

Inhalt

315 Editorial

FORSCHUNGSBERICHTE

- 317 Geographische Informationssysteme (GIS) und ihre Anwendung in den Sozialwissenschaften am Beispiel des Schweizer Umweltsurveys
Reto Meyer und Heidi Bruderer Enzler
- 347 Plagiate in studentischen Arbeiten
Eine empirische Untersuchung unter Anwendung des Triangular Modells
Julia Jerke und Ivar Krumpal
- 369 Zur Multiplen Imputation bei Item Non-response am Beispiel eines Analysemodells zum Abstimmungsverhalten bei Bürgerentscheiden
Martin Messingschlager und Marcel Preisig
- 397 Soziale Präsenz in Onlineumfragen
Alexander W. Schmidt-Catran und Katharina Hörstermann
- 433 „Kann es sein, dass ich das Beginndatum falsch erfasst habe?“
Reactive Dependent Interviewing bei der Erhebung von Ausbildungs- und Erwerbsverläufen junger Erwachsener
Melania Rudin und Christoph Müller

465 Autorinnen, Rezensentinnen, Gutachterinnen 2013

467 Information for Authors

Editorial

Liebe Leserin, lieber Leser,

mit dem dritten Heft der mda in diesem Jahr halten Sie die letzte deutschsprachige Ausgabe der Zeitschrift in den Händen. Wie bereits in Heft 1/2013 geschrieben, haben wir uns entschlossen, die mda beginnend mit dem achten Jahrgang 2014 vollständig auf englische Sprache umzustellen. Damit folgen wir der Entwicklung der Sozialwissenschaften, insbesondere aber der Survey Methodology, hin zu einer internationalisierten Wissenschaft.

Durch den Wandel der neuen „methods, data, analyses“ zu einer internationalen Zeitschrift erschließen wir einen neuen Autorenkreis und können Ihnen als Leserin und Leser in Zukunft ein noch breiteres Spektrum an Artikeln zur empirischen Sozialforschung anbieten. Gleichzeitig öffnen wir uns aber auch einem neuen, internationalen Leserkreis und machen so Ihre Forschung auch über die Sprachgrenzen hinaus zugänglich. Natürlich bieten wir auch weiterhin ein Forum für nationale Probleme der Surveyforschung, denn Umfragen finden immer in einem lokalen Kontext statt, der angemessen berücksichtigt werden muss.

Ich freue mich auf die weitere Entwicklung der mda ab 2014, auf Ihre Einreichungen, auf spannende Artikel und Diskussionen und einen erweiterten Leser- und Autorenkreis!

Mannheim, Dezember 2013

HENNING BEST

Geographische Informationssysteme (GIS) und ihre Anwendung in den Sozialwissenschaften am Beispiel des Schweizer Umweltsurveys

Geographic Information System (GIS) and its Application in the Social Sciences using the Example of the Swiss Environmental Survey

Reto Meyer und Heidi Bruderer Enzler

Zusammenfassung

Dem räumlichen Kontext wurde in den deutschsprachigen Sozialwissenschaften bislang nur beschränkt Beachtung geschenkt. Da geographische Informationssysteme (GIS) den Aufwand für die Erhebung und Analyse raumbezogener Daten erheblich verringern, unterstützen sie die Untersuchung einer Vielzahl verschiedener Forschungsfragen. Die große Bandbreite geografischer Funktionen ermöglicht das Erstellen ansprechender kartografischer Produkte, die das Entdecken räumlicher Muster erleichtern, und das Errechnen raumbezogener Prädiktoren, die in weitere Analysen einfließen können. Im ersten Teil des vorliegenden Artikels werden die Datenerfassung und -beschaffung sowie die Analyse-Grundfunktionen und Ausgabemöglichkeiten von GIS vorgestellt. Am Beispiel der Untersuchungen zur Umweltgerechtigkeit im Schweizer Umweltsurvey wird anschließend eine spezifische Anwendung beschrieben.

Abstract

To date, spatial context has only been taken into account very limitedly by German research in social sciences. As Geographic information systems (GIS) reduce the effort of collecting and analyzing spatial data substantially, they allow exploring a whole range of research questions. They facilitate the creation of visually appealing cartographic products that may help uncovering spatial patterns, and the computation of space-related predictors that can be used for further analyses. The first part of this article covers the topic of data acquisition as well as basic analytic functions and output capabilities of GIS. Afterwards, these are illustrated in a study of environmental justice based on the Swiss Environmental Survey.



1 Einleitung¹

Das Sozio-oekonomische Panel (SOEP) ist eine unentbehrliche Datenquelle der deutschen Sozial- und Verhaltensforschung, die es erlaubt, den sozialen Wandel in Deutschland seit Mitte der 1980er Jahre zu beschreiben und Veränderungsprozesse auf Basis von Mikrodaten zu analysieren.² Die Existenz dieser und ähnlicher Längsschnittdaten in anderen Ländern, beispielsweise die „Panel Study of Income Dynamics“ in den USA oder der „British Household Panel Survey“, legen den Schluss nahe, dass der zeitlichen Dimension in den Sozialwissenschaften allgemein große Bedeutung zugeschrieben wird. Dies geschieht unter anderem in der Hoffnung, durch das Verfolgen von Personen über die Zeit kausale Beziehungen aufdecken zu können. Der räumliche Kontext dagegen wurde in vielen sozialwissenschaftlichen Studien bislang nur marginal berücksichtigt, obwohl Opportunitätsstrukturen vor Ort das individuelle Handeln oft maßgeblich beeinflussen – sowohl aus theoretischer als auch empirischer Sicht (siehe Franzen 1997; Brimblecombe/Dorling et al. 1999; Shaw/Dorling et al. 1999; Best 2009). Wird dem räumlichen Kontext dennoch Beachtung geschenkt, so geschieht dies bisher in der Regel durch das Verwenden von Indikatoren für räumliche Zugehörigkeiten und Eigenschaften wie die Urbanität der Wohngemeinde, Sprachgebiete oder Stadtkreise. Während derartige Variablen durchaus ihre Berechtigung haben, kränken sie daran, dass sie sich in der Regel auf großflächige und/oder auf administrative Gebiete beziehen. Diese Grenzbeziehungen entsprechen nicht der lebensweltlichen Erfahrung der Menschen. Für diese ist wenig relevant, wo ein Stadtkreis endet, sondern vielmehr wo die nächsten Einkaufsmöglichkeiten, Arbeitsplätze oder Freizeiteinrichtungen sind. Geographische Informationssysteme (GIS) erlauben es, nicht nur mit den altbekannten Indikatoren zu arbeiten, sondern andere räumliche Einheiten wie Nachbarschaften zu berücksichtigen. So kann Befragten mit Hilfe eines geographischen Informationssystems beispielsweise ein (gewichteter) Mittelwert ihrer räumlichen Umgebung

- 1 Der Schweizer Umweltsurvey 2007 wurde durch den Schweizerischen Nationalfonds gefördert (Projekt: 100012-107835). Das Projekt wurde vom Bundesamt für Umwelt (BAFU), den Kantonen Basel-Stadt (Amt für Umwelt und Energie) und Zürich (Amt für Abfall, Wasser, Energie und Luft), den Zentralschweizer Kantonen (Umweltämter) und dem Umwelt- und Gesundheitsschutz der Stadt Zürich unterstützt. Ferner wurde das Projekt finanziell und bezüglich der Stichprobenauswahl vom Bundesamt für Statistik (BFS) gefördert. Das BFS hat keinen Einfluss auf die verwendete Methode und die daraus resultierenden Ergebnisse. Herzlicher Dank gilt Ulf Liebe und Christa Mühlemann für Anregungen und die Durchsicht des Aufsatzes und Kirk Ingold für die Unterstützung bei der Zuschreibung der Gebäudelärmdaten.
- 2 Eine Übersicht zu aktuellen Studien auf Basis der SOEP-Daten findet sich hier: <http://www.diw.de/soep> (07.06.2012).

zugeschrieben werden. Daneben eröffnen sich neue Möglichkeiten, Reisewege oder Entwicklungen in der Raumnutzung zu erfassen und zu analysieren.

Diese Art der Berücksichtigung raumbezogener Daten und deren kartografische Darstellung waren bis vor wenigen Jahren sehr aufwändig. So wurde beispielsweise im DEFECT-Projekt während der Stichprobenziehung mittels Random-Route-Verfahren einerseits für alle Adressen ein Beobachtungsbogen mit Angaben zur Klassifikation des Gebäudes und zu Umgebungsmerkmalen (z.B. Wald) ausgefüllt, andererseits wurden sämtliche begangenen Straßen für eine Nachkodierung gefilmt oder fotografiert (Schnell/Kreuter 2000). Zwar weniger zeitintensiv, aber vermutlich auch weniger genau dürfte die direkte Abfrage von räumlichen Gegebenheiten mittels eines Fragebogens sein, z.B. Informationen zur Wohnumgebung (Preisendörfer 1999) oder Abschätzungen von Distanzen (Franzen 1997).

Dank technischer und methodischer Fortschritte in den Geowissenschaften steht mit den geographischen Informationssystemen ein Instrument zur Verfügung, das die Arbeit zur Berücksichtigung und Analyse lokaler Begebenheiten stark vereinfacht und den Zeitaufwand erheblich verringert (Okabe 2006b; Goodchild 2010a). Ein Blick in die Literatur (z.B. Rigaux/Scholl et al. 2002; Meyer 2010) zeigt zwar, dass die Grundlagen von GIS nicht ganz einfach zu verstehen sind; für eine Mehrheit der Forschenden stellen diese Systeme aber dennoch eine erhebliche Erleichterung dar. So wie viele empirische Sozialforscher/-innen die Likelihood-Schätzung lediglich im einführenden Methodenkurs versuchen nachzuvollziehen, und für deren konkrete Anwendung lieber auf ein Statistikprogramm vertrauen, kann auch GIS als Werkzeug angesehen werden, das es erlaubt, vormals aufwändige und teilweise nur mit sehr spezifischem Fachwissen mögliche Analysen durchzuführen. Unerlässlich für eine korrekte Anwendung ist jedoch ein grundsätzliches Verständnis der entsprechenden Vorgänge und Ergebnisse in GIS; werden komplexere Verfahren eingesetzt, so ist schnell mehr Fachwissen nötig und es ist gegebenenfalls ratsam, GIS-Experten beizuziehen.

Noch finden GIS in den Geistes- und Sozialwissenschaften kaum Verwendung. Neuere Lehrbücher (Steinberg/Steinberg 2006; Lai/Mak 2007; DeMers 2009b; ESRI 2009a; Bodenhamer/Corrigan et al. 2010; Lloyd 2010), wissenschaftliche Aufsätze sowie Tagungen (beispielsweise das GeoForum 2012: „Geoinformation für den gesellschaftlichen Wandel neue Wertschöpfung mit Sozial- und Wirtschaftsdaten“) zur Anwendung von GIS in diesen Disziplinen deuten jedoch auf eine zunehmende Verbreitung des Instruments in Nicht-Geowissenschaften hin.

Einige ökonomisch orientierte Studien setzen GIS zur Bewertung von Häusern und Wohnumgebungen ein (mittels hedonischer Wohnpreisfunktion; Baranzini/Ramirez 2005; Asami/Gao 2006) oder zur Bestimmung optimaler Standorte für

öffentliche Anlagen und andere Einrichtungen (Germann/Seeland 2004; Kohsaka/Sekine 2006; Sorrentino/Meenar et al. 2008). In der Soziologie wird GIS vermehrt für die Untersuchung räumlicher Disparitäten und der Diskriminierung sozialer Gruppen (z.B. Hermann/Heye et al. 2005; Downey 2006; Baranzini/Caroline et al. 2008; Crowder/Downey 2010) oder zur Illustration sozio-struktureller Veränderungen eingesetzt (Schuler/Dessemontet et al. 2007). Gesundheitswissenschaftler/-innen berechnen die Erreichbarkeit von Anbietern medizinischer Versorgung mit Hilfe von GIS (vgl. Übersicht in Higgs 2004) oder setzen GIS ein, um die Ausbreitung von Umweltbelastungen (z.B. Briggs 2005; Heinrich/Gehring et al. 2005) und Krankheiten zu modellieren (Yang/Peng et al. 2007; Gómez-Rubio/Ferrándiz-Ferragud et al. 2009; Niyonsenga/Courteau et al. 2009). In der Kriminalitätsanalyse hofft man, durch GIS den Aktionsradius von Tätern eingrenzen zu können und ihnen dadurch eher auf die Schliche zu kommen (Spencer/Ratcliff 2005; van Schaaik/van der Kemp 2009). In der Geschichtswissenschaft und Archäologie wird unter Anwendung von GIS beispielsweise versucht, auf der Basis von Karten und Bildern historische Stätten zu visualisieren (Shimizu/Fuse 2006) oder die Subsistenzstrategien und Umweltbedingungen in Verbindung mit einem bestimmten Lebensstil zu bestimmen (u.a. Größe der Lebensgemeinschaften und Häuser; Tsumura 2006).³

Diese Beispiele verdeutlichen, wie vielfältig die Anwendungsgebiete von GIS bereits geworden sind. Mit dem vorliegenden Artikel möchten wir einen Beitrag dazu leisten, GIS für die deutsche Sozialwissenschaft zugänglicher zu machen und den Blick für potenzielle Anwendungen zu schärfen. Nach einem einführenden, allgemeinen Abschnitt zum Aufbau von GIS, zur Datenbeschaffung, zu Analyse-Grundfunktionen und Ausgabemöglichkeiten wird am Beispiel der Untersuchungen zur Umweltgerechtigkeit im Schweizer Umweltsurvey eine spezifische Anwendung beschrieben. Insbesondere gehen wir dabei der Frage nach, weshalb der mit unseren Daten berechnete Anteil Lärmbelasteter in der Schweiz geringer ausfällt als die offizielle Quote des Bundesamtes für Umwelt und welche Auswirkungen eine Änderung der Datenbasis auf die Ergebnisse hat. Der Ausblick enthält Überlegungen zum künftigen Einsatz von GIS und weitergehenden Analysemöglichkeiten für Daten mit Raumbezug. Der Beitrag schließt mit einer kurzen Linkliste, die auf GIS-Software, Lehrbücher, Datenquellen und Webseiten zu GIS (in den Sozialwissenschaften) verweist.

3 Für Hinweise auf aktuelle Beiträge vgl. <http://gisandscience.com/category/social-science> (07.06.2012)

2 Was ist ein geographisches Informationssystem?

Entgegen der vorherrschenden Meinung Mitte des 19. Jahrhunderts war der englische Arzt John Snow (1813–1858) der Ansicht, Cholera werde durch damals noch nicht entdeckte Organismen im Trinkwasser verursacht und nicht durch giftige Dämpfe aus Sümpfen, offenen Gräbern und Abfallgruben hervorgerufen (Frerichs 2001). Seine Hypothese versuchte Snow anhand verschiedener Untersuchungen zu stützen, u.a. mit einem Kartenausschnitt von London, in den er die Wohnadresse jedes Cholera-Toten einzeichnete. So konnte er zeigen, dass viele erkrankte Personen in der Nähe einer bestimmten Wasserpumpe wohnten oder dorthin kamen, um sich mit Wasser zu versorgen. Die Cholera-Rate des nahe gelegenen Armenhauses, das einen eigenen Brunnen hatte, war deutlich geringer, ebenso unter den Angestellten der lokalen Brauerei, da diese meist überhaupt kein Wasser tranken (Snow 1855; Frerichs 2001). Snow, der auch als Vater der (modernen) Epidemiologie gilt,⁴ dürfte einer der Ersten gewesen sein, der wissenschaftliche Untersuchungen durchgeführt hat, wie sie heute in einem geographischen Informationssystem üblich sind. Er hat verschiedene Arten von Daten (Brunnen, Todesfälle) über ihren jeweiligen Standort im Raum (Straßenkreuzung, Adresse des Wohnhauses) miteinander in Beziehung gesetzt und daraus neue Informationen abgeleitet.

In einem GIS können – wie auch in anderen computergestützten Informationssystemen (z.B. einer Datenbank oder einem Statistikprogramm) – Informationen erfasst, verwaltet, analysiert und ausgegeben werden. Die Stärke eines GIS liegt in den Verarbeitungs- und Präsentationsmöglichkeiten für räumliche Daten⁵. Diese zeichnen sich durch die Information zur Lage in einem Bezugssystem (Koordinatensystem) und zur geometrischen Form der Objekte aus (Carosio 2005; Okabe 2006b; Steinberg/Steinberg 2006; DeMers 2009a; Bill 2010; Lloyd 2010). Auf der Basis derartiger Informationen lassen sich Aussagen zu räumlichen Beziehungen zwischen Objekten ableiten (Topologie, beispielsweise Nachbarschaft oder Überlappung). Zudem können Untersuchungsobjekte nicht nur, wie in normalen Daten-

4 Weitergehende Informationen zu John Snow und seinen Untersuchungen (z.B. „Grand Experiment“) sowie die beschriebene Karte sind auf der John Snow-Webseite des Departments für Epidemiologie der UCLA zu finden: <http://www.ph.ucla.edu/epi/snow.html> (07.06.2012). Weitere Beispiele früher räumlicher Analysen sind in Steinberg/Steinberg (2006) beschrieben, z.B. zur sozialen Ungleichheit in den Slums von Chicago um 1893 oder die Verwendung des Bahnnetzes als Indikator für die Verbreitung der „Zivilisation“ um 1923.

5 Obwohl manche Autoren (Carosio 2005; Okabe 2006b) zwischen geografischen und räumlichen Daten unterscheiden, werden diese Begriffe im vorliegenden Beitrag als Synonyme verwendet.

banken gängig, über ein Schlüsselattribut (z.B. Familiennummer, Postleitzahl) verbunden werden, sondern auch über ihre Position im Raum.

Ein GIS, wie es für wissenschaftliches Arbeiten in der Regel verwendet wird, ist meist ein lokal installiertes Computerprogramm; beispielsweise die kommerzielle Software ArcGIS von ESRI. In den letzten Jahren wurden auch im Open-Source-Bereich verschiedene GIS-Pakete – beispielsweise Quantum GIS oder GRASS GIS – derart weiterentwickelt, dass sie mittlerweile für viele Anwendungszwecke gute Alternativen zu den käuflichen Programmen darstellen.⁶

Wie auch bei Statistikanwendungen müssen in dieses System vor einer Analyse Daten eingebunden werden. Dies bedeutet entweder, dass ein Datensatz lokal vorliegt, oder aber es wird ein Datensatz über das Internet lediglich eingebunden, nicht aber lokal gespeichert. Die Daten selbst sind typischerweise so organisiert, dass verschiedene Daten-Layer (Kartenschichten) übereinander gelegt und je nach Bedarf gemeinsam oder einzeln betrachtet und analysiert werden (ESRI 2009b). Die Basis bildet oft eine Grundlagenkarte, z.B. Landes- oder Gemeindegrenzen, auf die unterschiedliche thematische Karten gelegt werden, z.B. eine Straßenkarte, eine Karte mit Standorten von Bildungsstätten oder mit Informationen zum Anteil der Personen im Rentenalter.

Im Folgenden wird auf die in der Definition von GIS bereits erwähnten Prozesse der Erfassung oder Beschaffung von Daten sowie deren Verwaltung, Analyse und Ausgabe eingegangen.

2.1 Datenerfassung und Datenbeschaffung

Analog zur Differenzierung zwischen Primär- und Sekundärdaten in den Sozialwissenschaften können auch GIS-Daten selber erhoben werden oder man verwendet bereits erfasste Daten. Für eigene Datenerhebungen stehen heute neben der Vermessung vor Ort eine ganze Bandbreite unterschiedlicher Techniken zur Verfügung: z.B. Photogrammetrie (Satelliten- oder Luftbilder), Laser-Scanning aus Flugzeugen oder die Digitalisierung bereits vorliegender, analoger Karten (DeMers 2009a; Bill 2010). Neben weiteren Attributen zur Beschreibung der zu erfassenden Objekte sind deren räumlichen Eigenschaften zu ermitteln – die Geometrie sowie die Lage

6 Im Allgemeinen wird davon ausgegangen, dass sich die Erweiterung des Marktes um Open-Source-Produkte insofern positiv auf die Weiterentwicklung der kommerziellen Angebote auswirkt, als der Innovationsdruck steigt Mundhenke (2005). Mögliche negative Folgen liegen in einer vermehrten Implementierung proprietärer Schnittstellen (Brügge/Harhoff et al. 2004) – was im Falle der GIS-Anwendungen der aktuellen Forderung nach Interoperabilität entgegenläuft.

im Bezugssystem (Geokodierung oder Georeferenzierung). Aufzeichnungen aus einem Flugzeug oder Satelliten können beispielsweise aufgrund der Position und Ausrichtung des Aufnahmegeräts umgerechnet werden (DeMers 2009a; Bill 2010). Bei einer Erfassung vor Ort bildet das Global Positioning System (GPS) die Basis.

Liegen Objekte als Listen vor, z.B. Postanschriften, so können diese nachträglich georeferenziert werden, indem mittels einer Datenbank XY-Koordinaten zugeschrieben werden. Heute stehen teilweise geokodierte amtliche Adressdaten zum Abgleich zur Verfügung (siehe Anhang); alternativ liegen beispielsweise durch Anbieter von Online-Karten wie z.B. „Google Maps“ weitere Geokodierungsmöglichkeiten vor. Ohne vertieftes Wissen muss bei diesen Alternativen aber in der Regel jede Adresse einzeln eingegeben und ausgelesen werden. Es kann an dieser Stelle nicht beurteilt werden, wie genau und aktuell die Basisdaten dieser Kartendienste sind.

Voraussetzung für eine saubere Geokodierung über die Anschrift sind zunächst einmal korrekte Adressangaben, mittels derer die Koordinaten zugeschrieben werden. Sind diese falsch oder ungenau, können sie in der Datenbank nicht gefunden werden. Zwar können dank sogenannter Fuzzy-Algorithmen mittlerweile auch Adressen mit Schreibfehlern identifiziert werden, aber beispielsweise im Falle einer mündlichen Befragung lohnt es sich dennoch, die Interviewer entsprechend zu sensibilisieren. Sind Adressen trotzdem nicht direkt geokodierbar, so können als Approximation die Koordinaten eines bekannten, nahe gelegenen Objektes verwendet werden. Dies kann beispielsweise das nächstgelegene Haus, der Mittelpunkt der Wohnstraße oder des Postleitzahl-Gebietes sein.

Da die Frage der Güte der zu referenzierenden Adressen bedeutsam ist, möchten wir an dieser Stelle kurz auf einige Aspekte der Gewinnung der Adressen im Rahmen von Umfragen eingehen: Soll eine Adresse in einer guten Qualität telefonisch erhoben werden, so muss diese den Befragten geläufig sein, wie zum Beispiel die eigene Wohnadresse. Ist dies nicht der Fall (z.B. Arbeitsort oder Wohnort von Freunden), so ist eine schriftliche Befragung – auf Papier oder als Websurvey – vorzuziehen. Dies ermöglicht es den Befragten, bei Bedarf die genauen Angaben nachzusehen. Einen Schritt weiter ging das Schweizer Bundesamt für Statistik beim Mikrozensus 2005 zum Verkehrsverhalten in der Schweiz (Bundesamt für Statistik/Bundesamt für Raumentwicklung et al. 2007): In dieser telefonischen Erhebung wurden alle von den Befragten an bestimmten Tagen aufgesuchten Orte (z.B. Einkaufszentren, gastronomische Einrichtungen, Post, ÖV-Angebote oder Straßen) noch während des Telefoninterviews geokodiert (Bundesamt für Statistik/Bundesamt für Raumentwicklung et al. 2007). Hierfür stand den Interviewern eine GIS-Karte zur Verfügung, auf der sie die angegebenen Orte suchen konnten.

Die Karte ermöglichte es, gezielt Rückfragen zu stellen, beispielsweise bei ungenauen Ortsangaben. Durch Anklicken des Objektes (Haus, Haltestelle, etc.) auf der Karte konnten auch Orte ohne genaue Adressangabe geokodiert werden. Jedoch übersteigt die Implementierung eines solchen Systems und die Nutzungsrechte der hierfür notwendigen Daten das Budget der meisten Studien deutlich.

Da der Aufwand und damit auch die Kosten für direkte Erhebungen beträchtlich sind, wird oft auf bereits vorliegende Geodaten zurückgegriffen (für Verweise siehe Anhang). Hier stellt sich die Frage der Auffindbarkeit und des Zugangs zu Geodaten. In den letzten Jahren sind diesbezüglich große Veränderungen im Gange. Durch das Internet liegen Möglichkeiten vor, Geodaten sowie deren Metadaten ohne großen Aufwand zur Verfügung zu stellen. Metadaten sind Angaben zu Eigenschaften eines Datensatzes, wie beispielsweise die räumliche Ausdehnung und Auflösung, die Thematik, der Datenhalter oder Bezugsquellen. Bereits heute erleichtern Metadatenkataloge das Auffinden zumindest amtlicher Daten. Dennoch muss oftmals in den einzelnen Kommunen oder Bundesländern bei den zuständigen Stellen einzeln nachgefragt werden, ob Daten vorliegen und welche Qualität diese aufweisen.

Derzeit wird weltweit in vielen Ländern am Aufbau von nationalen Geodateninfrastrukturen gearbeitet, welche primär Geodaten der öffentlichen Hand erfassen. Das zentrale Element einer solchen Struktur ist ein Metadatenkatalog. In der EU wurde 2007 eine Richtlinie erlassen (2007/2/EC), welche den Aufbau ebensolcher Geodateninfrastrukturen verlangt, so dass zurzeit in vielen europäischen Ländern entsprechende Prozesse im Gange sind (Verweise auf entsprechende Webseiten finden sich im Anhang). Diese nationalen Strukturen sollen im Rahmen der EU-Initiative INSPIRE (Infrastructure for Spatial Information in Europe) zusammengeführt werden.⁷

Ob Geodaten zugänglich sind und wenn ja, in welcher Form, hängt in den meisten Fällen vom Dateneigner ab, der in der Regel auch für ihre Erhebung verantwortlich war. Für wissenschaftliche Zwecke dürfte die Chance für eine unentgeltliche oder zumindest kostengünstige Nutzung amtlicher Daten auf der Basis eines entsprechenden Datennutzungsvertrages vergleichsweise hoch sein. Teilweise werden Daten zwecks Wahrung des Datenschutzes vor der Herausgabe aggregiert, insbesondere wenn es sich um Angaben zu Einzelpersonen handelt (Arbeitslosigkeit, Einkommen, Gesundheit etc.). So liegen dann nicht Angaben zu einzelnen Personen vor, sondern beispielsweise für Gebiete des Ausmaßes 100 x 100 Meter.

7 Webseite der Europäischen Kommission zu INSPIRE: <http://inspire.jrc.ec.europa.eu> (07.06.2012)

Zunehmend können Geodaten in sogenannten Web-GIS über das Internet eingesehen werden.⁸ Dazu ist in der Regel keine spezialisierte GIS-Software nötig, sondern es genügt ein Internetbrowser mit entsprechenden Plug-Ins. Web-GIS eignen sich nur bedingt für den Einsatz in der Forschung, da sich die Nutzungsmöglichkeiten auf die vom Web-GIS zur Verfügung gestellten Daten und Funktionen beschränken (meist Anzeigen, Zoomen, Ausdrucken). Eine Kombination mit selber erhobenen Daten oder komplexere Analysen sind in der Regel nicht möglich. Damit sind diese GIS eher für konsultative Zwecke geeignet, um einen ersten visuellen Eindruck zu gewinnen.

Für die Forschung bedeutsamer sind sogenannte „Web Map Services“ (WMS) und „Web Feature Services“ (WFS). Dies sind standardisierte Schnittstellen⁹, die mittels einer lokal installierten GIS-Software den Zugriff auf Daten über das Internet ermöglichen. Sie sollen den Zugang und damit die Weiternutzung von Geoinformationen erleichtern. Konkret bedeutet dies, dass über Web Map Services Karten als Rasterbilder in die lokale Software eingebunden werden. Bei einer Anfrage an einen WMS-Server können die Auflösung, der räumliche Bezug sowie weitere Aspekte der Ausgabe angegeben werden, so dass das für den Anwender generierte „Bild“ optimal seinen Zweck erfüllt. Bei Web Feature Services dagegen werden nicht statische Bilder eingebunden, sondern bearbeitbare Vektordaten (vgl. unten für eine Erläuterung zu Datentypen). Dadurch sind mit Daten, die durch ein WFS-Angebot eingebunden werden, auch eigentliche GIS-Analysen möglich. Verweise auf verschiedene WMS- und WFS-Angebote der deutschsprachigen Länder finden sich im Anhang.

Speziell hinweisen möchten wir an dieser Stelle auf den Web Map Service des OpenStreetMap-Projekts (OSM)¹⁰, welches unentgeltlich topographische Karten zur Verfügung stellt. Jedoch existieren bei diesen Daten bislang noch Unsicherheiten bezüglich ihrer Vollständigkeit und Genauigkeit, weil sie von einer Vielzahl von Freiwilligen gesammelt werden und keiner zentralen Überprüfung unterworfen sind. Ludwig/Voss et al. (2010) kommen mit Blick auf das OSM-Straßennetz in Deutschland zum Schluss, dass seine Lage relativ genau ist, die Vollständigkeit jedoch regional stark schwankt, insbesondere zwischen Stadt und Land. Für die

8 Beispiele: Viewer der Nationalen Geodatenbasis Deutschland (<http://www.geoportal.de>, 07.06.2012), Modul zur Beobachtung des Umweltzustandes in der Schweiz (<http://umweltzustand.admin.ch>, 07.06.2012).

9 Weitere Informationen zu Standards wie WFS und WMS finden sich auf der Webseite des Open Geospatial Consortium: <http://www.opengeospatial.org/standards> (07.06.2012)

10 <http://www.openstreetmap.org> (07.06.2012).

Forschung dürfte es lohnend sein, die OSM im Auge zu behalten, da die Datenmenge stetig zunimmt und die Qualität steigt.

2.2 Datentypen

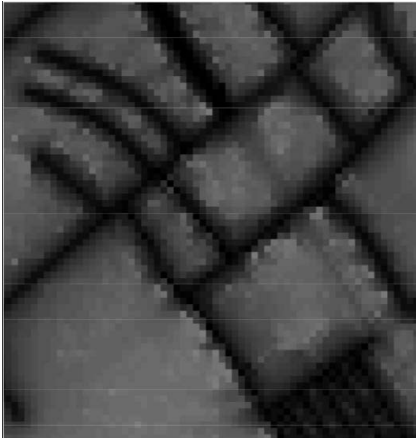
In einem GIS werden Daten in einer räumlichen Datenbank verwaltet, d.h. modelliert, strukturiert und gespeichert (Okabe 2006b). Diese kann, muss aber nicht, in die GIS-Software integriert sein. Die Verwaltungsmethoden sowie die Analysemöglichkeiten unterscheiden sich in Abhängigkeit des Datentyps (DeMers 2009a; Bill 2010).

Rasterdaten repräsentieren Merkmale (z.B. Lärmbelastung an einem bestimmten Standort) anhand von Pixeln, also Punkte oder Quadrate, die in einer Matrix aufgereiht sind. Bei der Betrachtung einer Bildaufnahme ist das Prinzip einfach verständlich: Ist das Bild stark vergrößert, sind einzelne Quadrate sichtbar (vgl. Abbildung 1, links), durch Verkleinern ergibt sich für das Auge aus gewisser Distanz betrachtet ein Gesamtbild. Jedem Pixel wird ein Attributwert zugeordnet, z.B. 60 für eine Lärmbelastung von 60dB(A) in einer Lärmkarte. Die Datenverwaltung geschieht demnach mittels einer Datenreihe von Zahlen, welche die Attributwerte und die Koordinaten der Pixel beschreiben (DeMers 2009a; Bill 2010).

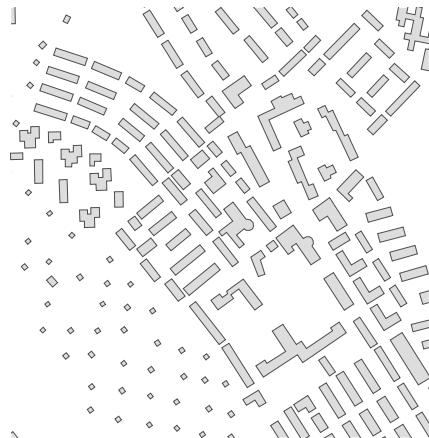
Vektordaten werden für den Betrachter durch Punkte, Linien(abschnitte) und Polygone (Vielecke) dargestellt (vgl. Abbildung 1, rechts). Abgespeichert sind diese geometrischen Elemente als Koordinaten von Punkten, die für sich stehen (Punkte) oder Start- und Endpunkte von Linienabschnitten sein können. Polygone sind als entgegen dem Uhrzeigersinn verbundene Punkte mit identischem Start- und Endpunkt abgelegt. Die Beziehungen zwischen diesen Objekten (z.B. welche Punkte eine Straße verbindet) und die Attribute derselben (z.B. Hauptstraße, Nebenstraße, Fußweg) sind in der Regel separat abgelegt. Die Verwaltung von Vektordaten ist damit anspruchsvoller als jene von Rasterdaten, da die topologischen, also die räumlichen Beziehungen zwischen Objekten von Interesse sind. Anhand solcher Daten lassen sich Fragen beantworten wie „Welche Linie schneidet mit einem gegebenen Linienabschnitt?“, „Welches Polygon schließt einen bestimmten Punkt ein?“ oder „Was ist die kürzeste Verbindung zwischen Punkt A und Punkt B?“ (DeMers 2009a; Bill 2010). In der sozialwissenschaftlichen Forschung könnten dies Fragen sein wie, welche Straßen ein Kind auf dem Weg zur Schule überqueren muss, im Einzugsgebiet welcher Stadt ein Befragter lebt oder ob jemand für den Arbeitsweg die kürzeste der möglichen Routen wählt.

Abbildung 1 Raster- und Vektordaten im Vergleich

Rasterdaten:



Vektordaten:



Quelle der Rasterdaten: SonBase, die GIS-Lärmdatenbank der Schweiz (BAFU 2009).

Quelle der Vektordaten: Gebäudekataster des Bundesamtes für Landestopografie (swisstopo 2011).

2.3 Analyse räumlicher Daten

Geographische Informationssysteme beinhalten Funktionen zur Analyse räumlicher und nicht-räumlicher Daten; für die Auswertung der nicht-räumlichen Daten stehen jedoch deutlich weniger Möglichkeiten zur Verfügung als in herkömmlichen Statistikprogrammen (DeMers 2009a). Zudem sind gerade diese Analysemethoden dem Leser bekannt, so dass sich die folgende Ausführung auf Operationen zur Manipulation räumlicher Daten beschränkt.

Ein vergleichsweise einfaches analytisches Verfahren ist das *kartographische Klassifizieren* (Steinberg/Steinberg 2006). Dadurch ist es möglich, eine anschauliche Übersicht zu bieten – zum Beispiel zur räumlichen Verteilung der Arbeitslosigkeit oder des Wähleranteils einer Partei. Je nach gewählter Kategorisierung fällt das Ergebnis anders aus, weshalb diese gut durchdacht sein sollte, wie dies auch in der „gewöhnlichen“ sozialwissenschaftlichen Datenanalyse der Fall ist.

Ein weiteres Set von Verfahren beschäftigt sich mit der *Messung geometrischer Größen*. Beispiele hierfür sind die Distanz zwischen zwei Punkten, die Länge einer Linie, die Fläche eines Polygons oder der Winkel einer Steigung (Okabe 2006b; Lloyd 2010). Anwendungsbeispiele wären die Berechnung der Distanz zwischen Wohn- und Arbeitsort oder einer Grundstücksfläche.

Zwei zentrale Funktionen, die zur Beantwortung einer großen Zahl an möglichen sozialwissenschaftlichen Fragen herangezogen werden können, sind das sogenannte „Buffern“ (Pufferzonen bilden) und die Überlagerung von Daten-Layern: Mit *Buffer-Funktionen* können neue Gebiete rund um ein Objekt ausgewiesen werden. Die zu wählende Distanz gibt die Ausdehnung des Gebietes an den Rändern des Objekts beginnend vor (Okabe 2006b; Lloyd 2010). Beispielsweise ergibt ein Buffer von 50 Metern um einen Funkmast (ein Punktobjekt) einen Kreis mit einem Durchmesser von etwa 100 Metern. Ein entsprechender Buffer um eine Autobahn oder ein Bahngelände (Linienobjekte) führt dagegen zu einer Trasse von gut 100 Metern Breite.

Bei einer *Überlagerung* („overlay“, Steinberg/Steinberg 2006) werden durch das Zusammenbringen von zwei Daten-Layern neue Informationen gewonnen. Geschieht dies nicht nur graphisch, sondern wird ein neuer Layer errechnet, so handelt es sich um eine „Verschneidung“. So könnte beispielsweise eine Karte der Einzugsgebiete von Schulen mit einem Liegenschaftskataster verschnitten werden. Ein anderes Beispiel wäre das Verschneiden einer Straßen- mit einer Landnutzungskarte. Im ersten Fall wären alle Liegenschaften im Einzugsgebiet bestimmter Schulen das Ergebnis, im zweiten Fall alle Straßen, die durch bestimmte Landnutzungszonen führen wie beispielsweise Wohngebiete.

Anschließend an eine Verschneidung (und ein eventuelles Buffern) kann zum Beispiel eine *räumliche Suche* durchgeführt werden. Dies ist der Überbegriff für Abfragen, die der Auswahl (Selektion) von Objekten dienen. Diese Operationen können sich auf die Thematik (Attribute), die Topologie (relative Lage von Objekten zueinander: Nachbarschaft, Überschneidung, Enthalten-Sein etc.) oder die Geometrie (Distanzen, Flächen etc.) der Objekte beziehen und werden oft mit Hilfe logischer Operatoren¹¹ ausgedrückt (Okabe 2006b; Lloyd 2010).

Eine mögliche Art der räumlichen Suche ist die Inklusions-Suche, mit der Punkte, Linien und Polygone gefunden werden können, die teilweise oder vollständig in einem gegebenen Polygon enthalten sind (Okabe 2006b; Lloyd 2010). Beispielsweise kann geprüft werden, welche Wohnhäuser im Einzugsgebiet bestimmter Schulen liegen, oder es kann innerhalb einer Region nach Arztpraxen, Flüssen oder Parkanlagen gesucht werden.

Die Distanz-Suche ermittelt Punkte, Linienabschnitte und Polygone, von denen jeweils zumindest ein Teil innerhalb einer gewissen Distanz von einem bestimmten geometrischen Element liegt (Okabe 2006b; Lloyd 2010). Diese Funk-

11 Beispiele: OR (Vereinigung: Straßen und/oder Wohnzonen), AND (Straßen in Wohnzonen) und NOT (Gegenstück: Wohnzonen ohne Straßen und Straßen außerhalb der Wohnzonen)

tion wird beispielsweise angewandt, wenn Ärzte gesucht werden, die innerhalb einer gewissen Minimaldistanz (z.B. 200 Meter) von einem bestimmten Punkt praktizieren, oder wenn alle Wohnhäuser gewählt werden sollen, die mehr als 300 Meter von der nächstgelegenen ÖV-Haltestelle entfernt liegen.

Die Schnittpunkt-Suche gibt Linienabschnitte oder Polygone aus, die mit einem gegebenen Element (Linie oder Polygon) schneiden, z.B. Flüsse, die mit einer Straße kreuzen, oder Schulwege, die ein Überqueren von stark befahrenen Straßen erfordern (Okabe 2006b; Lloyd 2010).

Bereits die Verknüpfung dieser Grundoperationen ermöglicht dem GIS-Anwender, Daten (Themenkarten) in vielfältiger Weise zu kombinieren. Für fortgeschrittene Analysemethoden, wie zum Beispiel die Modellierung von Oberflächen, die Analyse von Netzwerken oder räumliches Clustering sei an dieser Stelle auf weiterführende Literatur verwiesen (de Smith/Goodchild et al. 2009; DeMers 2009a).

2.4 Datenausgabe

Entsprechend dem Ausspruch „ein Bild sagt mehr als tausend Worte“ werden GIS-Daten oftmals als Grafik ausgegeben. Meist wird als Darstellungsform eine Karte gewählt, da dies gegenüber anderen Arten von Grafiken wie z.B. Balken- oder Kreisdiagrammen den Vorteil hat, dass räumliche Beziehungen explizit sichtbar werden (Goodchild/Janelle 2004). So gibt eine mit den jeweiligen Arbeitslosenraten versehene Karte der Bundesländer ein informativeres Bild als ein Balkendiagramm oder eine Tabelle, bei welchen der Betrachter zusätzliche Information zur Lage der einzelnen Bundesländer braucht, um z.B. Vergleiche mit benachbarten Bundesländern nachvollziehen zu können. Grundsätzlich ermöglicht die Darstellung als Karte eine sehr einfache und intuitive Art der räumlichen Datenanalyse (Goodchild/Janelle 2004). Hätte John Snow seine Cholera-Fälle nicht auf einer Karte illustriert, wären die Verantwortlichen der Stadt wohl kaum seiner Argumentation gefolgt und der Pump-Hebel des verseuchten Brunnens wäre womöglich nicht entfernt worden.

Unter Berücksichtigung verschiedener Charakteristiken (z.B. Skalenniveau) können die Merkmalsausprägungen der interessierenden Objekte in visuelle Variablen (Abstand, Größe, Form, Farbton, Helligkeit, Gruppierung) überführt werden, woraus sich ansprechende Kartenprodukte erstellen lassen. Als häufigste Kartentypen führt Okabe (2006b) folgende vier an: (1) Chloroplethen-Karten: Einfärbung von Mosaiksteinen (z.B. Bundesländer) nach proportionaler Intensität zum Attributwert (z.B. Arbeitslosenrate); (2) Karte mit Symbolen: Skalierung eines Symbols im Verhältnis zum Attributwert eines Objekts an einem repräsentativen Punkt,

Abbildung 2 Räumliche Verteilung der Wohnorte der Befragten des Schweizer Umweltsurvey vor dem Hintergrund einer Landeskarte



beispielsweise Kreise unterschiedlicher Größe proportional zur Arbeitslosenrate der Bundesländer; (3) Isarithmic- oder Kontur-Karten: ein Set von Linien (Isolinien, Höhenlinien), die gleiche Attributwerte miteinander verbinden; (4) Punkte-Karten: ein Set von Punkten, die den Ort eines Ereignisses, z.B. eines Delikts, repräsentieren. Abbildung 2 zeigt ein Beispiel für eine Punkte-Karte.

Neben der grafischen Darstellung können Daten, z.B. das Ergebnis einer Datenmanipulation, auch als Zahlen in Form einer Tabelle zur Weiterverarbeitung in einem anderen System, beispielsweise in einem Statistikprogramm, exportiert werden.

3 Anwendung von GIS im Schweizer Umweltsurvey

In diesem Kapitel wird eine Studie zu Umweltgerechtigkeit vorgestellt, die auf dem Schweizer Umweltsurvey basiert und die Verteilung von Luftschadstoff- und Lärmexpositionen auf verschiedene soziale Gruppen untersucht. Detaillierte Informationen zu den Thesen und Ansätzen der Umweltgerechtigkeitsforschung finden sich im Artikel von Diekmann/Meyer (2010), in welchem auch die Ergebnisse der Umweltgerechtigkeitsanalysen beschrieben sind.

Der Umweltsurvey ist eine Befragung, die auf einer zweistufigen Zufallsstichprobe aus der erwachsenen Wohnbevölkerung der Schweiz mit registrierten

Telefonanschlüssen basiert. Die ausgewählten Haushalte wurden zuvor angeschrieben und um die Mitarbeit an der Studie gebeten. Insgesamt wurde mit 3369 zufällig ausgewählten Haushalten ein telefonisches Interview geführt. An das telefonische Interview schloss sich eine schriftliche Nachbefragung an (vgl. Diekmann/Meyer 2008 für weitere Erläuterungen zur Erhebung).

Für die Zuschreibung von objektiven Umweltbelastungsdaten sowie für die Gewinnung weiterer Variablen mit räumlichem Bezug wurde in dieser Studie das geographische Informationssystem ArcGIS eingesetzt (Diekmann/Meyer 2010). Im Folgenden wird beispielhaft beschrieben, wie die Adressen der Befragten geokodiert wurden (Abschnitt 3.1) und den Personen Belastungswerte zugeschrieben wurden (Abschnitt 3.2). Es wird gezeigt, wie unterschiedliche Datengrundlagen zu verschiedenen zugeschriebenen Werten führen. In der vorliegenden Studie hat dies glücklicherweise zwar keine wesentlichen Auswirkungen auf die Ergebnisse; doch im Allgemeinen ist es wichtig, die Frage der Datenqualität nicht zu vernachlässigen. Anschließend wird in Abschnitt 3.3 mit GIS beispielhaft ein Maß zur Wohnlage berechnet und in die Analyse eingefügt.

3.1 Geokodierung im Schweizer Umweltsurvey

Damit den Befragten Belastungsdaten zugeschrieben werden konnten, waren verschiedene Schritte nötig. Zunächst mussten die Wohnorte der Befragten geokodiert werden. Dazu wurden in einem ersten Schritt die Adressen der Befragten gesichert – dies geschah einerseits im Rahmen einer Recherche auf Basis der Telefonnummern der Befragten, andererseits durch eine direkte Erhebung respektive Überprüfung am Ende des Telefoninterviews. Um allfälligen Bedenken des Datenschutzes vorzubeugen, wurde die Untersuchung in einem Informationsschreiben als eine Zusammenarbeit zwischen der Forschergruppe und dem zuständigen Befragungsinstitut (Link) angekündigt und es wurde in Übereinstimmung mit den Empfehlungen des Verbands Schweizer Marketing- und Sozialforscher, dem das Befragungsinstitut angehört, eine Frage gestellt, bei der die Befragten die Möglichkeit hatten, die Weiterverwendung der Adresse zu verweigern (Diekmann/Meyer 2008).

Anschließend wurden den Adressen XY-Landeskoordinaten zugeschrieben (Geokodierung), welche als Punkte in ein GIS eingelesen werden konnten. Dies erfolgte durch einen Abgleich der Adressen mit der Datenbank GeoPost der Schweizerischen Post (Post 2011). Diese beinhaltet Landeskoordinaten für sämtliche postalisch bedienten Gebäude der Schweiz. Diese Datenbank wird laufend aktualisiert und die definitiven Koordinaten werden per GPS-Messung auf die Gebäudemitte mit einer Genauigkeit von 1-2 Metern erfasst. Rund 96% der Punkte sind erfasst;

Tabelle 1 Genauigkeit der Geokodierung im Schweizer Umweltsurvey

Genauigkeit	N	Anteile %
Hausnummer	2.995	88,9
nächstgelegene Hausnummer	42	1,3
Mittelpunkt der Straße	112	3,3
Mittelpunkt des PLZ-Gebietes	211	6,3
Adresse nicht gefunden	9	0,3
Total	3.369	100,0

Anmerkung: Die Basis bilden die im Schweizer Umweltsurvey telefonisch erhobenen Adressen sowie die Geokodierungsdatenbank der Schweizerischen Post (GeoPost).

bei den restlichen 40% handelt es sich um provisorische Koordinaten, welche auf der Basis von Kartenmaterial und Angaben des Zustellboten über die Gegebenheiten vor Ort ermittelt wurden (Post 2005). Mit gut einem Schweizerfranken pro zugeordnetem Koordinatenpaar hat die Nutzung der „genauesten Gebäudeadress-Datensammlung der Schweiz“ (Post 2011) allerdings auch ihren Preis.

