Um die Erreichbarkeit von Zielpersonen zu erhöhen, erscheint es (ebenso wie für die Erreichbarkeit der Betriebe allgemein) angebracht, einen kontinuierlichen und zeitlich aufgesplitterten Anrufprozess zu gestalten. Mit entsprechender Personalausstattung und ggf. unterstützender Software ist dies realisierbar. Ein optimierter Anrufprozess (und ggf. eine verlängerte Feldphase) erhöht die Chance, die Ansprechpartner des Betriebs zu identifizieren und zu erreichen und somit auch die Ausfälle aufgrund von „Abbruch wegen erfolgloser Anrufversuche (oder Ende Feldzeit)“ zu reduzieren. Speziell geschulte Interviewer sollten Zustimmungen oder Ablehnungen von anderen als der Zielperson nicht akzeptieren.(vgl. auch Lynn/Sala 2004)

(2) Das Gatekeeper-Verhalten

Der Kontakt zum Ansprechpartner lief in der Regel über zentrale Telefondienste, die überwunden werden mussten. Die Anwahl von Telefonzentralen birgt zwei potentielle Risiken. Zum einen kann es zu unbemerkten, automatischen Weiterleitungen des Anrufs kommen. In diesem Fall kann ein Interviewer ohne dessen Wissen (und damit außerhalb jeglicher Kontrolle) zu einer anderen als der gesuchten betrieblichen Einheit gelangen. Hier gilt es, sich den Namen und die Adresse des Betriebs bestätigen zu lassen. Zum zweiten sind prinzipielle Ablehnungen und Weisungen, telefonische Anfragen von Befragungsinstituten nicht weiterzuleiten, an dieser Stelle schwer überwindbar („keine Auskunft am Telefon“). Abhilfe könnte z.B. das direkte Anwählen zufälliger Durchwahlstellen schaffen, welches mit einem entsprechenden Anstieg des Aufwands einherginge, der Einsatz entsprechend geschulter Interviewer oder eine schriftliche Vorankündigung, auf die verwiesen werden könnte.

(3) Das Auseinanderfallen von Entscheidungsgewalt und Teilnahmebereitschaft

Während in den genannten ersten zwei Punkten die Schwierigkeit der telefonischen Kontaktierung darin besteht, den korrekten Betriebsinformanten ausfindig zu machen, besteht das Problem dieses und des folgenden Aspekts darin, dass ein derartiger Ansprechpartner, der zugleich kenntnisreich und entscheidungskompetent sein muss, nicht immer in einer Person existiert. So zum Beispiel, wenn in manchen Betrieben die Entscheidungsgewalt zur Teilnahme nicht der Person obliegt, die vom Interviewer als Zielperson ausgemacht worden war und bereits ihre Teilnahmebereitschaft erklärte hatte (z.B. städtische Behörden, Großunternehmen mit einheitlicher Regelung, „Befugnis fehlt“, „Schulkonferenz notwendig“, „nur Anfragen der Sparkassen- und Finanzgruppe“, „zu genehmigen vom Städte- und Gemeindebund“, „GL stimmt nicht zu“). Wird die Teilnahmeerlaubnis durch eine andere (meist höhere) Stelle erteilt, sollte der Interviewer versuchen Kontakt mit dieser aufzunehmen, um die Erlaubnis selbst einzuholen und dies nicht der Zielperson überlassen.

(4) *Das Auseinanderfallen der Kenntnisse, die zur Beantwortung des Fragebogens notwendig sind, auf mehrere Personen*

Die Kompetenzen, die nötig sind, den Fragebogen auszufüllen, können ebenfalls auf mehr als eine Person verteilt sein. Sie können darüber hinaus auch über den einzelnen Betrieb hinaus verteilt sein, wie in den Fällen, in denen Personalabteilungen, die regelmäßig als Ansprechpartner für sozialwissenschaftliche Betriebsbefragungen dienen, zentral lokalisiert oder ausgelagert sind. Derartige Organisationsstrukturen sind mit forschenseitig zu beantwortenden Problemen behaftet. Wie ist sicherzustellen, dass der Fragebogen, wenn er zentral beantwortet wird, dennoch für die lokale Betriebsstätte ausgefüllt wird? Bestimmte (z.B. klimatische, kulturelle) Fragen, sind nur von vor Ort ansässigen Mitarbeitern beantwortbar. Auf diesen Umstand muss bereits am Telefon eingegangen werden, bspw. durch das Erfassen (Gespräche) mehrerer Ansprechpartner und das Aufsplitten des Fragebogens in entsprechende Einzelteile.

Tomaskevic-Dewey et al (1994) fanden heraus, dass die Informationsfragmentierung in Betrieben (von ihnen als fehlende *capacity* des Betriebs zur Teilnahme an einer Befragung bezeichnet) ebenso wie die fehlende Entscheidungsgewalt (*authority*) der identifizierbaren Personen eines Betriebs systematisch mit betrieblichen Faktoren zusammenhängt. In diesem Fall beeinträchtigen derartige Ausfälle die Generalisierbarkeit der Ergebnisse und sollten vermieden werden.

Die hinsichtlich der Identifizierbarkeit der Ansprechperson gemachten Beobachtungen sprechen insgesamt für den Einsatz professioneller, speziell geschulter Interviewer. Denn Interviewer sollten in der Lage sein, Zustimmungen oder Ablehnungen nur von der tatsächlichen Zielperson zu akzeptieren, Gatekeeper zu überwinden, und über besondere Strategien verfügen, für diejenigen Fälle, in denen die Teilnahmezusage von einer andere Stelle einzuholen ist oder der Fragebogen durch mehrere Personen ausgefüllt werden sollte.

4.4.2.3. *Zustimmung*

Hinsichtlich des Problems der Zustimmung gilt es, die Gründe für eine mögliche Zu- oder Absage näher zu betrachten. Die Formulierungen der Absagen lassen sich aufteilen auf die Kategorien „nein, prinzipiell nicht“, „nein, keine Zeit“ und „nein, kein Interesse“. 16 der 94 Ablehnungen entfielen auf „nein, prinzipiell nicht“.

„Nein, prinzipiell nicht“-Angaben lassen auf eine betriebsseitige Richtlinie schließen, die die Teilnahme an Umfragen jeglicher Art generell ausschließt. Der Repräsentant des Betriebs oder bereits der Gatekeeper müssten sich dieser Richtlinie widersetzen, um an der Befragung dennoch teilzunehmen. Die Einflussmöglichkeiten des einzelnen Forscherteams sind hier relativ gering einzuschätzen. Inwiefern „Nein, prinzipiell nicht“-Angaben durch Einsatz professioneller Telefon-Interviewer reduzierbar sind, ist offen. Denn die organisationsmethodische Forschung hat bislang keine Antworten geliefert dafür, welches die Ursachen sind für diese generell ablehnende Haltung von Betrieben gegenüber Befragungen, inwiefern diese systematisch die Stichproben verzerren, und ob es sich lohnt trotz ablehnender Haltungen der Betriebe, diese erneut bzw. mehrmals anzufragen.

Absagen aufgrund von Zeitmangel und fehlendem Interesse scheinen eher beeinflussbar zu sein. Denn diese kommen unter bestimmten Bedingungen zustande, die evtl. forschenseitig durch entsprechende Gestaltung des Designs manipulierbar sind. Ausfälle, die durch „Nein, keine Zeit“ begründet sind, könnten z.B. dadurch reduziert werden, dass besondere saisonale Schwankungen (Urlaubszeiten, Jahresabschluss), insofern diese branchenübergreifend übereinstimmen, berücksichtigt werden. Zudem könnten sich die saisonalen Auslastungsgrade zwischen einzelnen Abteilungen eines Betriebs unterscheiden. Im Gegensatz zu Ausfällen

aufgrund von Zeitmangel, beziehen sich durch „Nein, kein Interesse“ begründete Ausfälle weniger auf die Fähigkeiten als auf die Bereitschaft des Betriebs bzw. dessen Repräsentanten an einer bestimmten Befragung oder an Befragungen generell teilzunehmen. Das Interesse an einer bestimmten Befragung teilzunehmen ist beeinflussbar durch eine interessante Gestaltung des Themas der Studie und die Variation erhebungsspezifischer Merkmale, wie bspw. die Wahl des Erhebungsmodus, die Seriosität des Auftraggebers, explizite Anonymitätssicherungen und die Bereitstellung eines erzielbaren Nutzens der Teilnahme.

Abbildung 2 fasst den Prozess der Teilnahme eines Betriebs an einer Befragung anhand der Formel (1) zusammen. Formel (1) verdeutlicht, dass die endgültige Teilnahmeentscheidung (nach erfolgreicher Kontaktierung des Betriebs sowie des Ansprechpartners im Betrieb) bei dem jeweiligen Individuum liegt, welches als Betriebsinformant vorgesehen ist. Dessen Entscheidung ist sowohl von individuellen als auch von betrieblichen Faktoren abhängig (Formel 1.3.1.). Die betrieblichen Merkmale stellen als die situativen, strukturellen Bedingungen Restriktionen oder Opportunitäten für die individuelle Entscheidung dar. Sie beeinflussen das individuelle Kosten-/Nutzenverhältnis einer Teilnahme. (vgl. auch Schnabel 1997) Besitzt das Individuum bspw. keine Befugnis zur Teilnahme, wäre eine solche mit hohen individuellen Kosten des nonkonformen Verhaltens verbunden. Nutzengewinne entstünden z.B. dann, wenn durch das Forscherteam den Teilnehmern der Studie Wissensgewinne (in Form von für die Teilnehmer relevanten Informationen) ermöglicht werden oder andere incentives geboten würden.

Abbildung 2: Erklärungsansatz der Teilnahme von Betrieben an Umfragen

| |
|---|
| <p>(1.) Teilnahme_B = β₁ Erreichbarkeit_B + β₂ Identifikation_I + β₃ Teilnahme_I</p> <p>(1.1.) Erreichbarkeit_B = Existenz + Tel.Nr vorhanden + Feldphasenende</p> <p>(1.2.) Identifikation_I = β₁ Erreichbarkeit_I + β₂ Gatekeeper + β₃ Befugnis.vs.TB + β₄ Kenntnisse_I</p> <p>(1.3.) Teilnahme_I = Zustimmung_I + e</p> <p>(1.3.1.) Zustimmung_I = betriebl. Merkmale + individuelle Merkmale + Merkmale d. Erhebung = (β₁Zeit_B + β₂Interesse_B + β₃sonstige_B) + (β₄Zeit_I + β₅Interesse_I + β₆sonstige_I) + β₈Erhebungsmerkmale + ...</p> <p><i>Legende:</i> „Merkmal“_I: Merkmal des Informanten des Betriebs „Merkmal“_B: Betriebsmerkmal Identifikation_I: Identifikation des Ansprechpartners im Betrieb, siehe 4.4.2.2. Befugnis.vs.TB: Auseinanderfallen von Befugnis und Teilnahmebereitschaft Kenntnisse_I: ungenügende Kenntnisse des identifizierten Ansprechpartners e: Fehlerterm</p> |
|---|

Quelle: Eigene Zusammenstellung.

Zusammenfassend kann festgehalten werden, dass durch ein Forschungsdesign, welches den Bedürfnissen (Zeit, Erhebungsmodi) und Interessen (Thema) der Befragten entgegenkommt sowie dem Einsatz professioneller Telefonagenten die Zustimmungsquoten zur Teilnahme nach Ansicht der Autoren erhöht werden könnten. Dies würde den auf dieser Stufe vorliegenden Ausfall in Höhe von 35% Ablehnung möglicherweise reduzieren. Optimistisch stimmt in diesem Zusammenhang die Tatsache, dass der Umgang am Telefon lt. Interviewerberichten in den meisten Fällen höflich und freundlich war, selbst bei ablehnender Haltung gegenüber der Teilnahme.

4.4.3 Online- und postalische Befragung

Nach dem telefonischen Vorkontakt wurde der Ansprechpartner des jeweiligen Betriebs postalisch oder per Mail angeschrieben und über den Zweck und den Inhalt der Studie „Betriebliches Personalmanagement“ sowie den Auftraggeber aufgeklärt. In der Mail befand sich der Link zur Online-Umfrage während die Postsendung den schriftlichen Fragebogen ebenso wie einen adressierten und frankierten Rückumschlag beinhaltete. Nach ca. zwei Wochen gingen jeweils die erste und die zweite Erinnerung heraus, wobei der zweiten postalischen Erinnerung erneut der Fragebogen mit Rückumschlag beilag.