Im Umweltsurvey konnte bei der Geokodierung auf Basis der Daten von GeoPost die in Tabelle 1 beschriebene Genauigkeit erreicht werden. In etwa 11% der Fälle konnte keine exakte Zuschreibung vorgenommen werden, weil entweder keine Hausnummer vorlag oder für die angegebene Straße oder Hausnummer keine übereinstimmenden Angaben in der Datenbank gefunden werden konnten. In diesen Fällen wurden stellvertretend entweder die Koordinaten eines Nachbarhauses, des Mittelpunkts der Straße oder des Postleitzahlgebietes verwendet.

3.2 Raster- oder Vektordaten für die Lärmbelastung?

Auf Grundlage der XY-Koordinaten der Wohnadressen konnten den Befragten in einem GIS objektive Belastungswerte zugeschrieben werden. Bildlich gesprochen wurde hierfür der Punkte-Layer mit den Wohnorten der Befragten über verschiedene als Rasterdaten vorliegende Belastungskarten gelegt. Jedem Befragungspunkt wurde dann unter Verwendung der „Spatial Analyst Tool Box“ in ArcGIS der darunterliegende Rasterwert zugeordnet und für die weiteren Analysen exportiert. Diese Daten beschränken sich auf jene 2.961 Personen, bei denen sowohl eine hausnum-

mergenaue Geokodierung möglich war (2.995) und die zugleich bezüglich ihres Hauptwohnsitzes befragt wurden (3.322)¹².

Zunächst wurde geprüft, ob der Anteil der Personen mit Belastungswerten über den offiziellen Grenzwerten¹³ in unseren Daten mit den offiziellen Angaben des Bundesamtes für Umwelt (SAEFL 2003; 2004; BAFU 2009a) übereinstimmen. Es zeigte sich, dass dies für Stickstoffdioxid (15%) und Feinstaub (42%) der Fall ist (offizielle Angaben: 16% resp. 41%), nicht aber für den Anteil Lärmbelasteter (8% vs. 23% für Tages- und 4% vs. 13% für Nachtlärmwerte, siehe auch Tabelle 2 in Diekmann/Meyer 2010).

Ein möglicher Grund für die Diskrepanz beim Anteil Lärmbelasteter liegt in der Datenbasis der Lärmwerte. Zwar beruhen sowohl die amtlichen Daten und als auch jene des Umweltsurveys auf der SonBase, der GIS-Lärmdatenbank der Schweiz (BAFU 2009b). Darin werden die Lärmkarten aber in zwei verschiedenen Datentypen zur Verfügung gestellt, nämlich als flächendeckende Rasterdaten (Abbildung 1, links) und als Vektordaten (Abbildung 1, rechts). Die Vektordaten sind grundsätzlich identisch mit dem Gebäudekataster des Bundesamtes für Landestopografie (swisstopo 2011b), enthalten aber zusätzlich für jedes Gebäude einen Lärmwert. Während die im Umweltsurvey berichteten Ergebnisse auf den Rasterdaten basieren, beruhen die offiziellen Angaben des Bundesamtes für Umwelt auf den Vektordaten (BAFU 2009a). Um zu prüfen, ob dies die Ursache der Diskrepanz sein könnte, wurden den Befragungspunkten des Umweltsurveys mit einer Überlagerungs-Operation die Lärmwerte derjenigen Gebäude zugeschrieben, innerhalb derer ihre Koordinaten zu liegen kamen. Etwas überraschend konnte auf diese Weise nur 83% (2.460) der 2.961 Befragungspunkte ein Lärmwert zugeordnet werden (vgl. Tabelle 2). Für weitere 98 zugeordnete Gebäude ist gemäß Bundesamt für Umwelt anzunehmen, dass der Schallpegel geringer ist als 40dB(A), da für sie kein Lärmwert hinterlegt ist. Um für die verbleibenden Fälle ausschließen zu können, dass die Übereinstimmung zwischen Befragungspunkt und Gebäude nicht wegen kleinerer Ungenauigkeiten in der Geokodierung oder den Gebäudedaten ausgeblieben ist, wurde um jeden Befragungspunkt ein Buffer von fünf Metern (vgl. Abbildung 3) gezeichnet und erneut mit den Vektordaten der SonBase verschnitten. Dieses Vorgehen ermöglichte es, weitere 271 Fälle mit Gebäude-Lärmwerten zu versehen (vgl. Tabelle 2).

12 Übrige: Zweitwohnsitz oder Ferienwohnung.

13 Stickstoffdioxid NO₂: 30mg/m³; Feinstaub PM₁₀: 20 mg/m³ (Der Schweizerische Bundesrat 1985); Lärm Tag: 60 dB(A); Lärm Nacht: 50 dB(A) (beide Empfindlichkeitsstufe II; Der Schweizerische Bundesrat, 1986).

Tabelle 2 Zuschreibung der Lärmwerte auf der Basis der Vektordaten der SonBase

Genauigkeit der Übereinstimmung	N	Bemerkungen
Exakte Übereinstimmung des Befragungspunkts mit einem Gebäude des Katasters	2.460	} Diese Fälle fließen in die multivariante Analyse ein
Überschneidung des 5m-Buffers um des Befragungspunkt mit einem Gebäude des Katasters	271	
Exakte Übereinstimmung Befragungspunkt mit einem Gebäude des Katasters, für das kein Schallpegel hinterlegt ist	98	Hohe Wahrscheinlichkeit, dass das Gebäude nur geringem Straßenlärm ausgesetzt ist (<40 dB(A))
Keine Übereinstimmung oder Überschneidung	132	vg. Abbildung 2
Total	2.961	

Anmerkung: Die Basis bilden die im Schweizer Umweltsurvey telefonisch erhobenen Adressen sowie die Geokodierungsdatenbank GeoPost und die Gebäude-Lärmdaten der SonBase (Vektordaten).

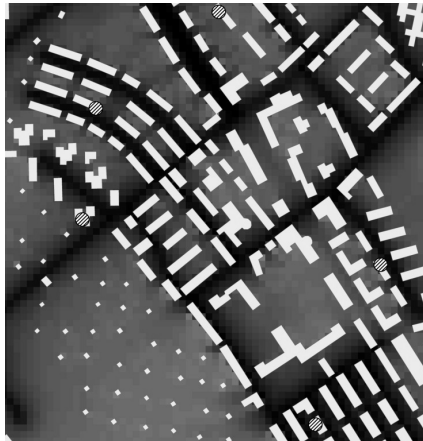
Abbildung 3 Vergleich Katasterdaten und Orthophoto (Luftbild)



Die beige Polygone repräsentieren die Gebäude des Katasters des Bundesamtes für Landestopografie (swisstopo 2011). Das Orthophoto stammt von map.search.ch. In Grau ist der 5-Meter-Buffer um einen ausgewählten Befragungspunkt ersichtlich.

Trotz dieser Bemühungen verblieben 132 Befragungspunkte ohne Zuordnung. Zur Aufklärung dieser Fälle wurden einzelne Adressen in den Webkarten von „Google Maps“ und „mapsearch.ch“ gesucht und mit dem Gebäudekataster verglichen. In allen der acht überprüften Fälle war auf dem Luftbild der Webkarten ein neues Gebäude oder eine Baugrube zu sehen (vgl. dazu Abbildung 3). Dies lässt sich

Abbildung 4 Zuordnung der Schallpegel auf Basis der Raster- und der Vektordaten



Anmerkungen: Rasterdaten: Je höher der Schallpegel desto dunkler die Einfärbung (Quelle: SonBase). Vektordaten: Die beige Polygone repräsentieren die Gebäude (Quelle: Gebäudekataster des Bundesamtes für Landestopografie (swisstopo 2011)). Die schraffierten Punkte repräsentieren die Wohnorte der Befragten, z.B. ganz oben im Gebäude in der Mitte.

so interpretieren, dass in der verwendeten und damals aktuellsten Ausgabe des Gebäudekatasters aus dem Jahr 2005 diejenigen Objekte gefehlt haben, die in den rund zwei Jahren bis zur Erhebung des Schweizer Umweltsurveys neu erstellt wurden (zur Qualität der Gebäudedaten vgl. auch swisstopo 2011b).

Auf der Grundlage jener 2.731 Fälle, denen auf Basis der Gebäudedaten (Vektordaten) ein Schallpegel zugeschrieben werden konnte, ergeben sich mit 23% tagsüber und 12% nachts nun vergleichbare Anteile Lärmbelasteter, wie sie offiziell vom Bundesamt für Umwelt kommuniziert werden (23% resp. 13%).

Nun stellt sich die Frage, welche Lärmwerte für weiterführende Analysen vorzuziehen sind. In den Vektordaten der SonBase wird dem ganzen Gebäude der Schallpegel an seiner lautesten Fassade zugeschrieben, unabhängig davon, ob eine Wohnung gegen den Hinterhof oder die Straßenseite ausgerichtet ist. In den Rasterdaten hingegen wird versucht, Abschirmungen Rechnung zu tragen. Dies illustriert Abbildung 4, in der die beiden Bilder aus Abbildung 1 übereinander gelegt sind. Auf der Rasterkarte sind die Innenhöfe heller eingefärbt (weniger Lärm), währenddessen in den Vektordaten für ein ganzes Gebäude der Schallpegel an der lautesten Fassade gilt (in Abbildung 4 sind die Gebäudelärmwerte farblich nicht unterschieden). Angesichts dieser Unterschiede könnte angenommen werden, dass die Rasterdaten genauer seien. Dies mag unter gewissen Umständen stimmen, kann aber für den vorliegenden Fall nicht beurteilt werden, denn die Lage der Punkte

in Abbildung 4 hängt ausschließlich von der XY-Koordinate der Geokodierung ab und gibt nicht die Stelle im Gebäude wieder, an der die befragte Person wohnt, d.h. unsere Daten blenden die Situation innerhalb des Gebäudes aus. Aufgrund der fehlenden Information zur genauen Lage der Wohnungen der Befragten im Gebäude kann damit nicht entschieden werden, ob die Realität mit den Gebäude- oder mit den Rasterdaten besser abgebildet wird.

Wie wirkt sich die Verwendung der Gebäude- gegenüber den Rasterdaten auf die Ergebnisse unserer Analyse zur sozialen Verteilung der Umweltbelastung aus? Um diese Frage zu beantworten, wurden für die Tages- und Nachtlärmwerte je zwei Regressionen gerechnet, einmal mit dem Schallpegel aus den Raster- und einmal aus den Vektordaten als abhängige Variable. Wie die Zahlen in den Spalten „Vektor“ und „Raster“ von Tabelle 3 zeigen, verändern sich die grundsätzlichen Befunde nicht. Die wichtigsten Effekte gehen in allen Modellen vom Urbanisierungsgrad und Hauseigentum aus,¹⁴ wobei die Stärke des Zusammenhangs und seine statistische Bedeutung in den Schätzungen auf der Basis der Vektordaten etwas grösser ausfallen als auf Grundlage der Rasterdaten. Diese Tendenz ist auch für die Effekte der sozialen Lage zu beobachten, insbesondere für die verschiedenen Ausländergruppen, bei denen mit den Vektordaten teilweise das allgemein geforderte Signifikanzniveau von 5% erreicht wird. Für die eigentlichen Ergebnisse der Umweltgerechtigkeitsanalysen sei auf den Artikel von Diekmann/Meyer (2010) verwiesen; eine ausführliche Besprechung würde hier den Rahmen sprengen.

14 Vgl. Tabelle 1 in Diekmann/Meyer (2010) für eine genauere Beschreibung der erklärenden Variablen.

Tabelle 3 Soziale Verteilung der Straßenlärmbelastung; Vektordaten und Rasterdaten im Vergleich, sowie Rasterdaten ergänzt um eine Variable zur Wohnlage

	Tag			Nacht		
	Vektor	Raster	Raster (erweitert)	Vektor	Raster	Raster (erweitert)
Schweizer/In	ref.	ref.	ref.	ref.	ref.	ref.
Westeuropa, Nordamerika	0,89 (1,62)	0,21 (0,35)	0,20 (0,34)	0,94+ (1,72)	0,29 (0,49)	0,28 (0,48)
Südeuropa	2,15** (3,35)	1,25+ (1,84)	1,04 (1,55)	2,21** (3,46)	1,54* (2,27)	1,36* (2,01)
Andere Staaten (Balkan, Osteuropa etc.)	2,26** (2,67)	1,51 (1,59)	1,35 (1,43)	2,47** (2,93)	1,59+ (1,67)	1,44 (1,52)
Höchste Anzahl Bildungsjahre im Haushalt (BFS 2007, in Zehnern)	-0,05 (-1,59)	-0,27 (-0,55)	-0,41 (-0,86)	-0,05+ (-1,67)	-0,57 (-1,17)	-0,70 (-1,45)
Äquivalenzeinkommen (monatlich in Tsd.)	-0,38 (-0,84)	-0,03 (-0,98)	-0,04 (-1,05)	-0,64 (-1,43)	-0,04 (-1,01)	-0,04 (-1,08)
Ländliches Gebiet	ref.	ref.	ref.	ref.	ref.	ref.
Agglomeration	1,98** (5,79)	2,10** (5,76)	1,56** (4,24)	2,36** (6,95)	2,44** (6,65)	1,96** (5,26)
Kleine oder mittlere Stadt	4,80** (11,35)	4,48** (9,95)	3,88** (8,54)	3,78** (9,01)	3,50** (7,72)	2,96** (6,46)
Großstadt	7,24** (18,49)	4,45** (10,48)	3,61** (8,25)	5,93** (15,26)	3,41** (7,96)	2,65** (6,00)
Deutsche Schweiz	ref.	ref.	ref.	ref.	ref.	ref.
Französische Schweiz	0,38 (1,10)	0,43 (1,18)	0,79* (2,17)	0,52 (1,52)	0,55 (1,49)	0,87* (2,37)
Italienische Schweiz	-0,34 (-0,45)	-0,61 (-0,91)	-0,62 (-0,93)	-0,14 (-0,19)	-0,41 (-0,61)	-0,42 (-0,63)
Haushaltsgröße	-0,37** (-3,56)	-0,32** (-2,83)	-0,28* (-2,51)	-0,31** (-2,99)	-0,26* (-2,30)	-0,23* (-2,02)
Alter (in Zehnern)	-0,24** (-2,96)	-0,14 (-1,61)	-0,14 (-1,63)	-0,24** (-3,08)	-0,14 (-1,64)	-0,14+ (-1,67)
Frau	-0,21 (-0,85)	-0,00 (-0,02)	-0,05 (-0,18)	-0,29 (-1,21)	-0,09 (-0,35)	-0,13 (-0,50)
Wohneigentümer	-1,72** (-6,09)	-1,55** (-5,19)	-1,33** (-4,48)	-1,58** (-5,64)	-1,45** (-4,83)	-1,26** (-4,20)
Wohnlage (Distanz zu Einkauf, Post und ÖV, in 100 Metern)			-0,09** (-6,82)			-0,08** (-6,09)
Konstante	55,61** (64,69)	50,59** (55,03)	52,50** (55,07)	44,44** (52,08)	39,51** (42,67)	41,22** (42,86)
Korrigiertes R-Quadrat	0,240	0,102	0,118	0,172	0,066	0,079
N	2.371	2.544	2.544	2.371	2.544	2.544

Anmerkungen: OLS-Regressionen, t-Werte in Klammern, + $p < 0,10$, * $p < 0,05$, ** $p < 0,01$. Die Modelle mit Rasterdaten entsprechen den Lärmschätzungen in Tabelle 6 und 16 in Diekmann/Meyer (2010).

Tabelle 4 Maß für die Zentralität der Wohnlage auf der Basis von Distanzmessungen in GIS

Variable	Mittelwert	Stdabw	N	Beschreibung
Distanz Zentrum (Meter)	628	405	2.961	Distanz zur nächsten Poststelle
Distanz ÖV-Haltestelle (Meter)	278	257	2.961	Distanz zur nächsten ÖV-Haltestelle: Bus, Tram oder Bahn
Distanz Einkaufen (Meter)	797	978	2.961	Distanz zum nächsten Einkaufsladen: Migros, Coop, Denner oder Volg
Total: Gesamtdistanz (Meter)	1.702	1.156	2.961	Geringste summierte Distanz zu allen Lageindikatoren: Zentrum, ÖV, Einkaufen

3.3 Die Wohnlage als Beispiel für eine mittels GIS generierte Variable

In GIS lässt sich eine große Zahl weiterer Prädiktoren entwickeln, die später beispielsweise in Regressionsanalysen verwendet werden können. So ließe sich in einer Analyse der Zufriedenheit mit der Wohnumgebung ein Maß für die Begrünung der Nachbarschaft oder die Verkehrsanbindung erstellen. Für die hier vorgestellte Umweltgerechtigkeitsstudie wurde ein Maß zur groben Abbildung der Wohnlage errechnet. Dieses basiert einerseits auf den Befragungspunkten aus dem Umweltsurvey (die Wohnorte), andererseits auf Punktdaten verschiedener Einrichtungen (Poststellen, ÖV-Haltestellen, Einkaufsläden: siehe Tabelle 4). Deren XY-Koordinaten sind in der Regel auf Anfrage erhältlich oder über die Adressen geokodierbar (vgl. weiter oben). Einfacher geht es bei Unternehmen wie beispielsweise der Supermarktkette Migros, die die Längen- und Breitengrade ihrer Standorte für den Import in Navigationssysteme im Internet zur Verfügung stellt (Migros 2011). Vor den eigentlichen GIS-Analysen mussten diese Längen- und Breitengrade in das gleiche Bezugssystem wie die übrigen Daten (Schweizer Landeskoordinaten) umgerechnet werden (swisstopo 2011a). Anschließend wurden die Luftliniendistanzen zur jeweils nächst gelegenen Einrichtung in ArcGIS mittels einer Proximity-Analyse berechnet und aufsummiert.

Wird die so erstellte Variable in die Regressionsmodelle zur Erklärung der Lärmbelastung eingeführt (siehe Tabelle 3), so zeigt sich beispielsweise, dass mit größerer Zentralität – wie bereits erwartet werden konnte – eine höhere Belastung einhergeht.

4 Ausblick

Dem räumlichen Kontext wurde in den deutschsprachigen Sozialwissenschaften bislang nur beschränkt Beachtung geschenkt. Nicht die mangelnde Relevanz entsprechender Forschungsfragen, sondern vielmehr der hohe Aufwand bei der Erhebung und Analyse raumbezogener Daten dürfte hierfür ausschlaggebend sein. Mit den geographischen Informationssystemen steht jedoch ein Instrument zur Verfügung, das diesen Aufwand erheblich verringert. Es erlaubt die Untersuchung einer Vielzahl verschiedener Forschungsfragen (z.B. Kidner/Higgs et al. 2003; Okabe 2006a). Die Bandbreite geografischer Funktionen ermöglicht auch Nicht-Geowissenschaftler/innen das Erstellen ansprechender kartografischer Produkte, die das Entdecken räumlicher Muster erleichtern, und das Errechnen raumbezogener Prädiktoren, die in weitere Analysen einfließen können. Auf diese Weise kann der Raumbezug menschlichen Handelns objektiver und situativ angepasster einbezogen werden als bislang üblich: objektiver, da die in GIS entwickelten Masse in der Regel nicht auf Beobachtung sondern auf physikalischen Messungen beruhen (wie Lärm, Distanzen etc.); situativ angepasster, weil man nicht an administrative Einheiten (wie Stadtteile, Bundesländer etc.) gebunden ist, sondern den relevanten Raum an die Fragestellung angepasst festlegen kann.

Wie in jedem Forschungsprozess üblich, sollte die Datenqualität reflektiert und dokumentiert werden. Im Rahmen der Untersuchungen zur Umweltgerechtigkeit im Schweizer Umweltsurvey wurden beispielsweise kritische Punkte ausgemacht wie die Genauigkeit der Adresserfassung zur Geokodierung, die Aktualität der Gebäudekataster sowie Mess- und Zuordnungsprobleme bei den Lärmdaten. Die Problematik der Datenqualität, deren Relevanz insbesondere beim Zusammenfügen von Daten aus verschiedenen Quellen deutlich wird, ist nicht grundlegend anders als jene Fragen der Objektivität, Validität und Reliabilität, die in den Sozialwissenschaften bereits heute zentral zur Beurteilung der Güte der Daten sind.

Zentrale Fragen zum Datenschutz bei der Verwendung von GIS-Daten sind bislang nicht eindeutig geklärt. Der „Rat für Sozial- und Wirtschaftsdaten“ bezeichnet eine Klärung und Harmonisierung der Regelungen als ein prioritäres Anliegen in Deutschland (RatSWD 2011).

Es ist anzunehmen, dass der räumliche Kontext in den Sozialwissenschaften erheblich an Bedeutung gewinnen wird. Als einer der Gründe ist die zunehmend leichtere Auffindbarkeit und teilweise unentgeltliche Zugänglichkeit von Geo-Daten anzuführen. Zudem hat die verfügbare Datenmenge beträchtlich zugenommen. Eine weitere Ursache ist in Open-Source-Angeboten zu suchen; einerseits erstellen Nutzer vermehrt kartographisches Material (z.B. OpenStreetMap), andererseits

liegen leistungsfähige, frei zugängliche Softwarepakete vor wie das umfangreiche GRASS GIS oder das einfacher bedienbare Quantum GIS. Diese ermöglichen auch Projekten mit einem beschränkten Budget mit Geo-Daten zu arbeiten.

Auch auf der Seite der Analyseverfahren sind Entwicklungen auszumachen. Komplexe geostatistische Verfahren werden zunehmend in herkömmliche GIS-Pakete integriert und damit einfacher zugänglich. So findet man bei der Literaturrecherche zu GIS vermehrt Hinweise auf Methoden der räumlichen Statistik und Ökonometrie (Anselin/Getis 2010; Goodchild 2010b). Den Ausgangspunkt bildet hierbei die Annahme, dass sich räumlich nahe stehende Objekte ähnlicher sind als weiter auseinandergelegene (Tobler 1970). Die daraus abgeleiteten Effekte der Abhängigkeit und der Heterogenität führen zu einer Verletzung der Unabhängigkeitsannahme, die vielen statistischen Tests zu Grunde liegt (Anselin/Getis 2010). Für allgemeine Bevölkerungsbefragungen wie die vorgestellte Studie dürften räumliche Interdependenzen jedoch kaum eine Rolle spielen, wie sie etwa für die gegenseitige Beeinflussung von Gebietseinheiten (z.B. Steuerwettbewerb zwischen den Kantonen der Schweiz) oder bei verbundenen Netzwerkknoten im sozialen Raum vermutet werden können (Anselin/Bera 1998; Anselin/Getis 2010). Kritischer könnte für solche Daten das räumliche Clustering von Objekten mit ähnlichen Untersuchungsmerkmalen sein (Homogenität von bzw. Heterogenität zwischen Gruppen, Goodchild 2010), dem man heute in Auswertungen in Statistikprogrammen lediglich mit Dummy-Variablen für verschiedene Gebietseinheiten begegnet. Sollten sich die Methoden der räumlichen Statistik als nützlich erweisen, dürfte dies auf eine ganze Reihe von Forschungsbereichen mit räumlich heterogen verteilten Untersuchungsmerkmalen einen Einfluss haben. Dies betrifft nicht zuletzt klassische Forschungsgebiete der Soziologie wie die Bildungsforschung, die Sozialstrukturanalyse oder die Arbeitsmarktforschung.

Literatur

- Anselin, L. und A. K. Bera, 1998: Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. S. 237-289 in: A. Ullah und D. E. A. Giles (Hg.): Handbook of Applied Economic Statistics. New York: Marcel Dekker
- Anselin, L. und A. Getis, 2010: Spatial Statistical Analysis and Geographic Information Systems. S. 35-48 in: L. Anselin und S. J. Rey (Hg.): Perspectives on Spatial Data Analysis. Heidelberg/Dordrecht/London/New York: Springer
- Asami, Y. und X. Gao, 2006: Effect of Environmental Factors on Housing Prices: Application of GIS to Urban-Policy Analysis. S. 211-228 in: A. Okabe (Hg.): GIS-Based Studies in the Humanities and Social Sciences. Boca Raton/London/New York: CRC Press, Francis and Taylor Group

- BAFU, 2009a: Lärmbelastung in der Schweiz. Schriftenreihe Umweltzustand. <http://www.bafu.admin.ch/publikationen/publikation/01036/index.html?lang=de> (17.06.2009).
- BAFU, 2009b: SonBase – Die GIS-Lärmdatenbank der Schweiz. Schriftenreihe Umweltwissen. <http://www.bafu.admin.ch/publikationen/publikation/01037/index.html?lang=de> (11.02.2011).
- Baranzini, A., S. Caroline, J. V. Ramirez und P. Thalmann, 2008: Do Foreigners Pay Higher Rents for the Same Quality of Housing in Geneva and Zurich? *Swiss Journal of Economics and Statistics* 144(4): 703-730.
- Baranzini, A. und J. V. Ramirez, 2005: Paying for Quietness. The Impact of Noise on Geneva Rents. *Urban Studies* 42(4): 633-646.
- Best, H., 2009: Kommt erst das Fressen und dann die Moral? Eine feldexperimentelle Überprüfung der Low-Cost-Hypothese und des Modells der Frame-Selektion. *Zeitschrift für Soziologie* 38(2).
- Bill, R., 2010: Grundlagen der Geo-Informationssysteme, Bd.1, Hardware, Software und Daten. Heidelberg: Wichmann.
- Bodenhamer, D. J., J. Corrigan und T. M. Harris (Hg.), 2010: *The Spatial Humanities. GIS and the Future of Humanities Scholarship*. Bloomington, Indianapolis: Indiana University Press.
- Briggs, D. J., 2005: The Role of Gis: Coping With Space (And Time) in Air Pollution Exposure Assessment. *Journal of Toxicology and Environmental Health, Part A* 68(13): 1243-1261.
- Brimblecombe, N., D. Dorling und M. Shaw, 1999: Mortality and migration in Britain, first results from the British Household Panel Survey. *Social Science and Medicine* 49: 981-988.
- Brügge, B., D. Harhoff, A. Picot, O. Creighton, M. Fiedler und J. Henkel, 2004: *Open-Source-Software – Eine ökonomische und technische Analyse*. Berlin: Springer.
- Bundesamt für Statistik, Bundesamt für Raumentwicklung und ecoplan, 2007: *Mobilität in der Schweiz. Ergebnisse des Mikrozensus 2005 zum Verkehrsverhalten*. Neuchâtel: Bundesamt für Statistik (bfs).
- Carosio, A., 2005: Einführung zum Nachdiplomkurs Raumbezogene Informationssysteme. Zürich: Institut für Geodäsie und Photogrammetrie, ETH Zürich.
- Crowder, K. und L. Downey, 2010: Interneighborhood Migration, Race, and Environmental Hazards: Modeling Microlevel Processes of Environmental Inequality. *American Journal of Sociology* 115(4): 1110-1149.
- de Smith, M. J., M. F. Goodchild und P. A. Longley, 2009: *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Leicester: Troubador.
- DeMers, M. N., 2009a: *Fundamentals of Geographic Information Systems*. Hoboken, New Jersey: John Wiley and Sons.
- DeMers, M. N., 2009b: *GIS for Dummies*. Indianapolis, Indiana: Wiley Publishing Inc.
- Der Schweizerische Bundesrat, 1985 Stand am 1. Januar 2009: Luftreinhalte-Verordnung. http://www.admin.ch/ch/d/sr/814_41/index.html (11.02.2010).
- Der Schweizerische Bundesrat, 1986 Stand am 1. Juli 2008: Lärmschutz-Verordnung. http://www.admin.ch/ch/d/sr/814_41/index.html (16.09.2009).
- Diekmann, A. und R. Meyer, 2008: *Schweizer Umweltsurvey 2007. Dokumentation und Codebuch*. Zürich: Professur für Soziologie, ETH Zürich.
- Diekmann, A. und R. Meyer, 2010: Demokratischer Smog? Eine empirische Untersuchung zum Zusammenhang zwischen Sozialschicht und Umweltbelastungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62(3): 437-457.
- Downey, L., 2006: Using Geographic Information Systems to Reconceptualize Spatial Relationships and Ecological Context. *American Journal of Sociology* 112(2): 567-612.

- ESRI (Hg.), 2009a: GIS Best Practice. Social Sciences. Redlands.
- ESRI, 2009b: What is GIS? in: ESRI (Hg.): GIS Best Practices. Social Sciences. Redlands, CA.
- Franzen, A., 1997: Umweltbewusstsein und Verkehrsverhalten. Empirische Analysen zur Verkehrsmittelwahl und der Akzeptanz umweltpolitischer Massnahmen. Chur/Zürich: Rügger.
- Frerichs, R. R., 2001: History, Maps and the Internet: UCLA's John Snow Site. *The Bulletin of the Society of Cartographers* 34(2): 3-7.
- Germann, C. und K. Seeland, 2004: Are urban green spaces optimally distributed to act as places for social integration? Results of a geographical information system (GIS) approach for urban forestry research. *Forest Policy and Economics* 6: 3-13.
- Gómez-Rubio, V., J. Ferrándiz-Ferragud und A. López-Quílez, 2009: Epidemiological Information Systems. S. 235-248 in: J. Pilz (Hg.): *Interfacing Geostatistics and GIS*. Berlin/Heidelberg: Springer.
- Goodchild, M. F., 2010a: Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science* 1: 3-22.
- Goodchild, M. F., 2010b: Whose Hand on the Tiller? Revisiting „Spatial Statistical Analysis an GIS“. S. 49-60 in: L. Anselin und S. J. Rey (Hg.): *Perspectives on Spatial Data Analysis*. Heidelberg/Dordrecht/London/New York: Springer.
- Goodchild, M. F. und D. G. Janelle, 2004: Thinking Spatially in the Social Sciences. S. 3-21 in: M. F. Goodchild und D. G. Janelle (Hg.): *Spatially Integrated Social Science*. Oxford/New York: Oxford University Press.
- Heinrich, J., U. Gehring, J. Cyrus, M. Brauer, G. Hoek, P. Fischer, T. Bellander und B. Brunekreef, 2005: Exposure to traffic related air pollutants: self reported traffic intensity versus GIS modelled exposure. *Occupational and Environmental Medicine* 62: 517-523.
- Hermann, M., C. Heye und H. Leuthold, 2005: Soziokulturelle Unterschiede in der Schweiz. Vier Indizes zu räumlichen Disparitäten, 1990-2000. *Statistik der Schweiz (BFS)*. Neuchâtel.
- Higgs, G., 2004: A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services. *Health Services & Outcomes Research Methodology* 5: 119-139.
- Kidner, D., G. Higgs und S. White (Hg.), 2003: *Socio Economic Applications in Geographical Information Science*. London/New York: Taylor and Francis.
- Kohsaka, H. und T. Sekine, 2006: Visualization for Site Assessment. S. 279-298 in: A. Okabe (Hg.): *GIS-Based Studies in the Humanities and Social Sciences*. Boca Raton/London/New York: CRC Press, Taylor and Francis Group
- Lai, P. C. und A. S. H. Mak (Hg.), 2007: *GIS for Health and the Environment. Development in the Asia-Pacific Region*. Lecture Notes in Geoinformation and Cartography. Berlin/Heidelberg/New York: Springer.
- Lloyd, C. D., 2010: *Spatial Data Analysis. An Introduction for GIS Users*. Oxford/New York: Oxford University Press.
- Ludwig, I., A. Voss und Krause-Traudes, 2010: Wie gut ist Open Street Map? Zur Methodik eines automatisierten objektbasierten Vergleiches der Strassennetze von OSM und NAVTEQ in Deutschland. *GIS.Science* 4: 148-158.
- Meyer, T. H., 2010: *Introduction to Geometrical and Physical Geodesy: Foundations of Geomatics*. Redlands: ESRI Press.
- Migros, 2011: Filialen und Öffnungszeiten. <http://www.migros.ch/de/services/filialsuche.html> (11.02.2011).
- Mundhenke, J., 2005: Ökonomische Eigenschaften von Software – Die Bedeutung von Open-Source-Software für den Wettbewerb auf Softwaremärkten in: M. Bärwolff, R. A.

- Gehring und B. Lutterbeck (Hg.): Open Source Jahrbuch 2005: Zwischen Softwareentwicklung und Gesellschaftsmodell. Berlin: Lehmanns Media.
- Niyonsenga, T., J. Courteau, C. Dean, A. Hemiari, G. Bénié und A. Vanasse, 2009: Geomatics, Epidemiology and BioStatistics: An Application to Acute Coronary Syndrome. S. 249-260 in: J. Pilz (Hg.): Interfacing Geostatistics and GIS. Berlin/Heidelberg: Springer.
- Okabe, A. (Hg.), 2006a: GIS-Based Studies in the Humanities and Social Sciences. Boca Raton/London/New York: CRC Press, Taylor and Francis Group.
- Okabe, A., 2006b: Introduction. S. 1-18 in: A. Okabe (Hg.): GIS-Based Studies in the Humanities and Social Sciences. Boca Raton/London/New York: CRC, Francis and Taylor
- Post, 2005: Die Post garantiert Qualität. Erfassung und Aktualisierung der GeoPost-Daten: Die Schweizerische Post: 2.
- Post, 2011: Coordinate: Genaueste Gebäudeadress-Datensammlung der Schweiz. <http://www.post.ch/post-startseite/post-geschaeftskunden/post-direct-marketing/post-direct-marketing-national/post-direct-marketing-adressen/post-direct-marketing-geo-post/post-direct-marketing-geopost-koordinaten.htm> (11.02.2011).
- Preisendörfer, P., 1999: Umwelteinstellungen und Umweltverhalten in Deutschland. Empirische Befunde und Analyse auf der Grundlage der Bevölkerungsumfragen 'Umweltbewusstsein in Deutschland' 1991-1998. Opladen: Leske + Budrich.
- RatSWD, 2011: Endbericht der AG „Georeferenzierung von Daten“. http://ratswd.de/Geodaten/downloads/RatSWD_Endbericht_Geo-AG.pdf (13.07.2012).
- Rigaux, P., M. Scholl und A. Voisard, 2002: Spatial Databases with Application to GIS. San Francisco/California: Morgan Kaufmann Publishers.
- SAEFL, 2003: Modelling of PM10 and PM2.5 ambient concentrations in Switzerland 2000 and 2010. Environmental Documentation No. 169, Air: Swiss Agency for the Environment Forests and Landscape.
- SAEFL, 2004: Modelling of NO2 and benzene ambient concentrations in Switzerland 2000 to 2020. Environmental Documentation No. 188, Air. Bern: Swiss Agency for the Environment Forests and Landscape.
- Schnell, R. und F. Kreuter, 2000: Das DEFECT-Projekt: Sampling-Errors und Nonsampling-errors in komplexen Bevölkerungstichproben. ZUMA-Nachrichten 47: 89-102.
- Schuler, M., P. Dessementet und D. Joye, 2007: Atlas des räumlichen Wandels der Schweiz: Bundesamt für Statistik BFS.
- Shaw, M., D. Dorling und N. Brimblecombe, 1999: Life chances in Britain by housing wealth and for the homeless and vulnerably housed. Environment and Planning A 31: 2239-2248.
- Shimizu, E. und T. Fuse, 2006: A Method for Visualizing the Landscapes of Old-Time Cities Using GIS in: A. Okabe (Hg.): GIS-Based Studies in the Humanities and Social Sciences. Boca Raton/London/New York: CRC Press, Francis and Taylor Group
- Snow, J., 1855: On the Mode of Communication of Cholera. London: John Churchill.
- Sorrentino, J. A., M. M. R. Meenar und B. J. Flamm, 2008: Suitable Housing Placement: A GIS-Based Approach. Environmental Management 42: 803-820.
- Spencer, C. und J. Ratcliff, 2005: GIS and Crime Mapping. Chichester, England: John Wiley and Sons, Ltd.
- Steinberg, S. J. und S. L. Steinberg, 2006: GIS, Geographic Information Systems for the Social Sciences. Investigating Space and Place. Thousand Oaks/London/New Delhi: Sage Publications.
- swisstopo, 2011a: Skripts für WGS84-<->CH1903. <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/software/products/skripts.html> (11.02.2011).

- swisstopo, 2011b: Vector25 <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/vector25.html> (11.02.2011).
- Tobler, W., 1970: A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46(2): 234-240.
- Tsumura, H. o., 2006: Site-Catchment Analysis of Prehistoric Settlements by Reconstructing Paleoenvironments with GIS. S. 175-190 in: A. Okabe (Hg.): *GIS-Based Studies in the Humanities and Social Sciences*. Boca Raton/London/New York: CRC Press, Taylor and Francis Group.
- van Schaaik, J. G. J. und J. J. van der Kemp (Hg.), 2009: *Real Crimes on Virtual Maps: The Application of Geography and GIS in Criminology*. Geospatial Technology and the Role of Location in Science, GeoJournal Library Springer Netherlands.
- Yang, K., S.-y. Peng, Q.-l. Xu und Y.-b. Cao, 2007: A Study on Spatial Decision Support Systems for Epidemic Disease Prevention Based on ArcGIS. S. 30-43 in: P. C. Lai und A. S. H. Mak (Hg.): *GIS for Health and the Environment. Development in the Asia-Pacific Region*. Berlin/Heidelberg/New York: Springer

Anschrift der Autoren

Reto Meyer
ETH Zürich
Soziologie
CLU D1
8092 Zürich
E-Mail: reto.meyer@soz.gess.ethz.ch

Heidi Bruderer Enzler
ETH Zürich
Soziologie
CLU D1
8092 Zürich
E-Mail: bruderer@soz.gess.ethz.ch

Anhang: Verweise auf Software, Webseiten und Lehrbücher

Auswahl von GIS-Programmen

- ArcGIS von Environmental Systems Research Institute (ESRI): <http://www.esri.com> (07.06.2012)
- GeoMedia von Intergraph: <http://www.intergraph.com> (07.06.2012)
- GRASS GIS der Open Source Geospatial Foundation (OSGeo): <http://grass.osgeo.org> (07.06.2012)
- Quantum GIS (QGIS) der Open Source Geospatial Foundation (OSGeo): <http://qgis.osgeo.org> (07.06.2012)
- Verweise auf weitere Open Source und frei zugängliche GIS-Softwareprojekte: <http://open-sourcegis.org> (07.06.2012)

Geo-Daten der deutschsprachigen Länder

- Infrastructure for Spatial Information in the European Community (INSPIRE): <http://inspire.jrc.ec.europa.eu> (07.06.2012)
- Geoportal.DE (Nationale Geodateninfrastruktur Deutschland sowie Web-GIS dazu): <http://www.geoportal.de> (07.06.2012)
- GeoDatenZentrum des Deutschen Bundesamts für Kartographie und Geodäsie: <http://www.geodatenzentrum.de> (07.06.2012)
- Verweise auf Karten verschiedenster Themen zu Deutschland (von der Karte zum Erdmagnetismus bis zur Spielkarten-Karte Deutschlands: „Welches Blatt wird gespielt?“): <http://www.hoeckmann.de/karten/> (07.06.2012)
- INSPIRE Österreich (Informationen zur nationalen Geodateninfrastruktur, Links auf verschiedene Geoportale in Österreich) <http://www.inspire.gv.at> (07.06.2012)
- Geoportal des Schweizerischen Bundes: <http://www.geo.admin.ch> (07.06.2012)
- Web-GIS des Schweizerischen Bundes: <http://map.geo.admin.ch> (07.06.2012)
- Swisstopo, das Geoinformationszentrum des Schweizerischen Bundes: <http://www.swisstopo.admin.ch> (07.06.2012)
- Metadatenkatalog für Geodaten der Schweiz (auch von Kantonen und privaten Datenhaltern): <http://www.geocat.ch/geonetwork/srv/deu/geocat> (07.06.2012)
- Nationale Geodateninfrastruktur des Fürstentums Liechtenstein: <http://www.llv.li/amtstsel/llv-tba-geodateninfrastruktur.htm> (07.06.2012)
- Geodatenportal der Landesverwaltung des Fürstentums Liechtenstein (Web-GIS): <http://geodaten.llv.li/geoshop/public.html> (07.06.2012)

Geokodierung in den deutschsprachigen Ländern

- Deutschland: Amtliche Hauskoordinaten (Bezirksregierung Köln, Abteilung Geobasis NRW, „Zentrale Stelle Hauskoordinaten und Hausumringe“, ZSHH): http://www.bezreg-koeln.nrw.de/brk_internet/organisation/abteilung07_produkte/liegenschaftsinformation/hausinformationen/hauskoordinaten/index.html (07.06.2012)
- Österreich: Adressregister, geführt vom Bundesamt für Eich- und Vermessungswesen: http://www.bev.gv.at/portal/page?_pageid=713,2167982&_dad=portal&_schema=PORTAL (07.06.2012)

- Schweiz: Eidgenössisches Gebäude- und Wohnungsregister (GWR): <http://www.bfs.admin.ch>, Einsicht über das Web-GIS unter <http://map.geo.admin.ch> (07.06.2012)
- Schweiz: GeoPost Coordinate: <http://www.post.ch/post-startseite/post-geschaeftskunden/post-direct-marketing/post-direct-marketing-national/post-direct-marketing-adresen/post-direct-marketing-geopost/post-direct-marketing-geopost-koordinaten.htm> (07.06.2012)
- Fürstentum Liechtenstein: Grundbuchplan: <http://www.llv.li/amtstellen/llv-tba-geodaten/infrastruktur.htm>, Einsicht über das Web-GIS unter <http://geodaten.llv.li/geoshop/public.html> (07.06.2012)
- Batchgeo, ein Online-Instrument zur Geokodierung von Adresslisten, das die Geocoding API von Google nutzt und daher aus rechtlichen Gründen die Adressen nur noch anzeigt, die Längen- und Breitengrade aber nicht mehr als Liste ausgibt: <http://www.batchgeo.com/> (07.06.2012)

Auswahl von GIS-Lehrbüchern

- Bill, R., 2010: Grundlagen der Geo-Informationssysteme, Bd.1, Hardware, Software und Daten. Heidelberg: Wichmann.
- Lloyd, C. D., 2010: Spatial Data Analysis. An Introduction for GIS Users. Oxford/New York: Oxford University Press.
- De Smith, M. J., M. F. Goodchild & P. A. Longley (2013). Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. Leicester: Troubador. Online: <http://www.spatialanalysisonline.com/HTML/index.html>

GIS in den Sozialwissenschaften

- Center for Spatially Integrated Social Science: <http://www.csiss.org> (07.06.2012)
- Blog von Matt Artz zu Neuigkeiten, Hilfsmittel und Kommentare zur wissenschaftlichen Anwendung der GIS-Technologie, darunter auch von Anwendungen in den Sozialwissenschaften, die als eigene Kategorie geführt wird: <http://gisandscience.com/category/social-science> (07.06.2012)
- Historical Geographical Information Systems (HGIS): Anwendung von GIS in den Geschichtswissenschaften: <http://www.hgis.org.uk/index.htm> (07.06.2012)

Plagiate in studentischen Arbeiten

Eine empirische Untersuchung unter Anwendung des Triangular Modells

Plagiarism in Student Papers

An Empirical Study Using the Triangular Model

Julia Jerke und Ivar Krumpal

Zusammenfassung

Sensitive Fragen in sozialwissenschaftlichen Erhebungen stellen Forscher immer wieder vor Probleme. Um beispielsweise niedrige Response-Raten und verzerrte Ergebnisse zu überwinden, schlagen Yu, Tian und Tang (2008) zwei vielversprechende neue Befragungstechniken vor: das Crosswise Modell (CM) und das Triangular Modell (TM). Diese beiden Modelle können als Alternative zur bekannten Randomized Response Technik (RRT) eingesetzt werden und überwinden die meisten Nachteile der RRT. Während das CM bereits empirisch erforscht wurde (Jann et al. 2012) liegen für das TM bisher noch keine empirischen Befunde vor. Wir stellen daher Ergebnisse aus einer an der Universität Leipzig durchgeführten Studie vor, in der das TM angewandt und der direkten Befragung gegenübergestellt wurde. Hierfür befragten wir Studenten, ob sie schon einmal plagiiert haben. Die Ergebnisse zeigen, dass sich in der TM-Kondition mehr

Abstract

Sensitive questions in social surveys often yield problems such as item nonresponse and response bias. Yu, Tian and Tang (2008) proposed two new models to overcome these problems: the crosswise model (CM) and the triangular model (TM). Both models can be seen as an alternative to the well-known randomized response technique (RRT) and are meant to overcome most of its drawbacks. While the CM has already been applied empirically (Jann et al. 2012), no evidence for the TM exists. We therefore present first results of an empirical survey conducted at the University of Leipzig comparing direct questioning and the TM. Students were asked to self-report whether they ever have plagiarised in one of their term papers. The results show that under the TM condition a higher percentage of students admit plagiarism than under direct questioning. This and the fact that the TM is very easy to implement show that the TM is



Studenten zu einem Plagiat bekennen als in der direkten Befragung. Dieses Resultat und die einfache Implementierung machen das TM zu einer vielversprechenden Alternative zur direkten Befragung, sowie zu anderen indirekten Befragungsmethoden, wie beispielsweise der RRT.

a promising alternative to direct as well as indirect questioning such as the RRT.

1 Einleitung

In sozialwissenschaftlichen oder medizinischen Befragungen werden häufig sensitive bzw. heikle Fragen, – etwa Fragen zu kriminellen Aktivitäten, zu sexuellen Präferenzen, zum illegalen Drogenkonsum, zum Einkommen oder zur Wahlbeteiligung, – gestellt. Das Stellen solcher Fragen verletzt zunächst die Norm der Privatheit. So werden bei Fragen nach der sexuellen Orientierung private Persönlichkeitsbereiche berührt, die ein Befragter nicht unbedingt mit anderen Personen teilen möchte. Zudem können heikle Fragen aber auch illegale oder strafrechtlich relevante Verhaltensweisen (z.B. Steuerhinterziehung, Plagiiere) des Befragten offenlegen. Bei einer ehrliche Beantwortung und Offenlegung von illegalen Verhaltensweisen gegenüber Dritten müsste der Befragte die Möglichkeit von unangenehmen Konsequenzen befürchten. Neben der Dimension der Privatheit und der strafrechtlichen Relevanz, tangieren heikle Fragen schließlich eine dritte Dimension, nämlich die der sozialen Erwünschtheit (Tourangeau/Yan 2007: 860)¹. Beispielsweise sind Fragen nach Einstellungen zu gesellschaftlichen Minderheiten wie Ausländern und Juden stark normativ konnotiert und führen häufig zu sozial erwünschtem Antwortverhalten bzw. Antwortverweigerungen (Krumpal 2012, 2013). Dabei ist das Problem der sozialen Erwünschtheit umso gravierender, je weniger anonym die Befragungssituation ist. Zusammenfassend lassen sich somit Fragen als ‚heikel‘ oder ‚sensitiv‘ definieren, wenn sie persönliche Details oder sozial unerwünschte Merkmale (Ver-

1 Auf Grundlage der bestehenden Forschungsliteratur können ‚Impression Management‘ sowie ‚Self-Deception‘ als mögliche Mechanismen angenommen werden, die sozial erwünschtes Antwortverhalten hervorbringen (Paulhus 1984; Paulhus/John 1998; Krumpal/Näher 2012). Während der Impression Management Mechanismus annimmt, dass der Befragte bewusst falsch antwortet, um vor einer tatsächlichen oder eingebildeten Öffentlichkeit ein vorteilhaftes Bild abzugeben, nimmt der Self-Deception Mechanismus an, dass der Befragte unbewusst falsch antwortet, indem er oder sie negative Aspekte des Selbst verdrängt und selbst an seine (falschen) positiven Selbstdarstellungen glaubt. In der vorliegenden Arbeit nehmen wir an, dass dem sozial erwünschten Antwortverhalten beide psychologischen Mechanismen zugrunde liegen.

haltensweisen oder Einstellungen), die gegen soziale Normen verstoßen und somit sanktioniert werden, nicht ausreichend anonym erfragen.

Wenn Befragte die Beantwortung unangenehmer Fragen verweigern oder keine wahrheitsgemäßen Antworten geben, dann kann das für die Validität der erhobenen Daten negative Konsequenzen haben. So leiden Umfragestudien zu sozial erwünschten Verhaltensweisen oder Einstellungen (aufgrund von ‚Overreporting‘) häufig unter systematischer Überschätzung und Studien zu sozial unerwünschten Verhaltensweisen oder Einstellungen (aufgrund von ‚Underreporting‘) unter systematischer Unterschätzung selbiger: *„Respondents consistently underreport some behaviors (the socially undesirable ones) and consistently overreport others (the desirable ones)”* (Tourangeau/Yan 2007: 863). Zudem können Antwortverweigerungen ähnliche Verzerrungen hervorrufen, nämlich dann wenn Befragte, die die Antwort verweigern, sich bezüglich des sensitiven Merkmals systematisch von denen unterscheiden, die an der Befragung teilnehmen.

Da direkte Befragungen zu sensitiven Themen häufig unter verzerrten Ergebnissen leiden, wurden in der Survey Forschung verschiedene Verfahren zur Reduzierung dieser Verzerrungen, wie etwa die Randomized Response Technik (RRT; Warner 1965), vorgeschlagen. Hierbei sollen über einen Zufallsmechanismus die Antworten anonymisiert und ein ehrlicheres Antwortverhalten erzielt werden. In einer Vielzahl von experimentellen Studien entlockte die RRT den Befragten mehr sozial unerwünschte Antworten als die konventionelle direkte Befragung (Lensvelt-Mulders et al. 2005). Allerdings wurde auch immer wieder die hohe Komplexität der RRT, wie etwa die Notwendigkeit eines Randomisierungsinstruments (z.B. Münzen, Würfel, etc.) und die direkte Konfrontation mit der heiklen Frage, moniert. Yu et al. (2008) haben deshalb zwei alternative Methoden für die Erhebung sensitiver Merkmale vorgeschlagen: das Crosswise Modell (CM) bzw. das Triangular Modell (TM). Diese neuen Befragungsverfahren sollen die Schwächen der RRT kompensieren und validere Resultate bei der Messung sensitiver Merkmale erzielen. Während das Crosswise Modell bereits empirisch evaluiert wurde und, im Vergleich zur direkten Befragung, gute Resultate erzielt hat (Coutts 2011; Jann et al. 2012), liegt zur Effektivität des TM bislang keine empirische Studie vor. Der nachfolgende Beitrag schließt diese Forschungslücke. In einem experimentellen Design wird das TM mit der direkten Befragungsmethode verglichen. Hierbei werden Studierende der Universität Leipzig zum heiklen Thema „Plagiate in Hausarbeiten“ befragt. Bisherige Befragungsstudien zu Plagiaten im universitären Kontext führten im deutschsprachigen Raum (Deutschland oder Schweiz) zu teilweise recht unterschiedlichen Prävalenzschätzungen. So reichten bei direkten Selbstauskünften die Anteile derjenigen Studierenden, die zugaben eine Haus- oder Abschlussarbeit teilweise plagiiert

zu haben von 3 % (Höglinger et al. 2013: 23) bzw. 8.1 % (Coutts et al. 2011: 755) bis zu 15,9 % (Preisendörfer 2008: 22) bzw. 22,9 % (Sattler 2007: 149). Direkte Fragen nach Vollplagiaten führten zu geschätzten Anteilen von 1,4 % (Preisendörfer 2008: 22) bis 2 % (Coutts et al. 2011: 755; Höglinger et al. 2013: 23). Diese variierenden Befunde sind aufgrund von Unterschieden in den Frageformulierungen, Stichprobensamensetzungen oder Erhebungsmodi nur eingeschränkt miteinander vergleichbar, können jedoch als grober Orientierungsrahmen dienen um die Ergebnisse der vorliegenden Studie besser in die Forschungslandschaft einzuordnen. Bei der vorliegenden Studie ist es von zentralem Interesse, ob unter Anwendung des TM mehr Studenten ein Plagiat offenbaren als bei einer klassischen direkten Befragung. Zudem sollen die Ergebnisse auch mit der bereits existierenden Studie zum CM (Jann et al. 2012) kontrastiert werden.

2 Befragungsmethoden im sensitiven Kontext

2.1 Die Randomized Response Technik (RRT)

Die RRT wurde erstmals im Jahr 1965 durch Stanley Warner vorgestellt und später vielfach weiterentwickelt (siehe beispielsweise Horvitz et al. 1967, Greenberg et al. 1969, Boruch 1971, Kuk 1990). Warners grundlegendes Ziel war es, die Anonymität des Befragten zu erhöhen. Dieser antwortet dazu nur mit einer vorher festgelegten Wahrscheinlichkeit auf die sensitive Frage: *„Essentially the method involves the device that [...] the interviewee responds with answers that furnish information only on a probability basis“* (Warner 1965: 63). Es wird also eine probabilistische Beziehung zwischen einer gegebenen Antwort und der heiklen Frage hergestellt. Die Umsetzung dieses Prinzips geschieht in Form eines Zufallsexperiments.

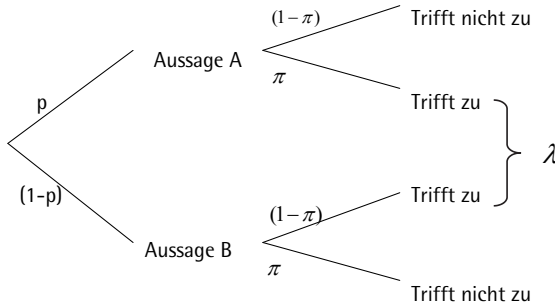
Die Individuen einer Grundgesamtheit lassen sich dahingehend unterscheiden, ob sie Träger eines sensitiven Merkmals X sind oder nicht. Gesucht ist dann der Anteil π der Merkmalsträger in der Grundgesamtheit, wobei $\pi = P(X = 1)$ auch als die Wahrscheinlichkeit, das sensitive Merkmal zu tragen, beschrieben werden kann. Dem Befragten werden dabei nachstehendem Schema folgend zwei Aussagen vorgelegt:

Aussage A: *Ich bin Träger des Merkmals X .*

Aussage B: *Ich bin nicht Träger des Merkmals X .*

Mithilfe eines Randomisierungsinstruments – durchgesetzt haben sich vor allem Würfel und Münze (Lensvelt-Mulders et al. 2005) – wird zufällig eine der beiden

Abbildung 1 Darstellung der Randomized Response Technik nach Warner (1965)



Anmerkung. p ist die Wahrscheinlichkeit mit der die Aussage A gewählt wird; π ist die gesuchte Wahrscheinlichkeit, das sensitive Merkmal X zu tragen; λ ist die Wahrscheinlichkeit, dass eine auf den Befragten zutreffende Aussage (ja' Antwort) gewählt wird

Aussagen ausgewählt. Die Auswahl im Warner Modell muss dabei zwei Bedingungen erfüllen: (1) die Auswahlwahrscheinlichkeiten der beiden Aussagen sind im Vorfeld bekannt und jeweils ungleich 0.5. (2) der Interviewer kennt den Ausgang des Zufallsexperimentes nicht, d.h. nur dem Befragten ist bekannt, welche der beiden Aussagen ausgewählt wurde. Er gibt dann lediglich an, ob die ausgewählte Aussage auf ihn zutrifft oder nicht. In Abbildung 1 ist das Befragungsprinzip schematisch dargestellt.

Aus diesem Baumdiagramm kann die Schätzformel für den Populationsparameter π über elementare wahrscheinlichkeitstheoretische Umformungen leicht abgeleitet werden. Wird der erwartete Anteil zutreffender Aussagen $\lambda = p\pi + (1 - p)(1 - \pi)$ durch $\hat{\lambda}$ – den beobachteten Anteil zutreffender Aussagen – ersetzt und umgeformt, dann ergibt sich der Schätzer von π (Warner 1965):

$$\hat{\pi} = \frac{\hat{\lambda} + p - 1}{2p - 1}$$

Die Varianz des Schätzers ist dann gegeben durch:

$$Var(\hat{\pi}) = \frac{\pi \cdot (1 - \pi)}{n} + \frac{p \cdot (1 - p)}{n \cdot (2p - 1)^2}$$

Die Varianz wird also in einen durch die Stichprobenziehung und einen durch die Randomisierung verursachten Anteil zerlegt (Warner 1965).

Lensvelt-Mulders et al. (2005) zeigen in einer Meta-Analyse, dass die RRT in sensitiven Befragungssituationen bessere Schätzungen sensitiver Merkmale liefert als die direkte Befragung. In sechs der ausgewerteten Studien lagen externe Validierungsdaten bezüglich des sensitiven Merkmals vor. Obwohl auch die RRT die tatsächliche Prävalenz der sensitiven Merkmale unterschätzte, lieferte sie deutlich validere Schätzungen als konventionelle direkte Befragungsmethoden. In weiteren von Lensvelt-Mulders et al. (2005) untersuchten Studien lagen keine Validierungsdaten vor und die RRT wurde nach dem Prinzip der „more is better“-Annahme (Umesh/Peterson 1991) anderen Befragungsmethoden gegenübergestellt. Diesem Prinzip folgend wird bei erwartetem ‚Underreporting‘ eine Methode umso valider bezeichnet, je höher die durch sie ermittelten Prävalenzraten des sensitiven Merkmals sind². Auch hier wurden mit der RRT validere Schätzungen ermittelt.

Durch die Einführung eines Randomisierungsinstruments in den Befragungsprozess folgen jedoch zahlreiche Probleme (Lensvelt-Mulders et al. 2005). Mit der erhöhten Varianz geht ein im Vergleich zur direkten Befragung größerer Standardfehler der RRT bei der Schätzung der Prävalenz sensibler Merkmale einher. Größere Stichproben können dieses Problem lösen, führen jedoch zu erhöhten Datenerhebungskosten. Ein weiterer Nachteil der RRT liegt in der Komplexität des Befragungsprozesses. Der Befragte muss nun nicht nur eine Antwort auf die ihm gestellte Frage finden, er muss sich auch stets die Regeln der RRT ins Gedächtnis rufen und zudem ein Randomisierungsinstrument (z.B. Münzen oder Würfel) zur Hand haben. Das kann zur kognitiven Überlastung und schließlich zu Beantwortungsfehlern führen. Ein Überblick über empirische Studien, bei denen die RRT im Vergleich zur direkten Befragung gleich hohe oder sogar deutlich niedrigere Anteilsschätzungen sozial unerwünschter Verhaltensweisen lieferte, findet sich bei Holbrook/Krosnick (2010).

- 2 Wenn Studien mit externem Außenkriterium aufgrund datenschutzrechtlicher Aspekte oder fehlender Existenz externer Validierungsdaten nicht möglich sind, dann werden häufig comparative, experimentelle Studien ohne externes Außenkriterium durchgeführt. In letzteren werden Hilfsannahmen getroffen, um Aussagen über die Validität einer Methode machen zu können. So wird angenommen, dass Befragte fälschlicherweise ein Verhalten leugnen können, aber niemals fälschlicherweise ein Verhalten zugeben werden was sie nicht begangen haben. Zudem wird in Abwesenheit von Validierungsdaten a priori angenommen, dass von der Norm abweichendes Verhalten systematisch unterberichtet wird. Gegeben diese Annahmen, werden höhere Anteilsschätzungen als valider eingestuft („more-is-better“ Annahme). Der Begriff der „Validität“ wird im Folgenden genau in diesem Sinne verwendet (vgl. auch Lensvelt-Mulders et al. 2005). Obwohl diese Annahmen in der vorliegenden Studie hoch plausibel sind, sind es letztendlich nur Annahmen und können eine „echte“ Validierungsstudie, also einen Vergleich von Antworten mit bekannten „wahren“ Werten, nicht ersetzen. Letzteres ist und bleibt der Goldstandard.

2.2 Eine Alternative zur RRT: Das Crosswise Modell (CM) und das Triangular Modell (TM)

Es werden zwei neue Befragungsmethoden vorgestellt, die einige Nachteile der RRT überwinden können (Yu et al. 2008): das Crosswise Modell (CM) bzw. das Triangular Modell (TM), welche in Anlehnung an Tian et al. (2009) im Folgenden unter dem Begriff der Non-Randomized Response (NRR) - Modelle zusammengefasst werden. Mathematisch weichen sie nur kaum von der RRT ab³; in ihrer Umsetzung unterscheiden sie sich jedoch erheblich von dieser. Maßgeblich ist dabei der Verzicht auf ein Randomisierungsinstrument (Tian et al. 2009). Yu et al. (2008: 255) fassen die Stärken der NRR-Modelle folgendermaßen zusammen: *„Neither model requires randomizing device, the models are easy to be implemented for both interviewer and interviewee, the interviewee does not face any sensitive questions, and both models can be applied to both face-to-face personal interviews and mail questionnaires.“* Analog zur zuvor vorgestellten RRT liegt auch bei den NRR-Modellen das Ziel darin, validere Schätzungen für ein sensitives Merkmal zu generieren. Dazu ist es notwendig, die Anonymität der Befragten zu erhöhen. Auf diese Weise sollen Vertrauen und Kooperationsbereitschaft der Befragten gesteigert werden.

Das Grundprinzip beider Modelle besteht darin, die Randomisierung mit Hilfe einer zweiten nicht-sensitiven, dichotomen Frage zu gewährleisten (Yu et al. 2008). Man betrachte erneut ein sensitives Merkmal X , wobei die Wahrscheinlichkeit für ein Individuum einer Population, dieses Merkmal zu tragen mit $\pi = P(X = 1)$ bezeichnet wird. Hinzu kommt nun ein zweites nicht-sensitives Merkmal Y , bei dem Unabhängigkeit von X vorausgesetzt wird. Mit $p = P(Y = 1)$ wird demnach die Wahrscheinlichkeit, das nicht-sensitive Merkmal zu tragen, bezeichnet. Diese Wahrscheinlichkeit wird als bekannt vorausgesetzt. Typischerweise wird dazu auf eine Geburtstagsfrage zurückgegriffen (wie z.B. „Sind Sie im Januar, Februar oder März geboren?“). Nachfolgend wird die Operationalisierung der NRR-Modelle in einer Befragungssituation mit Fragebogen erläutert. Ob und wie dies auf andere Befragungsmodalitäten (z.B. Befragungen via Online-Access-Panel oder telefonische Befragungen) übertragen werden kann, ist aufgrund mangelnder empirischer Basis in zukünftigen Studien zu untersuchen.

3 Um genau zu sein, ist der CM Schätzer mit dem klassischen RRT Modell von Warner (1965) formal identisch. Eine formale Darstellung des CM und des TM (d.h. Herleitung der Schätzformeln, Diskussion von asymptotischen Eigenschaften und Effizienzanalysen) mit Bezügen zu bisherigen RRT Designs findet sich in Yu et al. (2008).

Abbildung 2 Antwortschema des Crosswise Modells (Yu et al. 2008)

Crosswise		
	X=0	X=1
Y=0	□	■
Y=1	■	□

□ = {X = 0, Y = 0} ∪ {X = 1, Y = 1}

■ = {X = 0, Y = 1} ∪ {X = 1, Y = 0}

Anmerkung: X - sensitiv, Y - nicht-sensitiv; □ entspricht Antwortoption A, ■ entspricht Antwortoption B

2.2.1 Das Crosswise Modell (CM)

Dem Befragten werden die sensitive sowie die nicht-sensitive Frage in einem Block vorgelegt und er wird instruiert, eine kombinierte Antwort auf beide Fragen zu geben. Die dabei zu befolgenden Antwortregeln sind in Abbildung 2 dargestellt.

Der Befragte wird demnach gebeten Antwortoption A anzukreuzen, falls er beide Fragen gleichermaßen beantwortet, d.h. entweder beide mit „Ja“ oder beide mit „Nein“. Andernfalls, d.h. wenn er eine der beiden Fragen mit „Ja“, die andere mit „Nein“ beantwortet, soll er Antwortoption B ankreuzen. Zu keinem Zeitpunkt der Befragung sollen die beiden Fragen explizit einzeln beantwortet werden. Im CM sind beide Antwortoptionen als nicht-sensitiv zu betrachten. Da die Ausprägungen {X=0, Y=0} bei Option A bzw. {X=0, Y=1} bei Option B implizieren, dass der Befragte nicht Träger des sensitiven Merkmals ist, lassen sowohl Option A als auch Option B keinen eindeutigen Rückschluss darauf zu, welche Ausprägung der Befragte bezüglich des sensitiven Merkmals X hat. Yu et al. (2008) gehen davon aus, dass ein Erkennen dieses Mechanismus die Befragten motiviert, ehrlich auf die Fragen zu antworten, denn ihre Privatsphäre bleibt dadurch vollständig geschützt.

Ausgehend von der Annahme, dass alle Befragten die Methode korrekt anwenden, lässt sich durch einfache mathematische Überlegungen ein Schätzer für $\pi = P(X = 1)$, den Populationsanteil der Befragten, die das sensitive Merkmal tragen, herleiten. Es sei mit λ der unbeobachtete Anteil an „Option A“ - Antworten einer Grundgesamtheit bezeichnet. Dieser lässt sich innerhalb einer hinreichend großen Stichprobe durch den beobachteten Stichprobenanteil schätzen, also: $\hat{\lambda} = \frac{a}{n}$, wobei n die Anzahl aller Befragten in der Stichprobe und a die Anzahl der Befragten, die „Option A“ wählen, darstellen. Für den Erwartungswert des Anteils an „Option A“ - Antworten gilt: $\lambda = p\pi + (1 - p) \cdot (1 - \pi)$, wobei mit $p = P(Y = 1)$ die als bekannt vorausgesetzte Wahrscheinlichkeit, das nicht-sensitive Merkmal Y zu tragen, bezeichnet wird. Nun ergibt sich als Schätzer für π :

$$\hat{\pi}_C = \frac{\hat{\lambda} + p - 1}{2p - 1}$$

Offensichtlich muss p die Bedingung $p \neq \frac{1}{2}$ erfüllen. Für die Varianz des Schätzers gilt:

$$\text{Var}(\hat{\pi}_C) = \frac{\pi \cdot (1 - \pi)}{n} + \frac{p \cdot (1 - p)}{n \cdot (2p - 1)^2},$$

wobei folgende Formel für eine unverzerrte Schätzung der Varianz herangezogen werden kann:

$$\text{Var}(\hat{\pi}_C) = \frac{\hat{\pi}_C \cdot (1 - \hat{\pi}_C)}{n - 1} + \frac{p \cdot (1 - p)}{(n - 1) \cdot (2p - 1)^2}$$

An dieser Stelle wird der formale Bezug zur RRT nach Warner (1965) ersichtlich: Auch wenn beide Techniken in ihrer Implementierung grundverschieden sind, sind die statistischen Berechnungsformeln äquivalent.

Yu et al. (2008) liefern in ihrem Artikel theoretische Analysen und leiten statistische Schätzer für das CM her, führen jedoch keine eigene empirische Studie durch. In der ersten bisher bekannten Studie mit experimentellem Design zeigen Jann et al. (2012), dass das CM deutlich höhere Schätzungen sensitiven Verhaltens liefert als die direkte Befragung. Die Autoren befragten 474 schweizerische und deutsche Universitätsstudenten mittels selbstadministrierter Fragebögen zu Plagiaten in Hausarbeiten. Dazu wurde zwischen Teil- und Vollplagiaten unterschieden. Per Zufallsprinzip wurden die Studenten der direkten oder der CM-Befragung zugewiesen. Die Ergebnisse der Befragung zeigen ein klares Bild: 22,3% der CM-Befragten gaben demnach an, bereits einzelne Textstellen plagiiert zu haben. Dieser Wert liegt signifikant über dem der direkten Befragung: nur 7,3% der direkt befragten Studenten gaben ein Teilplagiat zu. Bei der Frage nach Vollplagiaten ergab sich ebenfalls eine Differenz zu Gunsten des CM (CM: 1,6 %, direkt: 1,0%), allerdings erwies sich diese nicht als signifikant. Die Autoren verweisen dabei auf die generell niedrige Prävalenz von Vollplagiaten in der Population. Insgesamt entlockte das CM in dieser Studie deutlich mehr sozial unerwünschte Antworten als eine direkte Befragung, allerdings müssen zukünftige Studien noch zeigen, ob sich die Ergebnisse auch auf andere Befragungskontexte und nicht-studentische Befragungspopulationen übertragen lassen (Jann et al. 2012).

Abbildung 3 Antwoortschema des Triangular Modells (Yu et al. 2008)

Triangular			
	X=0	X=1	
Y=0	□	■	$\square = \{X = 0, Y = 0\}$ $\blacksquare = \{X = 0, Y = 1\} \cup \{X = 1, Y = 0\} \cup \{X = 1, Y = 1\}$
Y=1	■	■	

Anmerkung: X - sensitiv, Y - nicht-sensitiv; □ entspricht Antwortoption A, ■ entspricht Antwortoption B.

2.2.2 Das Triangular Modell (TM)

Analog zum eben erläuterten Crosswise Modell werden dem Befragten sowohl eine dichotome sensitive als auch eine nicht-sensitive Frage in einem Block vorgelegt. Auch hier sieht es die Methode vor, eine kombinierte Antwort auf beide Fragen zu geben. An dieser Stelle unterscheiden sich die beiden Modelle, wie aus Abbildung 3 hervorgeht.

Der Befragte wählt Antwortoption A, wenn er beide Fragen mit „Nein“ beantwortet. Lautet seine Antwort auf mindestens eine der beiden Fragen „Ja“, das heißt also auf die sensitive oder auf die nicht-sensitive Frage (oder auf beide), so soll er Antwortoption B anzukreuzen. Auch hier ist es essentiell, dass die beiden Fragen nicht einzeln beantwortet werden. Somit erlauben auch im TM beide Antwortoptionen offensichtlich keine klaren Rückschlüsse darüber, ob der Befragte Träger des sensitiven Merkmals ist; sie sind also als nicht-sensitiv zu betrachten (Yu et al. 2008): Option A wählt der Befragte, wenn er weder Träger des sensitiven noch des nicht-sensitiven Merkmals ist $\{X=0, Y=0\}$. Option B enthält die Kategorie $\{X=0, Y=1\}$, welche impliziert, dass der Befragte nicht Träger des sensitiven Merkmals ist. Auch sie ist nicht-sensitiv. Somit bleibt die Privatsphäre des Befragten in jedem Fall geschützt.

Sei die Wahrscheinlichkeit $p = P(Y = 1)$, das nicht-sensitive Merkmal Y zu tragen, wieder als bekannt vorausgesetzt. Und sei zusätzlich mit λ der nicht beobachtbare Anteil der „Option A“ – Antworten in einer betrachteten Grundgesamtheit bezeichnet. Dieser kann analog zum CM durch den beobachteten Stichprobenanteil der „Option A“ – Antworten geschätzt werden: $\hat{\lambda} = \frac{a}{n}$, wobei n die Anzahl aller Befragten in der Stichprobe und a die Anzahl der Befragten, die „Option A“ wählen, darstellen. Für den Erwartungswert des Anteils an „Option A“ – Antworten gilt: $\lambda = (1 - p) \cdot (1 - \pi)$. Ein Umstellen der Gleichung führt zum Schätzer für $\pi = P(X = 1)$ (Yu et al. 2008):

$$\hat{\pi}_T = 1 - \frac{\hat{\lambda}}{1-p}$$

Für die Varianz gilt:

$$\text{Var}(\hat{\pi}_T) = \frac{\pi \cdot (1-\pi)}{n} + \frac{p \cdot (1-\pi)}{n \cdot (1-p)},$$

beziehungsweise als erwartungstreuen Varianzschätzer:

$$\text{Var}(\hat{\pi}_T) = \frac{\hat{\pi}_T \cdot (1-\hat{\pi}_T)}{n-1} + \frac{p \cdot (1-\hat{\pi}_T)}{(n-1) \cdot (1-p)}$$

2.2.3 Stärken der NRR-Modelle

Yu et al. (2008) folgend, beruhen sowohl die CM als auch die TM auf dem Randomisierungsprinzip der Paarung einer sensitiven mit einer nicht-sensitiven Frage und überwinden damit offensichtlich die aus der Notwendigkeit eines Randomisierungsinstruments (z.B. Münze, Würfel) resultierenden Probleme der RRT. Komplizierte Instruktionen werden somit überflüssig und Verzerrungen durch kognitive Verständnisprobleme minimiert. Es ist zudem mit geringeren Nonresponse-Raten zu rechnen. Yu et al. (2008) sehen in den NRR-Modellen einen weiteren Vorteil: bei der RRT wird der Befragte trotz Zufallsverschlüsselung weiter direkt mit der sensitiven Frage konfrontiert. Denn zu beachten ist, dass im oben vorgestellten Warner Modell neben der sensitiven Aussage auch ihre Negierung als sensitiv zu betrachten ist. Damit beantwortet der Befragte in jedem Fall eine ihm womöglich unangenehme Frage. Im Falle der NRR-Modelle wird er jedoch gebeten eine kombinierte Antwort auf die sensitive und eine nicht-sensitive Frage zu geben. Damit wird eine direkte Beantwortung des heiklen Items vermieden.

Im Vergleich der beiden NRR-Modelle miteinander weist das CM dem TM gegenüber einen Vorteil auf. Sowohl beim CM als auch beim TM werden zwar beide Antwortoptionen, Option A und Option B, als nicht-sensitiv betrachtet (Yu et al. 2008). Allerdings beinhaltet das CM für beide Optionen die Möglichkeit, dass der Befragte Träger des sensitiven Merkmals ist. Misstraut ein Befragter der Methode, gibt es für ihn keine offensichtliche Schutzstrategie, um sich mit Sicherheit als Nicht-Träger des sensitiven Merkmals darzustellen. Im Falle des TM findet er diese Strategie jedoch, indem er Option A wählt. Diese bedeutet, dass er weder das sen-

sitive noch das nicht-sensitive Merkmal trägt. Es besteht für den Befragten im TM prinzipiell die Möglichkeit, ausweichend zu antworten. Dagegen zeichnet sich das TM gegenüber dem CM für die meisten Parameterkombinationen durch einen effizienteren Schätzer aus⁴. Dadurch wird eine vergleichsweise geringere Stichprobengröße (und somit geringere Datenerhebungskosten) benötigt um einen gegebenen Standardfehler zu erreichen (Tian et al. 2011). Für das TM liegen bisher keinerlei empirische Befunde vor. Es ist folglich unklar, ob diese Befragungsmethode von den Befragten angenommen und korrekt umgesetzt wird. Diese Lücke schließt nun die vorliegende Studie.

3 Die Studie

In Anlehnung an die CM-Studie von Jann et al. (2012) wurden in der aktuellen Erhebung Leipziger Universitätsstudenten zum heiklen Thema „Plagiate in Hausarbeiten“ befragt. Es wurde sowohl ein Fragebogen mit dem TM als auch ein direkter Fragebogen implementiert. Für den Vergleich der beiden Befragungsmethoden wurde auf die bekannte „more-is-better“ Annahme zurückgegriffen. Unsere Hauptforschungshypothese ist, dass unter Anwendung des TM mehr Studierende ein Plagiat berichten, als in der direkten Befragung.

Nicht nur wegen der medialen Präsenz ist Plagiiere mehr als nur ein akademisches Tabuthema. Ein Student, dem ein Plagiat nachgewiesen wird, muss nicht nur mit sozialen und prüfungsrechtlichen Sanktionen seitens der Dozenten und der Mitstudierenden rechnen, ihm drohen im schlimmsten Fall auch strafrechtliche Konsequenzen, sobald der Betroffene, dessen Werk plagiiert wurde, Anzeige erstattet. Rechtliche Grundlage hierfür ist das Urheberrecht (*Gesetz über Urheberrecht und verwandte Schutzrechte*, Bundesministerium der Justiz 2008). So unterscheidet die „Satzung zur Sicherung guter wissenschaftlicher Praxis“ der Universität Leipzig (2002) umfänglich in arbeits- und dienstrechtliche, zivilrechtliche, akademische und strafrechtliche Konsequenzen. Es ist also zu erwarten, dass ein Student, der diese Konsequenzen ausschließen möchte, in einer direkten Befragung nicht angeben wird, bereits ein Plagiat eingereicht zu haben. Die Frage nach Plagiaten ist somit sensitiv. In unserer Erhebung unterscheiden wir zwischen Teil- und Vollpla-

4 Yu et al. (2008, S. 261) zeigen, dass die Varianz des CM-Schätzers stets größer ist als die des TM-Schätzers, solange die Wahrscheinlichkeit p , das nicht-sensitive Merkmal Y zu tragen, den Wert $2/3$ nicht übersteigt. Bei einem Wert für p von über $2/3$ hängt das Verhältnis der Varianzen zusätzlich von der Wahrscheinlichkeit π , das sensitive Merkmal zu tragen, ab.

giaten. Unter Teilplagiaten ist im Allgemeinen die Übernahme einzelner Textstellen aus einem fremden Werk zu verstehen, ohne die besagten Textstellen als Zitat zu kennzeichnen. Die unrechtmäßige Übernahme eines gesamten Werkes, sei es wörtlich oder paraphrasiert, erfüllt dagegen den Tatbestand des Vollplagiats.

3.1 Der Fragebogen

Im Rahmen eines experimentellen Designs wurden die Studierenden nach dem Zufallsprinzip entweder zur direkten Befragung oder zum TM zugewiesen. Mit Ausnahme des Experimentalteils, der die beiden sensitiven Fragen zu Teil- und Vollplagiaten enthält, sind die beiden Fragebogenversionen identisch. Insgesamt orientierte sich der Fragebogen an der CM-Studie von Jann et al. (2012). Der Wortlaut der beiden sensitiven Fragen ist wie folgt:

- **Teilplagiat:** *„Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Abschlussarbeit, Semesterarbeit etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne dies als Zitat zu kennzeichnen?“*
- **Vollplagiat:** *„Haben Sie schon einmal einen Großteil einer Arbeit durch eine fremde Person schreiben lassen oder eine fremde Arbeit (z.B. von www.hausarbeiten.de) als Ihre eigene ausgegeben?“*

Im direkten Fragebogen beantworten die Studenten die beiden Fragen direkt, in der TM-Kondition werden sie dagegen nicht direkt beantwortet. Stattdessen sind beide sensitive Fragen jeweils mit einer unabhängigen nicht-sensitiven Frage gepaart und die Befragten werden gebeten eine kombinierte Antwort zu geben. Die nicht-sensitiven Fragen lauten folgendermaßen:

- **Teilplagiat:** *„Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?“*
- **Vollplagiat:** *„Hat Ihr Vater in den Monaten Oktober, November oder Dezember Geburtstag?“*

Bei den beiden nicht-sensitiven Fragen haben wir einen Näherungswert für die Wahrscheinlichkeit einer „Ja“ Antwort ermittelt: ausgehend von einer Altersspanne von circa 18 bis 25 Jahren der befragten Studierenden und einem durchschnittlichen Alter von 24 Jahren bei der Geburt, sind die Eltern der befragten Studierenden mit hoher Wahrscheinlichkeit zwischen 1962 und 1969 geboren. Für diese Jahrgänge wurden beim Statistischen Bundesamt Deutschland die nach Monaten aufgeschlüsselten Geburtsraten erfragt und eine durchschnittliche Geburtenrate ermittelt. Die so ermittelte Wahrscheinlichkeit p_1 in den Monaten Januar/Februar/März geboren zu sein beträgt 0,26, die Wahrscheinlichkeit p_2 für die Monate

Oktober/November/Dezember hingegen 0,23.⁵ Abbildung 4 visualisiert die Implementierung des TM im Fragebogen der aktuellen Studie.

3.2 Stichprobe und Datenerhebung

Die Befragung fand schriftlich in der Vorlesung „Grundzüge der Soziologie“ statt. Sie wurde in den Wintersemestern 2009/2011 bzw. 2010/2011 durchgeführt. Es handelt sich hierbei um ein Convenience-Sample, d.h. die Vorlesung wurde nicht zufällig aus allen Vorlesungen in Leipzig ausgewählt. Neben weiteren Prüfungsleistungen, schreiben alle Studenten im Rahmen dieser Veranstaltung eine Hausarbeit. Der Abgabetermin der Hausarbeit lag im Semester jeweils einige Wochen vor Vorlesungsende. Die Befragungen fanden immer nach der Abgabe der Hausarbeiten statt. Somit hatte jeder Befragte potentiell eine Möglichkeit zu plagiierten. In der Studie wurden alle in der jeweiligen Vorlesung anwesenden Studenten befragt. Es wurden keinerlei monetäre oder sonstige materielle Anreize für die Teilnahme angeboten. Nach einer kurzen Vorstellung der Studie wurden die Fragebögen ausgeteilt. Während der Befragung wurden keinerlei Komplikationen oder Probleme beobachtet.⁶

Die Studenten wurden zufällig einer der beiden Befragungskonditionen – direkt versus TM – zugewiesen und füllten den Fragebogen selbst-administriert aus.⁷ Die Randomisierung fand im Verhältnis 1:3 statt.⁸ Das Verhältnis von 1:3 wurde mit dem Ziel gewählt, auch für die TM-Kondition einen akzeptablen Standardfehler zu erreichen. In der Stichprobe waren im Prinzip alle zur Modulabschlussklausur zugelassenen Studierenden in den betrachteten Jahrgängen (N=621). Von den 621 zur Modulabschlussklausur zugelassenen Studierenden waren 434 zur Befragung anwesend. 382 von ihnen füllten schließlich den Fragebogen aus. In der TM-Kondition wurden 281 Studierende befragt, den direkten Fragebogen füllten 101 Stu-

5 Diese Werte können Schwankungen unterliegen, allerdings sind sie als Näherungswerte besser als ein aus der Annahme einer Gleichverteilung resultierendes p von jeweils 0,25. Zudem wurden auch die Geburtsraten von 1960 bis 1974 überprüft. Es fanden sich die gleichen Wahrscheinlichkeiten.

6 Lediglich eine Studentin fragte nach, ob bei ihrem Fragebogen eine Seite fehlen würde. Sie hatte einen direkten Fragebogen zugeteilt bekommen, der eine Seite kürzer ist als der TM-Fragebogen.

7 Aufgrund der Randomisierung zu einer der beiden Befragungskonditionen ist eine hohe interne Validität sichergestellt. Es wird angenommen, dass sich die Ausfälle (sprich: die am Tag der Befragung nicht anwesenden Studenten) nicht systematisch zwischen den Experimentalgruppen unterscheiden.

8 Hierbei wurden die Fragebögen im Vorfeld dem Verhältnis entsprechend vorsortiert und dann in der Vorlesung der Reihe nach ausgeteilt.

Abbildung 4 Instruktion des Triangular Modells

<p>Im nun folgenden Teil der Befragung wende ich eine neuartige Befragungstechnik an, um Ihre Privatsphäre bei diesen für Sie möglicherweise unangenehmen Fragen zusätzlich zu schützen. Lesen Sie sich bitte zunächst die Instruktionen genau durch und beantworten Sie dann die Fragen.</p> <p>Es werden jeweils zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit <i>Ja</i> oder mit <i>Nein</i>), schreiben Sie dies aber <u>nicht</u> auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar nach folgenden Regeln:</p> <p>Lautet Ihre Antwort auf beide Fragen <i>Nein</i>, so setzen Sie Ihr Kreuz bei (A).</p> <p>Lautet Ihre Antwort auf mindestens eine der beiden Fragen <i>Ja</i>, so setzen Sie Ihr Kreuz bei (B).</p> <p>Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne. Mittels statistischer Verfahren kann ich aber berechnen, für wie viele Personen <u>insgesamt</u> die jeweils zweite Frage zutrifft.</p>
12) Block 1
<p>1. Frage: <i>Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?</i></p> <p>2. Frage: <i>Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Abschlussarbeit, Semesterarbeit, etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?</i></p> <hr/> <p>Wie lauten die Antworten auf die beiden Fragen?</p> <p><input type="radio"/> (A) auf beide Fragen <i>Nein</i></p> <p><input type="radio"/> (B) auf mindestens eine der beiden Fragen <i>Ja</i></p>
13) Block 2
<p>1. Frage: <i>Hat Ihr Vater in den Monaten Oktober, November oder Dezember Geburtstag?</i></p> <p>2. Frage: <i>Haben Sie schon einmal einen Großteil einer Arbeit durch eine andere Person schreiben lassen oder eine fremde Arbeit (z.B. von www.hausarbeiten.de) als Ihre eigene ausgegeben?</i></p> <hr/> <p>Wie lauten die Antworten auf die beiden Fragen?</p> <p><input type="radio"/> (A) auf beide Fragen <i>Nein</i></p> <p><input type="radio"/> (B) auf mindestens eine der beiden Fragen <i>Ja</i></p>

dierende aus. Die für beide Jahrgänge aggregierte Ausschöpfungsquote liegt bei 62%. Bezogen auf die anwesenden Studierenden beträgt die Response-Rate 88%.

4 Empirische Befunde

4.1 Soziodemographie in den Befragungskonditionen

Das Durchschnittsalter der befragten Studenten betrug 21,7 Jahre (mit einer Standardabweichung von 2,63 Jahren), 61% von ihnen waren weiblich. Über 50% befanden sich zum Befragungszeitpunkt im ersten Studienjahr. Über 97% der Studenten sind in Studiengängen der neuen Abschlüsse immatrikuliert, von denen sich etwa 43% zurzeit in einem Bachelor-Studium befinden. Alle Befragten haben bereits mindestens eine Hausarbeit geschrieben. Tabelle 1 listet ausgewählte Befragtenmerkmale aufgeschlüsselt nach den beiden Befragungskonditionen auf. Es wurden jeweils paarweise χ^2 -Unabhängigkeitstests zwischen der Befragungskondition und den Befragtenmerkmalen durchgeführt. Keiner der χ^2 -Werte ist auf dem 5% Niveau signifikant. Es bestehen keine systematischen Unterschiede zwischen den beiden Befragungskonditionen.

Tabelle 1 Soziodemographie in den beiden Befragungskonditionen

	Direkt (N=101)		Triangular (N=281)		χ^2 (p)
	n	%	n	%	
Geschlecht					
Männlich	38	37,6 %	110	39,2 %	0,063 (0,802)
Weiblich	62	61,4 %	169	60,1 %	
k.A.	1	1,0 %	2	0,7 %	
Alter					
18 – 21	56	55,5 %	151	53,7 %	1,476 (0,688)
22 – 25	39	38,6 %	107	38,1 %	
26 – 29	6	5,9 %	17	6,1 %	
30 und älter	0	0 %	4	1,4 %	
k.A.	0	0 %	2	0,7 %	
Studienjahr					
1	52	51,5 %	161	57,3 %	3,869 (0,276)
2	20	19,8 %	61	21,7 %	
3	28	27,7 %	52	18,5 %	
4	0	0,0 %	4	1,4 %	
6	1	1,0 %	0	0,0 %	
7	0	0,0 %	1	0,4 %	
k.A.	0	0,0 %	2	0,7 %	
Angestrebter Abschluss					
Bachelor	39	38,6 %	118	42,0 %	0,530 (0,767)
Master	58	57,4 %	153	54,4 %	
Diplom	1	1,0 %	3	1,1 %	
Magister	0	0,0 %	1	0,4 %	
Staatsexamen	1	1,0 %	2	0,7 %	
Sonstige	0	0,0 %	2	0,7 %	
k.A.	2	2,0 %	2	0,7 %	
Bisher geschriebene Hausarbeiten					
Keine	0	0,0 %	0	0,0 %	1,908 (0,385)
1 bis 2	57	56,4 %	180	64,0 %	
3 bis 4	19	18,8 %	46	16,4 %	
5 oder mehr	25	24,8 %	55	19,6 %	

Anmerkung: Item Nonresponse führt zu leichten Schwankungen in den Fallzahlen

Tabelle 2 Plagiatsraten nach Befragungskondition

	Direkt (N=101)	Triangular (N=280) ⁹	Differenz
Teilplagiat			
Anteil (in %)	9,9	18,0	8,1 (p = 0,104)
Standardfehler	3,0	4,0	4,9
95 % Konfidenzintervall	[4,1 , 15,7]	[10,2 , 25,7]	[-1,7 , 17,8]
Vollplagiat			
Anteil (in %)	0,0	0,8	0,8 (p = 0,805)
Standardfehler	0,0	3,3	3,3
95 % Konfidenzintervall	[0 , 0]	[-5,7 , 7,3]	[-5,7 , 7,4]

4.2 Plagiatsraten

Tabelle 2 zeigt die Prävalenzraten für Teil- und Vollplagiate aufgeschlüsselt nach den beiden Befragungskonditionen.

Die Differenz der geschätzten Anteile zwischen TM- und direkter Befragung beträgt für die Frage nach Teilplagiaten 8,1%. Es gaben in der TM-Kondition demnach mehr Studenten ein Teilplagiat zu als in der direkten Befragung (18,0% in der TM versus 9,9% in der direkten Version). Allerdings liegt das empirische Signifikanzniveau auf der Grundlage eines zweiseitigen Tests mit $p = 0,104$ über dem konventionellen Niveau von 5%¹⁰. Somit lassen sich in der vorliegenden Studie primär nur deskriptive Aussagen treffen. Für die Frage nach einem Vollplagiat unterscheiden sich die beiden Schätzungen nicht. In der direkten Befragung gab kein Student ein Vollplagiat an. In der TM-Kondition wurde ein Anteilswert von 0,8% geschätzt. Auch wenn dieser Wert leicht über dem der direkten Befragung liegt, weicht er nicht signifikant von 0% ab – leicht erkennbar am Konfidenzintervall. Für die Frage nach Vollplagiaten führt das TM demnach zu keiner nennenswert höheren Prävalenzschätzung als die direkte Befragung.

- 9 In der TM-Kondition verweigerte ein Student komplett die Beantwortung der Fragen nach Plagiaten.
- 10 Für den Test der statistischen Signifikanz wurde als Prüfgröße die Differenz der Plagiatsraten in der direkten und der TM-Kondition, gewichtet mit der Wurzel der gemeinsamen Varianz, gebildet:

$$t = \frac{|\hat{\pi}_d - \hat{\pi}_t|}{\sqrt{\frac{\hat{\pi}_d \cdot (1 - \hat{\pi}_d)}{n_d} + \text{Var}(\hat{\pi}_t)}}$$

wobei $\hat{\pi}_d$ der direkten und $\hat{\pi}_t$ der Triangular-Plagiatsrate entspricht (n_d ist die Stichprobengröße der direkten Befragung). Diese Prüfgröße entstammt einer t-Verteilung mit $n_d + n_t - 2$ Freiheitsgraden.

Tabelle 3 Plagiatsraten der aktuellen Triangular-Studie und der Crosswise-Studie von Jann et al. (2012)

	Crosswise-Studie (Jann et al. 2012)		Triangular-Studie (aktuelle Erhebung)	
	Direkt	Crosswise	Direkt	Triangular
N	96	310	101	280
Teilplagiat	7,3% (2,7)	22,3% (5,5)	9,9% (3,0)	18,0% (4,0)
Vollplagiat	1,0% (1,0)	1,6% (5,0)	0% (0,0)	0,8% (3,3)

Anmerkung: Standardfehler in Klammern.

5 Vergleich des TM und CM

Tabelle 3 kontrastiert die Ergebnisse der vorliegenden TM-Studie mit der CM-Studie von Jann et al. (2012).

Es zeigt sich, dass sowohl das CM mit 22,3% als auch das TM mit 18,0% für die Frage nach Teilplagiaten deutlich höhere Schätzungen erzielen als die direkte Befragung mit 7,3% bzw. 9,9%. Ein ebenso einheitliches Ergebnis zeigt sich für die Frage nach Vollplagiaten. Unter Anwendung dieser beiden Modelle konnten keine nennenswert höheren Raten ermittelt werden. Im direkten Vergleich des CM und des TM wird für die Frage nach Teilplagiaten eine Differenz von 4,3% zugunsten des CM ermittelt. Allerdings zeigt sich mit den deutlich geringeren Standardfehlern des TM ein (verglichen mit dem CM) klarer Vorteil bezüglich der statistischen Effizienz. Insgesamt ist nur ein grober und skizzenhafter Vergleich des CM mit dem TM möglich, da sie in verschiedenen Erhebungen und Stichproben evaluiert wurden¹¹. Auch wird in diesem Zusammenhang auf die Berechnung von Signifikanztests verzichtet, da sie nur wenig sinnvoll erscheinen. In einem nächsten Schritt könnte eine Anschlussstudie die CM und die TM in ein einheitliches Design implementieren und

11 Während in der vorliegenden TM-Studie ausschließlich Studierende der Universität Leipzig befragt wurden, wurden in der CM-Studie neben Leipziger Universitätsstudenten auch Studierende der ETH Zürich und der LMU München befragt. Zudem wurde bei der Ermittlung der Plagiatsraten in der CM-Studie eine Gleichverteilung für die nicht-sensitive Geburtstagsfrage angenommen. Für die Berechnungen lag die Wahrscheinlichkeit p , in den Monaten Januar/Februar/März geboren zu sein, demnach bei 0,25. Somit sind die beiden Studien nur bedingt miteinander vergleichbar und die vorliegende Darstellung dient lediglich einem groben Eindruck.

dann direkt miteinander kontrastieren (mit einer direkten Befragung als Kontrollgruppe).

6 Diskussion

In unserer Studie konnte gezeigt werden, dass unter Anwendung des TM deutlich mehr Studenten ein Teilplagiat zugeben als in einer direkten Befragung. Damit kann unsere Hauptforschungshypothese (mehr sozial unerwünschte Antworten unter anonymen Befragungsbedingungen) zumindest für die Frage nach Teilplagiaten bestätigt werden. Die beobachtete Differenz von 8,1% zwischen den beiden Konditionen verfehlt allerdings das konventionelle 5%-Signifikanzniveau. Somit lassen sich die vorliegenden Befunde nicht generalisieren und sind somit eher als explorativ und vorläufig zu betrachten. Für Vollplagiate konnte dagegen kein nennenswerter Unterschied zwischen den Experimentalkonditionen festgestellt werden. Die in der direkten und in der TM-Befragung ermittelten Anteilsschätzungen weichen statistisch nicht signifikant von Null und auch nicht voneinander ab.