Bei der Gestaltung des Fragebogens wurde überwiegend auf Fragen aus bestehenden Datensätzen wie dem deutschen IAB-Betriebspanel sowie dem US-amerikanischen National Organizations Survey (NOS) von 2002 zurückgegriffen. Die Orientierung am Betriebspanel erfolgte insb. hinsichtlich der Erfassung der „sozio-demographischen“ Merkmale der Betriebe. Hinzugefügt wurden Merkmale, die sich auf die Marktsituation, d.h. die relative Lage des Betriebs beziehen, Merkmale der Einkommens- und Mobilitätsstrukturen, personalpolitische Faktoren (wie Arbeitszeiten, Weiterbildung, Personalrekrutierung – größtenteils dem Betriebspanel entsprechend), Fragen zum Betriebsklima und die Erhebung von Daten über die für den Betrieb auskunftgebende Person. Eine Besonderheit des NOS wurde übernommen, indem einige Fragen jeweils getrennt für Beschäftigte der Kern- wie für Beschäftigte der Führungsbelegschaft gestellt wurden. Kernbelegschaft ist definiert als derjenige Teil der Belegschaft, der direkt an der Erstellung des Produkts oder der Dienstleistung beteiligt ist, der sogenannte operative Kern, exklusive der Stabstellen, Verwaltung oder der Personalabteilung.²³

4.4.3.1 Ergebnisse

Die Ausfallquote dieses spezifischen Schritts der Stichprobengenerierung ergibt sich durch das Nichtausfüllen des online- bzw. das Nichtzurücksenden des schriftlichen Fragebogens durch die Betriebe bis zum Ablauf der Feldphase. Die Ausfälle dieser Stufe belaufen sich auf 82% der Betriebe, die sich telefonisch zu einer Teilnahme bereit erklärt hatten. Davon entfällt ein kleiner Teil darauf, dass die Adresse fehlerhaft war und daher die (elektronische oder schriftliche) Post nicht zugestellt werden konnte. In der Literatur werden Ausschöpfungsquoten zwischen 10 und 80% berichtet.(vgl. etwa: Hartmann und Kohaut 2000; Lang 2005; Riedmann et al. 2006, Schmidt 2007; Schnabel 1997)

Tabelle 15: Ergebnisse Rücklauf der Betriebsbefragung

| | Online | Postalisch | N=161 |
|--|------------|------------|------------|
| <i>Ergebnisse</i> | 125 (100%) | 36 (100%) | 161 (100%) |
| Adresse falsch | 8 | 1 | 9 |
| Kein Rücklauf | 97 (60%) | 26 (72%) | 123 (76%) |
| <i>davon Fragebogen betrachtet, aber nicht beendet</i> | 43 | | |
| Rücklauf | 20 (16%) | 9 (25%) | 29 (18%) |

4.4.3.2 Mögliche Ursachen für Unit-Nonresponse

Folgt man den in Punkt 4.4.2.3. hinsichtlich Abbildung 2 gemachten Ausführungen zur Erklärung betrieblichen Nonresponse-Verhaltens sind nach erfolgreicher Kontaktierung des Betriebs und des Informanten für dessen Zustimmung und Teilnahme betriebliche Merkmale, individuelle Merkmale und Merkmale der Erhebung als Determinanten betrieblichen Nonresponse zu unterscheiden. Die Basis für die folgenden Erläuterungen bilden die betriebsseitig getätigten Absagegründe sowie das beobachtbare Verhalten der Betriebsinformanten.

²³ Im Unterschied zum NOS wurde diese Trennung innerhalb der jeweiligen Fragen vorgenommen, während die Forscher des NOS getrennte Frageblöcke vorgaben, und diese je einmal für die Kern- und einmal für die Gesamtheitsbelegschaft beantworten ließen.

Merkmale der Erhebung

(1) Wahl des Erhebungsmodus

Bei Betrachtung der nach Erhebungsmodus aufgespaltenen Rücklaufquoten fällt auf, dass der Rücklauf der postalischen Befragung neun Prozentpunkte höher liegt als derjenige der Onlinebefragung. Dies entspricht den Ergebnissen der Studie von Tarnai & Paxson bezüglich der Modus-Präferenzen US-amerikanischer Betriebe, welche den schriftlichen Modus klar gegenüber einer Befragung im Internet favorisieren. Welche Gründe hierfür anzuführen sind, ist eine offene Forschungsfrage. Evtl. wird ein schriftlich versandter Fragebogen als „offizieller“ eingeschätzt als eine Online-Befragung. Eine andere Erklärungsmöglichkeit könnte darin zu sehen sein, dass vor dem Hintergrund des heute üblichen hohen Ausmaßes an Emailverkehr die Wahrnehmung einer einzelnen Mail mit vergleichsweise unwichtigem Inhalt und unbekanntem Absender fraglich ist. Dieser Problematik sollte durch eine sehr kurzfristig nach dem telefonischen Vorkontakt versandte Einladungsmail begegnet werden, was allerdings in dieser Studie so gehandhabt wurde.

(2) Fragebogengestaltung

Interessant ist die Beobachtung, dass der Online-Fragebogen von 43 Betrieben zwar angeschaut (von 5 Betrieben sogar vollständig), aber nicht ausgefüllt wurde. Um diese Betriebe zu einer Teilnahme zu bewegen, könnte ein verändertes Fragebogendesign hilfreich sein. Wie ein Fragebogen in Organisationsbefragungen aussehen sollte, und ob sich Betriebe systematisch in ihren Präferenzen hinsichtlich Layout, Art der Fragestellung und Inhalt des Fragebogens unterscheiden, ist bisher eine noch kaum berücksichtigte methodische Forschungsfrage. (vgl. z.B. Phipps 1995)

(3) Inhalt des Fragebogens und Zweck der Studie

Aus der vorliegenden geringen Rücklaufquote (8 bzw. 10%) ebenso wie aus den gemachten Erfahrungen im Rahmen der Telefongespräche ist u.a. zu schließen, dass eine personalisierte, d.h. branchen- und betriebsspezifische - Ansprache anzuraten ist. Zu dem gleichen Schluss kommt auch Schnabel 1997, denn „ein Einheitsfragebogen als Konzept für alle Branchen und Betriebsgrößenklassen [gehe] an den Bedürfnissen und Fähigkeiten der Betriebe vorbei“ (Schnabel 1997:170) Die Unterschiede zwischen Betrieben (Groß- und Kleinbetriebe) und Branchen (öff. vs. nicht-öffentlicher Dienst, privatgewerblich oder gemeinnützig) können teilweise sehr groß sein. Titel wie Anschreiben einer bestimmten Ausrichtung können Betriebe oder ganze Branchen, auf die dies nicht zutrifft, abschrecken. Negative Rückmeldungen dieser Art kamen bspw. von einer Schule, einer kirchlichen Vereinigung und einer öffentlichen Behörde („Karrierechancen im öff. Dienst widerspricht sich“, die sich mit Titel („Betriebliches Personalmanagement“), Ausrichtung der Studie sowie dem im Anschreiben angepriesenen Nutzen einer Teilnahme (gestiegener Wettbewerb, Effizienzanforderungen) nicht identifizieren konnten. Eine gezielte (personalisierte) Ansprache birgt zudem die Möglichkeit, auf betriebs-/branchenspezifische Teilnahmehindernisse explizit einzugehen, dagegen zu argumentieren sowie betriebs-/branchenspezifische Teilnahmegründe zu lancieren. Schnabel 1997 nennt in diesem Zusammenhang die Förderung spezifischer Unternehmensziele durch die Teilnahme an einer Befragung und die Nennung einer Autoritätsperson der jeweiligen Branche.