Die Frage ist nun warum sich für das Item zu Vollplagiaten in der TM-Kondition kein nennenswerter Unterschied zur direkten Befragung zeigt. Dies könnte potentiell drei Ursachen haben: (1) Die Abgabe einer vollständig plagiierten Arbeit birgt ein sehr hohes Entdeckungsrisiko und ist in der Regel mit gravierenden Konsequenzen verbunden. Demzufolge sind die tiefen Schätzungen auf Grundlage des TM nicht auf ein Versagen der Methode zurückführbar, sondern auf die generelle Seltenheit von Vollplagiaten. (2) Die Befragten haben sich bei diesem besonders heiklen Item bewusst nicht an die Antwortregeln gehalten. Dies kann zum einen auf Befragte zutreffen, die bereits voll plagiiert haben und nun die sichere Antwortstrategie für den Selbstschutz – „Option A“ – wählen. Zum anderen wären auch Befragte denkbar, die nicht plagiiert haben, auf die jedoch das nicht-sensitive Item zutrifft. Diese könnten möglicherweise befürchten, dass ein Ankreuzen von „Option B“ fälschlicherweise als Vollplagiat interpretiert wird und würden deshalb aus Selbstschutz womöglich „Option A“ wählen. (3) Es kann nicht als gesichert gelten, dass tatsächlich auf 23,3 % der Studierenden das nicht-sensitive Merkmal zutrifft, das heißt, dass sie einen Vater haben, der in den Monaten Oktober, November oder Dezember Geburtstag hat. Der Wert ist daher zufälligen Schwankungen unterlegen, die erst bei sehr großen Stichproben an Einfluss auf die Schätzung der Anteilswerte verlieren. Zudem wird es durch die Einführung eines Randomisierungsprozesses so gut wie unmöglich bei kleinen Prävalenzen nahe Null zuverlässige und effiziente Schätzungen zu erreichen (vgl. Jann et al. 2012: 46).

Welches der oben angeführten Argumente für die Erklärung der kleinen Anteilsschätzungen beim Vollplagiats-Item zutreffend ist kann nicht ad hoc entschieden werden und sollte deshalb in zukünftigen Studien detaillierter untersucht werden. Auch könnte in einem nächsten Schritt die TM mit anderen indirekten Befragungsmethoden, wie beispielsweise der RRT, verglichen werden.

Literatur

- Boruch, R. F., 1971: Assuring confidentiality of responses in social research: a systematic analysis. *The American Psychologist* 26: 413-430.
- Bundesministerium der Justiz, 2008: Das Urheberrecht und verwandte Schutzrechte. Berlin: Bundesministerium der Justiz. <http://www.gesetze-im-internet.de/urhrg> (07.11.2013).
- Coutts, E., B. Jann, I. Krumpal und A.-F. Näher, 2011: Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions. *Jahrbücher für Nationalökonomie und Statistik* 231 (5+6): 749-760.
- Greenberg, B. G., A. A. Abul-Ela, W. R. Simmons und D. G. Horvitz, 1969: The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association* 64: 520-539.
- Holbrook, A. L. und J. A. Krosnick, 2010: Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity. *Public Opinion Quarterly* 74: 328-343.
- Horvitz, D. G., B. V. Shah und W. R. Simmons, 1967: The unrelated question randomized response model. S. 65-72 in: *Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: American Statistical Association.*
- Höglinger, M., B. Jann und A. Diekmann (2013): Sensitive Questions in Online Surveys: An Experimental Comparison of the RRT and the Crosswise Model. Unpublished Working Paper, ETH Zurich and University of Bern.
- Jann, B., J. Jerke und I. Krumpal, 2012: Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly* 76: 32-49.
- Krumpal, I. 2012: Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning. *Social Science Research* 41(6): 1387-1403.
- Krumpal, I. 2013: Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity* 47: 2025-2047.
- Krumpal, I. und A.-F. Näher, 2012: Entstehungsbedingungen sozial erwünschten Antwortverhaltens: Eine experimentelle Studie zum Einfluss des Wordings und des Kontexts bei unangenehmen Fragen. *Soziale Welt* 63: 65-89.
- Kuk, A.Y.C., 1990: Asking Sensitive Questions Indirectly. *Biometrika* 77: 436-438.
- Lensvelt-Mulders, G., J. Hox, P. van der Heijden und C. Maas, 2005: Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods & Research* 33: 319-348.
- Paulhus, D.L., 1984: Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46: 598-609.
- Paulhus, D.L. und O.P. John, 1998: Egoistic and Moralistic Biases in Self-Perception: The Interplay of Self-Deceptive Styles With Basic Traits and Motives. *Journal of Personality* 66: 1025-1060.

- Preisendörfer, P. 2008: Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu. *ETH Zurich Sociology Working Papers* 6, ETH Zurich, Chair of Sociology.
http://repec.ethz.ch/ets/papers/preisendoerfer_sensitive_questions.pdf (08.11.2013)
- Sattler, S. 2007: Plagiate in Hausarbeiten – Erklärungsmodelle mit Hilfe der Rational Choice Theorie. Hamburg, Kovač.
- Tian, G.-L., K.Ch. Yuen, M.-L. Tang und M.T. Tan, 2009: Bayesian non-randomized response models for surveys with sensitive questions. *Statistics and its interface* 2: 13-25.
- Tian, G.-L., M.-L. Tang, Z. Liu, M. Tan und N.-S. Tang, 2011: Sample size determination for the non-randomised triangular model for sensitive questions in a survey. *Statistical Methods in Medical Research* 20: 159-173.
- Tourangeau, R. und T. Yan, 2007: Sensitive Questions in Surveys. *Psychological Bulletin* 133: 859-883.
- Umesh, U. N. und R. A. Peterson, 1991: A Critical Evaluation of the Randomized Response Method: Applications, Validation and Research Agenda. *Sociological Methods Et Research* 20: 104-38.
- Universität Leipzig, 2002: *Satzung zur Sicherung guter wissenschaftlicher Praxis*. Leipzig: Universität Leipzig. <http://www.zv.uni-leipzig.de/forschung/satzung>. (07.11.2013).
- Warner, St. L., 1965: Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63-69.
- Yu, J.-W., G.-L. Tian und M.-L. Tang, 2008: Two new models for survey sampling with sensitive characteristic: Design and analysis. *Metrika* 67: 251-263.

Anschrift der Autoren

Julia Jerke (Korrespondenzautorin)
Universität Leipzig
c/o Ivar Krumpal, Institut für Soziologie
Beethovenstrasse 15
04107 Leipzig
E-Mail: julia.jerke@sozio.uni-leipzig.de

Ivar Krumpal
Universität Leipzig
Institut für Soziologie
Beethovenstrasse 15
04107 Leipzig
E-Mail: krumpal@sozio.uni-leipzig.de

Zur Multiplen Imputation bei Item Non-response am Beispiel eines Analysemodells zum Abstimmungsverhalten bei Bürgerentscheiden

Multiple Imputation Concerning Item Nonresponse Using the Example of an Analysis Model for Direct Democratic Voting Behaviour

Martin Messingschlager und Marcel Preisig

Zusammenfassung

In diesem Artikel werden Fragestellungen unterschiedlicher Ebenen behandelt. Zum einen befasst sich der Aufsatz mit dem politischen Verhalten bei direktdemokratischen Abstimmungen auf kommunaler Ebene. Konkret wird ein multivariates Logitmodell geschätzt, das themenspezifische Variablen als wichtigste Determinanten eines in Lichtenfels (Oberfranken) stattgefundenen Bürgerentscheids identifiziert. Die zweite Ebene, die zugleich den Kern der Arbeit darstellt, behandelt einen Vergleich von fall-reduzierendem Vorgehen gegenüber einer Imputation als Korrekturmethode auf Basis des realwissenschaftlichen Modells der ersten Fragestellungsebene. Die Verknüpfung der realwissenschaftlichen mit einer methodischen Fragestellung führt dabei vor Augen, wie wichtig ein sensibler Umgang mit fehlenden Daten – auch in den Sozialwissenschaften – ist. Der erhobene Daten-

Abstract

In this article, we deal with problems of two different natures. On the first level, the paper handles with direct democratic voting behavior on local level, concretely via logistic regression models which identify issue oriented variables as the most important determinants. The second level, which is the actual core of this article, discusses two approaches for dealing with missing values: case-reducing versus imputation of the missing data. The linkage between a real scientific and a methodological question shows the importance of a sensitive handling of missing data – not least in social sciences. Our sample is used for the comparison of Completes-Cases-Analyses with Multiple Imputation. By means of a procedure which combines the advantages of a simulation with those of real data we create a valid base for our comparison. The analytical and graphical results show clearly



satz dient als Ausgangspunkt für einen Vergleich der Complete-Cases-Methode und der Multiplen Imputation als Ergänzungsverfahren bei Item Nonresponse. Mit Hilfe eines Verfahrens, das die Vorteile von Simulationen mit denen realer Daten kombiniert, wird eine aussagekräftige Grundlage für den Methodenvergleich geschaffen. An den Ergebnissen, die sowohl grafisch als auch analytisch aufbereitet werden, wird deutlich, dass sich die Multiple Imputation in der vorliegenden Stichprobe zur Korrektur von Item Nonreponse im Vergleich zur Complete-Cases-Analyse empfiehlt.

that it's appropriate to use Multiple Imputation instead of using a Complete-Cases-Analysis.

1 Einleitung

Fehlende Werte stellen ein auch in den Sozialwissenschaften hinlänglich bekanntes Problem dar: So werden Datennutzer in nahezu allen Surveys damit konfrontiert, dass Befragte die Beantwortung einzelner Fragen, oder gar die gesamte Teilnahme, verweigern oder aus Unwissenheit keine Antworten geben.¹ Neben einer Verringerung der Fallzahl und dem damit verbundenen Informationsverlust stellt sich bei einem solchen Nonresponse ebenso die Frage, ob der Ausfall lediglich zufälliger oder eher systematischer Natur ist. Letzterer wiederum kann zu verzerrten Schätzparametern führen, was die Qualität der Datenbasis zusätzlich verringert.

Der vorliegende Beitrag nimmt sich dieses Problems an und dient einem Methodenvergleich zum Umgang mit Item Nonresponse. Dabei geht es in erster Linie darum, anhand eines konkreten Beispiels aus der Sozialwissenschaft zu zeigen, dass es z.B. in Form einer Multiplen Imputation (MI) eine relativ einfach anwendbare Alternative zur Complete-Cases-Analyse (CC) gibt.² Für den Methodenvergleich werden zunächst Analysemodelle auf Grundlage von multipel imputierten Datensätzen spezifiziert. Um die Sensitivität der berechneten Regressionsschätzer

- 1 Spricht man im Falle teilweiser Beantwortung von Item Nonresponse, wird beim sogenannten Unit Nonresponse hingegen kein einziges der interessierenden Merkmale erhoben (zur Übersicht vgl. Groves/Fowler/Couper/Leplowski/Singer/Tourangeau 2004).
- 2 Obwohl in der Literatur hinlänglich die Vorteile von MI gegenüber CC diskutiert worden sind, findet sich beispielsweise unter den offiziell registrierten Veröffentlichungen des ALLBUS nahezu keine andere Vorgehensweise als CC bei Item Nonresponse (vgl. Messingschlagler 2012). Ein Grund hierfür könnte sein, dass entsprechende Veröffentlichungen aus der Statistik evtl. nur unzureichend in den Sozialwissenschaften rezipiert werden.

zu testen, wird anschließend eine simulative Grundgesamtheit auf Basis der empirischen Stichprobe erzeugt, aus welcher wiederum 1000 Zufallsstichproben gezogen werden, die im Einzelnen imputiert und analysiert werden. Die gleiche Vorgehensweise wird ebenso für CC angewandt, also ohne eine Ergänzung der fehlenden Werte.³ Die vorgenommenen Vergleiche weisen auf eine weitaus höhere Stabilität (in Form einer geringeren Varianz der Schätzparameter und deren p-Werte) einer MI gegenüber der Verwendung von CC hin. Um dieses Ergebnis nochmals zu evaluieren, werden daraufhin informationsreichere Imputationsmodelle spezifiziert und zur Überprüfung der Stabilität herangezogen.

Als Datenbasis dient eine Erhebung, die im Vorfeld eines Bürgerentscheids im oberfränkischen Lichtenfels durchgeführt wurde. Hierbei wurde über die Frage abgestimmt, ob sich die Bürger für oder gegen die Einstellung eines Bauleitverfahrens zum Bau eines Fachmarktzentrums (FMZ) im Lichtenfelser Mainau-Gebiet aussprechen.

Der Abstimmung vorausgegangen war ein Beschluss des Stadtrates, der sich am 13. September 2010 mit 21:10 Stimmen für die Einleitung des Bauleitverfahrens entschieden hatte. Daraufhin gründete sich die Bürgerinitiative „Rettet das Herz der Lichtenfelser Innenstadt“, welche den Bau ablehnte und ein entsprechendes Bürgerbegehren initiierte. Der Initiative gelang es, 2.680 Unterschriften zu sammeln, was bei einer benötigten Anzahl von lediglich 1.326 Befürwortern zu einem erfolgreichen Begehren führte. Da sich der Stadtrat jedoch weiterhin für das Projekt aussprach, fand der entsprechende Bürgerentscheid am 20. März 2011 statt. Bei diesem entschieden sich fast 70 Prozent der Bürger (exakt: 5.602) für das Projekt, ca. 30 Prozent dagegen (exakt: 2.496) (vgl. Stadt Lichtenfels 2011).

Auf Grundlage theoretischer Überlegungen wird im Folgenden der Frage nachgegangen, welche Determinanten das Abstimmungsverhalten beim Bürgerentscheid beeinflusst haben könnten. Entsprechende Hypothesen werden im folgenden Abschnitt abgeleitet. Dem theoretischen Analyserahmen soll jedoch nur verhältnismäßig wenig Platz eingeräumt werden; im Fokus stehen die statistischen und methodischen Überlegungen zum systematischen Datenausfall und der Korrektur des Nonresponse – vor allem mit Blick auf die nur sehr geringe Datenmenge von 220 Beobachtungen.

3 Hierbei werden lediglich diejenigen Fälle für die Regressionsanalyse verwendet, welche für alle Merkmale des Modells beobachtete Werte aufweisen.

2 Theoretische Überlegungen

Bürgerbegehren und -entscheide (also durch Bürger initiierte Abstimmungen auf kommunaler Ebene) erfreuen sich seit geraumer Zeit einer immer größer werdenden Beliebtheit (vgl. Weixner 2006: 18; Mehr Demokratie e.V.). Umso erstaunlicher erscheint es da, dass sich die empirische politische Verhaltensforschung diesem Thema bislang nur marginal widmet. Stattdessen wird vornehmlich die Frage diskutiert, wie direktdemokratische Verfahren aus normativer und institutioneller Sichtweise zu werten seien.⁴

Auf Länderebene hingegen nimmt die Anzahl empirischer Arbeiten zu: Im Fokus zahlreicher Analysen stand hierbei u.a. die Volksabstimmung zum Nicht-raucherschutzgesetz in Bayern (vgl. Schoen 2011; 2012; Schoen/Glantz/Teusch 2011a; 2011b). Hierbei erwies sich eine modifizierte Version des klassischen Michigan-Modells zur Erklärung von Wahlverhalten (vgl. Campbell/Gutin/Miller 1954; Campbell/Gutin/Miller/Stokes 1960) als adäquater Theorierahmen zur Erklärung des Abstimmungsverhaltens: Laut klassischem Modell ließe sich die Wahlentscheidung mit Hilfe eines Kausalitätstrichters erklären, an dessen dünnem Ende letztlich die Wahlentscheidung stehe, die von politischen Einstellungen und Erfahrungen determiniert sei. Als effektvolle Determinanten werden hierbei die affektive Bindung an eine Partei (Parteiidentifikation) sowie die Bewertungen der Spitzenkandidaten (Kandidatenorientierung) und der parteipolitischen Policies (Sachfragenorientierung) spezifiziert, wobei der Parteiidentifikation (PID) ein besonderer Effekt zugebilligt wird: Diese stehe als langfristig ausgebildete Komponente kausal vor den beiden weiteren Einstellungen (zur weiteren Übersicht vgl. etwa Schoen/Weins 2005).

In einer modifizierten Variante des Michigan-Modells wird bei Schoen (2011; 2012) sowie Schoen/Glantz/Teusch (2011a; 2011b) nun der Sachfragenorientierung der vermeintlich stärkste Effekt zugeschrieben: Schließlich stand beim Volksentscheid in Bayern, anders als bei Wahlen, nicht die Entscheidung für oder gegen eine Partei, sondern die Meinung gegenüber einer Sachfrage im Mittelpunkt. Gleich-

4 Zur Frage, ob Bürger das benötigte Wissen zur Abgabe einer fundierten Meinung bei Volksentscheiden aufweisen, vgl. Oberreuter 2002; Decker 2005; Mendelsohn/Cutler 2000. Zum Argument, dass Entscheide selbst zu erhöhtem Wissen bei Bürgern führen, vgl. qualitative Arbeiten von Kirchgässner/Feld/Savioz 1999; Frey/Kirchgässner 1993. Für quantitative Arbeiten hierzu vgl. Smith 2002; Benz/Stutzer 2004; Christin/Hug/Scarini 2002. Zur Meinungsänderung durch Kampagneneffekte vgl. de Vreese/Semetko 2004; Joslyn/Haider-Merkel 2000; Kriesi 2002. Zum Effekt auf die Meinungsbildung, der durch Konsens auf der Elitenebene auftritt, vgl. Clarke/Kornberg 1996, Clarke/Kornberg/Stewart 2004; de Vreese 2006; Kriesi 2006; Selb/Kriesi/Hänggli/Marr 2009.

ches trifft ebenso auf das hier vorliegende Bürgerbegehren und den -entscheid zu, weshalb das modifizierte Michigan-Modell auch in dieser Arbeit das theoretische Fundament bildet.

Aus diesem Grund kann angenommen werden, dass sich referendumsspezifische Merkmale auch im Falle Lichtenfels vergleichsweise stark auf das Abstimmungsverhalten auswirken. Des Weiteren erscheint es aufgrund des vorausgegangenen Bürgerbegehrens plausibel, dass sich Bürger, die für das Bürgerbegehren unterschrieben haben, auch beim Entscheid eher für den Baustopp aussprechen. Wie sich die weiteren Einstellungen bzw. sozioökonomischen Merkmale auf das Abstimmungsverhalten auswirken, wird mit Hilfe empirischer Daten analysiert.

3 Zum verwendeten Datensatz

Als Datengrundlage dient eine CATI-Zufallsstichprobe von insgesamt $n=220$ Personen, die im Zeitraum vom 17. bis 24. Februar 2011, und somit etwa vier Wochen vor dem eigentlichen Bürgerentscheid, stattfand. Als Grundgesamtheit dient die wahlberechtigte Bevölkerung in Lichtenfels.

3.1 Operationalisierung

Da die Determinanten des Abstimmungsverhaltens beim Entscheid im vordergründigen Interesse stehen, entspricht die Abstimmungsentscheidung dem abhängigen Merkmal der späteren Analysemodelle. Hierzu wurden die Befragten zunächst um eine Stellungnahme hinsichtlich ihrer Teilnahmebereitschaft am Entscheid gebeten. Antworteten sie hierauf, dass sie wahrscheinlich bzw. sehr wahrscheinlich teilnehmen werden, wurden sie entsprechend nach ihrer Abstimmungsentscheidung gefragt. Gaben sie hingegen an, nur vielleicht oder wahrscheinlich nicht teilnehmen zu wollen, wurden die Befragten um eine hypothetische Entscheidung gebeten. Sowohl für diejenigen Befragten mit einer mehr oder minder sicheren als auch für solche mit einer eher unwahrscheinlichen Teilnahme wird im Falle einer (hypothetischen) Ablehnung des Baustopps (was ein Ja beim Entscheid bedeutete) beim abhängigen Merkmal eine 1 zugeordnet, anderenfalls eine 0. Weiß nicht- und Nicht-Angaben werden als fehlende Werte definiert; gleiches gilt für Befragte, die mit Sicherheit nicht am Entscheid teilnehmen werden.

Für die Imputations- und Analysemodelle werden auch alle weiteren Variablen⁵ dichotomisiert (siehe Anhang: Operationalisierung),⁶ wobei stets sowohl das Nichtwissen als auch fehlende Angaben als Missings definiert werden.⁷

Nach Zuweisung der fehlenden Werte bleiben nur noch vier Variablen übrig, die vollständig beobachtet werden, was einen vergleichsweise hohen Item Nonresponse bedeutet – ein mit Blick auf die Fallzahl des Analysemodells und den damit verbundenen großen Standardfehlern massives Problem. Ebenso ist die Frage zu stellen, ob dieser Ausfall unsystematisch ist, oder ob sich die jeweils Antwortenden von den Nicht-Antwortenden in den zu untersuchenden Merkmalen systematisch unterscheiden. Wäre dies der Fall, unterlägen die CC-Schätzer einer Verzerrung. Entsprechend widmet sich der folgende Abschnitt einer Deskription des Datenausfalls.

3.2 Deskription des Datenausfalls und Multiple Imputation

Tabelle 1 im Anhang beinhaltet die jeweiligen Anteile an fehlenden Werten innerhalb jeder Variable des Datensatzes.

Hiernach unterliegt der Datenausfall einer relativ großen Variation: Beträgt der Ausfall für eine ganze Reihe von Merkmalen nicht einmal 5%, liegt der Nonresponse in immerhin drei Variablen bei über 20% – darunter auch die abhängige Variable *Abstimmungsentscheidung* mit einem Anteil von 25,45%. Würden

- 5 Ein besonderes Augenmerk sei dabei auf die Erhebung der Parteineigung geworfen: Um diese abzubilden, wurde die prospektive Stadtratswahlentscheidung erhoben. Somit muss für die Messung des Effekts der Parteibindung eine Einschränkung gemacht werden: Der „klassische“ Indikator zur Messung der PID (siehe zum genauen Wortlaut etwa Schoen/Weins 2005: 209) wird in diesem Datensatz nicht verwendet, jedoch scheint die Operationalisierung der PID durch eine Verwendung der prospektiven Wahlentscheidung sinnvoll: Diese nämlich misst immerhin annäherungsweise die dauerhafte und affektive Bindung an eine politische Partei.
- 6 Die Dichotomisierung erfolgte vor allem wegen der Anzahl der zu schätzenden Parameter und den damit verbundenen Freiheitsgraden des späteren multivariaten Analysemodells. So weisen die meisten erhobenen Merkmale ein ordinales Skalenniveau auf. Um diese – ohne Informationsverlust – statistisch adäquat in die Regression einfließen zu lassen, wäre es notwendig, bei einer z.B. fünfstufigen Ordinalskala vier Dummyvariablen zu bilden. Aufgrund der geringen Fallzahl von 220 würde dies zu einer unverhältnismäßig hohen Anzahl an Modellparametern, wenigen Freiheitsgraden und großen Konfidenzintervallen führen. Da die dichotomen Variablen die empirische Wirklichkeit homomorph abbilden, scheint die vorgenommene Codierung also durchaus praktikabel – sowohl aus theoretisch inhaltlicher, methodischer als auch statistischer Sichtweise.
- 7 Eine Ausnahme hierbei bilden die Wissensfragen: Gaben die Befragten bei der Frage zum Bau des Casinos, des Media Marktes oder der Verkaufsfläche an, die Antwort nicht zu wissen, wird ihnen eine nicht-korrekte Antwort und damit eine 0 zugeordnet.

alle Variablen in ein multivariates Analysemodell ohne Imputation aufgenommen, ergäbe sich ein kumulierter Datenausfall von über 60%, wodurch nur noch 82 der ursprünglich 220 Fälle verwendet werden könnten.

Wie bereits angedeutet, ergibt sich durch den Datenausfall, neben der geringen Fallzahl, ein weiteres Problem: Es ist zu hinterfragen, ob die CC-Schätzer systematischen Verzerrungen unterliegen. Die Antwort hierauf können deskriptive und analytische Vergleiche liefern. Diese erfolgen beispielhaft anhand der Merkmale *Abstimmungsentscheidung* und *Teilnahmewahrscheinlichkeit* am Bürgerentscheid. Wie Tabelle 1 zu entnehmen ist, beträgt der Ausfall innerhalb der späteren abhängigen Variable *Abstimmungsentscheidung* 25,45%, der im Merkmal *Teilnahmewahrscheinlichkeit* dagegen nur 3,18%.

Ob der Ausfall in der Variable *Abstimmungsentscheidung* zufällig ist, lässt sich nun durch den Vergleich zweier Verteilungen des Merkmals *Teilnahmewahrscheinlichkeit* bestimmen: Die erste setzt sich aus *allen* Befragten zusammen, die eine Angabe zur Teilnahmewahrscheinlichkeit am Bürgerentscheid machten. Von diesen 213 Befragten gaben ca. 84% (exakt: 179 Befragte) an, am Bürgerentscheid teilnehmen zu wollen. Die zweite Verteilung für die Teilnahmewahrscheinlichkeit ergibt sich nun lediglich aus denjenigen Befragten, die auf die Frage zur Abstimmungsentscheidung eine Antwort gaben – und damit keine Missings darstellen. Der entsprechende Teildatensatz umfasst eine Fallzahl von nur noch 164 Befragten, wobei von diesen nun fast 93% angaben, an der Abstimmung teilnehmen zu wollen (siehe hierzu die Verteilungen in Abbildung 1).

Somit scheinen diejenigen Befragten, die keine Angabe zur Abstimmungsentscheidung machen, tendenziell seltener am Entscheid teilnehmen zu wollen. Sobald nämlich diese Stichprobeneinheiten aus der Verteilung zur Teilnahmewahrscheinlichkeit ausgeschlossen werden, erhöht sich der Anteil derjenigen, die beim Entscheid eine Stimme abgeben wollen. Diese Erkenntnis erscheint auch theoretisch plausibel: Befragte, die in der Erhebung (und damit vier Wochen vor dem eigentlichen Bürgerentscheid) dazu in der Lage sind, eine Entscheidung bezüglich der Abstimmung zu treffen, könnten am Bürgerentscheid tendenziell eher teilnehmen wollen. Gleichermäßen könnten – im Umkehrschluss – Personen, die sich eher gegen eine Teilnahme am Bürgerentscheid aussprechen, weniger dazu bereit sein, eine Abstimmungsentscheidung abzugeben.

Dieser Verdacht des systematischen Datenausfalls in der Variable zur Abstimmungsentscheidung wird durch einen Chi-Quadrat-Test bestätigt: Hierbei kann die Nullhypothese, wonach die ursprüngliche Verteilung der Teilnahmewahrscheinlichkeit und diejenige, welche sich lediglich aus dem Teildatensatz ergibt, aus einer identischen Grundgesamtheit stammen, bei einer Irrtumswahrscheinlichkeit

von 1% abgelehnt werden, womit ein rein zufälliger Datenausfall auszuschließen ist.

Auch für alle weiteren Merkmale kann mindestens eine Variable beobachtet werden, durch deren Verteilungsvergleich ein rein zufälliger Ausfallmechanismus ausgeschlossen werden kann.

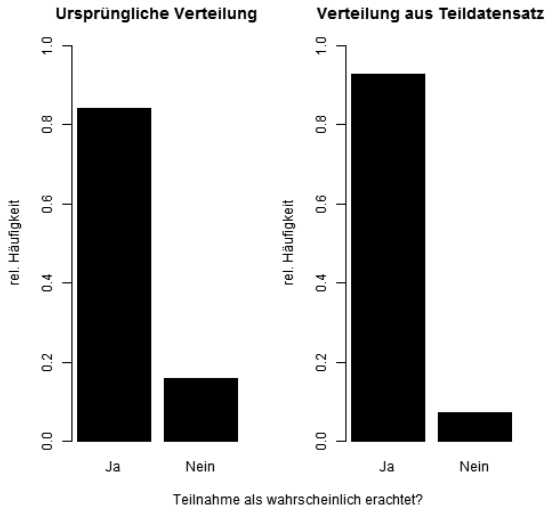
Aus diesem Grund wurden die fehlenden Werte des vorliegenden Datensatzes mit Hilfe einer MI⁸ ergänzt.⁹ Dabei entspricht die Spezifikation des Imputationsmodells derjenigen des Analysemodells (Tabelle 2¹⁰ im Anhang).¹¹

Hiernach scheint sich die Annahme, wonach besonders referendumspezifische Merkmale das Abstimmungsverhalten determinieren, zu bestätigen. Die Bewertung des vorhandenen Angebots sowie des Standortes als auch die Frage nach der Gefährdung von Arbeitsplätzen können die Entscheidung für oder gegen den Bau des FMZ signifikant erklären. Auch das Leisten einer Unterschrift wirkt sich auf das Abstimmungsverhalten aus, genau wie das referendumspezifische Wissen und ein hohes Alter.¹²

Um die Schätzung der MI besser evaluieren zu können, werden im Folgenden die Vor- bzw. Nachteile von MI- und CC-Schätzungen miteinander verglichen.

- 8 Eine exaktere Darstellung der Verfahrensannahmen erfolgt im nächsten Abschnitt, um die Vorteile gegenüber der Verwendung von CC-Schätzungen besser verdeutlichen zu können.
- 9 Hierzu wird für die Imputation des vorliegenden Datensatzes das *R-Paket Multivariate Imputation by Chained Equations* (kurz: *mice*) (vgl. van Buuren/Groothuis-Oudshoorn 2011) verwendet. Zur Ergänzung der fehlenden Werte wird ein Gibbs-Sampler (vgl. Geman/Geman 1984) mit 20 Iterationen verwendet, um insgesamt $m=10$ komplettierte Datensätze zu erzeugen.
- 10 Es sei darauf hingewiesen, dass allen Befragten, die hinsichtlich ihrer Teilnahmewahrscheinlichkeit angaben, mit Sicherheit nicht an der Abstimmung teilnehmen zu wollen (im erhobenen Datensatz gaben dies 16 Personen an), nach der Imputation erneut ein fehlender Wert für die Abstimmungsentscheidung zugewiesen wird, da diese Befragten explizit keine (hypothetische) Stimme abgeben. Dies trifft ebenso auf diejenigen Fälle zu, für welche innerhalb der Imputation in mindestens einem der 10 ergänzten Datensätze eine sichere Nichtteilnahme für die Teilnahmewahrscheinlichkeit imputiert wird. Im Falle der imputierten Datensätze, die als Datenbasis für Tabelle 2 dienen, ist dies für 21 Personen zutreffend, weshalb diese Merkmalsträger, trotz Imputation, für die Analyse nicht zur Verfügung stehen.
- 11 Tabelle 2 beinhaltet somit Regressionsparameter, die unter Verwendung des ursprünglichen und multipel imputierten Datensatzes geschätzt wurden. Diese Datenbasis wurde also nicht aus der später simulierten synthetischen Grundgesamtheit gezogen.
- 12 Da die statistischen und methodischen Verfahren im Vordergrund dieser Arbeit stehen, wird auf eine weitere politikwissenschaftliche Interpretation der Koeffizienten verzichtet.

Abbildung 1 Vergleich der ursprünglichen Verteilung des Merkmals *Subjektive Wichtigkeit* mit Verteilung aus Teilgesamtheit, die sich aus Antwortenden des Merkmals *Abstimmungsentscheidung* ergibt



4 Umgang mit fehlenden Werten

Im Allgemeinen läuft der Umgang mit fehlenden Werten auf zwei Strategien hinaus: Zum einen kann bei der Erhebung bereits von Anfang an auf eine Vermeidung fehlender Werte gezielt werden, zum anderen lassen sich diese aber auch nach ihrer Entstehung korrigieren (Post-Survey Adjustments); hierbei wird versucht, die durch die Missings verursachte Verzerrung zu reduzieren (vgl. de Leeuw/Hox/Dillman 2008: 10). Im vorliegenden Fall bleibt nur die letztgenannte Möglichkeit. Für die Frage, welches Verfahren im vorliegenden Fall bezüglich seiner Stabilität der Parameterschätzungen und p-Werte das zu Empfehlende ist, wird in dieser Arbeit der Blick auf zwei mögliche Vorgehensweisen geworfen, wobei eine von beiden – trotz häufiger praktischer Anwendung – nicht zur Familie der Korrekturmethode zählt und den Ausfallmechanismus damit weitestgehend ignoriert: die Eliminierung fehlender Werte und Verwendung der lediglich auf alle Fragen antwortenden Merkmalsträger. Wie bereits zu Beginn erwähnt, stellt gerade in den Sozialwissenschaften die CC-Methode das in der Regel angewandte Vorgehen dar. Auch wenn

in diesem Beitrag CC-Methoden mit MI verglichen wird, gäbe es natürlich noch viele andere Alternativen über MI hinaus.¹³

4.1 Multiple Imputation

Ohne auf die formale Theorie der MI (vgl. Rubin 1987; 1996) näher einzugehen, muss zumindest einer Annahme zur Modellierung eines adäquaten Imputationsmodells genauer nachgegangen werden: Der Ausfallmechanismus sollte – unter der Annahme, dass die unten folgenden Inferenzen lediglich auf die interessierenden Parameter des Analysemodells und nicht auf die Parameter des Antwortmechanismus abzielen – Completely At Random oder zumindest At Random sein. In diesem Falle würde ein sogenannter ignorierbarer Ausfallmechanismus vorliegen, weshalb dieser im Imputationsmodell nicht explizit modelliert werden müsste und die Modellspezifikation vereinfacht würde. Hierzu eine kurze Erläuterung (vgl. Rubin 1976; Little/Rubin 1987; 2002):

Der Ausfallmechanismus heißt „Missing Completely At Random“ (MCAR), wenn der Datenausfall rein zufällig ist. Ein Beispiel hierfür ist der Verlust von Fragebögen auf dem postalischen Weg vom Erhebungs- zum Analyseinstitut. Dieser Fall liegt jedoch praktisch nie vor, was auch für den vorliegenden Datensatz gilt (siehe obige Ausfalldeskription in Abschnitt 3.2).

Ist der Ausfall innerhalb eines Merkmals von den beobachteten Ausprägungen der anderen im Datensatz vorhandenen Variablen abhängig, nicht aber von der fehlenden Ausprägung des Merkmals selbst, heißt der Mechanismus „Missing At Random“ (MAR). Beispielsweise soll das Gewicht erhoben werden, wobei Frauen ihr Gewicht seltener angeben als Männer. Somit wäre der Ausfall in der Variable Gewicht vom Geschlecht, nicht jedoch zusätzlich vom Gewicht selbst abhängig. Mit anderen Worten wäre der Nonresponse vom Gewicht innerhalb des Merkmals Geschlecht wiederum zufällig.

Sollte der Ausfall einer Variable jedoch von der Ausprägung selbst abhängen, so ist der Ausfallmechanismus „Missing Not At Random“ (MNAR). In diesem Fall verkompliziert sich die Spezifikation des Imputationsmodells erheblich, da der Ausfallmechanismus nicht ignorierbar und damit explizit modelliert werden muss, auch wenn die spätere Inferenz lediglich auf die Parameter des Analysemodells abzielt. Auf das obige Beispiel angewandt hieße der Ausfall MNAR, wenn die Variable Gewicht gerade aufgrund des Gewichts einer Person nicht erhoben würde. Dies

13 Eine Zusammenfassung und Systematisierung von Korrekturmethode findet sich beispielsweise bei Rässler 2000, Göthlich 2007, Spieß 2008.

läge z.B. vor, wenn Befragte aufgrund eines sehr niedrigen oder hohen Gewichts die Angabe verweigerten.¹⁴

Ob die Missings im vorliegenden Datensatz MAR oder gar MNAR sind, lässt sich jedoch durch Tests nicht bestimmen: Zu messen, ob ein Befragter die Antwort auf eine Frage gerade wegen der von ihm generierten Antwort verweigert hat, würde voraussetzen, dass die nicht gegebene Antwort beobachtbar wäre.

Deswegen ist sowohl mit Blick auf die *Weiß nicht*-Kategorie als auch auf die Antwortverweigerung zu hinterfragen, welcher kognitive Prozess den Item Nonresponse verursacht haben könnte. Hierzu sei auf das Entscheidungsmodell für Item Nonresponse nach Beatty und Herrmann (2002) verwiesen. Demnach lässt sich die Response-Entscheidung eines Befragten in verschiedene Abschnitte zerlegen: Im ersten Schritt bedarf es einer verständlichen Interpretation der Fragestellung seitens des Befragten. Ist dieser nicht dazu in der Lage, den Sinn des Items zu verstehen, sei es ihm entsprechend auch nicht möglich, eine Angabe zu machen. Hat der Teilnehmer die Fragestellung hingegen verstanden, muss er darüber entscheiden, ob er den Aufwand betreibt, Eindrücke bzw. Wissen zum abgefragten Thema zu sammeln und zu verarbeiten. Entscheidet sich der Befragte dagegen, hängt der Nonresponse offensichtlich mit nicht ausreichend vorhandenen kognitiven Fähigkeiten oder einer fehlenden Motivation zusammen. Grund für den Nonresponse wäre somit nicht die generierte Antwort selbst, was auf einen MAR-Mechanismus hindeuten ließe. MNAR läge hingegen vor, wenn der Befragte zwar eine Antwort generieren wollte und könnte, diese aber aufgrund ihres Inhalts geheim hielte – beispielsweise aus sozialer Erwünschtheit. In diesem Fall wäre die Antwort selbst der Grund des Nonresponse.¹⁵

Auch weil die Messung nach einem MAR- oder MNAR-Mechanismus nicht möglich ist, wird für diese Arbeit die Annahme getroffen, dass die fehlenden Angaben bzw. *Weiß nicht*-Antworten einem MAR-Mechanismus unterliegen.

- 14 Diesem Ausfallmechanismus folgen, wie bereits oben dargestellt, auch die Missings in der Variable *Abstimmungsentscheidung*, die aufgrund der sicheren Nicht-Teilnahme am Bürgerentscheid definiert wurden. Der Ausfall in *Abstimmungsentscheidung* ist zwar von der Ausprägung des Merkmals *Abstimmungswahrscheinlichkeit* abhängig, jedoch innerhalb dieser nicht zufällig: Wenn ein Befragter *sicher nicht teilnehmen* wird, wird er mit Sicherheit auch keine Abstimmungsentscheidung treffen. Somit ist dieser Ausfall weder MCAR noch MAR.
- 15 Zu unterscheiden ist diese Form der Antwortverweigerung von der Geheimhaltung aufgrund fehlender Präzision: Hiermit gemeint ist der Eindruck des Befragten, dass die von ihm generierte Antwort qualitativ unzureichend begründet sei, weshalb er lieber keine Antwort abgibt. Auch in diesem Fall wäre der Antwortmechanismus MAR, da der Nonresponse nicht mit der Antwort selbst, sondern der unzureichenden Sicherheit des Befragten bei der Antwortvergabe zusammenhinge.

Dafür spricht erstens, dass der verwendete Fragebogen keinerlei Items bzw. Antwortkategorien enthält, die aufgrund ihrer Sensibilität – d.h. mit Blick auf soziale Erwünschtheit – einen vordergründigen MNAR-Ausfall vermuten ließen. Hinzu kommt, dass ein Großteil der kognitiven Prozesse auf dem Weg zur Item Non-response-Entscheidung auf einen MAR-Ausfall schließen lässt. Aufgrund dieser Annahmen kann die Modellierung des Imputationsmodells erheblich vereinfacht und durchgeführt werden.

4.2 Complete Cases

Das zweite Vorgehen stellt – wie schon erwähnt – keine Korrektur dar. Complete Cases eliminiert vielmehr alle Merkmalsträger, die auch nur hinsichtlich eines der ins Analysemodell aufgenommenen Merkmale fehlende Werte aufweisen. Dies führt zu einer Reduktion der für die Analyse verwendeten Fälle. Je nach Ausfallmuster kann selbst ein geringer Ausfall einzelner Variablen in einem multivariaten Modell zu einem hohen Gesamtausfall führen. Besonders in kleinen Stichproben wiegt jeder zusätzlich fehlende Merkmalsträger schwer, so dass sich schon aus ökonomischen Gründen eine CC-Analyse verbieten sollte. Die einzigen Ausfallmechanismen, unter denen CC unverzerrt schätzen kann, sind MCAR und ein besonderer Fall von MAR.¹⁶ Wie in Kapitel 3.2. gezeigt wurde, liegt innerhalb des verwendeten Datensatzes jedoch kein rein zufälliger Ausfall vor, wodurch die CC-Analyse in den nun folgenden Methodenvergleich mit zwei Vorbelastungen geht: einer unerfüllten Annahme zum Ausfallmechanismus und einer verfahrensimmanenten Reduktion der ohnehin kleinen Stichprobe.

5 Methodenvergleich und Simulation

Für einen validen Methodenvergleich ist es unabdingbar, eine Basis zu finden, die für beide miteinander verglichenen Vorgehensweisen zum Umgang mit fehlenden Werten die gleiche Ausgangslage bietet. Doch ein Methodenvergleich sollte noch mehr können: Veröffentlichungen zum Vergleich von Korrekturmetho-

16 Wenn lediglich die abhängige Variable eines Analysemodells fehlende Werte aufweist und die Wahrscheinlichkeit des Fehlens nur von den unabhängigen Variablen des Modells abhängt, dann führt die CC zu unverzerrten Schätzern im interessierenden Modell. Dass ein solcher MAR-Ausfall im hier verwendeten Datensatz vorliegt, erscheint jedoch aufgrund der ebenso in den erklärenden Variablen des Analysemodells auftretenden Missings als äußerst unplausibel.

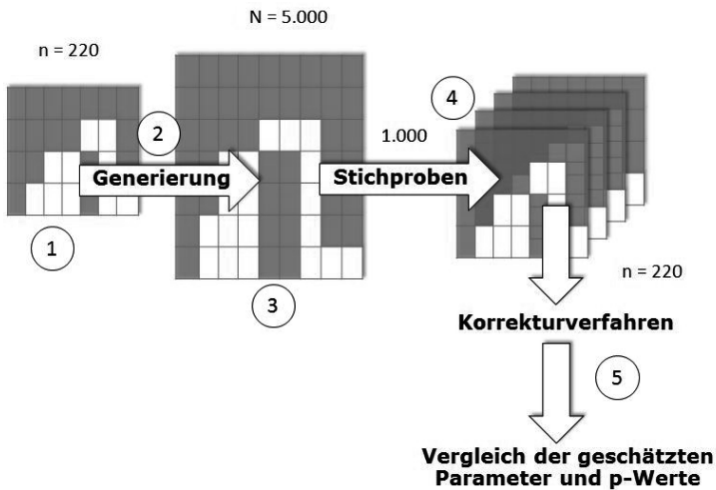
beinhalten entweder die bloße Durchführung der Methode anhand eines ausgewählten empirischen Datensatzes (vgl. Longford 2000) oder eine modellbasierte Simulationsstudie (vgl. Rässler 2000). Beide Optionen bringen jedoch Nachteile mit sich: Die einfache Durchführung mit einem realen Datensatz berücksichtigt zwar die Herausforderungen realer Datengegebenheiten, sie hat allerdings eine eingeschränkte Aussagekraft über den ausgewählten Datensatz hinaus. Simulationen hingegen müssen immer mit dem Makel des Künstlichen und weniger Komplexen leben, was eine Verallgemeinerung der Ergebnisse ebenfalls erschwert.

Für den folgenden Methodenvergleich wird ein vereinfachter Ansatz nach Messingschlager (2012) herangezogen (siehe hierzu Abbildung 2).¹⁷ Der erste Schritt zeigt den Datensatz in vereinfachter Form. Zu sehen sind gegebene und fehlende Daten (1). Der originale Datensatz wird nun via Bootstrapping vergrößert, indem Stichproben aus ihm gezogen und immer wieder an den ursprünglichen Datensatz angehängt werden (2). Damit erhält man Ziehung für Ziehung einen immer größeren Datensatz. Dies wird so lange wiederholt, bis eine Gesamtheit entsteht, die hinreichend groß ist, um Stichproben gegebener Größe ziehen zu können (Auswahlsatz deutlich unter 0,05). Konsequenterweise behält diese Gesamtheit annähernd die Zusammenhangsstruktur und das Ausfallmuster des ursprünglichen Datensatzes zur Befragung der Lichtenfelder Bürger vor dem Bürgerentscheid (3). Die neue Gesamtheit bildet nun eine valide Basis für den Methodenvergleich.¹⁸ Zu diesem Zweck werden für den simulativen Teil 1.000 Stichproben aus der neuen Gesamtheit gezogen. Die Stichprobengröße entspricht dabei jeweils genau derjenigen des Originaldatensatzes, was beispielsweise bei Bootstrapping angesichts der ohnehin sehr kleinen Stichprobe so nicht durchführbar wäre (4). Mit jeder Stichprobe wird das Modell zur Erklärung des Abstimmungsverhaltens beim Bürgerentscheid geschätzt, wobei CC als eliminierendes Vorgehen zum Einsatz kommt und

17 In der hier verwendeten Variante werden lediglich die Verteilungen der durch die Simulation geschätzten Parameterwerte anhand arithmetischer Mittel und Standardabweichungen miteinander verglichen. Bei Messingschlager (2012) wird hingegen ein Stresstest auf der Grundlage konstruierter wahrer Parameter durchgeführt.

18 Durch dieses Vorgehen wird sichergestellt, eine synthetische Grundgesamtheit zu generieren, deren Korrelationsstruktur derjenigen der empirisch erhobenen Stichprobe entspricht (Stichprobe von $n=220$ aus einer Grundgesamtheit von Wahlberechtigten, dann 1.000 Stichproben vom Umfang $n=220$ aus einer synthetischen Gesamtheit). Es geht an dieser Stelle nicht nur um die Schätzung von Parametern wie es z.B. bei Bootstrapping der Fall wäre, sondern darum, eine einheitliche Ausgangslage für den Methodenvergleich – und damit für die unten folgende Sensitivitätsanalyse – zu schaffen. Mit dieser synthetischen Grundgesamtheit ließen sich auch beliebig andere Analysen erstellen. Aufgrund der Übersichtlichkeit bleibt das Augenmerk jedoch auf dem logistischen Regressionsmodell.

Abbildung 2 Simulationsbeschreibung für den Methodenvergleich



MI als Imputationsverfahren dient (5).¹⁹ Die MI-Schätzung wurde dabei zunächst mit einem Modell unter Kongruenz von Imputations- und Analysemodell (1. MI-Simulation) durchgeführt. Anschließend wurde ein zweites Imputationsmodell spezifiziert (2. MI-Simulation). Dieses beinhaltet zusätzliche Merkmale²⁰ und stellt damit eine größere Informationsquelle für die Ergänzung der fehlenden Werte dar. Hierdurch sollte sich die Stabilität der späteren Parameterschätzungen weiter vergrößern (siehe zur Diskussion über *congeniality* Meng 1994 und Rubin 1996).

Der Vorteil dieses Vorgehens liegt zusammengefasst also darin, dass reale Daten den Ausgangspunkt für den Methodenvergleich bilden, indem ein simulativer Datensatz erzeugt wird, der nahezu gleiche Zusammenhangsstrukturen und identische Ausfallmuster besitzt.²¹ Somit wird eine valide Ausgangsbasis für den Methodenvergleich geschaffen.

- 19 Aus der Modellschätzung fallen allerdings auch hierbei diejenigen Merkmalsträger heraus, die geäußert haben, mit Sicherheit nicht am Bürgerentscheid teilnehmen zu wollen.
- 20 Zusätzliche Merkmale: Stellungnahme, ob FMZ eine Bereicherung darstellt; subjektive Wichtigkeit des Bürgerentscheids; Sympathieskalometer gegenüber Bürgermeisterin Bianca Fischer; Wohngegend in Lichtenfels.
- 21 Anders als bei einer Simulation kann der datengenerierende Mechanismus hier jedoch nicht mehr gesteuert werden. Es könnte sich also bei der Datengrundlage auch um einen Extremfall handeln, was jedoch aufgrund der vorgenommenen Datenerhebung (s.o.) unwahrscheinlich ist.

Die Ergebnisse der Schätzungen durch MI und CC werden im nächsten Kapitel grafisch an ausgewählten Beispielvariablen präsentiert.

6 Empirische Ergebnisse

Für alle exogenen Variablen wurden die jeweiligen 1.000 B-Parameter gemittelt (\bar{B}), deren Standardabweichungen (SD) ermittelt sowie die durchschnittlichen p-Werte (\bar{p}) mitsamt deren Standardabweichungen berechnet, welche den Tabellen 3 und 4 im Anhang zu entnehmen sind. Für drei ausgewählte Merkmale sollen die Ergebnisse exemplarisch erläutert werden:

- *Unterschrift bei Begehren* (1,36 % fehlende Werte), welches einen positiven Effekt auf die Abstimmungsentscheidung im obigen Analysemodell ausübte,
- *Kein Standort besser geeignet* (13,18 % fehlende Werte), welches sich negativ auf den Stopp des Baues auswirkte,
- *Prospektive Wahl der Freien Wähler* (26,81 % fehlende Werte), welche sich nur insignifikant auf den Stimmentscheid auswirkte.

Die Histogramme (Abbildungen 3–5) zeigen die Verteilungen der simulierten 1.000 B-Schätzer sowie deren p-Werte getrennt nach den CC- und den zwei MI-Modellen.

Betrachtet man für die Variable *Unterschrift bei Begehren* die Schätzwerteverteilung der Methoden, fällt zunächst die Breite der Verteilung von CC auf. Die Schätzwerte streuen mit einer Standardabweichung von 2,3649 gegenüber lediglich 0,7360 bei den Simulationen durch das erste MI-Modell. Bei der zweiten MI-Simulation reduziert sich die Standardabweichung der geschätzten Werte noch einmal leicht auf 0,7024. Mit Blick auf die Anzahl der durchgeführten Modellschätzungen wird zudem deutlich, dass nach der Eliminierung der fehlenden Werte 370 Stichproben gar keine Modellschätzungen mehr zuließen (CC).²² Die Interpretation der Schätzwerte wäre zudem sehr ungenau, da zwar ein Großteil der Werte einen positiven Zusammenhang zwischen der Unterschrift und dem Entscheid, jedoch ein nicht unbedeutender Teil der Schätzwerte einen negativen Effekt suggeriert. Bei den MI-Simulationen sind hingegen nur sehr wenige Ausreißer zu verzeichnen. Hieran zeigt sich die politikwissenschaftlich-interpretatorische Relevanz der vorgenommenen Simulation: Die berechneten Effekte der CC-Analysen erweisen sich

22 Schätzparameter werden dann als Extremwerte hinsichtlich ihrer Plausibilität ausgeschlossen, wenn sie Werte kleiner -10 oder größer +10 annehmen. Von den 1.000 simulierten Schätzparametern fallen also 370 CC-Parameter als Extremwerte aus dem Intervall -10 bis +10 heraus.

demnach als äußerst schwankend, wodurch sich keine stabilen Parameterschätzungen erzielen lassen.

Dies spiegelt sich auch in den Verteilungen für die p-Werte wider: Im Durchschnitt erreichen die beiden MI-Simulationen einen p-Wert von 0,0168 bzw. von 0,0135. Die Standardabweichungen der MI-Simulationen signalisieren große Stabilität im Vergleich zu den p-Werten, die aus den CC-Modellen hervorgehen. Unter Verwendung letzterer ist bei möglichen Aussagen zum Einfluss der Variable *Unterschrift* auf die Abstimmungsentscheidung große Vorsicht geboten, da sich die simulierten p-Werte als äußerst instabil erweisen.

Für die nächste Variable *Kein Standort besser geeignet* zeigen sich ganz ähnliche Ergebnisse: Die Standardabweichung der geschätzten Parameter liegt bei CC um ein Vielfaches höher als bei den MI-Simulationen – nämlich 2,1910 gegenüber 0,8664 bzw. 0,7037. Der Anteil der Stichproben unter CC, mit denen eine Modellschätzung überhaupt gelingt (konkret: 642), ist geringfügig höher als bei der ersten Variable *Unterschrift*. Die Abstände zwischen den beiden MI-Simulationen bleiben in etwa gleich, d.h. auch hier weist die zweite MI-Simulation eine minimal schmalere Schätzwerteverteilung auf.

Bei den p-Werten schwächt sich die Stabilität ab: Der durchschnittliche p-Wert für die geschätzten Parameterwerte liegt noch bei 0,1438 bzw. 0,1943. Im Durchschnitt sind unter Verwendung beider MI-Modelle demnach keine signifikanten Parameterschätzer für die Variable zur Standortbewertung zu beobachten.

Das letzte Beispiel anhand der Variable *Prospektive Wahl der Freien Wähler* zeigt die Grenzen der Schätzung bei derartigen Datenausfällen im multivariaten Raum durch MI auf: Die Streuung der geschätzten Parameterwerte durch MI (Standardabweichungen von 1,2204 bzw. 1,0260) ist zwar immer noch geringer als die der geschätzten Werte mittels CC (2,9513), jedoch nimmt die Breite der Verteilung derart zu, dass nicht mehr von einem hinreichend befriedigenden Ergebnis gesprochen werden kann.

Für alle Verfahren werden durchschnittlich keine signifikanten Schätzparameter angegeben, wobei wiederum zu berücksichtigen ist, dass bei CC nun gerade kaum mehr als die Hälfte der gezogenen Stichproben (konkret: 587) zu einer Modellschätzung führt, die realistische Schätzwerte annimmt.

Wie die Tabellen 3 und 4 weiter aufzeigen, ergeben sich unter den CC-Spezifikationen über alle Parameter hinweg größere Standardabweichungen als bei den MI-Modellen, was darauf hindeutet, dass die Schätzer einer hohen Ungenauigkeit unterliegen. Dieser Sachverhalt erhärtet sich mit Blick auf die insgesamt deutlich höhere Streuung der p-Werte. Die MI-Simulationen erweisen sich dem gegenüber für jedes Merkmal als die stabilere Schätzung, wobei die Aufnahme weiterer Varia-

Abbildung 3 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Unterschrift* bei *Begehren*

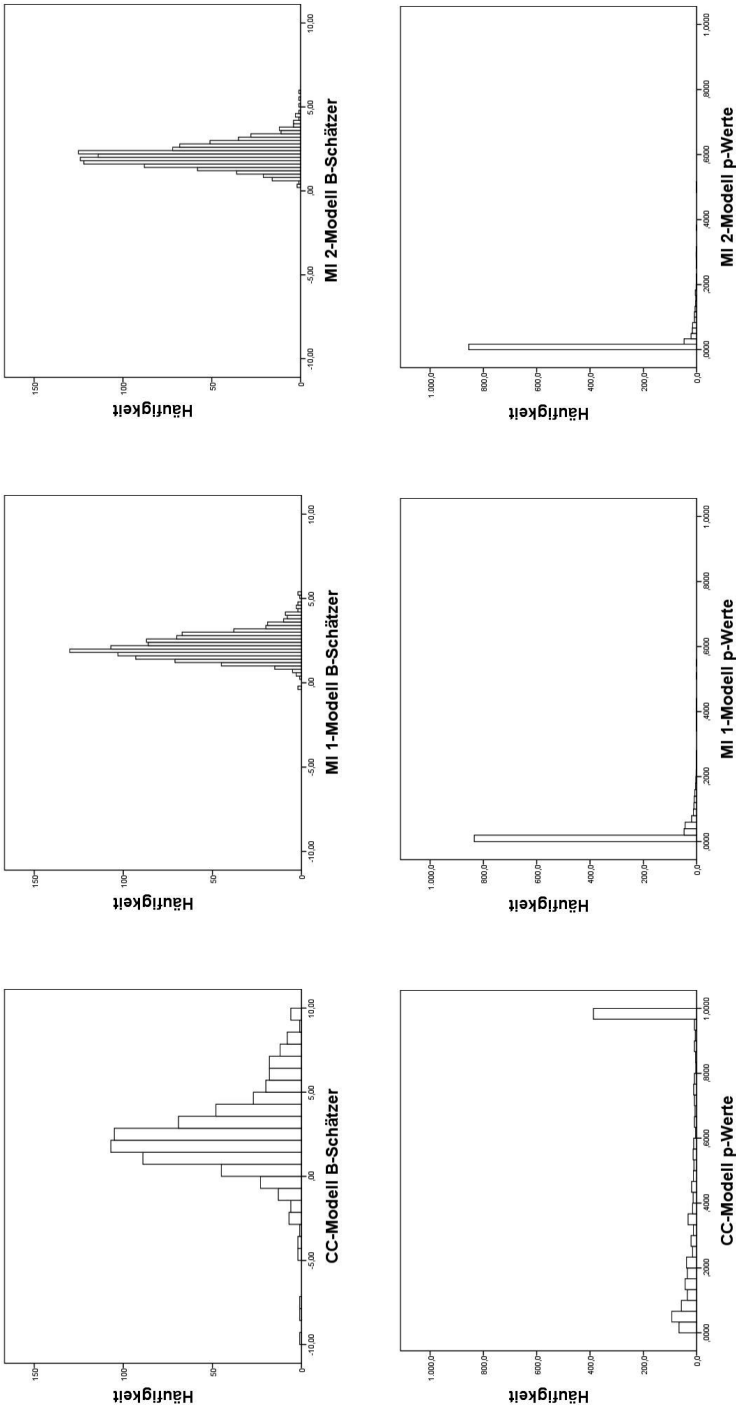


Abbildung 4 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Kein Standort besser geeignet*

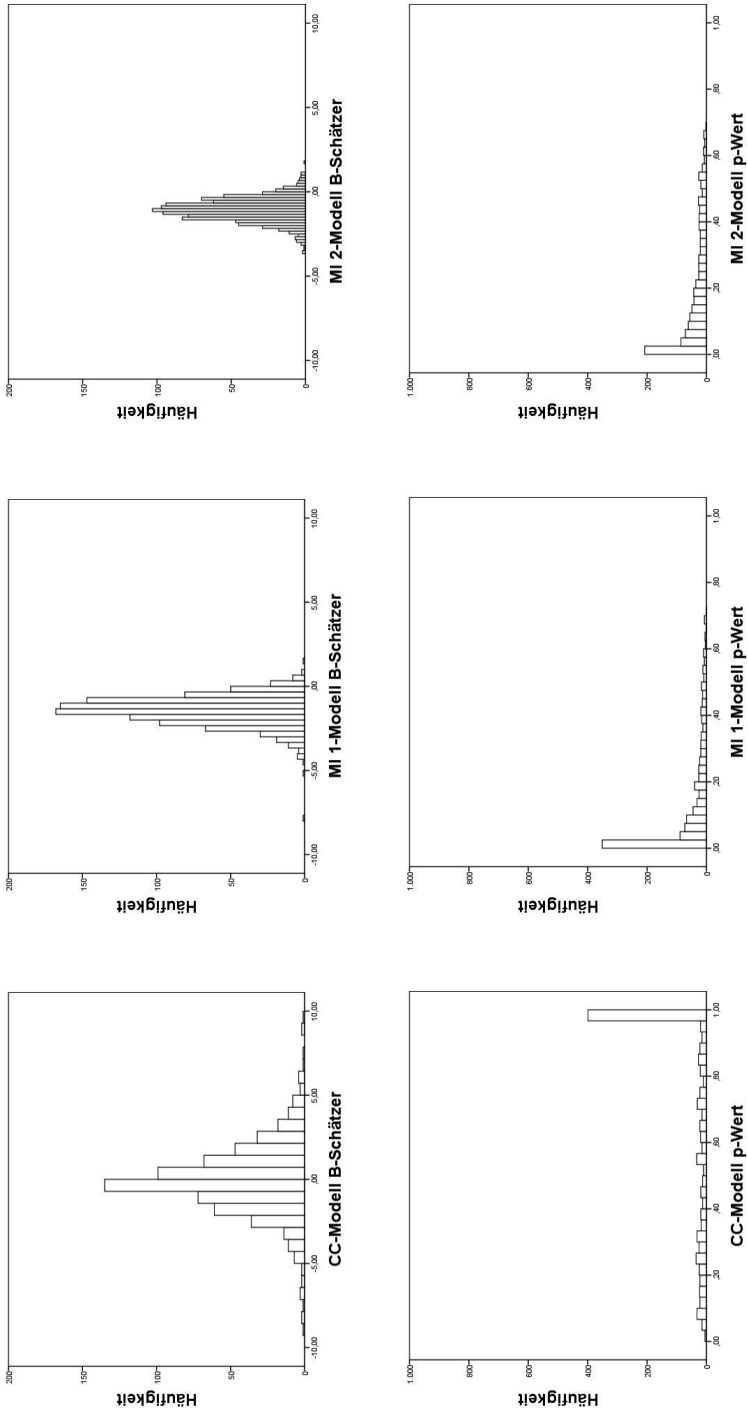
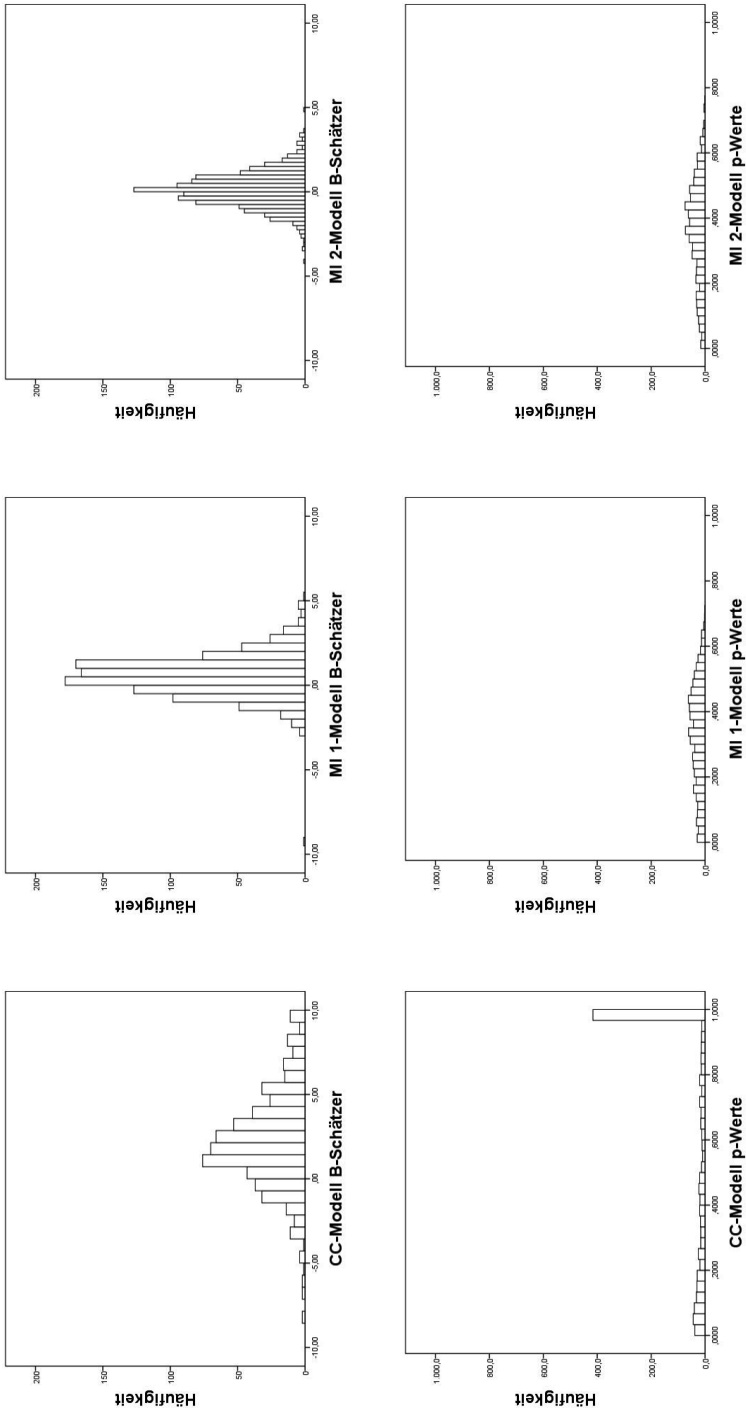


Abbildung 5 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Freie Wähler-Wahl* prospektiv



blen im zweiten Modell in nahezu allen Fällen zu genaueren Schätzungen der Parameter führt. Ebenso sinken nach der multiplen Ergänzung der fehlenden Werte die Standardabweichungen der p-Werte, was ebenfalls auf eine stabilere Schätzung hinweist.

7 Fazit und Empfehlungen

Somit lässt sich abschließend konstatieren, dass sich aus politikwissenschaftlicher Sicht besonders die referendumsspezifischen Merkmale auf die Abstimmungsentscheidung ausgewirkt haben. Demnach scheint die Stimmabgabe maßgeblich von der Haltung gegenüber dem in Lichtenfels bereits bestehenden Angebot, dem vorgesehenen Standort des FMZ sowie der vermuteten Arbeitsplatzgefährdung abzuhängen. Ebenso erwiesen sich die Teilnahme am vorangegangenen Bürgerbegehren, das referendumsspezifische Wissen und ein hohes Alter als signifikante Determinanten.

Die auf Grundlage des multipel imputierten Datensatzes berechneten Regressionsschätzer konnten mit Hilfe des Methodenvergleichs auf ihre Sensitivität hin untersucht werden. Die Ergebnisse deuten darauf hin, dass es sich für die Schätzung eines Modells zur Erklärung der Abstimmungsentscheidungen beim Bürgerentscheid in Lichtenfels angesichts der kleinen Stichprobe sehr empfohlen hat, eine MI gegenüber einer CC-Analyse vorzuziehen. Die Simulation zeigt, dass teilweise knapp die Hälfte aller Stichproben unter CC zu keinen annehmbaren oder adäquat interpretierbaren Schätzwerten führt. Dies erscheint umso wichtiger, da der Methodenvergleich so angelegt war, dass kein Verfahren durch das angewandte Vorgehen benachteiligt wird. Die Hinzunahme von Variablen über das Analysemodell hinaus verbessert durchgängig – wenn auch teilweise nur im geringen Maße – die Schätzungen unter MI weiter.

Die Ergebnisse des Methodenvergleichs ermutigen zum Nachdenken über das Problem fehlender Werte und den Umgang damit. Dass die bisher in den Sozialwissenschaften dominierende CC-Analyse im Sensitivitätsvergleich schlechter abschneidet als MI zeigt, dass es sich lohnt, bei der Datenanalyse auch im Umgang mit fehlenden Werten mehr Achtsamkeit an den Tag zu legen. Aufgrund der Implementation von MI in vielen Standardsoftwares bildet die Anwendung richtiger Korrekturmethode keine unüberwindliche Hürde mehr – wie hier ebenfalls gezeigt wurde.

Literatur

- Beatty, P. und D. Herrmann, 2002: To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. S. 71–85 in: R. M. Groves, D. A. Dillman, J. L. Eltinge und J. A. Roderick (Hg.): *Survey Nonresponse*. New York: John Wiley and Sons.
- Benz, M. und A. Stutzer, 2004: Are Voters Better Informed When They have a Larger Say in Politics? Evidence for the European Union and Switzerland. *Public Choice* 119, 31–59.
- Campbell, A., G. Gurin und W. E. Miller, 1954: *The Voter Decides*. Evanston/Illinois: Row, Peterson and Company.
- Campbell, A., P. E. Converse, W. E. Miller und D. E. Stokes, 1960: *The American Voter*. New York: Wiley.
- Christin, T., S. Hug und P. Sciarini, 2002: Interests and Information in Referendum Voting: An Analysis of Swiss Voters. *European Journal of Political Research* 41: 759–776.
- Clarke, H. D. und A. Kornberg, 1996: Choosing Canada? The 1995 Quebec Sovereignty Referendum. *Political Science and Politics* 29: 676–682.
- Clarke, H. D., A. Kornberg und M. C. Stewart, 2004: Referendum Voting as Political Choice: The Case of Quebec. *British Journal of Political Science* 34: 345–355.
- de Leeuw, E. D., J. J. Hox und D. A. Dillman, 2008: The Cornerstones of Survey Research. S. 1–17 in: E. D. de Leeuw, J. J. Hox und D. A. Dillman (Hg.): *International Handbook of Survey Methodology*. New York: Taylor & Francis.
- de Vreese, C. H., 2006: Political Parties in Dire Straits? Consequences of National Referendums for Political Parties. *Party Politics* 12: 581–598.
- de Vreese, C. H. und H. A. Semetko, 2004: News Matters: Influences on the Vote in the Danish 2000 Euro Referendum Campaign. *European Journal of Political Research* 43: 699–722.
- Decker, F., 2005: Die Systemverträglichkeit der direkten Demokratie. Dargestellt an der Diskussion um die Einführung von plebiszitären Elementen in das Grundgesetz. *Zeitschrift für Politikwissenschaft* 15: 1103–1147.
- Frey, B. S. und G. Kirchgässner, 1993: Diskursethik Politische Ökonomie und Volksabstimmungen. *Analyse & Kritik* 15: 129–149.
- Geman, S. und D. Geman, 1984: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–741.
- Göthlich, S. E., 2007: Zum Umgang mit fehlenden Daten mit großzahligen empirischen Erhebungen. S. 119–134 in: S. Albers, D. Klapper, U. Konradt, A. Walter und J. Wolf (Hg.): *Methodik der empirischen Forschung*. Wiesbaden: Gabler Verlag.
- Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer und R. Tourangeau, 2004: *Survey Methodology*. New York: John Wiley and Sons.
- Joslyn, M. R. und D. P. Haider-Markel, 2000: Guns in the Ballot Box: Information, Groups, and Opinion in Ballot Initiative Campaigns. *American Politics Research* 28: 355–378.
- Kirchgässner, G., L. P. Feld und M. R. Savioz, 1999: *Die direkte Demokratie: Modern, erfolgreich, entwicklungs- und exportfähig*. Basel/München: Helbing & Lichtenhahn.
- Kriesi, H., 2002: Individual Opinion Formation in a Direct Democratic Campaign. *British Journal of Political Science* 32: 171–185.
- Kriesi, H., 2006: Role of The Political Elite in Swiss Direct-Democratic Votes. *Party Politics* 12: 599–622.
- Little, R. J. A. und D. B. Rubin, 1987: *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

- Little, R. J. A. und D. B. Rubin, 2002: *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.
- Longford, N. T., 2000: Multiple Imputation in an International Database of Social Science Surveys. *ZA-Information* 46: 72–95.
- Mehr Demokratie e.V.: Übersicht pro Bundesland. <http://www.mehr-demokratie.de/5970.html> (31.01.2012)
- Mendelsohn, M. und F. Cutler, 2000: The Effect of Referendums on Democratic Citizens: Information, Politicization, Efficacy and Tolerance. *British Journal of Political Science* 30: 669–698.
- Meng, X. L., 1994: Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 9: 538–573.
- Messingschlager, Martin (2012): *Fehlende Werte in den Sozialwissenschaften – Analyse und Korrektur mit Beispielen aus dem ALLBUS*. Bamberg, University of Bamberg Press.
- Oberreuter, H., 2002: Direkte Demokratie und die repräsentative Verfassung der Bundesrepublik Deutschland. *Zeitschrift für Politik* 49: 290–305.
- Rässler, S., 2000: Ergänzung fehlender Daten in Umfragen. *Jahrbücher für Nationalökonomie und Statistik* 220: 64–94.
- Rubin, D. B., 1976: Inference and Missing Data. *Biometrika* 63: 581–592.
- Rubin, D. B., 1987: *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.
- Rubin, D. B., 1996: Multiple Imputation After 18+ Years. *Journal of the American Statistical Association* 91: 473–489.
- Schoen, H., 2011: Eine eindeutige Entscheidung nach resonanzarmer Kampagne. Überlegungen zum Volksentscheid über den Nichtraucherschutz in Bayern. *Zeitschrift für Politikberatung* 3: 457–461.
- Schoen, H., 2012: Wählen und abstimmen – zwei Seiten einer Medaille? Eine Analyse am Beispiel des Volksentscheids zum Nichtraucherschutz in Bayern. *Politische Vierteljahrszeitschrift – Sonderheft* 45: 514–535.
- Schoen, H. und C. Weins, 2005: Der sozialpsychologische Ansatz zur Erklärung von Wahlverhalten. S. 187–242 in: J. W. Falter, H. Schoen (Hg.): *Handbuch Wahlforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schoen, H., A. Glantz und R. Teusch, 2011a: Abstimmungskampf, Informationsvermittlung und Stimmenscheidung beim Volksentscheid über den Nichtraucherschutz in Bayern. S. 295–320 in: L. P. Feld, P. M. Huber, O. Jung, C. Welzel und F. Wittreck (Hg.): *Jahrbuch für direkte Demokratie 2010*. Baden-Baden: Nomos.
- Schoen, H., A. Glantz und R. Teusch, 2011b: Raucher oder Nichtraucher – ist das die Frage? Wahlbeteiligung und Abstimmungsverhalten beim Volksentscheid über das Rauchverbot in Bayern. *Zeitschrift für Parlamentsfragen* 42: 492–502.
- Selb, P., H. Kriesi, R. Hänggli und M. Marr, 2009: Partisan Choices in a Direct-Democratic Campaign. *Political Science Review* 1: 155–172.
- Smith, M. A., 2002: Ballot Initiatives and the Democratic Citizen. *The Journal of Politics* 64: 892–903.
- Spieß, M., 2008: *Missing-Data-Techniken. Analyse von Daten mit fehlenden Werten*. Hamburg: Lit.
- Stadt Lichtenfels, 2011: Klares Ergebnis beim Bürgerentscheid. <http://www.lichtenfels-city.de/index.php?id=7777,589> (20.09.2012)
- van Buuren, S. und K. Groothuis-Oudshoorn, 2011: mice: Multivariate Imputation by Chained Equations in R. *Journal Of Statistical Software* 45: 1–67.
- Weixner, B. M., 2006: Direkte Demokratie in den Bundesländern. *Aus Politik und Zeitgeschichte* 10: 18–24.

Anschriften der Autoren

Martin Messingschlager
Otto-Friedrich-Universität Bamberg
Fakultät Sozial- und Wirtschaftswissenschaften
Lehrstuhl für Statistik und Ökonometrie
96052 Bamberg
E-Mail: martin.messingschlager@uni-bamberg.de

Marcel Preisung
Otto-Friedrich-Universität Bamberg
Fakultät Sozial- und Wirtschaftswissenschaften
Lehrstuhl für Statistik und Ökonometrie
96052 Bamberg
E-Mail: marcel.preising@uni-bamberg.de

Anhang

Operationalisierung:

- *Unterschrift bei Begehren:* Misst dichotom, ob die Befragten beim bereits durchgeführten Bürgerbegehren eine Unterschrift geleistet hatten (1=Ja, 0=Nein).
- *Subjektive Wichtigkeit des Bürgerentscheids:* Ausprägungen wichtig und sehr wichtig zur 1 zusammengefasst, die restlichen drei Ausprägungen zur 0.
- Drei Wissensfragen, auf die die Befragten mit *Ja* oder *Nein* bzw. *Weiß nicht* antworten konnten: 1. ob im FMZ ein Spielcasino und 2. ein *Media Markt* geplant seien. 3. ob sich die *Verkaufsfläche* in Lichtenfels durch das FMZ verdoppele. Die korrekte Antwort lautet jeweils *Ja*. Bei richtiger Antwort wird die Ausprägung 1 zugeordnet, bei falscher oder keiner Antwort eine 0. Hiernach Bildung eines Index *Wissen*, welcher angibt, wie viele der drei Fragen von den Teilnehmern korrekt beantwortet wurden.
- Sieben sachfragenorientierte Merkmale, gemessen via fünfstufiger Likert-Skalen. Ein erhöhter Skalenpunkt bedeutet immer eine positivere Einstellung zum FMZ (bzw. eine minder negative). Aussagen: 1. „Das FMZ wird Arbeitsplätze in der Innenstadt gefährden.“, 2. „Es gibt in Lichtenfels Standorte, die sich besser für das FMZ eignen als die Mainau.“, 3. „Das FMZ ist eine Bereicherung für Lichtenfels.“, 4. „Lichtenfels braucht kein Geld von auswärtigen Investoren.“, 5. „Die Stadtverwaltung und die Bürgermeisterin haben eine klare Vorstellungen darüber, wie es mit dem Einzelhandel in Lichtenfels weitergehen soll.“, 6. „Die Bürgerinitiative „Rettet das Herz der Lichtenfelser Innenstadt“ vertritt einseitig die Interessen der Einzelhändler in Lichtenfels.“, sowie 7. „Ich bin zufrieden mit dem bestehenden Angebot des Einzelhandels in Lichtenfels.“. Es werden jeweils die beiden Kategorien, welche eine eher oder voll positive Haltung zum FMZ widerspiegeln, zur 1 zusammengefasst. Befragten, die eine der restlichen Ausprägungen besetzen, wird eine 0 zugeordnet.²³
- *Parteieneigung:* siehe Fußnote 5.
- *Bianca Fischer (CSU):* Ursprünglich fünfstufige Sympathiebewertung. Die beiden höchsten Merkmalswerte werden zur 1, die restlichen drei Kategorien zur 0 zusammengefasst.

23 Auf die Erzeugung eines Index wurde verzichtet, da die gemessenen Indikatoren keine gemeinsame Dimension abbilden. Zur Messung der Reliabilität wurde Cronbachs α für verschiedene Merkmalskombinationen berechnet. Dieses fiel jedoch in jedem Fall zu gering für eine Indexbildung aus (geringer als 0,5).

- Soziodemographische Merkmale: *Geschlecht* (Mann=1, Frau=0), *Alter* (eingeteilt in drei Klassen <35, ≥35 & ≤65 sowie >65, wobei jüngste Befragte als Referenzkategorie), höchster *Schulabschluss* (eingeteilt in Abitur oder Hochschulabschluss=1, restliche Abschlüsse=0) sowie der *Wohnort* in Lichtenfels, in dem der Befragte lebt. Alle Befragten aus dem Stadtzentrum zur 1, diejenigen aus nördlichen Stadtteilen – und somit relativ nahe am Mainaugebiet Wohnende – zur 2, sowie Personen aus dem Süden Lichtenfels' zur 3 zusammengefasst.

Tabelle 1 Anteil des Datenausfalls in den verwendeten Merkmalen (aufsteigend)

Merkmal	Anteil fehlender Werte (%)
Wissen	0
Geschlecht	0
Angebot nicht ausreichend	0,09
Unterschrift bei Begehren	1,36
Keine Arbeitsplatzgefährdung	1,82
Bürgerentscheid Teilnahmewahrscheinlichkeit	3,18
Alter	3,64
Hohe Bildung	4,09
Einseitige Vertretung durch Bürgerinitiative	12,73
Kein Standort besser geeignet	13,18
Abstimmungsentscheidung	25,45
Wahl prospektiv	26,81
Investorengelder benötigt	28,64
Kumuliert	62,72

Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 2 Multivariate Effekte auf die Wahrscheinlichkeit, für den Stopp des FMZ zu votieren (Datensatz mit Hilfe einer Multiplen Imputation ergänzt)

Merkmal	\hat{B}	$\sqrt{s^2}$
Unterschrift bei Begehren	1,65***	(0,59)
Wissen	-0,92***	(0,32)
Angebot nicht ausreichend	-1,21**	(0,57)
Kein Standort besser geeignet	-1,49***	(0,60)
Keine Arbeitsplatzgefährdung	-1,35***	(0,51)
Investorengelder benötigt	-0,51	(0,61)
Einseitige Vertretung durch BI	-0,33	(0,56)
CSU-Wahl prospektiv	-0,18	(0,73)
SPD-Wahl prospektiv	0,24	(0,79)
Grünen-Wahl prospektiv	0,57	(1,21)
Freie Wähler-Wahl prospektiv	0,69	(1,11)
Geschlecht (Mann)	-0,65	(0,53)
Hohe Bildung	0,41	(0,62)
Mittleres Alter	0,19	(0,70)
Hohes Alter	1,93**	(0,88)
Konstante	3,47***	(1,27)
N		199
Nagelkerke R ²		0,58

Signifikanzniveau: *: p=0,10, **: p=0,05, ***: p=0,01.

Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 3 Übersicht: Parameter der Verteilungen der geschätzten B-Werte

Merkmal	CC-Modell			MI 1-Modell			MI 2-Modell		
	\bar{B}	SD	N	\bar{B}	SD	N	\bar{B}	SD	N
Unterschrift bei Begehren	2,4825	2,3649	630	2,1513	0,7360	999	2,1149	0,7024	1000
Wissen	-1,3958	1,6992	700	-0,8621	0,3529	1000	-0,7240	0,3310	1000
Angebot nicht ausreichend	-2,2884	2,3433	634	-1,1114	0,6462	1000	-0,9708	0,5866	1000
Kein Standort besser geeignet	-0,0475	2,1910	642	-1,4374	0,8664	1000	-1,0615	0,7037	1000
Keine Arbeitsplatzfähigkeit	-4,1296	2,2835	577	-1,7020	0,8094	1000	-1,5751	0,6917	1000
Investorengelder benötigt	-1,2416	2,5987	617	-0,9855	0,7685	1000	-0,6425	0,6532	1000
Einseitige Vertretung durch BI	-1,6423	2,1564	639	-0,4856	0,7071	1000	-0,5826	0,6504	1000
CSU-Wahl prospektiv	1,6440	3,1126	635	-0,1903	0,9701	1000	-0,6714	0,7691	1000
SPD-Wahl prospektiv	4,7500	2,7539	499	-0,0824	1,0028	1000	-0,2513	0,8393	1000
Grüne-Wahl prospektiv	2,6563	3,2990	372	1,3548	1,9807	998	0,2385	1,3309	1000
Freie Wähler-Wahl prospektiv	2,2501	2,9513	587	0,5625	1,2204	1000	0,1174	1,0260	1000
Geschlecht (Mann)	-1,2727	2,2564	651	-0,5811	0,6953	1000	-0,4297	0,6132	1000
Hohe Bildung	2,5671	2,3720	612	0,5319	0,8119	1000	0,4301	0,7037	999
Mittleres Alter	2,1549	2,9270	594	0,5198	0,9284	1000	0,3105	0,8027	1000
Hohes Alter	4,4181	2,8009	525	2,8320	1,4045	996	2,5235	1,1820	998
Konstante	1,3793	3,7776	538	3,3542	1,7332	998	3,1573	1,5809	999

Angaben sind arithmetische Mittel und Standardfehler der jeweils 1.000 simulierten B-Schätzer. N entspricht der Anzahl an Parametern, die im Intervall von -10 bis +10 liegen. Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 4 Übersicht: Parameter der Verteilungen der geschätzten p-Werte

Merkmal	CC-Modell		MI 1-Modell		MI 2-Modell	
	\bar{p}	SD	\bar{p}	SD	\bar{p}	SD
Unterschrift bei Begehren1	0,5475	0,4065	0,0168	0,0485	0,0135	0,0425
Wissen	0,5125	0,4023	0,0857	0,1345	0,1062	0,1460
Angebot nicht ausreichend	0,5135	0,4261	0,1549	0,1820	0,1810	0,2044
Kein Standort besser geeignet	0,6849	0,3325	0,1438	0,1709	0,1943	0,1831
Keine Arbeitsplatzgefährdung	0,4426	0,4655	0,0631	0,1195	0,0611	0,1174
Investorengelder benötigt	0,6237	0,3718	0,2225	0,1900	0,2897	0,1906
Einseitige Vertretung durch BI	0,5611	0,4056	0,3130	0,2118	0,3007	0,2165
CSU-Wahl prospektiv	0,6621	0,3576	0,3190	0,1690	0,2827	0,1830
SPD-Wahl prospektiv	0,5075	0,4442	0,3345	0,1679	0,3540	0,1785
Grüne-Wahl prospektiv	0,7772	0,3408	0,3715	0,1376	0,3706	0,1342
Freie Wähler-Wahl prospektiv	0,6412	0,3729	0,3247	0,1632	0,3583	0,1587
Geschlecht (Mann)	0,5966	0,3772	0,3011	0,2205	0,3428	0,2219
Hohe Bildung	0,5487	0,4031	0,3386	0,2515	0,3650	0,3038
Mittleres Alter	0,5901	0,3968	0,3541	0,2246	0,4109	0,2209
Hohes Alter	0,5159	0,4441	0,0729	0,1332	0,0754	0,1391
Konstante	0,6606	0,3614	0,1086	0,1566	0,1050	0,1473

Angaben sind arithmetische Mittel und Standardfehler der jeweils 1.000 simulierten p-Werte.

Quelle: Datensatz zur Befragung in Lichtenfels.

Soziale Präsenz in Onlineumfragen

Social Presence in Online Surveys

Alexander W. Schmidt-Catran und Katharina Hörstermann

Zusammenfassung

Onlinebefragungen zählen mittlerweile zum Standardrepertoire vieler Forschungsanstalten, einige methodische Fragen bleiben bis dato jedoch noch unbeantwortet. Diese Studie beschäftigt sich mit der Frage, ob durch die Simulation sozialer Präsenz das Antwortverhalten der Befragten beeinflusst werden kann. Mit einem experimentellen Design werden Effekte der sozialen Präsenz auf Nonresponse sowie sozial erwünschtes Antwortverhalten untersucht. Dabei wird davon ausgegangen, dass die Implementierung sozialer Präsenz den Vorteil einer erhöhten Teilnahmebereitschaft und den Nachteil von sozial erwünschtem Antwortverhalten mit sich bringen kann. Der inhaltliche Schwerpunkt des Experiments liegt auf Einstellungen zu Sexismus, Geschlechterrollen und der Berufstätigkeit von Frauen. Zur Simulation der sozialen Präsenz werden Fotografien verwendet. Zusätzlich zum Effekt der sozialen Präsenz wird untersucht, ob das Geschlecht und die Attraktivität der auf den Fotos gezeigten Personen einen Einfluss auf das Antwortverhalten haben. Bei diesem Effekt handelt es sich um eine Art „Gender-of-Interviewer“-Effekt. Bezüglich der Teilnahme- und Antwortbereitschaft konnten keine positiven Effekte nachgewiesen werden. Die Ergebnisse hin-

Abstract

Today, online surveys belong to the standard instruments of most survey research institutes, but some methodical questions are still unanswered. This study deals with the question of whether the simulation of social presence has an effect on the response style. Using an experimental design, the effects of social presence on non-response and socially desirable answers are tested. We expect social presence to lower non-response but to induce socially desirable answers. Topics of the survey are attitudes towards sexism, gender roles and the participation of women in the labor market. Social presence is simulated by pictures of the "investigators" of the study. In addition to the effect of social presence, the study investigates whether the gender and the attractiveness of the person on the pictures have an effect on the answers. This effect might be called a "gender-of-interviewer"-effect. Concerning the non-response rate no positive effect of social presence could be found. The results with regard to social desirability and a possible gender-of-interviewer-effect are ambiguous but show significant differences between the control group and our experimental groups.



sichtlich möglicher Interviewereffekte in Form von sozial erwünschtem Antwortverhalten sind nicht eindeutig, es tauchen aber systematische Unterschiede zwischen den Experimental- und der Kontrollgruppe auf.

1 Einleitung

Umfragedaten bilden seit Jahrzehnten einen der wichtigsten Bestandteile der sozialwissenschaftlichen Forschung. In den letzten Jahren kam es dabei durch veränderte technologische, ökonomische und gesellschaftliche Rahmenbedingungen zu einer Anpassung in den Methoden der Umfrageforschung. Generell ist die Entwicklung durch eine Technisierung und insbesondere Computerisierung geprägt, die sich in einer weiteren Ausdifferenzierung möglicher Erhebungsarten niederschlug (Weichbold 2005; Gerich 2009). Die neuen Erhebungsmethoden ebneten unter anderem den Weg zur Implementierung multimedialer Elemente und lassen das Ausmaß möglicher Interaktion zwischen Interviewer und Befragten sowie den Einsatz experimenteller Stimuli variabler gestalten. Die Auswirkungen dieser Neuerungen auf die Qualität der Daten sind bis dato jedoch noch nicht vollständig entschlüsselt (Couper 2005).

In diesem Beitrag soll der Frage nachgegangen werden, wie sich die Simulation sozialer Präsenz in Onlineumfragen auf das Antwortverhalten der Befragten auswirkt. Ein möglicher positiver Effekt ist die Erhöhung der in Onlineumfragen üblicherweise geringen Teilnahmebereitschaft. Ein erwarteter negativer Effekt ist die Tendenz zu sozial erwünschtem Antwortverhalten. Über mögliche Interviewereffekte bei Einstellungsfragen in Face-to-Face-Befragungen wurde bereits umfangreich geforscht (u.a. Steinert 1984; Reinecke 1993; Haunberger 2006) und ihr Auftreten gilt mittlerweile als gesichertes Faktum. Bei telefonischen Befragungen entfallen zwar sichtbare Interviewermerkmale, nichtsdestotrotz führen Stimmcharakteristika des Interviewers zu Verzerrungen bei Einstellungsfragen (Klein/Kühhirt 2010). Das Auftreten klassischer Interviewereffekte bei Onlineumfragen wird bisweilen ausgeschlossen, da keine direkte Interaktion zwischen Interviewer und Befragten existiert.

Die Studie orientiert sich an einem Experiment von Tourangeau et al. (2003), das sich derselben Thematik widmete, jedoch methodische Schwächen aufweist, die eine eindeutige Interpretation der Ergebnisse nicht ermöglichen. Diese Schwächen werden in der vorliegenden Studie berücksichtigt. Zusätzlich untersuchen wir

neben Gender-of-Interviewer-Effekten erstmals Effekte der Interviewerattraktivität auf das Antwortverhalten von Befragten.

2 Theoretischer Hintergrund

Personen werden durch die Anwesenheit anderer Personen in ihrem Verhalten beeinflusst. Im Rahmen der Umfrageforschung sind es häufig die Interviewer selbst, die das Antwortverhalten der Befragten systematisch beeinflussen (Diekmann 2004). Die Auswirkungen auf die Antwort- und Teilnahmebereitschaft der Befragten und die Korrektheit der Antworten sind daher nicht selten ein zentrales Kriterium bei der Wahl der „richtigen“ Befragungsform. Wie sich die hier verhandelten Modi persönliches Interview sowie Onlinebefragung bezüglich beider Merkmale unterscheiden, soll deshalb kurz erläutert werden.

Das persönliche Interview wurde einst von Rene König (1972) – einem der Begründer der modernen Sozialforschung in Deutschland – als „Königsweg“ der Sozialforschung bezeichnet. Es knüpft an die alltägliche Situation des Fragestellens an, ist gleichzeitig aber auch eine künstliche und asymmetrische Interaktion unter Fremden. Sowohl Interviewsituation, als auch Interviewerverhalten und Konstruktion des Fragebogens können dabei einen Einfluss auf das Teilnahme- und Antwortverhalten des Befragten haben. Die Entscheidung zur Teilnahme an der Befragung ist im Rahmen jeder Befragungsmethode das Ergebnis einer Kosten-Nutzen-Abwägung, welche sich über die Thematik der Befragung, die Verwendung eines populären Sponsors, wiederholte Kontaktversuche oder auch schriftliche Ankündigung positiv beeinflussen lässt (Groves et al. 2004). Allgemein hat die Teilnahmebereitschaft stark unter den sich veränderten gesellschaftlichen Rahmenbedingungen der letzten Jahrzehnte gelitten. Neben demographischen Änderungen führten auch Phänomene wie zunehmende Mobilität, Veränderungen in der Bedeutung informationeller Selbstbestimmung und zunehmende Furcht vor Haustürkäufen und Kriminalität zu einer abnehmenden Bereitschaft an Umfragen teilzunehmen (Evans/Mathur 2005; Porst et al. 1998).

Gegenüber einigen anderen Befragungsmodi hat das persönliche Interview durch die direkte Interaktion zwischen Interviewer und Befragten den Vorteil, dass Befragte eher bereit sind an der Umfrage teilzunehmen, bzw. die Hemmschwelle höher ist, gegenüber dem Interviewer die Teilnahmeverweigerung zu kommunizieren. Die Ausschöpfungsquote ist daher bei den interaktiven Befragungsformen am höchsten (Groves et al. 2004). Auch bei der Antwortverweigerungsquote erzielen interaktive Befragungsformen die besten Werte (ebd.). Interviewer können bei Ver-

ständnisfragen unterstützen, auf die korrekte Bearbeitung des Fragebogens achten, bei fehlenden Angaben unmittelbar nachfragen und um Vollständigkeit der Angaben bitten. Zudem besteht auch hier wiederum eine gewisse Hemmschwelle, dem Interviewer gegenüber die Antwortverweigerung zu kommunizieren.

Das persönliche Interview hat aber nicht nur Vorteile. So kann es sein, dass sich der Befragte durch die Anwesenheit einer anderen Person dazu verleitet fühlt, sich dieser gegenüber in ein positives Licht zu rücken. In der Umfrageforschung schlägt sich das darin nieder, dass die Befragten Antworten geben, von denen sie glauben, sie trafen eher auf Zustimmung als die korrekte Antwort, bei der sie soziale Ablehnung befürchten. Im Fachterminus wird von „sozialer Erwünschtheit“ oder auch „sozial erwünschten Antwortverhalten“ gesprochen. Da der Befragte dem Fragenden beim persönlichen Interview unmittelbar gegenüber sitzt, wird er eher geneigt sein, seine Antworten in die Richtung des sozial Erwünschten zu korrigieren (Heerwegh 2009). Insbesondere bei Fragen mit besonders heiklem oder sensiblen Inhalt passt der Interviewte seine Antwort eventuell dem sozial Erwünschten an oder er verweigert die Angabe. Zusammenfassend kann also festgehalten werden, dass Befragungsmodi mit einem hohen Grad an sozialer Interaktion den Vorteil von geringeren (Item- und Unit-) Nonresponsequoten haben, sie allerdings auch anfällig für sozial erwünschtes Antwortverhalten sind.¹

Mit der Verbreitung des Internets übernahmen viele Markt- und Sozialforschungsinstitute die Onlineumfrage in ihr Repertoire und mittlerweile gelten internetbasierte Befragungen als die am weitesten verbreitete Befragungsform (ADM 2010). Neben ihren geringen Kosten sind die hohe zeitliche Flexibilität und die multimedialen Gestaltungsmöglichkeiten Vorteile gegenüber dem klassischen persönlichen Interview (Evans/Mathur 2005; Schnell 2012). Aufgrund des relativ anonymen Umfelds sind Onlinebefragungen zudem weniger anfällig für soziale Erwünschtheit und erzielen bei sensiblen Themen höhere Responsequoten (Holbrook/Green/Krosnick 2003). Probleme der Onlineforschung liegen jedoch unter anderem in ihrer Repräsentativität, den Möglichkeiten der Stichprobenziehung und der niedrigen Ausschöpfungsquote (Evans/Mathur 2005). Hinzu kommt, dass das einfache „Wegklicken“ einer Umfrage am Computer ein geringeres Hemmnis darstellt als die Kommunikation der Teilnahmeverweigerung gegenüber einem Interviewer. Die geringe soziale Kontrolle und das anonyme Umfeld mindern somit zwar das Risiko sozial erwünschten Antwortens und der Antwortverweigerungen bei heiklen Fragen im

1 Formale Modelle des Antwortprozesses, die den Entstehungsprozess und die Determinanten sozialer Erwünschtheit beinhalten, finden sich bei Groves et al. (2004), Tourangeau et al. (2000) sowie Krosnick und Alwin (1987).

Vergleich zu persönlichen Befragungen, bringen aber das Risiko erhöhter Nonresponsequoten mit sich.