Die Kriterien, die sich als relevante Unterscheidungsmerkmale zwischen Betrieben erwiesen haben, welche in einer Ansprache zu berücksichtigen wären, sind (mindestens) öffentlicher vs. nicht-öffentlicher Dienst, privatgewerblich vs. gemeinnützig tätige Organisation, Klein-, Mittel-, Großbetrieb und diversifizierte vs. zentralisierte Organisationsstruktur.

Individuelle Merkmale des Betriebsinformatanten

Zusammenhänge zu individuellen oder betrieblichen Merkmalen sind auf Basis vorhandener Daten (fehlender Informationen der Nonresponse-Betriebe) nicht messbar. Jedoch wurden in dieser Studie am Ende des Fragebogens einige Fragen den betrieblichen Informanten betreffend gestellt, die zumindest Aussagen über deren Merkmalsverteilungen erlauben. Erfasst wurden Alter, Bildungsabschluss, Jobbezeichnung, Tätigkeitsgebiete und Dauer der Betriebszugehörigkeit sowie die vorher eingenommene Stelle der für den Betrieb stellvertretend antwortenden Person. Nach Durchsicht dieser Antworten zeigt sich, dass die betrieblichen Informanten i.d.R. eine hohe Position im Personalbereich ausführen. Die kürzeste Betriebszugehörigkeit belief sich auf 4 Monate (2 Personen), ansonsten waren die Personen bereits zwischen 14 und 354 Monaten im Betrieb beschäftigt. Das Ausbildungsniveau der Auskunftsggeber ist durchgehend hoch bis sehr hoch, keiner der befragten Personen war noch „in Ausbildung“.

Möglich ist, dass die Position (und Qualifikation) dieser Person ebenso wie die Dauer der Betriebszugehörigkeit nicht nur die Teilnahmebereitschaft, sondern auch maßgeblich die Kenntnisse (Teilnahmefähigkeit) bestimmen und somit letztlich die Validität der Angaben beeinflussen. Würde dazu übergegangen werden, regelmäßig informantbezogene Merkmale zu erheben, wären mögliche Zusammenhänge mit dem Responseverhalten oder Einflüsse auf die Ergebnisse nachträglich zumindest kontrollierbar.

Scheinzusagen während des telefonischen Vorkontakts

Letztlich und unabhängig von den oben angeführten Nonresponse-Ursachen könnten die hohen Ausfälle zwischen der Stichprobe der Zustimmung signalisierenden Betriebe und derjenigen der tatsächlich teilnehmenden Betriebe mit der oben beschriebenen Problematik der Identifizierbarkeit eines kompetenten und entscheidungsbefugten Ansprechpartners im Betrieb zusammenhängen. Konnte telefonisch nicht mit der Zielperson selbst gesprochen werden, sind die von Dritten erhaltenen Zustimmungen eventuell wenig valide. Gleiches gilt, wenn die Zielperson zwar die Kompetenz aber nicht die ausreichende Befugnis besaß, einer Teilnahme zuzustimmen. Hinweise hierfür finden sich auch in einzelnen Absagegründen der Betriebe, die lauteten „Anfrage wurde weitergeleitet, weil selbst nicht zuständig“ und „Entscheidung der GL“.

Insgesamt ist festzuhalten, dass in dieser Phase der Erhebung durch eine Optimierung des Kontaktprozesses (personalisierte Ansprache, zeitnahe Erinnerung, Interessantes Themenfeld, Incentives, verlängerte Feldphase) ein höherer Rücklauf möglich sein sollte.

4.5 Stichprobenvergleich

Die jeweiligen Ausfälle der einzelnen Phasen des Erhebungs- und des Generierungsprozesses der Betriebsstichprobe (verknüpften Datensatzes) sind letztlich v.a. dann problematisch, wenn diese systematisch und nicht zufällig erfolgen. Hierüber kann jedoch in der Regel keine Auskunft gegeben werden, weil dazu genau die Daten notwendig wären, die gerade nicht vorhanden sind: es ist nicht bekannt, *wer* verweigert! Die Möglichkeit besteht, die resultierenden Datensätze anhand vorhandener Angaben zu vergleichen, um zu beurteilen, ob die Datensätze in dieser Hinsicht als „repräsentativ“ gelten können.

Hierzu wurde zunächst die Gesamtstichprobe der Teilnehmer der SOEP-Befragung (N=1023) mit der Stichprobe der Befragten (Stichprobe A, N=370) und der Stichprobe nach Ausschluss der Nonresponse-Fälle, d.h. der Stichprobe der Betriebskontaktangaben (Stichprobe B, N=323), anhand ihrer Personenmerkmalsverteilungen verglichen. Die Ergebnisse sind in den

Abbildungen 3a) bis c) aufgeführt. Zu sehen ist, dass sich die Stichproben in den üblichen soziodemographischen Merkmalen der Personen nur geringfügig unterscheiden.

Abbildung 3: Vergleich der soziodemographischen Merkmale der Personen in den unterschiedlichen Stichproben

Abbildung 3a): Vergleich von Geschlecht, Alter und Wohnort West-/Ostdeutschland

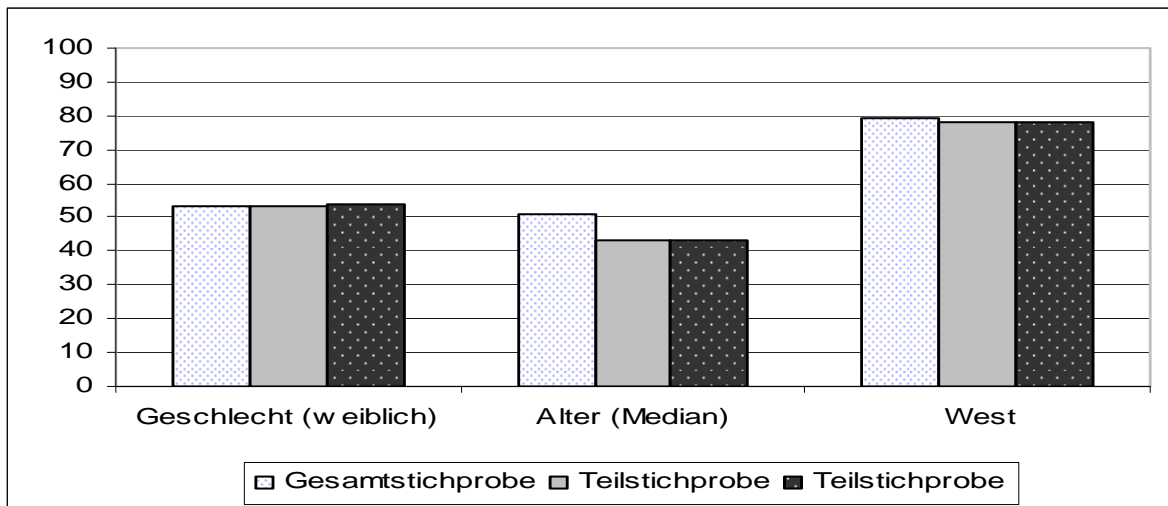


Abbildung 3b): Vergleich Schulabschluss

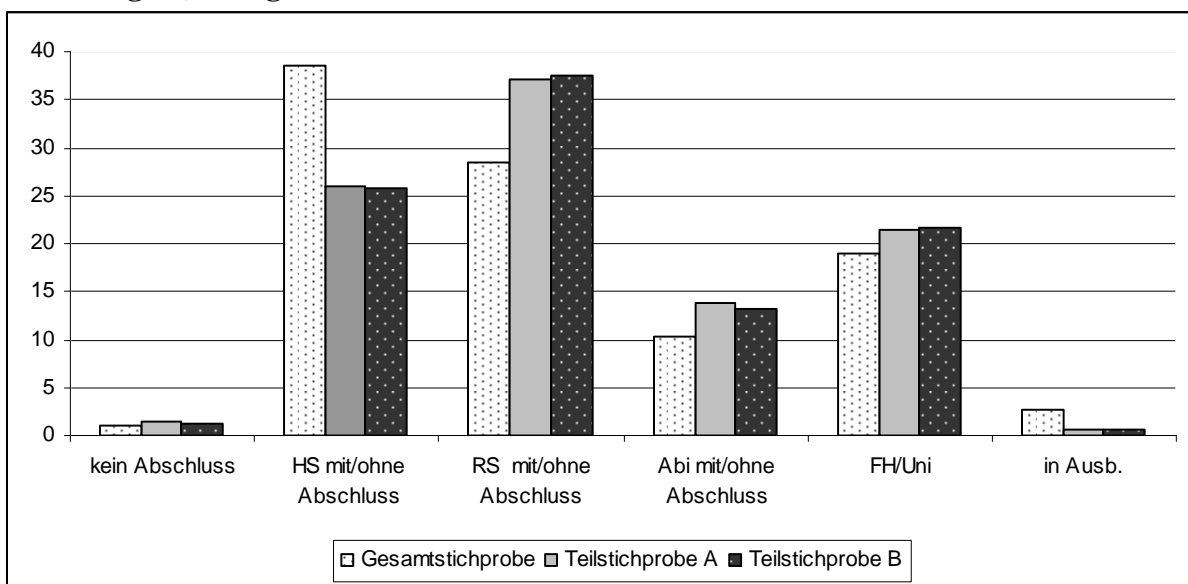
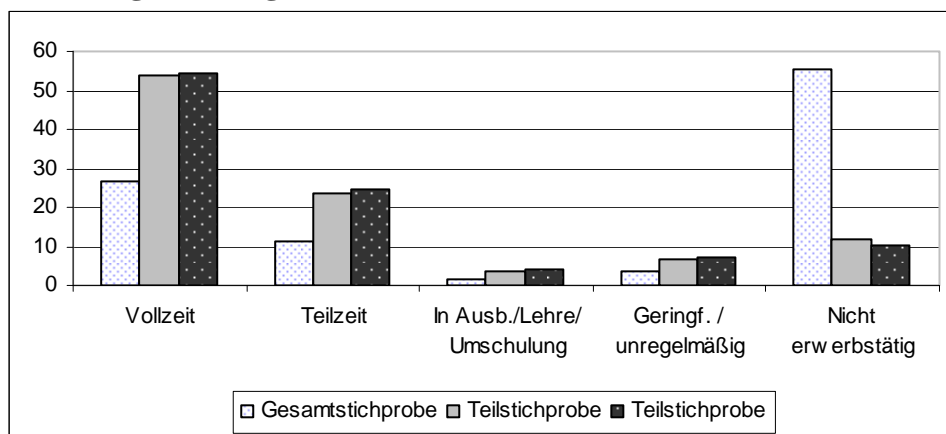


Abbildung 3c): Vergleich Erwerbsstatus



Ein Vergleich der Betriebsstichprobe mit einem externen Kriteriumsdatensatz (z.B. IAB-Betriebspanel) wurde aufgrund der geringen Anzahl der Betriebsangaben (N=29) nicht durchgeführt.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag zeigt, dass hinsichtlich der Erstellung verknüpfter Datensätze unterschiedliche Vorgehensweisen mit ihren spezifischen Vor- und Nachteilen anwendbar sind. Hinsichtlich der hier näher betrachteten und untersuchten Vorgehensweise der employee-first-method gibt es bisher in Deutschland keine entsprechenden Referenzstudien insbesondere in Bezug auf Antwortbereitschaft und -fähigkeit der befragten Personen sowie die Selektivität des so generierbaren Betriebsdatensatzes. Es konnte gezeigt werden, dass erstens, die Angaben der Personen hinreichend vollständig und fehlerfrei sind. Zweitens, konnte mittels der Methode der Datenedition relativ einfach eine beachtliche Verbesserung der Güte der durch offene Abfrage erhaltenen Betriebskontaktadressen erreicht werden. Die Güte und Anzahl der Nennungen der Betriebskontaktdaten ebenso wie die Anzahl der teilnehmenden Betriebe kann durch vielfältige Verbesserungen in den unterschiedlichen Phasen des Generierungsprozesses gesteigert werden. Erste Hinweise und Anregungen hierzu wollte der vorliegende Beitrag liefern.

Die Ergebnisse, gemeinsam mit der beschriebenen Notwendigkeit angesichts der sich durch LEE-Datensätze eröffnenden Analysepotentiale, weisen darauf hin, die Arbeit in diesem Bereich fortzusetzen. Ziel muss es sein, durch weitere methodische Forschung, einen systematischen Einblick zu gewähren über die „best practice“ der Erstellung verknüpfter Datensätze – in Abhängigkeit der jeweils vorliegenden Fragestellungen. Während die empirische Tatsache einer Bedeutungsverschiebung in Richtung „Organisationsgesellschaft“ in die theoretischen Konzepte bereits hinreichend Eingang gefunden hat, (und auch seit längerem statistische Analysemodelle bereitstehen, mit denen derartige Zusammenhänge abgebildet werden können), fehlt es bislang an der hinreichenden Verfügbarkeit entsprechender Datensätze, welche die Mesoebene enthalten. Diese Tatsache gilt es in der methodischen Forschung zukünftig verstärkt zu berücksichtigen. Insgesamt ermutigen die Ergebnisse der Pilotstudie einen ähnlichen Weg auch in der Haupterhebung des SOEP frühestens für das Erhebungsjahr 2011 in Auge zu fassen (vgl. Anger et al. 2008).