Die Social-Interface-Theorie geht davon aus, dass über die Simulation sozialer Präsenz in Onlineumfragen ähnliches Antwortverhalten evoziert werden kann wie in Formaten mit direkter Kommunikation. Nass et al. (1999) berichten in ihrer Studie, dass die Tendenz zu Geschlechterstereotypen bei Onlinebefragungen bereits durch kleine Schlüsselreize wie z.B. Computerstimmen erhöht wird. Auf der Basis weiterer Experimente schlussfolgerten Nass und seine Kollegen (Fogg/Nass 1997; Nass/Fogg/Moon 1996; Nass/Moon/Carney 1999; Nass/Moon/Green 1997; Nass/Moon 2000), dass sich Personen gegenüber der grafischen Benutzeroberfläche genauso verhalten wie gegenüber sozialen Akteuren. In diesem Sinne gehen Sundar und Nass (2000) davon aus, dass Heuristiken und Regeln der sozialen Interaktion auf die Mensch-Computer-Interaktion übertragen werden. Sproull et al. (1996) verglichen in ihrem Experiment die Antworten auf verschiedene Items in Abhängigkeit davon, ob die Items von einem sprechendem Gesicht vorgelesen werden oder ob der Befragte die Items vom Bildschirm abliest. Sie kamen zu dem Ergebnis, dass die Probanden sich gegenüber dem sprechenden Gesicht „in a more positiv light“ (S.116) präsentierten. Sind die Annahmen der Social-Interface-Theorie korrekt, dann ist damit zu rechnen, dass der Nutzen der Simulation sozialer Präsenz in Form von geringeren Nonresponsequoten von einem höheren Risiko sozialer Erwünschtheit flankiert wird.

Aus bisheriger Forschung ist ferner bekannt, dass Befragte bei Geschlechterthematiken gegenüber weiblichen Interviewern tendenziell dazu neigen, häufiger Partei für Frauen zu ergreifen als gegenüber männlichen Interviewern. Dies wurde vor allem bezüglich der wahrgenommenen Diskriminierung von Frauen (Ballou/DelBoca 1980; Kane/Macaulay 1993), Einstellungen zu Geschlechterrollen (Ballou/DelBoca 1980; Grimes/Hansen 1984; Lueptow/Moser/Pendleton 1990; Kane/Macaulay 1993) und der Unterstützung von Frauenrechten (Kane/Macaulay 1993; Huddy et al. 1997) festgestellt. Steht der thematische Schwerpunkt der Umfrage in keinem Zusammenhang mit dem Geschlecht des Befragten, sind die Ergebnisse bezüglich eines Gender-of-Interviewer-Effekts inkonsistent (vgl. auch Klein/Kühhirt 2010). Couper et al. (2001) testeten ausgehend von der Social-Interface-Theorie, ob sich Gender-Of-Interviewer-Effekte auch in Onlineumfragen mit sozialer Präsenz nachweisen lassen. Als experimentellen Stimulus präsentierten sie den Befragten an einigen Stellen im Fragebogen entweder das Bild eines Forschers oder einer Forscherin sowie einen persönlichen Begrüßungstext. Sie konnten zeigen, dass die Befragten bei Items bezüglich ihrer Einstellung zu Geschlechterrollen gegenüber einer Forscherin signifikant pro-feministischere Einstellungen äußerten

als gegenüber einem Mann. Als möglichen Erklärungsansatz verweisen die Autoren auf die „polite stranger“-Hypothese, einer Ableitung des Race-of-Interviewer-Effekts, nach der Personen aus Höflichkeitsgründen möglichst vermeiden, negative Gruppenstereotype gegenüber einer Person dieser Gruppe zu artikulieren (mehr Informationen zum Race-of-interviewer-Effekt u.a. in Athey et al. 1960; Schuman/Converse 1971).

Tourangeau et al. (2003) verwendeten ebenfalls Bilder eines Forschers oder einer Forscherin als experimentellen Stimulus, wobei das Studien-Logo als Kontrollgruppe diente. Die Autoren fanden keine Hinweise auf verringerte Abbruchquoten in den Experimentalgruppen. Die Annahmen bezüglich sozial erwünschter Antworten konnten im Hinblick auf geschlechtsspezifische Thematiken nachgewiesen werden. So gaben die Befragten signifikant pro-feministischere Antworten, wenn sie das Foto einer Frau gezeigt bekamen. Fuchs (2009) verstärkte den experimentellen Stimulus und verwendete statt Bildern des Interviewers kurze Videosequenzen, in denen der Interviewer bzw. die Interviewerin dem Befragten die Fragen vorlasen. Die gefundenen Gender-of-Interviewer-Effekte waren mit denen eines klassischen Face-to-Face-Interviews vergleichbar. Aus unserer Sicht sind die Ergebnisse der zuvor präsentierten Studien hinsichtlich des Gender-of-Interviewer-Effekts jedoch nur mit Vorsicht zu interpretieren, da pro Experimentalgruppe nur jeweils ein Bild bzw. Video eines Interviewers verwendet wird und die verwendeten Bilder im Vorfeld nicht auf physische Attraktivität des Interviewers getestet wurden und es somit nicht ausgeschlossen werden kann, dass gefundene Effekt tatsächlich durch Attraktivität erklärt werden können.

Nicht eindeutig geklärt werden konnte bis jetzt, ob entweder Männer oder Frauen generell stärker auf das Geschlecht des Interviewers reagieren. Während Kane und Macaulay (1993) keinen Unterschied in der Anfälligkeit für Gender-of-Interviewer-Effekte zwischen den Geschlechtern feststellen konnten, berichteten z.B. Whelchel (1987) sowie Ballou und Del Boca (1980) von einer stärkeren Anfälligkeit seitens der männlichen Befragten. Argentino et al. (1977) und Lueptow et al. (1990) hingegen belegten eine stärkere Beeinflussbarkeit der weiblichen Befragten. Ein Blick in die Literatur zum Thema Geschlecht und Konversation legt die Vermutung nahe, dass durchaus geschlechtsspezifische Unterschiede beim Gender-of-Interviewer-Effekt erwartet werden können. Die Gesprächsführung von Männern charakterisiert sich im Allgemeinen durch ein dominantes Auftreten, häufige Gesprächsunterbrechungen und dem Bestreben nach Gesprächskontrolle. Frauen streben hingegen viel stärker danach, im Gespräch zu einem Konsens zu gelangen (Athenstaed/Hass/Schwab 2004; Carli/Bukatko 2000; Kollock/Blumstein/Schwartz 1985; Leaper/Robnett 2011; Spencer/Drass 1989; Zimmerman/West 1975). Als mög-

liche Erklärungen werden in der Literatur zwei Ansätze verhandelt: Zum einen die These, nach der sich das unterschiedliche Verhalten in einem Gespräch aus Status- oder Machtgefällen zwischen Männern und Frauen ergibt (vgl. Johnson 1994; Lakoff 1973; Wood/Dindia 1998) und zum anderen die Annahme, dass die Ursachen in einer geschlechtsspezifische Sozialisation zu finden sind. Die weibliche Geschlechterrolle impliziere dabei ein eher emotionales und soziales Verhaltensmuster, während die männliche stärker einem instrumentellen und wenig emotionalen Verhalten entspreche (Geis 1993; Hall/Briton 1993). Die Meta-Analysen von Twenge (2001) sowie Leaper und Robnett (2011) untersuchten, ob sich die Geschlechterunterschiede im Kommunikationsverhalten durch veränderte gesellschaftliche Rollenbilder im Zeitverlauf verringert haben. Während Twenge (2001) noch von einer Annäherung der geschlechtsspezifischen Stile berichtet, konnten Leaper und Robnett (2011) keine Entwicklung feststellen.

Ferner ist bekannt, dass Gender-of-Interviewer-Effekte vor allem bei geschlechtsheterogenen Gesprächspartnern auftreten (Tu/Liao 2007). Unterhalten sich Männer und Frauen, passen sich beide Gesprächspartner an das Verhalten des jeweiligen Gegenübers an, wobei in Einklang mit den zuvor genannten Theorien das Entgegenkommen der Frauen stärker ausgeprägt ist (Athenstaedt/Haas/Schwab 2004; Deaux/LaFrance 1998). Im Falle gleichgeschlechtlicher Gesprächspartner lassen sich verstärkt geschlechtsstereotypische Verhaltensmuster finden (Athenstaedt/Haas/Schwab 2004; Fitzpatrick/Mulac/Dindia 1995). Laut Aries (1976) sowie Dindia und Allen (1992) präferieren Frauen weibliche Gesprächspartner, da sie in der gleichgeschlechtlichen Konversation größere Freiheiten und Autonomie verspüren. Landis et al. (1973) sowie Tu und Liao (2007) konnten für beide Geschlechter zeigen, dass bei gleichgeschlechtlichen Interviewpartnern die soziale Distanz geringer ausgeprägt war, wodurch die Kommunikation erleichtert wurde und die Antworten weniger sozial erwünscht ausfielen. Die Ergebnisse von Fuchs' (2009) personalisierter Onlineumfrage zeigen, dass Gender-of-Interviewer-Effekte nicht nur in geschlechtsheterogenen, sondern auch bei geschlechtshomogenen Gesprächskombinationen auftreten können, je nachdem welche geschlechtsstereotypischen Vorstellungen durch die Frage aktiviert wurden.

Neben dem Geschlecht des Interviewers sollte sich auch dessen Attraktivität auf das Antwortverhalten des Befragten auswirken. Aus der Attraktivitätsforschung ist bekannt, dass attraktive Personen mehr Aufmerksamkeit auf sich ziehen („Attractive Attention Boost“, siehe u.a. Maner et al. 2003; Mulford et al. 1998), ihnen getreu der Logik „what is beautiful is good“ (unberechtigterweise) prinzipiell positivere Persönlichkeitseigenschaften zugeschrieben werden („Attractiveness Stereotype“, siehe u.a. Dion/Berscheid/Walster 1972) und objektiv gegebenes Fehl-

verhalten leichter vergeben wird („Attractiveness Glamour Effect“, siehe auch Basili 1981; Grammer 2002). Des Weiteren dürfen attraktive Menschen stärker darauf hoffen, in konkreten Handlungssituationen eine bessere Behandlung und mehr Respekt durch andere Menschen zu erfahren („Attractiveness Treatment Advantage“, Bian 1997; Dabbs/Stokes 1975; Wilson/Eckel 2006). Studien, die sich direkt mit dem Einfluss von Attraktivität des Interviewers auf das Antwortverhalten von Befragten beschäftigen, liegen noch nicht vor. Jedoch lassen sich einige Erkenntnisse aus anderen Studien auf die Interviewsituationen übertragen: So konnte beispielsweise gezeigt werden, dass attraktivere Personen bessere Verkaufszahlen erzielen als unattraktive Verkäufer (DeShields/Kara/Kaynak 1996; Reinigen/Kernan 1993). Im Kontext der Umfrageforschung kann man folglich vermuten, dass es attraktiveren Personen besser gelingt, Befragungspersonen für ihre Umfrage zu gewinnen und die Responsequoten zu steigern. Besondere Aufmerksamkeit sollte dabei attraktiven Personen des anderen Geschlechts gezollt werden. Sowohl Männer als auch Frauen sind evolutionsbedingt gewillt, insbesondere mit attraktiven Menschen Beziehungen einzugehen, da diesen hohe Gesundheit und Fertilität zugeschrieben wird (Maner et al. 2003). Sozial erwünschtes Antwortverhalten sollte daher mit der Attraktivität des Interviewers zunehmen, insbesondere wenn Interviewer und Befragter unterschiedlichen Geschlechtern angehören. Gutierrez et al. (1999), Buunk et al. (2007) sowie Maner et al. (2009) weisen jedoch darauf hin, dass im Falle gleichgeschlechtlicher Gesprächspartner vor allem Frauen negativ auf attraktive Gegenüber reagieren, da sie in ihnen potenzielle Konkurrentinnen sehen. Diese Befunde legen nahe, dass weibliche Befragte gegenüber einer attraktiven Interviewerin weniger Entgegenkommen zeigen sollten.

2.1 Hypothesen

Auf Grundlage der zuvor skizzierten theoretischen und empirischen Befunde werden folgende Forschungshypothesen aufgestellt:

Response

Onlineumfragen weisen in der Regel geringere Responsequoten auf als Face-to-Face-Befragungen, da die Kosten eines Interviewabbruchs oder einer Antwortverweigerung deutlich geringer sind. Durch die Simulation sozialer Präsenz in Onlinebefragungen sollte die Umfrage zum einen motivierender und interessanter auf den Befragten wirken und zum anderen im Sinne der Social-Interface-Theorie die Hemmschwelle für eine Antwortverweigerung erhöhen.

H_{1a}: Die Simulation sozialer Präsenz führt zu verringerten Nonresponsequoten und zu einer Effizienzsteigerung von Antwortaufforderungen.

Effekte der Attraktivität

Sozialpsychologische Untersuchungen legen nahe, dass die Attraktivität einer Person einen Einfluss auf das Verhalten der Interaktionspartner hat. Attraktiven Personen werden positive Eigenschaften zugeschrieben und dies führt in der Regel dazu, dass ihnen gegenüber mehr Entgegenkommen gezeigt wird. Der Wunsch einer attraktiven Person zu helfen, sollte sich daher in einer Steigerung der Responsequoten niederschlagen.

H_{1b}: Der positive Effekt der Simulation von sozialer Präsenz auf die Responsequoten wird verstärkt, wenn die gezeigte Person als attraktiv bewertet wird.

Soziale Erwünschtheit

Durch die Simulation sozialer Präsenz in Onlineumfragen wird die Gefahr von sozial erwünschtem Antwortverhalten erhöht, denn entsprechend der Social-Interface-Theorie werden die für das persönliche Interview geltenden Gesprächsnormen auf den Kontext von Websurveys projiziert. Es wird davon ausgegangen, dass egalitäre Einstellungen zu Geschlechterrollen sowie eine kritische Einstellung zu sexistischem Verhalten dem allgemein Anerkannten und sozial Erwünschten entsprechen. Männer und Frauen sollten somit unter der Bedingung sozialer Präsenz Sexismus stärker verurteilen und egalitärere Rollenvorstellungen bekunden.

H_{2a}: Bei Einstellungsfragen verleitet die fingierte soziale Präsenz zu einer Editierung der Antworten in Richtung des sozial Erwünschten.

Aus der Gesprächsanalyse ist ferner bekannt, dass Männer und Frauen in Gesprächen unterschiedliche Verhaltensmuster an den Tag legen. Frauen vertreten einen weniger dominanten Gesprächsstil und suchen eher den Konsens mit ihrem Gesprächspartner. Somit sollten sie generell anfälliger sein für sozial erwünschtes Antwortverhalten bei der Simulation sozialer Präsenz in Onlineumfragen.

H_{2b}: Die Simulation sozialer Präsenz bewirkt bei Frauen eine stärkere Anpassung der Antworten als bei Männern.

Gender-of-Interviewer-Effekt

Das Geschlecht des Interviewers ist bei der Erhebung von Einstellungen zu Geschlechtsthematiken von entscheidender Bedeutung für das Ausmaß sozial erwünschten Antwortverhaltens. Bei weiblichen Interviewern werden eher egalitäre Einstellungen hinsichtlich Geschlechterrollen als sozial erwünscht assoziiert als bei männlichen Interviewern.

H_{3a}: Gegenüber einem weiblichen Interviewer werden stärker egalitäre Einstellungen geäußert als gegenüber einem männlichen.

Das Ausmaß des Gender-of-Interviewer-Effekts wird vom Geschlecht des Befragten determiniert. Analog zu Hypothese H_{2b} wird angenommen, dass Frauen aufgrund der stärkeren Konsensorientierung ihre Antworten stärker editieren als Männer.

H_{3b}: Das Geschlecht des Befragten moderiert das Ausmaß des Gender-of-Interviewer-Effekts. Frauen editieren ihre Antworten stärker als Männer.

Einfluss der Attraktivität des Interviewers

Attraktiven Personen schenkt man mehr Aufmerksamkeit und man ist ihnen eher behilflich als weniger attraktiven. Simuliert man soziale Präsenz in Onlineumfragen mittels eines Fotos, besteht die Gefahr, dass das Antwortverhalten der Probanden von der physischen Attraktivität des Interviewers beeinflusst wird. Man möchte einer attraktiven Person, vor allem des anderen Geschlechts, gefallen, was sich verstärkend auf die Tendenz zu sozial erwünschtem Antwortverhalten auswirkt. Wird allerdings eine Frau von einer attraktiven Frau interviewt, manifestiert sich in dieser Dyade eine Konkurrenzsituation. Diese gefühlte Konkurrenz wird dazu führen, dass die Probandin gegenüber der attraktiven Interviewerin weniger gefällig und bewusst entgegen des sozial Erwünschten reagiert.

H₄: Je attraktiver der Interviewer/die Interviewerin, desto stärker ist der Effekt sozialer Erwünschtheit. Dies gilt insbesondere bei Interviewern des anderen Geschlechts. Eine Ausnahme bildet hier die Konstellation attraktive Interviewerin und weibliche Befragte.

3 Methode, Daten und Operationalisierung

3.1 Methode

Insgesamt wurden drei verschiedene Varianten der Onlinebefragung generiert, darunter zwei Experimentalgruppen und eine Kontrollgruppe. In den Fragebögen der Experimentalgruppen wurden Fotos und ein Anschreiben der/s Forscherin/s eingefügt. Das Anschreiben erhält einen persönlichen Charakter, indem der Forscher/die Forscherin sich kurz vorstellt und betont, wie wichtig die Ergebnisse der Studie für sie/ihn sind (vgl. Abbildung 1). Dies erfolgt mit der Intention, innerhalb der Onlineumfrage eine Interaktion zwischen Interviewer und Befragten zu simulieren, wie man sie bei persönlichen Interviews oder auch Telefonbefragungen vorfindet. Damit soll eines der zentralen Probleme von Onlinebefragungen – die geringen Responsequoten – umgangen werden. Der Eindruck sozialer Kontrolle soll dadurch verschärft werden, dass im Falle einer fehlenden Angabe ein Hinweis für den Befragten eingeblendet wird, auf dem erneut das Foto der/s Forscherin/s zu sehen ist verbunden mit der Bitte, auf alle Fragen zu antworten, um die Studie nicht zu gefährden (vgl. Abbildung 2).

Um den erwarteten Gender-of-Interviewer-Effekt zu testen, wurden zwei Varianten der Experimentalgruppe erstellt: Variante 1 zeigt das Bild eines männlichen Interviewers, in Variante 2 wurde ein Foto einer weiblichen Interviewerin verwendet.² Um den Attraktivitätseffekt zu messen, wurden in beiden Experimentalgruppen jeweils vier Fotos von unterschiedlich attraktiven Interviewern verwendet.³ Die Attraktivität der Interviewer auf den Fotos wurde zuvor von Personen, die nicht an der Studie teilnahmen, bewertet. Blickwinkel, Kleidung und Frisur der Personen auf den Bildern waren dabei immer identisch. Alle Personen hatten zudem ungefähr das gleiche Alter (Mitte 20 bis Mitte 30). Pro Experimentalgruppe wurden eine besonders attraktive Person, zwei durchschnittlich attraktive und eine weniger attraktive Person ausgewählt (siehe Tabelle A1 im Appendix).⁴ Da aus Forschungen der Sozialpsychologie bekannt ist, dass die Attraktivitätsbewertung einer

2 Da wir an Gender-of-Interviewer-Effekten interessiert sind, verwenden wir im Folgenden weiterhin den Begriff Interviewer anstelle des Begriffs Forscher. Tatsächlich haben sich die Personen auf den Fotos aber als Urheber der Studie ausgegeben (vgl. Abbildung 1).

3 An dieser Stelle gilt unser besonderer Dank Joël Binckli (Universität zu Köln) und Ulrich Rosar (Universität Düsseldorf) für die Bereitstellung der Fotos.

4 Die Fotos wurden von insgesamt 13 männlichen und 13 weiblichen Personen bewertet. Nach Henss (1992) reicht diese Zahl aus um ein äußerst stabiles Attraktivitätsranking zu erreichen.

Abbildung 1 Personalisierter Begrüßungstext der Experimentalgruppe Mann (anonymisiert)

0% ausgefüllt

Herzlich Willkommen!

Mein Name ist



Vielen Dank, dass sie sich die Zeit nehmen, an meiner Studie teilzunehmen. In der folgenden Umfrage möchte ich gerne ihre Meinung zu den **Themen Sexismus, Geschlechterrollen und Geschlechtergleichstellung** erfahren.


Sie können mit ihrer Teilnahme einen wichtigen Beitrag zur wissenschaftlichen Bewertung dieses Themas leisten. Die Befragung wird **ca. 10 Minuten** dauern. Ihre Daten werden anonym erhoben und von mir streng vertraulich behandelt, d.h. ihre Daten werden von mir nicht an Dritte weitergegeben und ein Rückschluss auf ihre Person wird zu keinem Zeitpunkt möglich sein.

Bitte lesen Sie die folgenden Fragen sorgfältig durch und beantworten Sie jede davon so spontan und ehrlich wie möglich. Beachten Sie, dass es bei der Beantwortung der Fragen keine richtigen oder falschen Antworten gibt. Ich bin an ihrer persönlichen Meinung interessiert.

Mit freundlichen Grüßen

Abbildung 2 Antwortaufforderung in der Experimentalgruppe Mann (anonymisiert)

5% ausgefüllt



Sie haben meine Frage nicht beantwortet. Für meine Studie ist es sehr wichtig, dass Sie möglichst alle Fragen beantworten.

Ich möchte Sie daher bitten, zu überlegen, ob sie den "Zurück"-Button nutzen wollen, um ihre Antwort noch einmal zu verändern. Andernfalls klicken Sie bitte auf "Weiter".

Person ein sehr guter Proxy für andere positive Merkmalszuschreibungen – wie u.a. Kompetenz und Intelligenz – einer Person ist, kann davon ausgegangen werden, dass wir nicht nur die Attraktivität des Interviewers kontrollieren, sondern auch die Einschätzung von Intelligenz und Kompetenz berücksichtigen. Unterschiedliche Bewertungen zwischen gleichattraktiven Personen verschiedenen Geschlechts sollten somit nur auf den Effekt des Geschlechts und nicht auf unterschiedliche Kompetenz- oder Intelligenzeinstufungen zurückzuführen sein.

Als Kontrollgruppe wurde eine „klassische“ Variante der Umfrage geschaltet, in der die Universität zu Köln als Initiator der Befragung genannt wird und ihr Siegel anstelle eines Fotos angezeigt wird.

Der thematische Schwerpunkt der Befragung liegt auf den Einstellungen der Befragten zu Geschlechterrollen und -gleichstellung sowie der zum Zeitpunkt der Befragung aktuellen Sexismus-Debatte, ausgelöst durch Äußerungen des FDP-Politikers Rainer Brüderle gegenüber einer Journalistin. Die Einstellungen zu Geschlechterrollen wurden mit Hilfe von Items aus den ALLBUS-Wellen 1994 und 2008 operationalisiert. Zur Sexismusdebatte wurden zwei Items des Umfrageinstituts YouGov übernommen und zwei Items selbst konstruiert (siehe Tabelle A2 im Appendix).

3.2 Daten und Operationalisierung

Die Daten wurden mit Hilfe des Socio-Scientific-Panels⁵ erhoben, einem nicht-repräsentativen Pool von Interviewpartnern zur Unterstützung der wissenschaftlichen Forschung an Universitäten und Hochschulen. Für die Identifikation eines kausalen Effektes im Experiment ist die Repräsentativität des Panels keine notwendige Voraussetzung, da durch die zufällige Zuweisung der Befragten in Kontroll- und Experimentalgruppen auf mögliche Einflüsse durch Drittvariablen kontrolliert wird. Unterschiede zwischen Experimental- und Kontrollgruppe bei der Reaktion auf den Stimulus können somit nicht durch unterschiedliche Stichprobenszusammensetzungen erklärt werden.⁶ Für unser Experiment wurden 4.700 Personen angeschrieben

5 <https://www.socisurvey.de/panel/researchers.php>. An dieser Stelle möchten wir uns herzlich bei Dominik Leiner, dem Betreiber des Socio-Scientific-Panels, für die Unterstützung unserer Forschung und die Bereitstellung von Panelmitgliedern bedanken.

6 Von den Befragten entfallen 34,1% auf die Kontrollgruppe, 32,7% auf die Experimentalgruppe Mann und 33,2% auf die Experimentalgruppe Frau. Bezüglich soziodemographischer Merkmale unterscheiden die Gruppen sich nicht. Das durchschnittliche Alter in der Stichprobe entspricht 34,8 (SD = 13,4). 51% der Befragten haben einen Hochschulabschluss; 29,5% haben die Hochschulreife (Abitur); 5% der Befragten haben Fachabitur; 5% haben eine abgeschlossene Lehre; 4% besitzen die mittlere Reife;

und zur Teilnahme an der Befragung eingeladen. 1.507 Personen öffneten den Link, 1.405 begannen den Fragebogen und 1.268 schlossen das Interview ab.

Mit Hilfe von Hauptkomponentenanalysen konnten insgesamt vier Indizes auf Basis der abgefragten Items generiert werden (Tabellen A3 und A5 im Appendix). Die Einstellungen zur Sexismus-Debatte wurden mit vier Items gemessen (Tabelle A2, Items SE1-SE4). Die Items konnten zu einem Index zusammengefasst werden, da alle Items auf einer einzigen Hauptkomponente laden (Eigenvalue 2,258). Bei dem Sexismus-Index handelt es sich im Folgenden um einen mittelwertzentrierten Summenindex, wobei die verwendeten Items mit den Faktorladungen aus der Hauptkomponentenanalyse gewichtet wurden (Mean=0; SD=1,50; Min=-3,23; Max=3,20; Cronbach's Alpha der vier Items=0,74). Er korreliert positiv mit der Aussage, dass die Sexismus-Debatte übertrieben sei und dass Frauen normales Verhalten oft als sexuelle Belästigung überinterpretieren. Der Index misst also die Einstellung, dass Sexismus kein ernsthaftes gesellschaftliches Problem sei (im Folgendem „Kein Sexismusproblem“-Index).

Die Items aus den ALLBUS-Wellen können zu insgesamt drei Indizes zusammengefasst werden (Tabelle A3 im Appendix). Die Items GR1.4 und GR3.3 mussten aus der Analyse ausgeschlossen werden, weil sie auf keiner Komponente eindeutig laden. Nach Ausschluss dieser Items ergibt sich eine klare Faktorstruktur, drei Hauptkomponenten haben einen Eigenvalue größer 1. Analog zum Sexismus-Index wurden auf Basis der Faktorladungen drei gewichtete und mittelwertzentrierte Summenindizes gebildet. Tabelle A4 zeigt die Korrelation der drei Indizes auf. Auf der ersten Hauptkomponente laden stark ideologisch ausgerichtete Items, die Aussagen gegen die Gleichstellung/-berechtigung von Mann und Frau enthalten. Am stärksten korreliert das Item „Für eine Frau ist es wichtiger, ihrem Mann bei seiner Karriere zu helfen, als selbst Karriere zu machen“ (im Folgendem „Gegen Gleichstellung“-Index; Mean=0; SD=1,85; Min=-1,93; Max=9,41; Cronbach's Alpha der sechs Items=0,84). Auf der zweiten Komponente laden Items, die sich auf die negativen Konsequenzen der Berufstätigkeit von Frauen beziehen. Die Items befassen sich insbesondere mit den Konsequenzen für das Familienleben und die Kindererziehung. Am stärksten lädt das Item „Ein Kleinkind wird sicherlich darunter leiden, wenn seine Mutter berufstätig ist“ (im Folgendem „Anti-Berufstätigkeit“-Index; Mean=0; SD=1,70; Min=-2,67; Max=5,43; Cronbach's Alpha der vier Items=0,82). Durch die dritte Komponente werden Items erfasst, die sich auf die positiven

Fortsetzung Fußnote 6

die restlichen Befragten verteilen sich auf andere Abschlüsse (noch Schüler, Hauptschulabschluss, kein Abschluss, Promotion). Junge und hochgebildete Personen sind in unserem Sample, wie in Onlinebefragungen üblich, also überrepräsentiert.

Konsequenzen der Berufstätigkeit von Frauen beziehen. Am stärksten lädt das Item „Einen Beruf zu haben ist das beste Mittel für eine Frau, um unabhängig zu sein“ (im Folgendem „Pro-Berufstätigkeit“-Index; Mean=0; SD=1,26; Min=-4,85; Max=2,45, Cronbach's Alpha der drei Items=0,51⁷).

4 Ergebnisse

Die Einstellungen zu Geschlechterrollen stehen seit einigen Jahrzehnten im Fokus verschiedener wissenschaftlicher Disziplinen (vgl. u.a. Höllinger 1991; Kasten 1996; BMFSFJ 2010). Theorie und Empirie sind folglich reichlich vorhanden, weshalb sich die nun folgende Darstellung der Ergebnisse primär auf die Auswertung des Experiments konzentrieren wird. Der erste Abschnitt beschreibt den Effekt der Personalisierung auf die Responseraten. Es folgen die Ergebnisse bezüglich der Effekte sozialer Erwünschtheit.

4.1 Responseraten

Ein erwarteter Vorteil der Personalisierung von Onlineumfragen ist eine höhere Motivation auf Seiten der befragten Personen. Hierdurch sollte die Datenqualität verbessert und die Responsequoten erhöht werden. Unit-Nonresponse wird im Folgenden als Befragungsabbruch auf der Begrüßungsseite oder bei der Einstiegsfrage definiert. Betrachtet man den Anteil an Unit-Nonresponse in Kontroll- und Experimentalgruppen, findet man jedoch den entgegengesetzten Effekt (siehe Tabelle 1). In der Kontrollgruppe brachen 2,71% der Teilnehmer die Befragung ab, in der Experimentalgruppe 4,75%. Der Unterschied zwischen beiden Gruppen ist allerdings nicht signifikant (Chi²-Test, Chi²(1)=3,37; p=0,066). Den Fragebogen beendeten in der Kontrollgruppe 93,53% der Teilnehmer, 88,55% in den Experimentalgruppen. Dieser Unterschied ist signifikant (Chi²-Test, Chi²(1)=8,88; p=0,003). Unsere Ergebnisse sprechen also gegen Hypothese *H_{1a}*.

Hinsichtlich der Teilnahmebereitschaft ebenfalls nicht bestätigt werden konnte die Annahme, dass attraktiven Interviewern mehr Unterstützung zuteilwird als weniger attraktiven Personen. Die Unit-Nonresponserate beträgt bei attraktiven Interviewern 8,04% (vgl. Tabelle 2). Das ist mehr als das Doppelte der Unit-

7 Obwohl der Reliabilitätskoeffizient hier nur knapp über dem minimal akzeptablen Niveau liegt, wollen wir den Index aus der Analyse nicht ausschließen. Dazu sei darauf hingewiesen, dass wir in diesem Artikel nicht primär an einer inhaltlichen Analyse der einzelnen Einstellungsdimensionen interessiert sind.

Tabelle 1 Unit- und Item-Nonresponse – Kontroll- vs. Experimentalgruppen

	Kontrollgruppe	Experimentalgruppen	Signifikanz
Unit-Nonresponse	2,71% (n=479)	4,75% (n=926)	Chi ² (1)=3,37 p=0,066
Anteil beendeter Fragebögen	93,53% (n=479)	88,55% (n=926)	Chi ² (1)=8,88 p=0,003
Item-Nonresponse	0,94% (n=479)	0,97% (n=926)	Chi ² (11)=13,5 p=0,257 ^a
Item-Nonresponse Einkommen	4,42% (n=452)	6,81% (n=852)	Chi ² (1)=2,98 p=0,084
Reduktion Item-Nonresponse Einkommen nach Antwortaufforderung	-51,22% (n=41)	-31,76% (n=85)	Chi ² (1)=4,44 p=0,035

Anmerkungen: Pearson Chi²-Tests auf Unabhängigkeit (Kontingenztafel). ^a Der Chi²-Test wurde mit den Variablen Zahl der Antwortverweigerungen und Experimentalgruppe durchgeführt.

Tabelle 2 Unit-Nonresponse und Abbruchquoten nach Attraktivität

	Attraktivität des Interviewers			Signifikanz
	niedrig	mittel	hoch	
Unit-Nonresponse	5,12% (n=215)	3,08% (n=487)	8,04% (n=224)	Chi ² (2)=8,41 p=0,015
Anteil beendeter Fragebögen	88,37% (n=215)	89,12% (n=487)	87,53% (n=224)	Chi ² (2)=0,40 p=0,817
Item-Nonresponse	0,65% (n=215)	1,10% (n=487)	0,99% (n=224)	Chi ² (18)=17,4 p=0,498 ^a
Item-Nonresponse Einkommen	6,06% (n=198)	6,19% (n=452)	8,91% (n=202)	Chi ² (2)=1,85 p=0,396
Reduktion Item-Nonresponse Einkommen nach Antwortaufforderung	-33,33% (n=18)	-33,33% (n=42)	-28,00% (n=25)	Chi ² (3)=0,24 p=0,891

Anmerkungen: Pearson Chi²-Tests auf Unabhängigkeit (Kontingenztafel). ^a Der Chi²-Test wurde mit den Variablen Zahl der Antwortverweigerungen und Experimentalgruppe durchgeführt

Nonresponserate bei durchschnittlich attraktiven Interviewern (3,08%). Interviewer mit geringer Attraktivität kommen auf Unit-Nonresponseraten von 5,12%. Der Zusammenhang zwischen Attraktivität des Interviewers und Unit-Nonresponse ist signifikant (Chi²-Test, Chi²(2)=8,41; p=0,015). Eine logistische Regressionsanalyse der Unit-Nonresponserate zeigt, dass diese bei hoch-attraktiven Interviewern signifikant höher ist als bei durchschnittlich attraktiven Interviewern (p=0,005). Zwischen durchschnittlich und weniger attraktiven Interviewern besteht dagegen kein signifikanter Unterschied (p=0,193). Der Anteil vollständig beendeter Inter-

views unterscheidet sich nicht signifikant nach der Attraktivität des Interviewers (Chi^2 -Test, $\text{Chi}^2(2)=0,40$; $p=0,817$). Unsere Ergebnisse legen also nahe, Hypothese H_{7b} zu verwerfen.

Auch für die Item-Nonresponsequoten wurden positive Effekte durch die Simulation einer Interviewsituation erwartet. Item-Nonresponse wird hier als Antwortverweigerungen (keine Angabe) definiert. Befragte, die den Fragebogen komplett abgebrochen haben, sind ab dem Moment des Abbruchs aus der Analyse ausgeschlossen. Der Anteil nicht beantworteter Items liegt in der Kontrollgruppe bei 0,94% und in den Experimentalgruppen bei 0,97% (vgl. Tabelle 1). Der Unterschied zwischen beiden Gruppen ist nicht signifikant (Chi^2 -Test, $\text{Chi}^2(11)=13,5$; $p=0,257$). Bezüglich der Attraktivität des Interviewers finden sich keine signifikanten Unterschiede im Antwortverweigerungsverhalten (vgl. Tabelle 2).

Insgesamt liegen sowohl Unit- als auch Item-Nonresponse im gesamten Sample auf einem sehr geringen Niveau. Im folgenden Abschnitt wird daher eine gesonderte Analyse der Item-Nonresponserate für die Einkommensfrage durchgeführt, da Fragen nach dem Einkommen generell durch hohe Verweigerungsraten gekennzeichnet sind und sich somit besonders für eine Untersuchung des Effektes der Antwortaufforderung eignen. In der Kontrollgruppe liegt die Item-Nonresponsequote bei 4,42% und in den Experimentalgruppen bei 6,81% (vgl. Tabelle 1). Die Simulation sozialer Präsenz in der Onlinebefragung führt somit, entgegen der Annahme, zu einer höheren Antwortverweigerung. Der Unterschied zwischen den beiden Gruppen ist jedoch nicht signifikant (Chi^2 -Test, $\text{Chi}^2(1)=2,98$; $p=0,084$). Die Aufforderung zur Eingabe nach einer Antwortverweigerung war in der Kontrollgruppe signifikant erfolgreicher als in den Experimentalgruppen (Chi^2 -Test, $\text{Chi}^2(1)=4,44$; $p=0,035$). In der Kontrollgruppe konnten durch die Aufforderung 51,22% der Verweigerer zu einer Antwort motiviert werden (vgl. Tabelle 1). In der Experimentalgruppe konnte der Anteil der Antwortverweigerungen nur um 31,76% reduziert werden.

Betrachtet man die Item-Nonresponseraten für die Einkommensfrage getrennt nach der Attraktivität des Interviewers, zeigt sich erneut, dass höhere Attraktivität keinen positiven Einfluss auf die Nonresponseraten hat (vgl. Tabelle 2). Tendenziell liegen die Verweigerungsraten bei attraktiveren Interviewern höher (8,91%) als bei durchschnittlich (6,19%) oder weniger attraktiven Interviewern (6,06%). Auch die Aufforderung zur Angabe des Einkommens ist bei attraktiveren Interviewern tendenziell weniger erfolgreich. Die Unterschiede in den Nonresponseraten sind allerdings nicht signifikant (vgl. Tabelle 2).

Insgesamt kann festgehalten werden, dass keine der in H_{1a} erwarteten positiven Effekte der Personalisierung nachgewiesen werden konnten. Zwischen

Experimental- und Kontrollgruppen zeigten sich kaum Unterschiede in Bezug auf Item- oder Unit-Nonresponse. Jene Indikatoren mit signifikanten Unterschieden sprechen jedoch gegen die Implementierung sozialer Präsenz. Auch Hypothese H_{1b} , die behauptet, dass attraktivere Interviewer/innen höhere Responseraten erzielen, konnte nicht bestätigt werden.

4.2 Soziale Erwünschtheit

Der folgende Abschnitt untersucht den erwarteten negativen Effekt in Form von sozial erwünschtem Antwortverhalten. Es wird getestet, ob sich die Angaben zur Einstellung gegenüber Geschlechterrollen signifikant danach unterscheiden, ob dem Befragten das Bild einer Person präsentiert wird. Zuerst wird eine Betrachtung unabhängig vom Geschlecht des Interviewers vorgenommen. Im darauf folgenden Abschnitt wird untersucht, inwiefern die Antworten sich zwischen männlichen und weiblichen Interviewern unterscheiden. Dabei wurden die vier oben beschriebenen Indizes mit Hilfe von OLS-Regressionsmodellen untersucht. Da die Verteilungsannahmen der OLS-Regressionen, insbesondere die Annahme heteroskedastischer Fehlerterme, nicht in allen Fällen vollständig zutreffen, berechnen wir alle Modelle mit robusten Standardfehlern.⁸

Vergleich Kontroll- und Experimentalgruppe

Um die Hypothesen H_{2a} und H_{2b} zu testen, wurde eine dichotome Variable zur Unterscheidung von Kontroll- und Experimentalgruppe verwendet (vgl. Tabelle 3). In einem ersten Schritt wurden Regressionsmodelle geschätzt, die nur das Geschlecht des Befragten kontrollieren. Dabei zeigt sich in allen Regressionsmodellen, dass Frauen im Vergleich zu Männern grundsätzlich egalitäre Einstellungen zur Rollenverteilung haben bzw. dass sie Sexismus kritischer betrachten (vgl. Modelle M1, M4, M7, M10 in Tabelle 3). In einem zweiten Schritt wird die Zugehörigkeit zur Experimentalgruppe in die Regressionsmodelle aufgenommen. Die Zugehörigkeit zu einer der Experimentalgruppen hat dabei keinen signifikanten Effekt auf das Antwortverhalten (vgl. Modelle M2, M5, M8, M11 in Tabelle 3). Dieses Ergebnis spricht gegen Hypothese H_{2a} , nach welcher unter der Bedingung sozialer Präsenz eine Editierung der Antwort in Richtung egalitärer Rollenvorstellungen stattfindet.

8 Alle Regressionsmodelle wurden auf Kollinearitätsprobleme überprüft. Der durchschnittliche Varianz-Inflations-Faktor (VIF) ist in keinem Modell höher als 5. Die VIFs der einzelnen Variablen sind in allen Fällen kleiner als 10 und liegen damit unterhalb des von Kutner et al. (2004) vorgeschlagenen Grenzwertes.

Tabelle 3 Regressionsanalysen Kontroll- vs. Experimentalgruppe

Variable	Index „Kein Sexismusproblem“			Index „Gegen Gleichstellung“			Index „Anti-Berufstätigkeit“			Index „Pro-Berufstätigkeit“		
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
<i>Befragtingeschlecht</i>												
Mann	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Frau	-0,750***	-0,751***	-1,046***	-0,733***	-0,733***	-1,076***	-0,730***	-0,730***	-0,881***	0,643***	0,643***	0,574***
<i>Gruppe</i>												
Kontrollgruppe	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Experimentalgruppe	0,076	-0,197	0,455**	0,087	-0,229	0,529*	-0,012	-0,151	0,233	0,011	-0,053	0,107
Experimentalgr.XFrau												
Konstante	0,482***	0,433***	0,609***	0,456***	0,400***	0,604***	0,454***	0,461***	0,551***	-0,392***	-0,399***	-0,358***
<i>Statistiken</i>												
n	1.248	1.248	1.248	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215
R ²	0,062	0,062	0,068	0,037	0,038	0,042	0,044	0,044	0,045	0,063	0,063	0,063
F	86,726	43,332	31,226	44,810	23,570	18,273	54,055	27,050	18,210	79,854	39,890	26,820

Anmerkungen: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ (robuste Standardfehler). Index „Kein Sexismusproblem“: vgl. Tabelle A2 und A5; Indizes „Gegen Gleichstellung“, „Anti-Berufstätigkeit“ und „Pro-Berufstätigkeit“: vgl. Tabellen A3 und A2.

Im nächsten Schritt wurden Modelle geschätzt, in denen die Zugehörigkeit zur Experimentalgruppe mit dem Geschlecht des Befragten interagiert (vgl. Modelle M3, M6, M9, M12 in Tabelle 3). Diese Modelle legen offen, dass Männer und Frauen ihre Antworten in unterschiedliche Richtungen editieren. Für alle vier Indizes gilt, dass Männer in der Experimentalgruppe tendenziell egalitäre Einstellungen vertreten als in der Kontrollgruppe. Dies entspricht den in Hypothese H_{2a} formulierten Erwartungen einer Anpassung an das sozial Erwünschte. Die Unterschiede zwischen Experimental- und Kontrollgruppe im Antwortverhalten von Männern sind allerdings in keinem Modell signifikant. Im Antwortverhalten von Frauen finden sich signifikante Unterschiede zwischen Experimental- und Kontrollgruppe für die zwei Indizes „Kein Sexismusproblem“ und „Gegen Gleichstellung“. In Bezug auf den ersten Index haben Frauen in der Experimentalgruppe durchschnittlich einen um 0,26 Skaleneinheiten höheren Wert als Frauen in der Kontrollgruppe (T-Test für lineare Kombinationen; $t=2,19$; $p=0,029$; robuste Standardfehler). In Bezug auf den Index „Gegen Gleichstellung“ antworten Frauen in der Experimentalgruppe durchschnittlich um 0,30 Skaleneinheiten höher als Frauen in der Kontrollgruppe (T-Test für lineare Kombinationen; $t=2,37$; $p=0,018$; robuste Standardfehler).

Im Vergleich zur Kontrollgruppe editieren Frauen in der Experimentalgruppe ihre Antworten in Richtung anti-egalitärer Rollenvorstellungen. Dieses Ergebnis ist sehr verwunderlich und erklärt gleichzeitig warum kein Effekt sozialer Erwünschtheit im vorherigen Modell zu finden ist: Männer und Frauen editieren ihre Antworten in entgegengesetzte Richtungen unter der Bedingung sozialer Präsenz. Erwartet wurde, dass das „sozial Erwünschte“ unabhängig vom Geschlecht des Befragten eine egalitäre Einstellung ist. Unklar ist, ob Frauen hier entgegen des sozial Erwünschten antworten oder ob Frauen es als sozial erwünscht empfinden, nicht auf eine egalitäre Rollenverteilung und eine Verurteilung von Sexismus zu beharren. Tiefgreifende Interpretation dieser Ergebnisse sind aber nicht Thema dieses Beitrags. Es soll nur festgehalten werden, dass Frauen ihre Antworten unter der Bedingung sozialer Präsenz signifikant verändern. Dies gilt für zwei von vier getesteten Indizes. Für Männer konnten keine signifikanten Unterschiede im Antwortverhalten festgestellt werden.

Hypothese H_{2a} konnte folglich nicht bestätigt werden. Trotzdem zeigen sich unter der Bedingung sozialer Präsenz signifikante Unterschiede im Antwortverhalten von Frauen. Dieses Ergebnis spricht für den in Hypothese H_{2b} ausgedrückten Zusammenhang: Frauen sind stärker anfällig für den Effekt sozialer Präsenz.⁹

9 Weiterhin wurde getestet, ob das Hinzufügen der Interviewerattraktivität zu einer Verbesserung des Modellfits führt. Dies konnte für keines der Modelle bestätigt werden.

Vergleich der Kontrollgruppe mit den Experimentalgruppen Mann und Frau

Dieser Abschnitt befasst sich mit dem in den Hypothesen H_3 bis H_4 beschriebenen Gender-of-Interviewer-Effekt. Wie bei der vorherigen Analyse werden OLS-Regressionsmodelle mit robusten Standardfehlern verwendet (vgl. Tabelle 4). Da die Modelle aufgrund der Dreifach-Interaktion zwischen Befragtingeschlecht, Interviewergeschlecht und Attraktivität des Interviewers teilweise sehr komplex sind, stehen im Appendix zusätzliche deskriptive Tabellen zur Verfügung (siehe Tabellen A6 - A9). Die Werte in den Zellen dieser Tabellen lassen sich mit den Parametern der geschätzten Regressionsmodelle replizieren (Modelle M15, M18, M21, M24 in Tabelle 4). Alle berichteten Hypothesentests sind auf Basis der Regressionsmodelle durchgeführt worden.

Hypothese H_{3a} geht von der Annahme aus, dass männliche und weibliche Befragte gegenüber einer weiblichen Interviewerin egalitäre Rollenvorstellungen äußern als gegenüber einem männlichen Interviewer. Die entsprechenden Regressionsanalysen zeigen jedoch, dass es zwischen den beiden Experimentalgruppen keine signifikanten Unterschiede gibt (vgl. Tabelle 4, Modelle M13, M16, M19, M22). Die Antworten der Befragten aus den beiden Experimentalgruppen unterscheiden sich ebenfalls nicht signifikant von den Antworten in der Kontrollgruppe. In den vorherigen Analysen deutete sich jedoch bereits an, dass Männer und Frauen unter der Bedingung sozialer Präsenz ihre Antworten in unterschiedliche Richtungen editieren. Folglich war zu erwarten, dass keine Unterschiede zwischen der Kontroll- und den beiden Experimentalgruppen auftreten, solange die Effekte geschlechtsunabhängig modelliert werden.

In einem nächsten Schritt wurden die Regressionsmodelle daher um die Interaktion des Interviewergeschlechts mit dem Geschlecht des Befragten erweitert (Tabelle 4, Modelle M14, M17, M20, M23). In Bezug auf Hypothese H_{3a} zeigt sich, dass es zwischen den beiden Experimentalgruppen keine signifikanten Unterschiede gibt, weder für männliche noch für weibliche Befragte (F-Test für lineare Restriktionen, für alle vier Indizes gilt $p > 0,05$; robuste Standardfehler). Hypothese H_{3a} kann also nicht bestätigt werden.

Betrachtet man die Punktschätzer unabhängig von ihrer Signifikanz, ergibt sich ein uneinheitliches Bild. So haben männliche Befragte für den Index „Gegen Gleichstellung“ tendenziell geringere Werte in der Experimentalgruppe Frau (0,347) als in der Experimentalgruppe Mann (0,403). Hier kann also eine erwartungskonforme Tendenz zu egalitäreren Einstellungen in der Experimentalgruppe Frau beobachtet werden. In Bezug auf den Index „Pro-Berufstätigkeit“ tritt eine gegenläufige Tendenz auf. Männer stimmen den Items dieses Indizes in der Experimentalgruppe Frau (-0,512) weniger zu als in der Experimentalgruppe Mann (-0,306). Dies

Tabelle 4 Regressionsanalysen Kontrollgruppe vs. Experimentalgruppe Mann vs. Experimentalgruppe Frau

Variable	Index „Kein Sexismusproblem“				Index „Gegen Gleichstellung“				Index „Anti-Berufstätigkeit“				Index „Pro-Berufstätigkeit“			
	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24				
<i>Befragtingeschlecht</i>																
Mann	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.			
Frau	-0,751***	-1,046***	-1,046***	-0,733***	-1,076***	-1,076***	-0,729***	-0,881***	-0,881***	0,643***	0,574***	0,574***	0,574***			
<i>Gruppe</i>																
Kontrollgruppe	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.			
Experimentalgruppe Mann	0,033	-0,202	-0,045	0,089	-0,201	-0,059	-0,081	-0,242	-0,047	0,027	0,052	0,052	-0,095			
Experimentalgruppe Frau	0,118	-0,191	-0,293	0,086	-0,257	-0,354	0,055	-0,064	-0,177	-0,005	-0,154	-0,154	0,012			
FrauXExperimentalgruppe Mann	0,393*	0,215			0,484	0,530		0,268	0,203		-0,039	-0,039	0,074			
FrauXExperimentalgruppe Frau	0,515**	0,635**			0,573*	0,702*		0,199	0,305		0,249	0,249	0,037			
<i>Interviewerattraktivität</i>																
hoch			-0,166			-0,123			-0,359				0,183			
durchschnittlich			Ref.			Ref.			Ref.				Ref.			
niedrig			-0,440			-0,442			-0,448				0,408			
hochXInterviewer= Frau			0,393			0,668			0,746				-0,554			
niedrigXInterviewer= Frau			0,610			0,272			0,499				-0,679*			

Variable	Index „Kein Sexismusproblem“				Index „Gegen Gleichstellung“				Index „Anti-Berufstätigkeit“				Index „Pro-Berufstätigkeit“			
	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28
hochXInterviewer= FrauXFrau			-0,249			-0,595										0,550
niedrigXInterviewer= FrauXFrau			-0,234			0,076										0,281
hochXInterviewer= MannXFrau			0,557			-0,028										-0,251
niedrigXInterviewer= MannXFrau			0,157			-0,205										-0,202
Konstante	0,433***	0,609***	0,609***	0,400***	0,604***	0,604***	0,461***	0,551***	0,551***	-0,399***	-0,358***	-0,358***				
<i>Statistiken</i>																
n	1.248	1.248	1.248	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215	1.215
R ²	0,063	0,068	0,076	0,038	0,042	0,051	0,045	0,046	0,051	0,063	0,066	0,072				
F	28,945	18,821	7,933	15,733	10,980	5,220	18,445	11,202	5,029	26,639	16,821	7,519				

Anmerkungen: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ (robuste Standardfehler). Index „Kein Sexismusproblem“: vgl. Tabelle A2 und A5; Indizes „Gegen Gleichstellung“, „Anti-Berufstätigkeit“ und „Pro-Berufstätigkeit“: vgl. Tabellen A3 und A2.

entspricht einer Tendenz zu sozial unerwünschtem Antwortverhalten. Bei weiblichen Befragten können ebenfalls solche Tendenzen identifiziert werden: So haben weibliche Befragte für den Index „Kein Sexismusproblem“ durchschnittlich höhere Werte in der Experimentalgruppe Frau (-0,114) als in der Experimentalgruppe Mann (-0,246). Frauen verurteilen Sexismus also weniger stark gegenüber einer Interviewerin und verhalten sich damit entgegen den Erwartungen (H_{3a} , H_{3b}).

Vergleicht man das Antwortverhalten von männlichen und weiblichen Befragten in den beiden Experimentalgruppen mit dem Antwortverhalten in der Kontrollgruppe, ergibt sich ein anderes Bild. Für männliche Befragte lassen sich keine signifikanten Unterschiede zwischen der Kontrollgruppe und den beiden Experimentalgruppen feststellen (F-Test für lineare Restriktionen, für alle vier Indizes gilt $p > 0,05$; robuste Standardfehler). Tendenziell lässt sich jedoch erkennen, dass männliche Befragte in beiden Experimentalgruppen eher egalitäre Antworten geben als in der Kontrollgruppe. Für weibliche Befragte ergibt sich in zwei Fällen ein signifikanter Unterschied zwischen Kontrollgruppe und Experimentalgruppe. Frauen haben für den Index „Kein Sexismusproblem“ signifikant höhere Werte in der Experimentalgruppe Frau (-0,114) als in der Kontrollgruppe (-0,437; Differenz[Kontrollgruppe – Gruppe Frau]=-0,323; F-Test für lineare Restriktionen: $F(1,1242)=6,01$; $p=0,01$; robuste Standardfehler). Ein identisches Muster beobachten wir in Bezug auf den Index „Gegen Gleichstellung“. So haben Frauen für diesen Index in der Kontrollgruppe durchschnittlich einen Wert von -0,473. In der Experimentalgruppe Frau haben sie einen durchschnittlichen Wert von -0,157 (Differenz[Kontrollgruppe – Gruppe Frau]=-0,316; F-Test für lineare Restriktionen: $F(1,1209)=4,82$; $p=0,028$; robuste Standardfehler).

Es lässt sich resümieren, dass Männer tendenziell egalitäre Meinungen vertreten, wenn sie einer der beiden Experimentalgruppen angehören. Wider Erwarten lässt sich aber nicht erkennen, dass dieser Effekt in der Experimentalgruppe Frau größer wäre als in der Experimentalgruppe Mann. Für die Gruppe der weiblichen Befragten zeigt sich, dass sie dazu tendieren, ihre Antworten in Richtung anti-egalitärer Rollenvorstellung zu editieren (im Vergleich zur Kontrollgruppe). Vor allem gegenüber einer weiblichen Interviewerin scheinen Frauen geneigt, anti-egalitäre Einstellungen zu vertreten. Im Vergleich zur Gruppe der männlichen Befragten sind die Unterschiede zwischen Experimental- und Kontrollgruppe bei Frauen deutlich größer. Ähnlich wie in der oben berichteten Analyse legen die Ergebnisse also nahe, dass Frauen eher geneigt sind, ihre Antworten unter der Bedingung sozialer Präsenz zu editieren. Dieses Ergebnis spricht für Hypothese H_{3b} . Unbeantwortet bleibt weiterhin, wie die beobachteten Tendenzen im Antwortverhalten erklärt werden

können. Insbesondere die Tendenz von Frauen, gegenüber anderen Frauen anti-egalitäre Einstellungen zu vertreten, widerspricht den Annahmen.

Effekte der Interviewerattraktivität

In einem letzten Schritt wurde geprüft, ob die Effekte sozialer Präsenz von der Attraktivität des Interviewers moderiert werden (vgl. Tabelle 4, Modelle M15, M18, M21, M24). Dementsprechend beinhalten die Regressionsmodelle einen Interaktionsterm mit drei Faktoren. Grundsätzlich kann festgehalten werden, dass das Hinzufügen der Attraktivitätsvariablen und der entsprechenden Interaktionen in keinem Fall zu einer signifikanten Modellverbesserung führt (LR-Test; $\chi^2(8)$; in allen Fällen gilt $p > 0,05$).

Insgesamt lässt sich nur in einem Fall ein eindeutiges Muster im Antwortverhalten erkennen.¹⁰ In Bezug auf den Index „Kein Sexismusproblem“ nimmt der Mittelwert bei den weiblichen Befragten mit der Attraktivität des Interviewers kontinuierlich zu. Der Unterschied im Antwortverhalten bei einem unattraktiven Mann (-0,550) und einem attraktiven Mann (0,124) ist dabei signifikant (F-Test für lineare Restriktionen: $F(1, 1234) = 6,22$; $p = 0,01$; robuste Standardfehler). Im Vergleich zu einem unattraktiven Interviewer bezeichnen Frauen Sexismus also seltener als ein Problem, wenn der Interviewer attraktiv ist. Dies entspricht den Erwartungen, dass mit zunehmender Attraktivität stärker sozial erwünscht geantwortet wird (H_4). Zusammenfassend kann festgehalten werden, dass Attraktivitätseffekte nur sehr bedingt auftauchen. Signifikante Unterschiede im Antwortverhalten konnten nur in der Konstellation weibliche Befragte und männlicher Interviewer festgestellt werden. Frauen sehen Sexismus umso weniger kritisch, je attraktiver der männliche Interviewer ist.

5 Zusammenfassung und Diskussion

In dieser Studie wurde der Effekt sozialer Präsenz in Onlineumfragen auf verschiedene Nonresponseraten untersucht. Bezüglich der Teilnahmebereitschaft konnten signifikante Unterschiede in der Anzahl der vollständig bearbeiteten Interviews sowie der Effizienz von Antwortaufforderungen ermittelt werden. Die Ergebnisse widersprechen dabei den Erwartungen eines positiven Effekts sozialer Präsenz. Wie auch Tourangeau et al. (2003), die bezüglich der Teilnahmebereitschaft und

10 Für die folgenden Betrachtungen liefern die Tabellen A6–A9 im Appendix eine einfache Übersicht, da die referierten Zahlen aus den Regressionstabellen nur schwer ersichtlich sind.

Abbruchquoten keinerlei signifikante Unterschiede aufzeigen konnten, ziehen wir den Schluss, dass die Implementierung sozialer Präsenz nicht den erwarteten positiven Effekt auf die Teilnahmebereitschaft hat. Hinsichtlich der Responsequoten sprechen die Befunde somit gegen die Implementierung von sozialer Präsenz in Onlinebefragungen.

Neben einem Einfluss auf die Teilnahmebereitschaft wurde geprüft, ob die Simulation sozialer Präsenz von negativen Effekten flankiert wird. Es wurde vermutet, dass die Simulation sozialer Präsenz zu sozial erwünschtem Antwortverhalten führt. In einigen Fällen konnten tatsächlich signifikante Abweichungen zwischen den zwei Experimentalgruppen und der Kontrollgruppe identifiziert werden. Frauen haben sich dabei im Ganzen stärker anfällig für den Stimulus gezeigt als Männer. Insgesamt entsprechen die beobachteten Antwortmuster aber nur teilweise den Erwartungen. Das Antwortverhalten von weiblichen Befragten gegenüber einer weiblichen Interviewerin ist hierbei besonders unerwartet: Im Vergleich zur Kontrollgruppe haben weibliche Befragte teilweise signifikant weniger egalitäre Auffassungen vertreten, wenn sie von einer Frau interviewt wurden. Unsere Ergebnisse widersprechen damit den Analysen von Couper et al. (2001) und Tourangeau et al. (2003), die von erwartungskonformen Gender-of-Interviewer-Effekten berichten.

Mögliche Erklärungen könnten sein, dass ein Effekt sozial unerwünschten Antwortverhaltens vorliegt oder, dass die Definition von Egalitarismus als sozial erwünscht unzutreffend ist. So könnte es beispielsweise sein, dass die befragten Frauen befürchten, eine extrem kritische Einstellung gegenüber sexistischem Verhalten könne als Überempfindlichkeit gedeutet werden. Eine Tendenz zu sozial unerwünschtem Antwortverhalten ließe sich möglicherweise auch durch eine subjektive Konkurrenzsituation zwischen der Befragten und der Interviewerin erklären (Gutierrez/Kenrick/Partch 1999; Bruunk/Massar/Dijkstra 2007; Maner et al. 2009). Jedoch präsentierte sich auch bei männlichen Befragten eine Tendenz zu sozial unerwünschtem Antwortverhalten.

Obwohl wir einen weit weniger subtilen experimentellen Stimulus verwendeten als einige Studien der Social-Interface-Theorie, ist die Verwendung von Bildern als experimenteller Stimulus möglicherweise auch nicht hinreichend stark genug, um durchgängig die erwarteten Effekte zu beobachten. Couper et al. (2001) sowie Tourangeau et al. (2003) führen als mögliche Ursache für das Ausbleiben der Effekte sozialer Präsenz in Onlineumfragen eben jene zu geringe Intensität des Stimulus an. Für diese Interpretation sprechen auch die relativ starken Gender-of-Interviewer-Effekte, die Fuchs (2009) durch den Einsatz von Videosequenzen evozieren konnte. Allerdings sind auch bei Fuchs (2009) die Ergebnisse nicht vollständig konsistent mit den formulierten Erwartungen.

Die vorliegende Studie lässt somit einige Fragen unbeantwortet und bietet Raum für weitere Forschungen. Für unsere Forschungsfrage ist die Interpretation der teilweise kontraintuitiven Ergebnisse jedoch nicht von Bedeutung. Ausschlaggebend ist die Beobachtung von signifikant abweichendem Antwortverhalten unter der Bedingung sozialer Präsenz, um eine eindeutige Schlussfolgerung zu ziehen: Die Implementierung von sozialer Präsenz mit Hilfe von Fotografien kann zu Verzerrungen im Antwortverhalten führen. Folglich sprechen die Ergebnisse auch im Hinblick auf die vermuteten negativen Effekte gegen eine Implementierung von sozialer Präsenz in Onlineumfragen.

Literatur

- ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., 2010: Online Dokument. <http://www.adm-ev.de/index.php?id=startseite> (01.03.2012).
- Athenstaedt, U., E. Haas und S. Schwab, 2004: Gender Role Self-Concept and Gender-Typed Communication Behavior in Mixed-Sex and Same-Sex Dyads. *Sex-Roles* 50(1/2): 37-52.
- Argentino, C., M. Aline, C. Kidd und K. Bogart, 1977: The Effect of Experimenter's Sex and Subject's Sex on the Attitudes toward Women on Fraternity, Sorority, and Mixed-dormitory Residents. *Journal of Community Psychology* 5: 186-188.
- Aries, E., 1976: Interaction Patterns and Themes of Male, Female, and Mixed Groups. *Small Group Behavior* 7: 7-18.
- Athey, K. R., J. E. Coleman, A. P. Retiman und J. Tang, 1960: Two Experiments Showing the Effects of the Interviewer's Racial Background on Responses to Questionnaires Concerning Racial Issues. *Journal of Applied Psychology* 44: 562-566.
- Ballou, J. und F. K. Del Boca, 1980: Gender Interaction Effects on Survey Measures in Telephone Interviews. Paper presented at the annual meeting of the American Association of Public Opinion Research, Mason, Ohio.
- Bassili, J. N., 1981: The attractiveness stereotype: Goodness or glamour? *Basic and Applied Social Psychology* 2: 235-252.
- Berger, J., S. J. Rosenholtz und M. Jr. Zelditch, 1980: Status organizing processes. *Annual Review of Sociology* 6: 479-508.
- Bian, F., 1997: The effects of attractiveness on helping behavior. Claremont: Harvey Mudd College.
- BMFSFJ, 2010: Die Rolle annehmen? In der Rolle bleiben? Neue Rollen leben? Einstellungen und Vorstellungen von Frauen und Männern mit und ohne Zuwanderungsgeschichte zur Gleichberechtigung. Bundesministerium für Familie, Senioren, Frauen und Jugend. Drensteinfurt: Claßen Druck.
- Buunk, A.P., K. Massar und P. Dijkstra, 2007: A social cognitive evolutionary approach to jealousy: The automatic evaluation of one's romantic rivals. S.213-228 in: J. P. Forgas, M. G. Haselton und W. von Hippel (Hg.): *Evolution and the social mind*, New York: Psychology.
- Carli, L. L. und D. Bukatko, 2000: Gender, communication, and social influence: A developmental perspective. S. 295-332 in: T. Eckes und H. M. Trautner (Hg.): *The developmental social psychology of gender*, Mahwah, NJ: Erlbaum.

- Couper, M. P., 2005: Technology Trends in Survey Data Collection. *Social Science Computer Review* 23: 486-501.
- Couper, M.P., R. Tourangeau und D. M. Steiger, 2001: Social Presence in Web Surveys. *Computer Human Interaction* 3(1): 412-417.
- Dabbs, J. M. und N. A. Stokes, 1975: Beauty is power: The use of space on the sidewalk. *Sociometry* 38: 551-557.
- Deaux, K. und M. LaFrance, 1998: Gender. S. 788-827 in: D. T. Gilbert, S. T. Fiske und G. Lindzey (Hg.): *The handbook of social psychology*, Vol. 1, Boston: McGraw-Hill.
- DeShields, O.W., A. Kara und E. Kaynak, 1996: Source effects in purchase decisions: The impact of physical attractiveness and accent of salesperson. *International Journal of Research in Marketing* 13: 89-101.
- Dindia, K. und M. Allen, 1992: Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin* 112: 106-124.
- Diekmann, A., 2004: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. 11. Auflage, Reinbek: Rowohlt.
- Dion, K.K., E. Berscheid und E. Walster, 1972: What is beautiful is good. *Journal of Personality and Society Psychology* 24: 285-290.
- Evans, J.R. und A. Mathur, 2005: The value of online surveys. *Internet research* 15(2): 195-219.
- Fitzpatrick, M.A., A. Mulac und K. Dindia, 1995: Gender preferential language use in spouse and stranger interaction. *Journal of Language and Social Psychology* 14: 18-39.
- Fogg, B.J. und C. Nass, 1997: Silicon sycophants: the effect of computers that flatter. *International Journal of Human-Computer Studies* 46: 551-561.
- Fuchs, M., 2009: Gender-of-Interviewer-Effects in a Video-Enhanced Web Survey. Results from a Randomized Field Experiment. *Social Psychology* 40(1): 37-42.
- Geis, F., 1993: Self-fulfilling prophecies: A social psychological view of gender. S. 9-54 in: A. E. Beall und R. J. Sternberg (Hg.): *The psychology of gender*, New York: Guilford.
- Gerich, J., 2009: Multimediale Elemente in der Computerbasierten Datenerhebung. Der Einfluss Auditiver und Visueller Elemente auf das Antwortverhalten in Befragungen. S. 107-130 in: M. Weichbold, J. Bacher und C. Wolf (Hg.): *Umfrageforschung. Herausforderungen und Grenzen*. Wiesbaden: Verlag für Sozialwissenschaften (9. Sonderheft der Österreichischen Zeitschrift für Soziologie).
- Grammer, K., 2002: *Signale der Liebe. Die biologischen Gesetze der Partnerschaft*. Frankfurt a. M.: dtv.
- Grimes, M. D. und G. L. Hansen, 1984: Response Bias in Sex-Role Attitude Measurement. *Sex Roles* 10: 67-72.
- Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer und R. Tourangeau, 2004: *Survey Methodology*. Hoboken: Wiley.
- Guerrero, L. K., 1997: Nonverbal interactions with same-sex friends, opposite-sex friends, and romantic partners: Consistency or change? *Journal of Social and Personal Relationships* 14: 31-58.
- Gutierrez, S. E., D. T. Kenrick und J. J. Partch, 1999: Beauty, dominance, and the mating game: Contrast effects in self-assessment reflect gender differences in mate selection. *Personality and Social Psychology Bulletin* 25: 1126-1134.
- Hall, J. A. und N. J. Briton, 1993: Gender, nonverbal behavior, and expectations. S. 276-295 in: P. M. Blanck (Hg.): *Interpersonal expectation theory: Research and applications*, New York: Cambridge University Press.
- Haunberger, S., 2006: Das standardisierte Interview als soziale Interaktion: Interviewereffekte in der Umfrageforschung. *ZA-Information* 58: 23-58.

- Henss, R., 1992: „Spieglein, Spieglein an der Wand...“ Geschlecht, Alter und physische Attraktivität. Weinheim: Psychologie Verlags Union.
- Heerwegh, D., 2009: Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research* 21(1): 111-121.
- Holbrook, A. L., M. C. Green und J. A. Krosnick, 2003: Telefon versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly* 67: 79-125.
- Höllinger, F., 1991: Frauenerwerbstätigkeit und Wandel der Geschlechterrollen im internationalen Vergleich. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 43: 753-771.
- Huddy, L., J. Billig, J. Bracciodieta, L. Hoeffler, P.J. Moynihan und P. Pugliani, 1997: The Effect of Interview Gender on the Survey Response. *Political Behaviour* 19: 197-220.
- Johnson, C., 1994: Gender, legitimate authority, and conversation. *American Sociological Review* 59: 122-35.
- Johnson, C., J. Clay-Warner und S. J. Funk, 1996: Effects of authority structures and gender on interaction in same-sex task groups. *Social Psychology Quarterly* 59: 221-236.
- Kane, E. W. und L. J. Macaulay, 1993: Interviewer Gender and Gender Attitudes. *Public Opinion Quarterly* 57: 1-28.
- Kasten, H., 1996: Weiblich - männlich. Geschlechterrollen und ihre Entwicklung. Berlin/Heidelberg: Springer-Verlag.
- Klein, M. und M. Kühhirt, 2010: Sozial erwünschtes Antwortverhalten bezüglich der Teilung häuslicher Arbeit. Die Interaktion von Interviewergeschlecht und Befragtenmerkmalen in Telefoninterviews. *Methoden Daten Analysen* 4(2): 79-104.
- König, R., 1972: Das Interview. 7. Auflage, Köln: Kiepenheuer.
- Kollock, P., P. Blumstein und P. Schwartz, 1985: Sex and Power in Interaction. *Conversation Privileges and Duties*. *American Sociological Review* 50: 34-46.
- Krosnick, J. und D. Alwin, 1987: An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly* 51: 201-219.
- Kutner, M. H., C. J. Nachtsheim und J. Neter, 2004: *Applied Linear Regression Models*. 4. Auflage, New York: McGraw-Hill.
- Lakoff, R., 1973: Language and woman's place. *Language in Society* 2: 45-79.
- Landis, J. R., D. Sullivan und J. Sheley, 1973: Feminist Attitudes as Related to Sex of the Interviewer. *The Pacific Sociological Review* 16(3): 305-314.
- Leaper, C. und R. D. Robnett, 2011: Women Are More Likely Than Men to Use Tentative Language, Aren't They? A Meta-Analysis Testing for Gender Differences and Moderators. *Psychology of Women Quarterly* 35(1): 129-42.
- Lueptow, L.B., S.L. Moser und B.F. Pendleton, 1990: Gender and Response Effects in Telephone Interviews about Gender Characteristics. *Sex Roles* 22: 29-42.
- Maner, J. K., D. T. Kenrick, D. V. Becker, A. W. Delton, B. Hofer, C. J. Wilbur und S.L. Neuberg, 2003: Sexually Selective Cognition: Beauty Captures the Mind of the Beholder. *Journal of Personality and Social Psychology* 85(6): 1107-1120.
- Maner, J. K., S. L. Miller, D. A. Rouby und M. T. Gailliot, 2009: Intrasexual vigilance: The Implicit Cognition of Romantic Rivalry. *Journal of Personality and Social Psychology* 97(1): 74-87.
- Mulford, M., J. Orbell, C. Shatto und J. Stockard, 1998: Physical attractiveness, opportunity and success in everyday exchange. *American Journal of Sociology* 103: 1565-1593.
- Nass, C., B. J. Fogg und Y. Moon, 1996: Can computers be teammates? *International Journal of Human-Computer Studies* 45: 669-678.

- Nass, C. und Y. Moon, 2000: Machines and mindlessness: social responses to computers. *Journal of Social Issues* 56(1): 81-103.
- Nass, C., Y. Moon und P. Carney, 1999: Are people polite to computers? Responses to computers interviewing systems. *Journal of Applied Social Psychology* 29: 1093-1110.
- Nass, C., Y. Moon und N. Green, 1997: Are machines gender neutral? Gender-stereotypic responses to computers. *Journal of Applied Social Psychology* 27: 864-876.
- Porst, R., S. Ranft und B. Ruoff, 1998: Strategien und Maßnahmen zur Erhöhung der Ausschöpfungsquoten bei sozialwissenschaftlichen Umfragen: ein Literaturbericht. ZUMA-Arbeitsbericht 1998/07.
- Reinecke, J., 1993: Explanations of Social Desirability and Interviewer Effects. S. 315-337 in: D. Krebs und P. Schmidt (Hg.): *New Directions in Attitude Measurements*, Berlin: Walter de Gruyter.
- Reinigen, P. H. und J. B. Kernan, 1993: Social perception and interpersonal influence: Some consequences of the physical attractiveness stereotyp in personal selling setting. *Journal of Consumer Psychology* 2(1): 25-28.
- Schuman, H. und J. Converse, 1971: The Effects of Black and White Interviewers on Black Responses in 1968. *Public Opinion Quarterly* 35: 44-68.
- Schnell, Rainer, 2012: *Survey-Interviews. Methoden standardisierter Befragungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Spencer, W. J. und K. A. Drass, 1989: The Transformation of Gender into Conversational Advantage. A Symbolic Interactionist Approach. *Sociological Quarterly* 20: 363-383.
- Spender, D., 1985: *Man Made Language*. Boston: Routledge & Kegan Paul.
- Sproull, L., M. Subrami, S. Kiesler, J. H. Walker und K. Waters, 1996: When the interface is a face. *Human-Computer Interaction* 11: 97-124.
- Steinert, H., 1984: Das Interview als soziale Interaktion. S. 17-59 in: H. Meulemann und K.-H. Reuband (Hg.): *Soziale Realität im Interview: Empirische Analyse methodischer Probleme*, Frankfurt: Campus.
- Sundar, S. S. und C. Nass, 2000: Source orientation in human-computer interaction: programmer, networker, or independent social actor? *Communication Research* 27(6): 683-703.
- Tooke, W. und W. Camire, 1991: Patterns of deception in intersexual and intrasexual mating strategies. *Ethology and Sociobiology* 12: 345-364.
- Tourangeau, R., L. J. Rips und K. Rasinski, 2000: *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R., M. P. Couper und D. M. Steiger, 2003: Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior* 19: 1-24.
- Tu, S.-H. und P.-S. Liao, 2007: Social Distance, Respondent Cooperation and Item Nonresponse in Sex Survey. *Quality & Quantity* 41: 177-199.
- Twenge, J. M., 2001: Changes in Women's Assertiveness in Response to Status and Roles: A Cross-Temporal Meta-Analysis, 1931-1993. *Journal of Personality and Social Psychology* 81: 133-145.
- Weichbold, M., 2005: *Touchscreen-Befragungen. Neue Wege in der empirischen Sozialforschung*. Frankfurt am Main: Lang.
- Whelchel, N. L., 1987: Sex-based Respondent-Interviewer Interaction Effects on Gender-sensitive Questions. Paper presented at the 1987 meetings of the American Association for Public Opinion Research, Hershey, PA.
- Wilson, R. K. und C. C. Eckel, 2006: Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly* 59: 189-202.

- Wood, J. T. und K. Dindia, 1998: What's the difference? A dialogue about differences and similarities between women and men. S.19-39 in: D. J. Canary und K. Dindia (Hg.): Sex differences and similarities in communication, Mahwah, N.J: Erlbaum.
- Zimmerman, D. H. und C. West, 1975: Sex roles, Interruptions and Silence in Conversation. S. 105-129 in: B. Thorne und N. Haley (Hg.): Language and Sex. Difference and Dominance, Rowley, MA: Newbury.

Anschriften der Autoren

Alexander W. Schmidt-Catran
Universität zu Köln
Institut für Soziologie und Sozialpsychologie
Lehrstuhl für Empirische Sozial- und
Wirtschaftsforschung
Albertus-Magnus-Platz
50923 Köln
E-Mail: alex@alexanderwschmidt.de

Katharina Hörstermann
Universität zu Köln
Institut für Soziologie und Sozialpsychologie
Lehrstuhl für Empirische Sozial- und
Wirtschaftsforschung
Albertus-Magnus-Platz
50923 Köln
E-Mail: hoerstermann@wiso.uni-koeln.de

Anhang

Tabelle A1 Attraktivitätsranking der verwendeten Fotos

Foto (Attraktivität)	Gruppe Mann	Gruppe Frau
1 (mittel)	3,63	3,67
2 (mittel)	3,78	3,74
3 (niedrig)	3,05	3,07
4 (hoch)	4,80	4,85
Total	3,82	3,83

Anmerkungen: Mittelwert aller uns zur Verfügung gestellten 35 Fotos: 3,77 (Std.Abw.=0,77).
Bewertet wurden die Fotos von jeweils 13 männlichen und 13 weiblichen Personen.

Tabelle A2 Übersicht Item-ID und genauer Wortlaut der abgefragten Items

Item ID	Item Wortlaut	Quelle	Mean	Median	SD	n
GR1.1	Eine berufstätige Mutter kann ein genauso herzliches und vertrauensvolles Verhältnis zu ihren Kindern finden, wie eine Mutter, die nicht berufstätig ist.	ALLBUS 2008	4,46	5	0,94	1.309
GR1.2	Ein Kleinkind wird sicherlich darunter leiden, wenn seine Mutter berufstätig ist.	ALLBUS 2008	2,32	2	1,25	1.306
GR1.3	Alles in allem: Das Familienleben leidet darunter, wenn die Frau voll berufstätig ist.	ALLBUS 1994	2,48	2	1,31	1.304
GR1.4	Das Familienleben leidet oft, weil Männer sich zu sehr auf ihre Arbeit konzentrieren.	ALLBUS 1994	3,62	4	0,97	1.300
GR1.5	Es ist für ein Kind sogar gut, wenn seine Mutter berufstätig ist und sich nicht nur auf den Haushalt konzentriert.	ALLBUS 2008	3,27	3	1,18	1.301
GR2.1	Einen Beruf zu haben ist ja ganz schön, aber was die meisten Frauen wirklich wollen, ist ein Heim und Kinder zu haben.	ALLBUS 1994	1,99	2	1,04	1.294
GR2.2	Hausfrau zu sein ist genauso erfüllend wie gegen Bezahlung zu arbeiten.	ALLBUS 1994	2,41	2	1,16	1.280
GR2.3	Es ist nicht gut, wenn der Mann zu Hause bleibt und sich um die Kinder kümmert und die Frau außer Haus berufstätig ist.	ALLBUS 1994	1,60	1	0,93	1.306
GR2.4	Die Aufgabe des Mannes ist es, Geld zu verdienen, die der Frau, sich um Haushalt und Familie zu kümmern.	ALLBUS 1994	1,49	1	0,85	1.306
GR2.5	Für eine Frau ist es wichtiger, ihrem Mann bei seiner Karriere zu helfen, als selbst Karriere zu machen.	ALLBUS 2008	1,50	1	0,82	1.303

Item ID	Item Wortlaut	Quelle	Mean	Median	SD	n
GR2.6	Eine verheiratete Frau sollte auf eine Berufstätigkeit verzichten, wenn es nur eine begrenzte Anzahl von Arbeitsplätzen gibt, und wenn ihr Mann in der Lage ist, für den Unterhalt der Familie zu sorgen.	ALLBUS 2008	1,53	1	0,92	1.303
GR2.7	Es ist für alle Beteiligten viel besser, wenn der Mann voll im Berufsleben steht und die Frau zuhause bleibt und sich um den Haushalt und die Kinder kümmert.	ALLBUS 2008	1,52	1	0,86	1.304
GR3.1	Einen Beruf zu haben ist das beste Mittel für eine Frau, um unabhängig zu sein.	ALLBUS 1994	4,35	5	0,84	1.301
GR3.2	Der Mann und die Frau sollten beide zum Haushaltseinkommen beitragen.	ALLBUS 1994	4,10	4	0,89	1.301
GR3.3	Die meisten Frauen müssen heutzutage arbeiten, um ihre Familie zu unterstützen.	ALLBUS 1994	3,99	4	0,92	1.291
SE1	Die Diskussion um Sexismus in Deutschland ist längst überfällig und muss in aller Ernsthaftigkeit geführt werden.	YouGov 2013	3,30	3	1,32	1.310
SE2	Ich finde es wichtig, Grenzüberschreitungen zu thematisieren. Man muss aber nicht aus jedem misslungenen Herrenwitz gleich eine Sexismus-Debatte machen.	YouGov 2013	4,13	5	1,16	1.321
SE3	Frauen provozieren durch freizügige Kleidung Männer bewusst zu sexistischem Verhalten.	eigene Konstruktion	2,53	2	1,32	1.317
SE4	Frauen interpretieren harmloses Verhalten häufig übertriebenermaßen als sexuelle Belästigung.	eigene Konstruktion	2,63	3	1,18	1.314

Anmerkungen: Alle Items wurden mit einer fünf-stufigen Likert-Skala abgefragt (stimme gar nicht zu, stimme eher nicht zu, weder noch, stimme eher zu, stimme voll und ganz zu).

Tabelle A3 Hauptkomponentenanalyse der abgefragten Items (GR1-GR3)

Item	Komponente			Unerklärte Varianz
	1	2	3	
GR1.1	0,004	-0,452	-0,0155	0,427
GR1.2	0,017	0,513	0,061	0,256
GR1.3	0,022	0,494	0,124	0,339
GR1.5	0,114	-0,479	0,167	0,334
GR2.1	0,320	0,075	0,002	0,546
GR2.2	0,081	-0,034	-0,519	0,527
GR2.3	0,436	-0,045	0,180	0,469
GR2.4	0,432	0,015	-0,018	0,325
GR2.5	0,465	-0,146	-0,127	0,370
GR2.6	0,364	0,052	-0,010	0,464
GR2.7	0,390	0,098	-0,032	0,284
GR3.1	0,023	0,054	0,657	0,389
GR3.2	0,000	-0,123	0,450	0,538
Eigenvalue	3,299	2,732	1,554	

Anmerkungen: n=1.242; schiefwinkelig-rotierte Komponenten (oblique promax rotation); erklärte Varianz=59,49%.

Tabelle A4 Korrelation der Hauptkomponenten aus Tabelle A3

	1	2	3
1	1		
2	0,591	1	
3	-0,321	-0,415	1

Tabelle A5 Hauptkomponentenanalyse der Items zur Sexismus-Debatte (SE)

Item	Ladung	Unerklärte Varianz
SE1	-0,509	0,416
SE2	0,473	0,494
SE3	0,472	0,498
SE4	0,543	0,335
Eigenvalue	2,258	

Anmerkungen: n=1.291; erklärte Varianz=56,45%.

Tabelle A6 Index „Kein Sexismusproblem“

Attraktivität	Experimentalgruppe							Total
	Uni-Logo	Mann			Frau			
	trifft nicht zu	niedrig	mittel	hoch	niedrig	mittel	hoch	
Mann	0,609 (n=176)	0,124 (n=43)	0,564 (n=79)	0,397 (n=35)	0,486 (n=39)	0,315 (n=79)	0,542 (n=43)	0,482 (n=494)
Frau	-0,437 (n=260)	-0,550 (n=57)	-0,267 (n=133)	0,124 (n=54)	-0,158 (n=50)	-0,095 (n=139)	-0,118 (n=61)	-0,268 (n=754)
Total	-0,015 (n=436)	-0,260 (n=100)	0,043 (n=212)	0,231 (n=89)	0,124 (n=89)	0,053 (n=218)	0,155 (n=104)	0,000 (n=1.291)

Anmerkungen: Angegeben sind die Mittelwerte des Index „Kein Sexismusproblem“ (vgl. Tabelle A2 und Tabelle A5).

Tabelle A7 Index „Gegen Gleichstellung“

Attraktivität	Experimentalgruppe							Total
	Uni-Logo	Mann			Frau			
	trifft nicht zu	niedrig	mittel	hoch	niedrig	mittel	hoch	
Mann	0,604 (n=172)	0,103 (n=40)	0,545 (n=80)	0,422 (n=34)	0,081 (n=41)	0,250 (n=77)	0,796 (n=41)	0,456 (n=485)
Frau	-0,473 (n=253)	-0,649 (n=56)	-0,002 (n=127)	-0,152 (n=53)	-0,218 (n=51)	-0,125 (n=130)	-0,174 (n=60)	-0,277 (n=730)
Total	-0,037 (n=425)	-0,335 (n=96)	0,209 (n=207)	0,072 (n=87)	-0,085 (n=92)	0,015 (n=207)	0,220 (n=101)	0,000 (n=1.242)

Anmerkungen: Angegeben sind die Mittelwerte des Index „Gegen Gleichstellung“ (vgl. Tabelle A2 und Tabelle A3 [Komponente 1]).

Tabelle A8 Index „Anti-Berufstätigkeit“

Attraktivität	Experimentalgruppe							Total
	Uni-Logo	Mann			Frau			
	trifft nicht zu	niedrig	mittel	hoch	niedrig	mittel	hoch	
Mann	0,551 (n=172)	0,057 (n=40)	0,505 (n=80)	0,146 (n=34)	0,426 (n=41)	0,375 (n=77)	0,762 (n=41)	0,453 (n=485)
Frau	-0,329 (n=253)	-0,591 (n=56)	-0,173 (n=127)	-0,311 (n=53)	-0,200 (n=51)	-0,201 (n=130)	-0,175 (n=60)	-0,276 (n=730)
Total	0,027 (n=425)	-0,321 (n=96)	0,089 (n=207)	-0,133 (n=87)	0,079 (n=92)	0,013 (n=207)	0,205 (n=101)	0,000 (n=1.242)

Anmerkungen: Angegeben sind die Mittelwerte des Index „Anti-Berufstätigkeit“ (vgl. Tabelle A2 und Tabelle A3 [Komponente 2]).

Tabelle A9 Index „Pro-Berufstätigkeit“

	Experimentalgruppe							Total
	Uni-Logo	Mann			Frau			
Attraktivität	trifft nicht zu	niedrig	mittel	hoch	niedrig	mittel	hoch	
Mann	-0,358 (n=172)	-0,044 (n=40)	-0,453 (n=80)	-0,270 (n=34)	-0,617 (n=41)	-0,346 (n=77)	-0,718 (n=41)	-0,392 (n=485)
Frau	0,216 (n=253)	0,401 (n=56)	0,195 (n=127)	0,127 (n=53)	0,274 (n=51)	0,265 (n=130)	0,443 (n=60)	0,252 (n=730)
Total	-0,016 (n=425)	0,216 (n=96)	-0,055 (n=207)	-0,028 (n=87)	-0,123 (n=92)	0,037 (n=207)	-0,028 (n=101)	0,000 (n=1.242)

Anmerkungen: Angegeben sind die Mittelwerte des Index „Pro-Berufstätigkeit“ (vgl. Tabelle A2 und Tabelle A3 [Komponente 3]).

„Kann es sein, dass ich das Beginndatum falsch erfasst habe?“

Reactive Dependent Interviewing bei der Erhebung von Ausbildungs- und Erwerbsverläufen junger Erwachsener

“Are you sure you remembered the start-date correctly?”

Survey Data Collection of Educational and Labour Pathways of Young Adults using Reactive Dependent Interviewing

Melania Rudin und Christoph Müller

Zusammenfassung

Eine der Herausforderungen bei der Erhebung von Lebensverlaufsdaten im Panel-design besteht darin, die übergreifende Konsistenz der Daten aus den verschiedenen Erhebungswellen zu gewährleisten. „Dependent Interviewing“ ist eine Möglichkeit, diese Schwierigkeit zu verringern, indem man sich auf Daten bezieht, die in früheren Wellen erhoben wurden und diese Informationen für die Formulierung der Interviewfragen und die Ablaufsteuerung des Erhebungs-Instruments benutzt. In der achten Erhebungswelle der Längsschnittstudie TREE (Transitions from Education to Employment) wurden Formen von „Reactive Dependent Interviewing“ verwendet, um den Episodenanschluss an bereits vorhandene Lebensverlaufsdaten vorzunehmen. Gestützt auf Kenntnisse der Kognitionspsychologie wurden weitere Plausibilisierungs-

Abstract

One of the challenges in the collection of life course-data in panel-design-surveys is to guarantee overall consistency of the data from different annual surveys. “Dependent Interviewing” is one possibility to reduce these difficulties, by drawing on data collected at previous waves to phrase questions or direct respondents through the Survey-instrument. For the eighth wave of the longitudinal survey TREE (Transitions from Education to Employment) forms of “Reactive Dependent Interviewing” were used to ensure the connection to existing episodes of previous waves. Based on findings from cognitive psychology, further modules of plausibilisation were integrated in the CATI-instrument: a visual presentation of life course-data on the screen for interviewers and consistency checks for gaps – periods without activity – as well as



Module ins CATI-Instrument integriert: Eine graphische Darstellung der Verlaufsdaten für das Interviewpersonal, sowie Prüfmodule für Lücken und Überschneidungen von Tätigkeiten. Dieser Artikel beschreibt anhand einer Datenauswertung von Rohdaten die Verwendung und Nützlichkeit der entwickelten Werkzeuge.

overlapping activities. This article presents results of the first analysis of survey raw data and describes the application and usefulness of the developed CATI-tools.

1 Herausforderungen bei der Erhebung von Ausbildungs- und Erwerbsverläufen junger Erwachsener in der Schweiz

Der Übergang von der Schule in den Beruf gestaltet sich in der Schweiz wie auch in anderen Ländern heute weniger gradlinig als für frühere Generationen von Schulabgängern (Buchholz et al. 2009; OECD 2009: 375–376). Das Ziel des Schweizer Jugendlängsschnitts TREE (Transitions from Education to Employment) ist es, diese Übergänge zu analysieren (TREE 2008)¹. Die Ergebnisse der Längsschnittuntersuchung ermöglichen es, Muster bezüglich der Ausbildungs- und Erwerbsverläufe zu identifizieren und zu quantifizieren, indem detaillierte Verlaufsdaten einer Schulabgängerkohorte erhoben und ausgewertet werden (Keller/Hupka-Brunner/Meyer 2010). Die Panelstudie beruht auf einer Stichprobe von rund 6000 Jugendlichen, die im Jahr 2000 an der Schweizer PISA-Erhebung teilgenommen hatten und im selben Jahr die obligatorische Schulzeit beendeten. Sie wurden in den Jahren 2001 bis 2007 jährlich mittels einer Kombination von computergestützten Telefoninterviews (Computer Assisted Telephone Interview: CATI) und schriftlichen Ergänzungsfragebogen zu ihren Ausbildungs- und Erwerbsverläufen befragt. Nach einer Unterbrechung von drei Jahren wurde im Jahr 2010 die achte Erhebungswelle durchgeführt. Die Herausforderung für diese Erhebungswelle bestand einerseits darin, Ausbildungs- und Erwerbsverläufe für rund drei Jahre möglichst widerspruchsfrei und lückenlos zu erfassen, andererseits auch die Konsistenz zu den in den vorhergehenden Panelwellen erhobenen Lebensverlaufsdaten zu gewährleisten. Zudem sollte das Erhebungsinstrument die Erinnerung und Datierung von Ereignissen und

1 Die Schweizer Jugendlängsschnittstudie TREE (Transitionen von der Erstausbildung ins Erwerbsleben, www.tree.unibas.ch) läuft seit 2000 und wurde bisher durch den Schweizerischen Nationalfonds, die Universität Basel, die Bundesämter für Berufsbildung und Technologie bzw. Statistik sowie die Kantone Bern, Genf und Tessin finanziert.

Zeitabschnitten für die Befragten erleichtern, dadurch den nachträglichen Editions- aufwand verringern und die Datenqualität erhöhen. Die Beschreibung der entwickelten Werkzeuge für das CATI-Instrument der achten TREE-Erhebungswelle und deren Nutzen in Bezug auf Datenqualität und Editions-aufwand sind Inhalt dieses Artikels².

Lebensverlaufsdaten weisen immer komplexe Strukturen auf und sowohl die Datengenerierung als auch die (nachträgliche) Datenedition gestalten sich aufwendiger als bei Querschnittsdaten. Besonders akzentuiert ist dies der Fall, wenn Lebensverläufe im Paneldesign erhoben werden (Solga 2001). Die Edition der erhobenen Daten beinhaltet dann die Korrektur von inkonsistenten Angaben für die aktuelle Erhebungswelle und eine Prüfung der Konsistenz mit Tätigkeiten, die bereits in früheren Wellen erfasst wurden. Bei der Datenaufbereitung werden die Angaben aller Wellen miteinander abgeglichen. Wenn eine Tätigkeit – beispielsweise eine Erwerbsarbeit – sich über den Zeitraum von mehreren Erhebungswellen erstreckt, dann wird sie zwar meistens in mehreren Wellen berichtet; es kommt aber oft vor, dass einzelne Angaben zu dieser Tätigkeit sich zwischen den verschiedenen Wellen unterscheiden – beispielsweise das Beginndatum der Tätigkeit, die Berufsbezeichnung oder der Name des Betriebs³.

Das Problem beim Erstellen von Lebensverlaufsdaten im Paneldesign mit einer nachträglichen Datenedition besteht nicht nur im grossen Aufwand, sondern insbesondere darin, dass sich selten generalisierbare Regeln ableiten lassen, welche der vorhandenen Informationen valide sind (Matthes/Reimer/Künster 2007). Werden Lebensverlaufsdaten nach der Erhebung editiert, besteht die Gefahr, dass ungewöhnliche Verläufe „geglättet“, Daten also fälschlicherweise korrigiert werden (Matthes/Reimer/Künster 2007). Weil ein Projektziel der TREE-Studie darin besteht, ein Augenmerk auf „kritische“ Transitionen zu haben (also Ausbildungsabbrüche, -unterbrüche, -wechsel und Repetitionen abzubilden), müssen unkorrekte „Glättungen“ der Daten unbedingt vermieden werden. Unterlässt man jedoch die Kor-

- 2 Die Konzeption und Erarbeitung des hier vorgestellten Erhebungsinstruments beruht zu grossen Teilen auf der Leistung von Christina von Rotz und David Edlmann. Die Autorin und der Autor danken ihnen für die Arbeit.
- 3 Beispiel: In der einen Erhebungswelle wird eine Erwerbstätigkeit als Detailhandelsangestellte seit Juli 2007 im Modegeschäft X an der Bahnhofstrasse in Zürich angegeben. In der nächsten Welle wird eine Erwerbstätigkeit als Detailhandelsangestellte seit Oktober 2007 im Modegeschäft Y an der Bahnhofstrasse in Zürich berichtet. Denkbar ist, dass es sich um zwei verschiedene Erwerbstätigkeiten handelt. Es ist aber auch möglich, dass es sich um die gleiche Erwerbstätigkeit handelt, sich die befragte Person im Beginndatum geirrt hat, und es sich um das gleiche Modegeschäft handelt, welches aufgrund eines unterschiedlichen Vorgehens bei der Stichwortsuche anders codiert wurde.

rekturen, kann dies zu Fehlern in den Ergebnissen führen, wenn die Inkonsistenzen auf Erinnerungsfehlern oder Missverständnissen beruhen (Matthes 2007). Auch dieses Problem ist für die TREE-Studie zentral, besteht doch der Fokus auf der Darstellung, Quantifizierung und Analyse sowohl von gradlinigen wie auch von diskontinuierlichen Verläufen.