6 Literaturverzeichnis

- Abowd, John M. und Francis Kramarz, 1999: The analysis of labor markets using matched employer-employee data. S. 2629-2709 in O. Ashenfelter and D. Card (Hrsg.), *Handbook of Labour Economics*, Vol 3. Amsterdam: North Holland.
- Alda, Holger, Stefan Bender und Hermann Gartner, 2005: The linked employer-employee dataset of the IAB (LIAB). IAB Discussion Paper No. 6/2005.
- Anger, Silke, et al. 2008: Zur Weiterentwicklung von SOEPsurvey und SOEPservice. In: *Vierteljahrshefte zur Wirtschaftsforschung* 77(3): 157-177.
- Bayard, Kimberly, Joel Elvery, Judith Hellerstein and David Neumark, 2002: The 1990 Decennial Employer-Employee Dataset. Working Papers from U.S. Census Bureau, Center for Economic Studies. www.ces.census.gov/index.php/ces/cespapers?down_key=101660
- Bender, Stefan, Joachim Wagner und Markus Zwick, 2007: KombiFiD – Kombinierte Firmendaten für Deutschland. University of Lueneburg Working Papers in Economics No. 60.
- Bender, Stefan, Ralf Himmelreicher, Sylvia Zühlke & Markus Zwick 2008: ForschungsDatenZentren - Eine informationelle Infrastruktur mit erfreulichem Datenzugang. In: Gabriele Rolf, Markus Zwick & Gert G. Wagner (Hrsg.); *Fortschritte der informationellen Infrastruktur in Deutschland*. Baden-Baden: Nomos, S. 147-158.
- Brose, Hanns-Georg, Martin Diewald, Anne Goedicke 2006: Beschäftigungsverhältnisse und sozialer Tausch. Wechselwirkungen zwischen Arbeitsangebot und Arbeitsnachfrage. Kurzdarstellung des gleichnamigen DFG-Projekts. <http://www.uni-due.de/beata/projekt.pdf>
- Capelli, Peter, 2001: The National Employer Survey: Employer Data on Employment Practices. *Industrial Relations* 40: 635-647.
- Calderwood, Lisa & Carli Lessof 2009: Enhancing Longitudinal Surveys by Linking to Administrative Data. in: Peter Lynn (Ed.). *Methodology of Longitudinal Surveys*. Chichester: John Wiley & Sons, S. 55-72.
- Falk, Martin und Viktor Steiner, 2001: Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen, in: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.), *Wege zu einer besseren informationellen Infrastruktur*. Baden-Baden: Nomos.
- Cohen, E., 2000: Equitable classrooms in changing society, in: Hallinan, Maureen ed.: *Handbook of the sociology of education*, New York: Kluwer Academic / Plenum Publishers.
- Fischer, Gabriele, Florian Janik, Dana Müller, Alexandra Schmucker: The IAB-Establishment Panel- from Sample to Survey to Projection, in: *FDZ-Methodenreport* No.01/2008
- Gamoran, A., Secada, W., Marret C., 2000: The Organizational Context of Teaching and Learning, in: Hallinan, Maureen ed.: *Handbook of the sociology of education*, New York: Kluwer Academic / Plenum Publishers.
- Gupta, Nina, Jason D. Shaw, John E. Delery, 2000: Correlates of Response Outcomes among Organizational Key Informants; in: *Organizational Research Methods* 2000; 3; 323
- Hamermesh, Daniel, 2008: Fun with matched firm-employee data: progress and road maps; in: *Labour Economics*, Vol 15, Issue 4, p.662-672.
- Hartmann, Josef und Susanne Kohaut, 2000: Analysen zu Ausfällen (Unit-Nonresponse) im IAB-Betriebspanel, in: *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 33: 609-618.
- Hauptmanns, P., Rainer G. Saurwein, 1992: Probleme bei schriftlichen Betriebsbefragungen, *ZUMA-Nachrichten* 31, Jg. 16: 70-86.
- Jäckle, Annette, Emanuela Sala, Stephen P. Jenkins und Peter Lynn, 2004: Validation of Survey Data on Income and Employment. The ISMIE Experience. ISER Working Papers, Number 2004-14. Colchester.
- Kaiser, Ulrich & Joachim Wagner (2008): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten. *Perspektiven der Wirtschaftspolitik* 9(3): 329-349.
- Kalleberg, Arne L., David Knoke, Peter V. Marsden, Joe L. Spaeth (Hrsg.), 1996: *Organizations in America*. Thousand Oaks: Sage Publications.
- Kmec, Julie A., 2003: Collecting and Using Employer-Worker Matched Data. *Sociological Focus* 36: 81-96.
- Lang, Reinhard, 2005: Organizational Survey. S. 85-114 in Stefan Kühl, Petra Strodtholz und Andreas Taffertshofer (Hrsg.), *Quantitative Methoden der Organisationsforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lee, Valerie, 2000: School size and the organization of Secondary schools, in: Hallinan, Maureen ed.: *Handbook of the sociology of education*, New York: Kluwer Academic / Plenum Publishers.

- Lengfeld, Holger (2008): Klasse, Organisationsstruktur und soziale Ungleichheit. Habilitationsschrift Berlin: Freie Universität Berlin.
- Lynn, P. Sala, E., 2004: The Contact and Response Process in Business Surveys: Lessons From a Multimode Survey of Employers in the UK. ISER Working Paper, Number 2004-12
- Phipps, P.A., Butani, S.J., Chun, Y.I., 1995: Research on Establishment-Survey Questionnaire Design. In: Journal of Business & Economic Statistics, July 1995, Vol.13, No.3, pp. 337-346.
- Reimer, M. und R. Kuenster, 2004: Linking Job Episodes from Retrospective Surveys and Social Security Data: Specific Challenges, Feasibility and Quality of Outcome. Arbeitspapier Nr. 8 des Projekts Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Berlin: Max-Planck-Institut für Bildungsforschung.
- Riedmann, Arnold, Harald Bielenski; Teresa Szczurowska und Alexandra Wagner, 2006: Working time and work-life balance in European companies. Dublin: European Foundation for the Improvement of Living and Working Conditions.
- Rosemann, Martin, 2006: Auswirkungen datenverändernder Anonymisierungsverfahren auf die Analyse von Mikrodaten. Institut für Angewandte Wirtschaftsforschung.
- Schmidt, Werner, 2007: Arbeitsbeziehungen und Sozialintegration in Industriebetrieben mit Beschäftigten deutscher und ausländischer Herkunft. Industrielle Beziehungen 14: 334-356.
- Schnabel, Annette, 1997: Teilnahmeverhalten bei Unternehmensbefragungen, in: Arbeit, Heft 2, Jg. 6 (1997), S.154-172
- Smith, Tom W., Arne L. Kalleberg, Peter V. Marsden, 2004: National Organizations Survey (NOS), 2002. Ann Arbor: ICPSR.
- Spaeth, Joe L. und Diane P. O'Rourke, 1996: Design of the National Organizations Study. In Arne L. Kalleberg, David Knoke, Peter V. Marsden and Joe L. Spaeth: Organizations in America, Thousand Oaks, Sage: 23-44.
- Stephan, Gesine, 2001: The Lower Saxonian Salary und Wage Structure Survey-Linked Employer-Employee Data from Official Statistics. Schmollers Jahrbuch 121: 267-274.
- Struck, Olaf 2005: Betrieb und Arbeitsmarkt. In: Martin Abraham und Thomas Hinz: Arbeitsmarktsoziologie. Wiesbaden: VS-Verlag für Sozialwissenschaften, S. 169-198.
- Tarnai, J., Paxson, M.C., 2004: Survey Mode Preferences of Business Respondents.
<http://www.amstat.org/Sections/Srms/Proceedings/>
- Tomaskovic-Devey, D., Leiter, J., Thompson, S., 1994: Organizational Survey Response. In: Administrative Science Quarterly 39: 439-457.
- Troske, Kenneth R., Kimberly Bayard, Judith Hellerstein, David Neumark, 2000: The New Worker Establishment Characteristics Database. Proceedings of the Second International Conference on Establishment Surveys. Alexandria, VA: American Statistical Association, 2000: 981-990
- Wharton, Amy, 1994: Structure and Process: Theory and Research on Organizational Stratification. In: Knotterus, D., Prendergast C. (eds.), Current Perspectives in Social Theory. 119-148
- Willimack, D.K., Dalzell, J.L., 2006: An Examination of Non-Contact as a Source of Nonresponse in a Business Survey. <http://www.amstat.org/Sections/Srms/Proceedings/>
- Wirth, Heike, 2003: Angriffsszenarien auf wirtschaftsstatistische Einzeldaten - ein Überblick, S. 11-24 in: Forum der Bundesstatistik, Band 42. Stuttgart: Metzler-Poeschel.

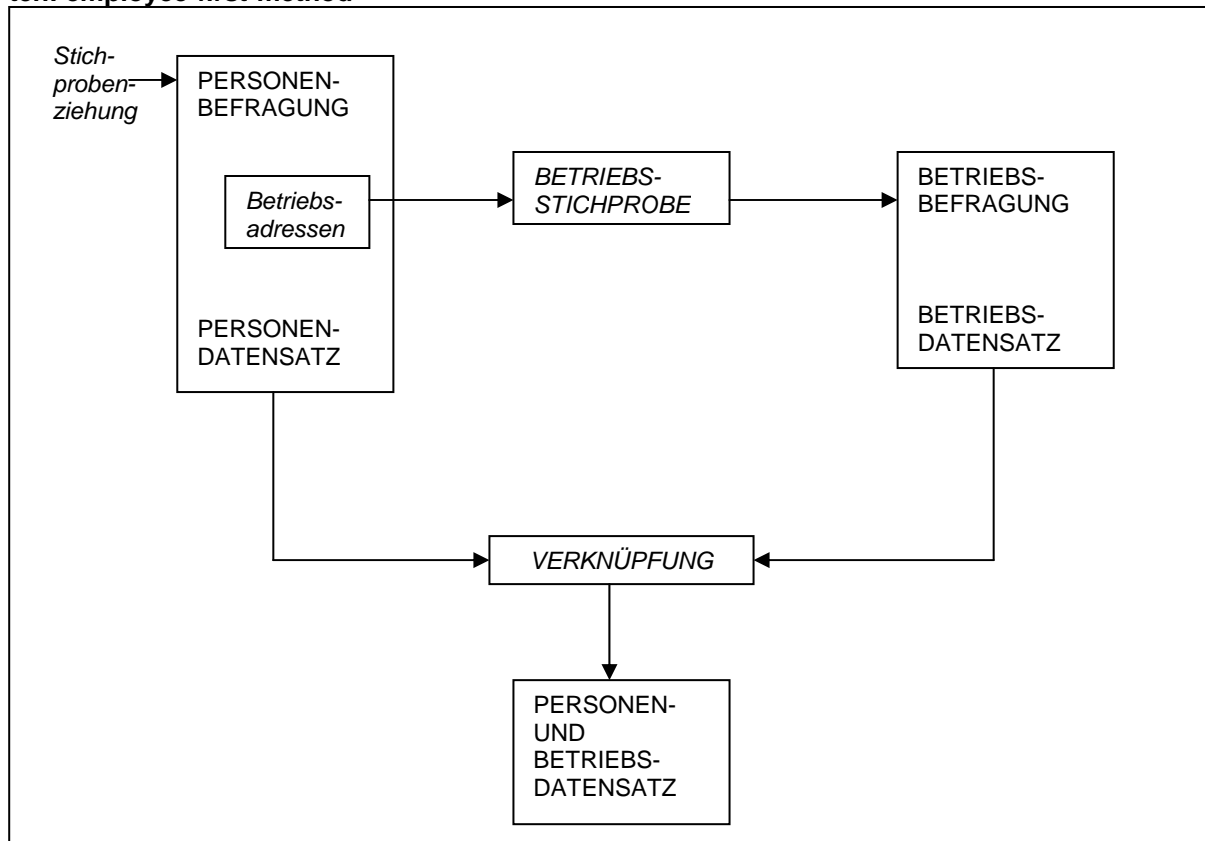
7 Anhang

Anhang 1: Informationstext, der im Falle einer Antwortverweigerung vorgelesen wurde:

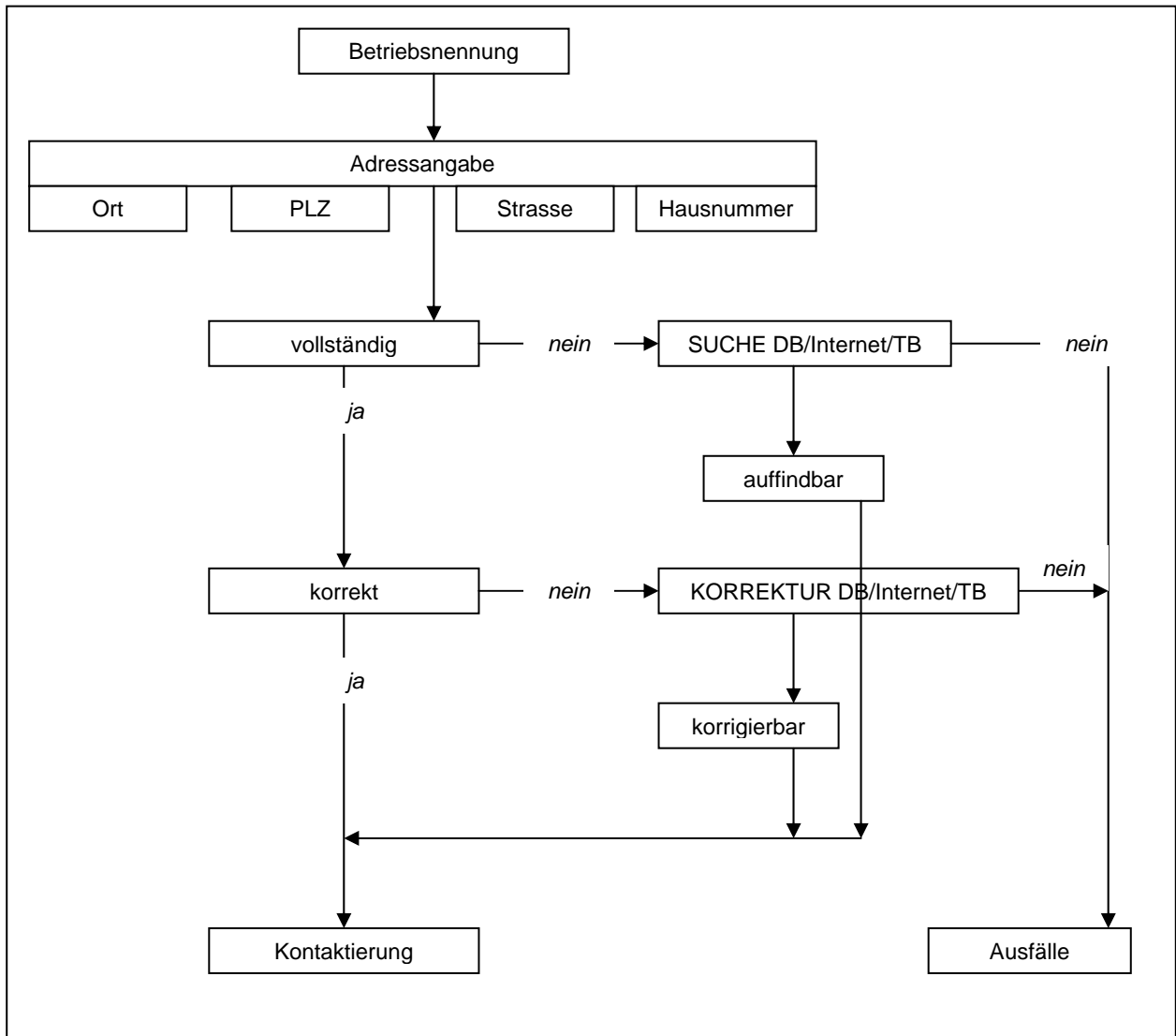
Ziel dieser Studie ist es, einen umfassenden statistischen Einblick in die Einkommens- und Beförderungschancen in Deutschland zu erhalten. Dabei interessiert es nicht, welche konkreten Einkommens- oder Beförderungschancen eine einzelne Person hat, sondern allein wie sich die Situation für unterschiedliche Berufsgruppen insgesamt darstellt. Dazu soll in einer weiteren Erhebung eine Unternehmensbefragung durchgeführt werden. Die Betriebe werden dazu vom renommierten Deutschen Institut für Wirtschaftsforschung (DIW Berlin) kontaktiert und nach ihrer ökonomischen Situation, ihrer Personalpolitik sowie den bestehenden Regelungen zur Entlohnung und Beförderung befragt.

Dabei ist einmal sichergestellt, dass Ihr Name im Zuge der Unternehmensbefragung nicht genannt wird. Ihr (früherer)Arbeitgeber wird also in keinem Fall erfahren, dass wir über Sie Kontakt mit ihm aufgenommen haben. Zweitens werden in dieser Unternehmensbefragung Ihrem (frühere)Arbeitgeber keinerlei Information gegeben, die ihm Rückschlüsse auf Ihre Person ermöglichen könnten. Sie könnte also absolut sicher sein, dass Ihre Anonymität in jedem Fall gewahrt bleiben wird.

Anhang 2: Prozess der Generierung und Bereitstellung verknüpfter Personen- und Betriebsdaten: employee-first-method



Anhang 3: Prozess der Edition der Betriebsadressen



Anhang 4: Abfrage des Betriebskontakts im SOEP-Pretest 2007

Split 1:

Ziel dieser Studie ist es, einen umfassenden statistischen Einblick in die Einkommens- und Beförderungschancen in Deutschland zu erhalten. Dabei interessiert es nicht, welche konkreten Einkommens- oder Beförderungschancen eine einzelne Person hat, sondern allein, wie sich die Situation für unterschiedliche Berufsgruppen insgesamt darstellt. Dazu soll in einer weiteren Erhebung eine Unternehmensbefragung durchgeführt werden. Die Betriebe werden dazu vom renommierten Deutschen Institut für Wirtschaftsforschung [DIW Berlin] kontaktiert und nach ihrer ökonomischen Situation, ihrer Personalpolitik sowie den bestehenden Regelungen zur Entlohnung und Beförderung befragt. Dabei ist einmal sichergestellt, dass Ihr Name im Zuge der Unternehmensbefragung nicht genannt wird. Ihr (*früherer*) Arbeitgeber wird also in keinem Fall erfahren, dass wir über Sie Kontakt mit ihm aufgenommen haben. Zweitens werden in dieser Unternehmensbefragung Ihrem (*früheren*) Arbeitgeber keinerlei Informationen gegeben, die ihm Rückschlüsse auf Ihre Person ermöglichen könnten. Sie können also absolut sicher sein, dass Ihre Anonymität in jedem Fall gewahrt bleiben wird.

68. Wie heißt die Firma, die Behörde, bei der Sie derzeit beschäftigt sind (*zuletzt beschäftigt waren*)?

Befragter verweigert die Auskunft und zwar weil:

69. Wo befindet sich die Arbeitsstelle (*bei der Sie zuletzt beschäftigt waren*)?

Ort:

Postleitzahl:

Straße:

Befragter verweigert die Auskunft und zwar weil:

Split 2:

68. Wie heißt die Firma, die Behörde, bei der Sie derzeit beschäftigt sind (*zuletzt beschäftigt waren*)?

Nur falls Auskunft verweigert, folgt der Infotext. Sonst weiter mit Frage 69!

68a. Wären Sie eventuell doch bereit, den Namen der Firma, der Behörde, bei der Sie derzeit beschäftigt

sind (*zuletzt beschäftigt waren*) zu nennen?

Ja Nein *Sie springen auf Frage 70/74!*

69. Wo befindet sich die Arbeitsstelle (*bei der Sie zuletzt beschäftigt waren*)?

Ort:

Postleitzahl:

Straße:

Anhang 5: Fragebogen der nachgelagerten Betriebsbefragung
[Betriebsfragebogen_soeppretest2007.pdf](#)