Trotz dieser erschwerenden Bedingungen weisen Panelstudien im Vergleich zu reinen Retrospektivbefragungen für die Erfassung von Verlaufsdaten entscheidende Vorteile auf (Solga 2001). Das Erinnerungsproblem ist kleiner, weil die Erhebungszeitpunkte näher bei den Ereignissen liegen (Solga 2001: 297) und es können zu jedem Erhebungszeitpunkt Einschätzungen der subjektiven Situation erhoben werden, was mit einer Retrospektivbefragung nicht möglich wäre. Der vielleicht grösste Vorteil des Paneldesigns besteht darin, dass die Gefahr einer „geglätteten“ Erzählung des Verlaufs⁴ kleiner ist (Solga 2001: 307). Allerdings werden auch in Panelbefragungen die Tätigkeiten eines bestimmten Zeitraums retrospektiv abgefragt. Insbesondere bei jüngeren Befragten ist es gängig, dass mehrere Tätigkeiten parallel ausgeübt werden und sich die einzelnen Episoden somit zeitlich überschneiden. Bei der nachträglichen Datenedition von Retrospektivdaten lässt sich oft nicht ableiten, welche Überschneidungen plausibel sind und wo eine Korrektur von Inkonsistenzen angebracht wäre. Eine weitere Herausforderung besteht darin, Zeitperioden ohne Angaben einer Tätigkeit (Lücken) bei der Datenerfassung zu vermeiden.

2 „Dependent Interviewing“ als unterstützende Erhebungstechnik bei Panelstudien

Wenn Differenzen zwischen den Angaben der aktuellen und der früheren Befragung direkt während des Interviews besprochen werden, bietet dies eine Möglichkeit, die Konsistenz zwischen den Angaben verschiedener Erhebungswellen zu verbessern und die Erinnerungsarbeit der Befragten zu erleichtern. Dazu werden Techniken von „Dependent Interviewing“ (DI) verwendet. Beim DI bezieht man sich auf Daten, die in früheren Erhebungswellen erfasst wurden und verwendet diese Informationen für die Formulierung der Fragen und die Ablaufsteuerung des aktuellen Befragungsinstruments (Uhrig/Sala 2011). DI kann entweder „proaktiv“ (PDI) oder „reaktiv“ (RDI) ausgestaltet werden (Lynn/Sala 2006). Beim „Proactive Dependent Interviewing“ (PDI) präsentiert man Daten von vorhergehenden Erhe-

4 durch die befragte Person

bungswellen den Befragten und bittet sie, diese zu bestätigen, zu korrigieren oder ausgehend vom letzten Status mit der Berichterstattung fortzufahren. Beim „Reactive Dependent Interviewing“ (RDI) werden dem Befragten zuerst unabhängig von bereits vorhandenen Informationen Fragen gestellt. Die Informationen aus den Vorwellen werden dann genutzt, um die Angaben der aktuellen Befragung mit denen aus früheren Erhebungswellen zu vergleichen (Uhrig/Sala 2011:334). So kann beispielsweise während des Interviews geklärt werden, ob es sich bei der aktuellen Erwerbstätigkeit um dieselbe handelt, die in der letzten Welle bereits erfasst wurde. Falls es sich um die gleiche Tätigkeit handelt, die Angaben zu einzelnen Variablen jedoch voneinander abweichen, können diese Differenzen ebenfalls während des Interviews angesprochen und bereinigt werden. Heute wenden die meisten grossen Längsschnittstudien Formen von DI an (Uhrig/Sala 2011), wobei vor allem PDI zum Einsatz kommt⁵. DI hat sich bewährt, um die Wahrscheinlichkeit von unechten Transitionen aufgrund unterschiedlicher Angaben in den einzelnen Wellen zu verringern (Jäckle und Lynn 2007, Hale und Michaud 1995) und um Schätzfehler für die Anzahl der Stellenwechsel aufgrund abweichender Angaben zur Erwerbstätigkeit zu reduzieren (Lynn/Sala 2006; Mathiowetz/McGonagle 2000).

Für das CATI-Instrument der achten TREE-Erhebungswelle wurde der Beschluss gefasst, Formen von „Dependent Interviewing“ zu verwenden. „Proactive Dependent Interviewing“ hat den Vorteil, dass die Befragten ausgehend von der Situation zum Zeitpunkt der letzten Befragung ihre darauffolgenden Tätigkeiten entlang der Zeitachse berichten können⁶. Ein Vorteil von „Reactive Dependent Interviewing“ besteht darin, dass sie zuerst unbeeinflusst ihre aktuelle Situation schildern können und zu Beginn des Interviews nicht gleich mit Erinnerungsschwierigkeiten konfrontiert werden. Dieser Aspekt schien bei dieser Erhebung aufgrund der längeren abgefragten Zeitspanne (3 Jahre) besonders relevant. Deshalb wurde für das TREE-Erhebungsinstrument die Anwendung von „Reactive Dependent Interviewing“ beschlossen. Zudem waren Bedenken vorhanden, dass die Anwendung von PDI stärker zum Glätten von Verläufen einladen kann. Diese Gefahr besteht insbesondere, wenn Befragte angeben, immer noch die gleiche Tätigkeit auszuüben wie beim letzten Interview. Auch wenn dieses Risiko durch einen gut strukturierten CATI-Ablauf und spezifische Fragen nach Wechselereignissen erheblich verringert werden kann, wurde ein Vorteil von RDI darin gesehen, zuerst unabhängig

5 Z.B. British Household Panel Survey, Sozio-oekonomisches Panel (SOEP) (Deutschland), Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (LV-Pamel 71).

6 Das Erinnern entlang des sequentiellen Erinnerungspfads wird dadurch erleichtert (vgl. Kapitel 3).

die aktuellen Tätigkeiten aufzunehmen und erst danach einen Abgleich mit den Informationen aus der Vorwelle vorzunehmen.

3 Weitere CATI-Werkzeuge zum Erleichtern des autobiografischen Erinnerns

Ein zentrales Problem bei der Erhebung von Lebensverlaufsdaten ist, dass es den Befragten häufig schwer fällt, sich an Ereignisse zu erinnern und diese zu datieren (Reimer 2001; Reimer 2005; Matthes/Reimer/Künster 2007). Diese Erinnerungsschwierigkeiten sind eine Ursache für Inkonsistenzen zwischen den Daten der einzelnen Erhebungswellen, unplausible Überschneidungen mehrerer Tätigkeiten oder Lücken in den Lebenslaufdaten. In Anlehnung an Befunde aus der Kognitionspsychologie und Erfahrungsberichte aus anderen grossen Längsschnitterhebungen wurden deshalb zusätzliche CATI-Werkzeuge für die achte TREE-Erhebungswelle konzipiert⁷. Im Folgenden werden einige Ergebnisse der Kognitionspsychologie festgehalten und Konsequenzen daraus für die Erarbeitung des Erhebungsinstruments abgeleitet.

3.1 Erkenntnisse der Kognitionspsychologie und Konsequenzen für die Ausgestaltung des Erhebungsinstruments

Verschiedene Publikationen⁸ von M. Reimer, K. Drasch, B. Matthes, A. Dürnberger und R. Künster haben die Befunde der Kognitionspsychologie zur Erinnerung autobiografischer Ereignisse in Bezug auf Datenerhebungen in Anlehnung an Barsalou (1988), Conway (1996) sowie Conway und Pleydell-Pearce (2000) dargestellt.

In der Kognitionspsychologie werden vor allem drei Leistungen unterschieden, die beim Erinnern von autobiographischen Ereignissen nötig sind (Reimer 2001: 15; Dürnberger/Drasch/Matthes 2011: 8):

- die Erinnerung, dass ein bestimmtes Ereignis stattgefunden hat – „erinnern dass“
- die Erinnerung an konkrete Details und Zusammenhänge des Ereignisses – „erinnern wie/wo/warum“
- die Erinnerung an die korrekte zeitliche Verortung des Ereignisses – „erinnern wann“.

7 In Anlehnung u.a. an Matthes/Reimer/Künster 2007, Dürnberger/Drasch/Matthes 2011,

8 Reimer 2001; Reimer 2005; Reimer, Matthes 2006; Matthes/Reimer/Künster 2007; Drasch/Matthes 2009; Dürnberger/Drasch/Matthes 2011.

Ohne die erste Leistung – die Erinnerung, dass ein Ereignis stattgefunden hat – sind die beiden anderen Erinnerungsprozesse nicht möglich. Die drei Erinnerungsleistungen sind deshalb als hierarchischer Prozess zu betrachten (Dürnberger/Drasch/Matthes 2011). Details zu einem Ereignis werden im zweiten Schritt rekonstruiert. Die Datierung eines Ereignisses ist der dritte Schritt in der Erinnerungsleistung und Forschungsbefunde deuten darauf hin, dass das Erinnern eines Ereignisses und seine zeitliche Einordnung unabhängig voneinander ablaufen (Reimer 2001: 17; Wagenaar 1986).

Diese Befunde legen es für die Erarbeitung des Erhebungsinstruments nahe, zuerst nach dem Vorkommen eines Ereignisses oder einer Tätigkeit zu fragen (z.B. einer Phase von Arbeitslosigkeit) und danach die konkreten Details dieser Tätigkeit zu erfassen. Sie zeigen auch, dass die Angabe eines exakten Datums oder Zeitraums eine schlechte Hilfe für das Erinnern eines Ereignisses darstellt (Reimer 2001; Conway 1996; Wagenaar 1986). Die Befunde zeigen drittens, dass insbesondere für die Erinnerung des Beginn- und Enddatums einer Tätigkeit eine zusätzliche Unterstützung durch das CATI-Instrument sinnvoll ist.

Anhand drei verschiedener Erinnerungspfade können Ereignisse aus dem autobiografischen

Gedächtnis abgerufen werden (Dürnberger/Drasch/Matthes 2011; Barsalou 1988; Conway/Pleydell-Pearce 2000):

- Hierarchische Erinnerungspfade: Hier verläuft die Erinnerung vom Ereignis selbst zu dessen Eigenschaften („top-down“). Die Strukturen der Erinnerungsorganisation sind themengebunden und verlaufen vom Allgemeinen zum Speziellen (z.B. meine Erwerbstätigkeit → meine Zeit an der Stelle X → Arbeitspensum an der Stelle X).
- Parallele Erinnerungspfade: Hier handelt es sich um themenübergreifende Erinnerungspfade zwischen Lebensbereichen, die sowohl gleichzeitige als auch unmittelbar aufeinander folgende Ereignisse miteinander verknüpfen (z.B. meine Zeit an der Stelle X ↔ Sprachkurs im Zusammenhang mit der Arbeit an der Stelle X).
- Sequenzielle Erinnerungspfade: Diese verlaufen entlang einer kausal-temporären Abfolge von Ereignissen innerhalb eines Lebensbereichs und verbinden oft länger andauernde Ereignisse miteinander (z.B. meine Zeit an der Stelle X → meine darauffolgende Zeit an der Stelle Y).

Um auf Interviewfragen zu vergangenen Ereignissen zu antworten, durchsuchen Menschen das gesamte Gedächtnis entlang der drei verschiedenen Erinnerungspfade (Matthes/Reimer/Künster 2007). Die Kombination eines modularen Vorgehens nach Themenbereichen und eines sequentiellen Vorgehens entlang einer Zeitachse bei der CATI-Erhebung von Lebenslaufdaten hat sich in mehreren Panel-

erhebungen⁹ bewährt, auch weil dadurch das Erinnern über den sequentiellen Erinnerungspfad erleichtert wird.

Weil für die achte Erhebungswelle der TREE-Befragung entschieden wurde, „Reactive Dependent Interviewing“ anzuwenden, wird das Erinnern über den sequentiellen Erinnerungspfad für die Befragten am Anfang erschwert. Sie berichten zuerst ihre aktuelle Situation und werden dann mit den Angaben aus den vorhergehenden Erhebungswellen konfrontiert. Umso bedeutsamer ist es, das Erinnern über die hierarchischen und parallelen Erinnerungspfade durch die Ausgestaltung des Erhebungsinstruments zu erleichtern.

Weil exakte Datumsangaben gemäss der Gedächtnispsychologie nicht in die Struktur des autobiografischen Erinnerns eingebettet sind (Reimer 2005: 41), wenden Menschen zur Vergabe von Kalenderdaten Inferenzstrategien an (Matthes/Reimer/Künster 2007). Eine dieser Inferenzstrategien ist die Erinnerung über den Kontext. Anhand hierarchischer, themenübergreifender und sequentieller Abrufpfade werden verschiedene Ereignisse miteinander verbunden und zeitliche Verknüpfungen wie „früher als“, „danach“ und „seither“ benutzt, um Ereignisse datieren zu können (Matthes/Reimer/Künster 2007; Huttenlocher/Hedges/Prohaska1988). Oft wird das erinnerte Detail mit anderen Erinnerungen abgeglichen, um sicher zu gehen, dass es übereinstimmend mit weiteren wichtigen Ereignissen genannt wurde (Matthes/Reimer/Künster 2007; Huttenlocher/Hedges/Bradburn1990; Larsen/Thomson 1995; Loftus und Marburger 1983). Neben der Inferenzstrategie über den Kontext gibt es individuelle Erinnerungsstrategien, welche stark variieren (Matthes/Reimer/Künster 2007).

Diese Befunde legten nahe, das Erhebungsinstrument so auszugestalten, dass möglichst rasch und flexibel auf Kontextangaben zurückgegriffen werden kann.

3.2 CATI-Werkzeuge zur Erleichterung des Erinnerns über parallele Erinnerungspfade

Die Förderung des Erinnerns über den parallelen Erinnerungspfad stellt komplexe Anforderungen an das Erhebungsinstrument. Zusätzliche Hilfestellungen und oft auch ein Abweichen von strikt standardisierten Fragen sind nötig, um den persönlichen Kontext der befragten Person sinnvoll zu involvieren. Insgesamt zeigen Forschungsergebnisse, dass eine kontextgestützte Abfrage zu weniger Lücken

9 Erhebung des LV-Panel 71 „Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland“ (Matthes/Reimer/Künster 2007), und IAB-ALWA Leben und „Arbeiten im Wandel“ (Drasch/Matthes 2009).

und Fehlern in der biografischen Erinnerung führt (Dürnberger/Drasch/Matthes 2011; Drasch/Matthes 2009). Durch das Einbinden vorhandener Informationen in die Formulierung der Fragen kann eine bessere Kontextualisierung gewährleistet werden¹⁰. Auch eine flexible Unterstützung durch das Interviewpersonal kann beim Erinnern über den parallelen Erinnerungspfad helfen. Um diese anbieten zu können, muss das Interviewpersonal intensiv geschult werden und während des Interviews Informationen zur Biografie der befragten Person zur Verfügung haben. Diese Informationen werden durch die Verwendung von biografischen Kalendern (auch „calendar instruments“¹¹) geleistet. Bei der Verwendung von biografischen Kalendern enthält das Erhebungsinstrument eine automatische Visualisierung von Verläufen. Diese basiert auf bereits erhobenen Daten und wird während des Interviews laufend mit den neuen Angaben ergänzt. Die Interviewer/innen können den Befragten dadurch Hilfestellungen für die Erinnerungsarbeit liefern (Glasner/van der Vaart 2009). Biografische Kalender haben sich als nützlich erwiesen, um die Datenqualität von Retrospektivbefragungen zu verbessern (Belli/Lee/Staford/Chou 2004; Glasner/van der Vaart 2009). Die Darstellung von Verlaufsdaten ermöglicht es zudem, Korrekturmodule in die Befragung zu implementieren, so dass eine Behebung von Inkonsistenzen – beispielsweise in Form von sich zeitlich überschneidenden Tätigkeiten – bereits während des Interviews in Rücksprache mit der befragten Person möglich ist.

4 Aufbau des Erhebungsinstruments für die TREE-Befragung 2010

Die langjährige Erfahrung bei der Erhebung und Edition der TREE-Verlaufsdaten lieferte die Grundlage zur Erweiterung des Erhebungsinstruments für die achte Welle. Aufbauend darauf sowie auf Erfahrungsberichten anderer Panelerhebungen wurden die folgenden Beschlüsse für die Ausarbeitung des CATI-Instruments gefasst:

- Im Interview werden als erstes die aktuellen Tätigkeiten erfasst. Der Episodenanschluss an die vorhandenen Daten geschieht dann anhand von „Reactive Dependent Interviewing“.

10 Eine Frageformulierung kann dann beispielsweise lauten: „Ihre Arbeit als <<Berufsbezeichnung>> im Betrieb <<Betriebsname>> dauerte gemäss Ihren Angaben bis <<MM.YYYY>>. Haben Sie danach noch eine andere Stelle gehabt?“

11 Eine Übersicht der verwendeten Termini findet sich in Glasner/van der Vaart (2009).

- Die Verlaufsdaten werden am Bildschirm visualisiert, damit eine flexible Unterstützung der Erinnerungsarbeit der Befragten durch die Interviewer/innen möglich ist.
- Anhand „Proactive Dependent Interviewing“ werden die Enddaten der Tätigkeiten erfasst, die zum Zeitpunkt der letzten Befragung noch aktuell waren und inzwischen beendet wurden.
- Zeitliche Inkonsistenzen – in Form von sich überschneidenden Tätigkeiten und Zeitperioden ohne Aktivität – sollen während des Interviews erkannt und visualisiert werden, damit eine Korrektur der Daten – falls nötig – bereits während des Interviews möglich ist.
- Das Ablaufschema des CATI-Instruments basiert auf einem konsequent modularen Vorgehen, indem so lange nach weiteren Tätigkeiten eines Themenbereichs – beispielsweise Erwerbstätigkeit – gefragt wird, bis keine weiteren Tätigkeiten in diesem Bereich mehr genannt werden.

Im Folgenden wird der grundsätzliche Aufbau des CATI-Fragebogens beschrieben (Abschnitt 4.1) und danach auf die einzelnen Werkzeuge des Erhebungsinstruments eingegangen (Abschnitte 4.2 bis 4.5).

4.1 Aufbau des CATI-Instruments

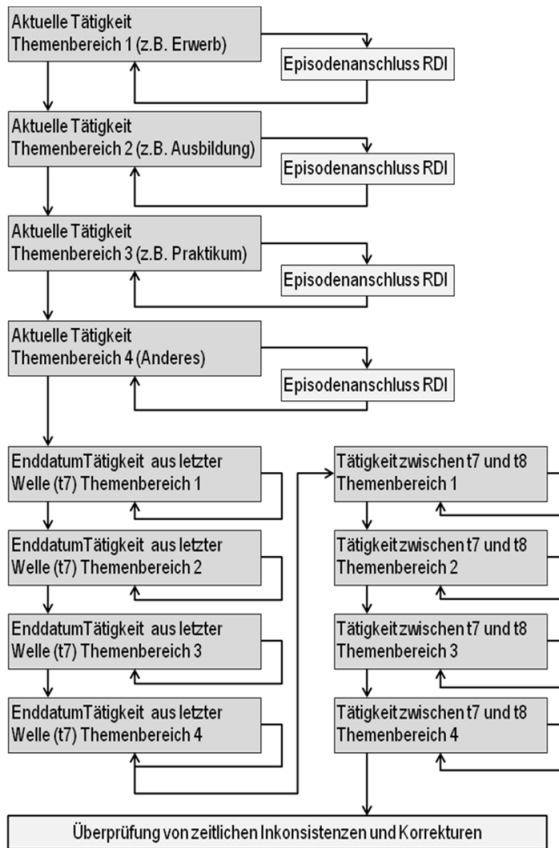
Das Interview begann mit der Frage nach der aktuellen Haupttätigkeit und dem Erfassen der Details zu dieser Episode.¹² In Abbildung 1 ist der Ablauf des CATI-Interviews schematisch dargestellt.

Generell wurden immer nach dem Erinnern einer Tätigkeit die Details dazu erfragt¹³. Nach der Erfassung der aktuellen Haupttätigkeit wurden alle aktuellen Tätigkeiten zu diesem Themenbereich – beispielsweise Erwerbstätigkeit – erhoben. Falls das Beginndatum der aktuellen Tätigkeiten vor dem Zeitpunkt der letzten Befragung lag, wurde anhand von RDI der Anschluss an die bestehenden Episoden geprüft. Als nächstes wurden alle aktuellen Tätigkeiten der anderen Themenbereiche erhoben und auch hier anhand RDI der Episodenanschluss abgeklärt (vgl.

12 Das Ziel war, einen für die Befragten möglichst angenehmen Einstieg in das Interview zu ermöglichen, indem sie mit der Auskunft zu dem Lebensbereich beginnen können, der aktuell für sie zentral ist. Damit sollten zu Beginn des Interviews auch Erinnerungsschwierigkeiten vermieden werden, die demotivierend wirken können.

13 Pro Tätigkeit wurde eine Episode erfasst. Nach dem Erinnern an die Tätigkeit wurden die zentralen Variablen der Episoden erfasst, bestehend aus Angaben zum Beginn- und Enddatum, zur Art der Tätigkeit, zur Berufs- oder Ausbildungsbezeichnung sowie zu den Erwerbs- und Ausbildungsbetrieben. Zusätzlich wurden pro Episode weitere Detailangaben erhoben, wie z.B. die Anzahl wöchentlicher Arbeitsstunden, oder bei abgeschlossenen Tätigkeiten der Grund der Beendigung.

Abbildung 1 Schematische Darstellung des CATI-Ablaufs



Abschnitt 4.3). Es wurde jeweils so lange nach weiteren Tätigkeiten des Bereichs gefragt, bis die Befragten keine zusätzlichen Episoden dieses Bereichs erinnerten¹⁴. Erst dann wurde in ein nächstes thematisches Modul gewechselt.

Das Erhebungsinstrument enthielt vier Module: „Erwerb“, „Ausbildung“, „Praktikum“ und „Anderes“. Im Interview wurde immer mit dem Bereich begonnen, dem die Haupttätigkeit der befragten Person zugeordnet wurde. Dadurch unterschied sich die konkrete Reihenfolge der Module zwischen den einzelnen Inter-

14 Die Fragen lauteten beispielsweise: „Gehen Sie aktuell noch einer weiteren Erwerbstätigkeit nach?“

views¹⁵. Unabhängig davon, welche Episoden bereits erfasst waren, wurden immer alle Module durchlaufen. Das bedeutet, dass beispielsweise immer danach gefragt wurde, ob die Person aktuell ein Praktikum absolviert, auch wenn bereits Erwerbstätigkeiten, Ausbildungen oder beides berichtet wurden¹⁶. In jedem Modul konnten beliebig viele Episoden erfasst werden¹⁷.

Im zweiten Teil des Interviews wurde für die Episoden, die in der letzten Befragung genannt wurden und nicht mit den aktuellen Tätigkeiten verbunden werden konnten, mittels PDI ein Enddatum gesetzt (vgl. Abschnitt 4.4). Als letztes wurden Tätigkeiten erfasst, die in der Zwischenzeit seit der letzten Befragung ausgeübt worden sind, zum aktuellen Zeitpunkt aber bereits wieder beendet waren. Auch hier wurde ein konsequent modulares Vorgehen angewendet, indem solange nach weiteren Tätigkeiten eines Themenbereichs gefragt wurde, bis die Nachfrage mit „nein“ beantwortet wurde.

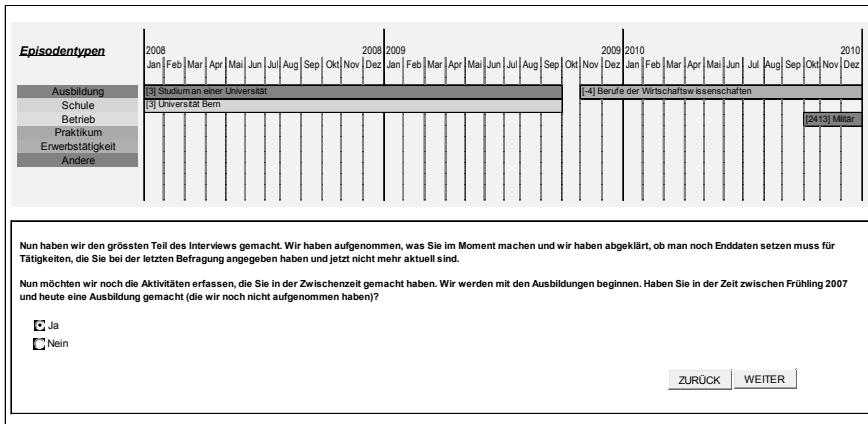
Nachdem alle Tätigkeiten erfasst waren, wurden zeitliche Inkonsistenzen am Bildschirm angezeigt, damit jedenfalls unkorrekte Angaben in Rücksprache mit der befragten Person korrigiert werden konnten (siehe Abschnitt 4.5).

4.2 Graphische Darstellung der Verlaufsdaten

Um eine flexible Unterstützung der Erinnerungsarbeit zu ermöglichen, wurden die Verläufe graphisch am CATI-Bildschirm dargestellt. Die Interviewer/innen konnten sich kurz vor dem Interview einen Überblick über die Verlaufsdaten verschaffen, die in den letzten zwei Wellen erhoben worden waren. Die Visualisierung war zudem unerlässlich, um zum Schluss des Interviews zeitliche Inkonsistenzen zu thematisieren und in Absprache mit den Befragten noch während des Interviews Korrekturen vorzunehmen. Der CATI-Bildschirm war zweigeteilt, wobei die obere Zone dazu diente, die verschiedenen Episoden über die Zeitachse hinweg darzustellen. Unterhalb des Diagramms befand sich der Bereich zur Dateneingabe. Links neben

- 15 Dieses Vorgehen hat sich in den früheren TREE-Erhebungswellen bewährt. Theoretisch ist es denkbar, dass durch die unterschiedliche Reihenfolge der Module Folgeeffekte entstehen. Dadurch, dass systematisch nach Tätigkeiten in allen Themenbereichen gefragt wird, dürfte diese Gefahr jedoch gering sein.
- 16 Auch wurden alle Personen gefragt, ob sie aktuell auf Stellensuche seien. Auch nach Phasen von Arbeitslosigkeit wurden – abgesehen von Personen mit vollem Erwerbsspensum – alle Personen gefragt (mit einer entsprechenden Einleitung, dass Arbeitslosigkeit auch parallel zu einem Erwerb oder einer Ausbildung möglich ist), weil es in der Schweiz auch möglich ist, sich auf einer regionalen Arbeitsvermittlungsstelle als teilzeitarbeitslos zu melden.
- 17 Die Programmierung ermöglichte maximal 50 Episoden pro Modul (was nicht ausgeschöpft wurde).

Abbildung 2 Screenshot; graphische Darstellung der Verlaufsdaten



Hinweise zum Screenshot: Die beiden langen Balken oben links stehen für ein Studium (Ausbildungsbezeichnung und Ausbildungsinstitution), das bereits in der letzten Erhebung angegeben wurde, und für welches in der aktuellen Erhebung ein Enddatum gesetzt wurde. Der Balken rechts steht für eine neu erfasste Erwerbstätigkeit. Der unterste Balken steht für eine Militärepisode, die in der aktuellen Erhebung neu erfasst wurde.

der Zeitachse erklärt eine Legende die sogenannten „Episodentypen“, jeder mit einer spezifischen Farbe versehen. Alle Episoden wurden durch einen Balken auf die Zeitachse geplottet. Die abgebildete Einheit ist die Monatsdauer und die Episoden waren mit der groben Bezeichnung der Tätigkeit sowie der Episodenummer (in rechteckigen Klammern) beschriftet¹⁸. Auf Grund des Platzmangels innerhalb der Balken wurden für jede Episode Zusatzinformationen in Form eines Popup-Fensters angezeigt. Jede neu erfasste Episode erschien sofort als Balken am Bildschirm. Abbildung 2 zeigt einen schematischen Screenshot¹⁹, auf welchem Episoden aus der aktuellen und aus früheren Erhebungswellen zu sehen sind.

Die Interviewer/innen konnten mit Hilfe der graphischen Darstellung Unterstützung bei der Erinnerungsarbeit leisten; insbesondere bei der zeitlichen Verortung von Episoden. Wenn eine befragte Person beispielsweise Mühe bekundete, ihre Phase von Arbeitslosigkeit nach Abschluss der Ausbildung zeitlich genau zu verorten, konnte der/die Interviewer/in das Enddatum der Ausbildung ablesen und somit die Erfassung eines konsistenten Verlaufs erleichtern.

18 Neu erfasste Episoden wurden mit einem „-“ vor der Episodenummer gekennzeichnet.

19 Die Original-Screenshots sind schlecht lesbar, weil grosse Bildschirme mit hoher Auflösung für die Befragung verwendet wurden. Bei den Darstellungen in Abbildung 2 bis Abbildung 7 handelt es sich um nachgezeichnete Screenshots. Im Anhang befindet sich ein Original eines Screenshots.

4.3 Erfassung der aktuellen Situation und Datenabgleich mit Reactive Dependent Interviewing

Für alle aktuellen Episoden, deren Beginndatum vor dem Zeitpunkt der letzten Erhebung lag, wurde anhand RDI überprüft, ob die Angaben mit denen aus der letzten Erhebungswelle übereinstimmten. Das Ziel dieses Vorgehens bestand einerseits darin, den nachträglichen Aufwand für die Datenedition zu verringern, indem Episoden bereits während des Interviews zusammengeführt wurden. Insbesondere diente das Verfahren aber dazu, Unstimmigkeiten zwischen den Angaben aus einzelnen Erhebungswellen während des Interviews zu thematisieren und wenn möglich zu beheben.

Für Erwerbstätigkeiten wurden die Angaben von neuen Episoden mit denen aus den letzten Erhebungswellen anhand dreier Angaben verglichen:

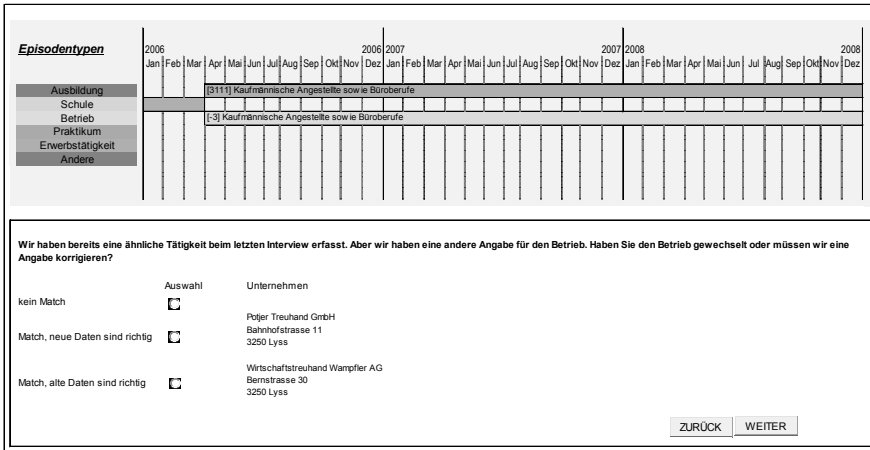
- Beginndatum
- Berufsbezeichnung
- Betriebsbezeichnung (BUR-Nummer²⁰)

Wenn die Angaben für alle drei Variablen übereinstimmten, wurden diese automatisch zu einer Episode verbunden. Wenn nur zwei der drei Angaben übereinstimmten, wurde eine Frage eingeblendet, um die Abweichung bei der einen Variable mit der befragten Person zu klären. Dabei war ein Teil der Frage standardisiert vorgegeben. Die Frage lautete beispielsweise „Wir haben bereits eine ähnliche Tätigkeit beim letzten Interview erfasst. Aber wir haben eine andere Angabe für den Betrieb. Haben Sie den Betrieb gewechselt oder müssen wir eine Angabe korrigieren?“ Anhand der visualisierten Verläufe auf dem Bildschirm konnten die Interviewer/innen weitere Hilfestellungen anbieten. Danach bestanden zwei Möglichkeiten: Die Episoden konnten als einzelne unterschiedliche Tätigkeiten stehen gelassen werden (kein Match), oder die Tätigkeit aus der Erhebung 2010 konnte einer bereits bestehenden Episode zugeordnet werden (Match). Bei einem Match gab es ausserdem die Möglichkeit, zu unterscheiden, ob die alten Angaben aus der letzten Erhebung als korrekt befunden wurden (Match alte Daten), oder ob eine alte Information mit den neu gemachten Angaben überschrieben werden sollte (Match neue Daten)²¹.

20 Die Betriebe wurden anhand der BUR-Nummer erfasst. Das Betriebs- und Unternehmensregister BUR umfasst alle Unternehmen und Betriebe des privaten und öffentlichen Rechts, die in der Schweiz domiziliert sind. Das Bundesamt für Statistik stellt das BUR für Erhebungszwecke zur Verfügung.

21 Insbesondere für Angaben zum Beginndatum einer Tätigkeit ist denkbar, dass die Informationen aus früheren Wellen valider sind, da der ältere Erhebungszeitpunkt näher am Ereignis liegt. Aus diesem Grund wurden Datumsangaben aus früheren Wellen nicht

Abbildung 3 Screenshot; Episodenanschluss mit RDI



Hinweise zum Screenshot: Die Person gab in der Befragung im Jahr 2007 (t7) an, als kaufmännische Angestellte im Betrieb X zu arbeiten und zwar seit April 2006 (Episodennummer 3111). In der Erhebung im Jahr 2010 (t8) gab die Person ebenfalls an, seit April 2006 als kaufmännische Angestellte zu arbeiten, jedoch im Betrieb Y (Episodennummer -3).

Abbildung 3 zeigt einen Screenshot für die Situation mit gleicher Berufsbezeichnung, gleichem Beginndatum und unterschiedlichen Angaben für den Betrieb.

Wenn lediglich für eine der drei Variablen eine Übereinstimmung bestand, wurde nicht nachgefragt, ob es sich um die gleiche Tätigkeit handelt. Bei dieser Konstellation wurden parallele Episoden generiert. Inkonsistenzen aufgrund sich überschneidender Episoden wurden jedoch am Ende des Interviews aufgegriffen (Kapitel 4.5).

Für die Erfassung der Berufstätigkeit wurde eine detaillierte Liste des Bundesamtes für Statistik verwendet, welche rund 18.000 Berufsbezeichnungen enthält. Insofern ist es nicht erstaunlich, dass es in den einzelnen Wellen zu unterschiedlichen Codierungen eines Berufs kommen kann – beispielsweise Maurer oder Baumaurer, Primarlehrer oder Volksschullehrer. Bei Unstimmigkeiten dieser Art wurde in der Regel eine der Berufsbezeichnungen für beide Wellen übernommen. Das Interviewpersonal wurde aber intensiv daraufhin geschult, bei Änderungen der beruflichen Position die Episoden nicht zu verbinden, sondern die Episoden mit den

Fortsetzung Fußnote 21

effektiv aus den Daten gelöscht, wenn die Befragten darauf bestanden, dass das Datum überschrieben werden sollte, sondern die Angaben bekamen in den Daten lediglich eine Markierung „gelöscht“.

unterschiedlichen Berufsbezeichnungen einzeln stehen zu lassen (beispielsweise Maurer und Maurerpolier, Primarschullehrer und Berufsschullehrer). Das Interviewpersonal hatte die Anweisung, im Zweifelsfall die beiden Episoden einzeln stehen zu lassen, denn dass Tätigkeiten im gleichen Berufsfeld mit unterschiedlichen beruflichen Positionen als einzelne Episode erhoben werden, ist eine essenzielle Voraussetzung, um Karriereprozesse zu erfassen.

Auch bei Ausbildungen wurde anhand RDI der Episodenanschluss geprüft. Für Ausbildungsepisoden wurden die Angaben zweier Variablen verglichen: dem Beginndatum und der Bezeichnung der Ausbildung. Wenn die Ausbildungsbezeichnung und die Fachrichtung in beiden Wellen übereinstimmten, wurde eine Frage eingeblendet, um zu klären, ob es sich um die gleiche Ausbildung handelte, die bei der letzten Erhebung genannt wurde²². Detailangaben zur Ausbildung wurden danach ergänzt. So wurden Schulwechsel und Wechsel des Ausbildungsbetriebs erfasst, indem sogenannte „Subepisoden“ für Schulen und Ausbildungsbetriebe erhoben wurden. Auch Praktika im Rahmen einer Ausbildung wurden als „Subepisoden“ erfasst.

Die Verwendung von RDI bringt das Risiko mit sich, dass die Befragten verunsichert oder verärgert reagieren, wenn man sie mit den widersprüchlichen Angaben konfrontiert. Das Interviewpersonal wurde daraufhin geschult, in diesen Situationen nicht zu insistieren, bis Unstimmigkeiten aufgehoben waren, sondern im Zweifelsfall beide Episoden stehen zu lassen. Allenfalls konnten bestehende Unstimmigkeiten am Ende des Interviews beim Überprüfungsmodul von Überschneidungen nochmals aufgegriffen werden oder es musste bei der nachträglichen Datenedition eine Entscheidung getroffen werden²³.

4.4 Erfassung von Enddaten mit „Proactive Dependent Interviewing“

Tätigkeiten, welche die Befragten bei der letzten Erhebung noch ausübten, im Jahr 2010 jedoch nicht mehr, mussten im Prinzip in der Zeit zwischen den beiden Erhebungen geendet haben. Für diese Tätigkeiten wurde anhand von PDI das Enddatum erfasst. Die Frageblöcke in diesem zweiten Teil des Interviews enthielten auch Fragen zu Schul- und Ausbildungsbetriebswechseln, Abschlusserfolgen, Noten und

22 Als „Matchkriterium“ musste sowohl die Bezeichnung der Ausbildung (beispielsweise Studium an einer Universität) als auch die Fachrichtung (beispielsweise Psychologie) übereinstimmen.

23 Hinweise über die Nützlichkeit des Episodenanschlusses mit RDI und die Reaktionen der Befragten finden sich in Abschnitt 5.1.

Abbildung 4 Screenshot; Abfrage von Enddaten

The screenshot displays a data entry interface for tracking activities over time. The top section is a grid with columns for months from January 2006 to December 2008. The rows represent different activity types: Ausbildung (Education), Schule (School), Betrieb (Company), Praktikum (Internship), Erwerbstätigkeit (Employment), and Andere (Other). A legend on the left identifies the activity types: [1] Studium an einer Universität, [2] Universität Bern, [3] Studium an einer Universität, and [4] Berufe der Wirtschaftswissenschaften. A callout box points to a specific entry in the 'Ausbildung' row for January 2007, stating: '[3] Studium an einer Universität, Anfang: 10/2005, Ende: offen, Volkswirtschaftslehre'. Another callout points to an entry in the 'Berufe der Wirtschaftswissenschaften' row for October 2008, stating: '[4] Berufe der Wirtschaftswissenschaften, [2] [3] Müller'. Below the grid, a text box asks: 'Sie haben bei der letzten Befragung angegeben, dass Sie eine Ausbildung "Studium an einer Universität / Volkswirtschaftslehre" machen (Graphik [3]). Bis wann haben Sie diese Ausbildung gemacht?' with dropdown menus for 'Januar' and '2007'. At the bottom right are buttons for 'ZURÜCK' and 'WEITER'.

Hinweise zum Screenshot: Es wird nach dem Enddatum der Tätigkeit gefragt, die bei der letzten Befragung angegeben und nicht mit einer aktuellen Tätigkeit zu einer Episode verbunden werden konnte.

Misserfolge bei Ausbildungen sowie zu Gründen für die Beendigung von Erwerbstätigkeiten. Durch die Verwendung von vorhandenen Angaben in der Frageformulierung wollten wir Missverständnisse vermeiden. Es war uns beispielsweise wichtig, dass verschiedene Erwerbstätigkeiten im gleichen Betrieb nicht zu einer Episode verbunden wurden und dass Berufsaufstiege zuverlässig erfasst wurden. Aus diesem Grund wurden die Berufsbezeichnung und der Betriebsname in den Fragetext eingebunden²⁴.

4.5 Visualisierung von Inkonsistenzen und Korrektur der Daten während des Interviews

Gerade in der Altersgruppe, der die TREE-Befragten angehören²⁵, kommen mehrere parallele Ausbildungs- und Erwerbstätigkeiten oft vor²⁶. Auch Wechselereignisse und Episoden ohne Ausbildungs- oder Erwerbstätigkeit sind verbreitet. Eine Übersicht über die Aktivitäten in den verschiedenen Bereichen zu behalten, ist trotz einem guten modularen Aufbau und den bereits diskutierten Hilfestellungen sowohl

24 Eine der Fragen für Erwerbstätigkeiten lautete beispielsweise: „Sie haben bei der letzten Befragung angegeben, dass Sie als <<Berufsbezeichnung>> im Betrieb <<Betriebsname>> arbeiten. Bis wann haben Sie als <<Berufsbezeichnung>> im Betrieb <<Betriebsname>> gearbeitet?“ Das Interviewpersonal wurde auch für diesen Teil der Erhebung dahingehend geschult, bei Änderungen der beruflichen Position eine neue Episode zu erfassen.

25 Die TREE-Befragten waren bei der Erhebung im Jahr 2010 etwa 26 Jahre alt.

26 Beispielsweise mehrere Erwerbstätigkeiten parallel zu einem Studium.

für die befragten Personen als auch für das Interviewpersonal anspruchsvoll. Aus der Kognitionspsychologie weiss man, dass insbesondere die zeitliche Verortung von Episoden aufgrund von Erinnerungsproblemen schwierig ist. Es ist deshalb gut möglich, dass Verläufe lückenhaft erfasst wurden, oder dass (aus der Sicht des Forschenden) Inkonsistenzen aufgrund sich überschneidender Tätigkeiten auftreten. Aus diesem Grund wurde ein Korrekturmodul in das CATI-Instrument integriert, in welchem Inkonsistenzen aufgegriffen wurden und falls nötig eine Datenedition in Absprache mit der befragten Person vorgenommen werden konnte. Dieses Korrekturmodul enthielt drei Komponenten: Eine Überprüfung der kumulierten Anzahl Arbeitsstunden aller aktuellen Tätigkeiten²⁷, ein Prüfmodul für Überschneidungen/Parallelität von Tätigkeiten und ein Prüfmodul für Lücken²⁸. Nach jeder Veränderung der Daten (Neuerfassung von Episoden, Löschen von Episoden und Veränderung von Datumsangaben) wurden die Prüfmodule für Überschneidungen und Lücken nochmals durchlaufen. Dieser Prüfprozess dauerte so lange, bis sämtliche Inkonsistenzen aufgelöst oder durch die befragte Person bestätigt waren. Dieser iterative Prozess garantiert, dass bei der Edition der Daten im Prüfmodul nicht neue Inkonsistenzen geschaffen wurden.

Überprüfung der Anzahl aktuell geleisteter Arbeitsstunden

Falls mehrere aktuelle Tätigkeiten erfasst wurden und die Summe der geleisteten Wochenarbeitsstunden grösser als 65 war, wurde eine Frage eingeblendet, um diese Summe mit der befragten Person zu besprechen²⁹. Danach bestand die Möglichkeit, diese Angabe zu bestätigen, die Anzahl der Arbeitsstunden der einzelnen Episoden anzupassen oder Episoden zu löschen³⁰. Bei der Überprüfung der Anzahl Stunden wurden Arbeitsstunden für Erwerbstätigkeiten, Praktika und Ausbildungen berücksichtigt.

27 Die Überprüfung der aktuell geleisteten Arbeitsstunden erfolgte nach der Erfassung aller aktuellen Tätigkeiten.

28 Die Prüfmodule für Lücken und Überschneidungen wurden am Schluss des Interviews durchgeführt.

29 Die durchschnittliche Wochenarbeitszeit in der Schweiz für das Jahr 2010 beträgt 41.6 Stunden. http://www.bfs.admin.ch/bfs/portal/de/index/themen/03/02/blank/data/07.html#parsys_14124

30 Wenn Episoden gelöscht wurden, für welche bereits in früheren Erhebungswellen Angaben erfasst wurden, dann wurden die Angaben nicht effektiv aus den Daten gelöscht, sondern die Episode in einer zusätzlichen Variable als gelöscht markiert. Dadurch wird vermieden, dass valide Angaben aus früheren Erhebungswellen aufgrund von Missverständnissen nicht mehr in den Daten enthalten sind. Dieses Verfahren gilt für das Löschen von Episoden in allen Prüfmodulen.

Abbildung 5 Screenshot; Prüfmodul für die Anzahl aktuell geleisteter Arbeitsstunden

Episodentypen

Episodentyp	2008	2008	2009	2010																					
	Jan	Feb	Mar	Apr	Ma	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Jan	Feb	Mar	Apr	Ma	Jun	Jul	Aug	Sep	Okt	Nov	Dez	
Ausbildung																									
Schule																									
Betrieb																									
Praktikum																									
Erwerbstätigkeit																									
Andere																									

Stimmt so nicht, Stundenangaben müssen angepasst werden
 Stimmt so nicht, Episode muss gelöscht werden
 Stimmt so

ZURÜCK WEITER

Hinweise zum Screenshot: Die Person gab aktuell zwei Tätigkeiten an: Eine Erwerbstätigkeit als Volkswirtin und ein Praktikum als Volkswirtin. Das Prüfmodul wurde initiiert, weil die Anzahl aktuell geleisteter Arbeitsstunden grösser als 65 war. Es könnte sein, dass es sich um dieselbe Tätigkeit handelt, die irrtümlicherweise zweimal erfasst wurde.

Prüfmodul für Überschneidungen

Wenn sich Erwerbstätigkeiten während mindestens dreier Monate überschneiden, wurde der betreffende Zeitbereich durch das Prüfmodul für Überschneidungen am Bildschirm gelb hinterlegt und eine Frage wurde eingeblendet³¹. Die Situation musste zwischen den Interviewpartnern besprochen werden. In der standardisiert vorgegebenen Frageformulierung wurden die vorhandenen Informationen berücksichtigt. Die eingeblendete Frage lautete beispielsweise „Ich sehe nun, dass sich Ihre Erwerbstätigkeit als Werbefachmann und Ihre Praktikumsstelle als Journalist zwischen September 2008 und Juni 2010 überschneiden. Haben wir das richtig erfasst oder müssen wir etwas korrigieren?“ Die Interviewer/innen hatten die Möglichkeit, ergänzend zur standardisiert vorgegebenen Frage Informationen aus der graphischen Darstellung und dem Pop-up-Fenster zu berücksichtigen, um die Tätigkeiten, die sich überschneiden, genauer zu benennen. Im Prüfmodul für Überschneidungen bestanden drei Handlungsmöglichkeiten:

- Die Überschneidung von Episoden konnte als korrekte Angabe bestätigt werden.
- Datumsangaben von erfassten Episoden konnten angepasst werden.
- Episoden konnten gelöscht werden.

³¹ Ebenfalls angezeigt wurden Überschneidungen von Berufslehren und Erwerbstätigkeiten, sowie von Praktika und Erwerbstätigkeiten.

Abbildung 6 Screenshot; Prüfmodul für Überschneidungen

Episodentypen

Episodentyp	2008	2009	2010
Ausbildung			
Schule			
Betrieb			
Praktikum			
Erwerbstätigkeit			
Andere			

Ich sehe nun, dass sich [-4] Journalist und [-5] Werbefachleute vom 01.07.2008 bis zum 30.09.2010 überschneiden. Haben wir das richtig erfasst oder müssen wir etwas korrigieren?

Stimmt so nicht, Datumsangaben müssen angepasst werden
 Stimmt so nicht, Episode muss gelöscht werden
 Stimmt so

ZURÜCK WEITER

Hinweise zum Screenshot: Die Erwerbsepisode als Werbefachmann (Episodenummer -5) und die Praktikusepisode als Journalist (Episodenummer -4) überschneiden sich zwischen Juli 2008 und September 2010. Der Bereich der Überschneidung wird markiert.

Abbildung 6 zeigt einen Screenshot für die Überschneidung einer Erwerbstätigkeit und eines Praktikums.

Prüfmodul für Lücken

Wenn in einem Zeitraum von mindestens drei Monaten keine Tätigkeit erfasst war, wurde der relevante Zeitabschnitt durch das Prüfmodul für Lücken gelb hinterlegt und eine Frage wurde eingeblendet. Es gab drei Möglichkeiten, die Inkonsistenz zu beheben: Falls die befragte Person nicht angeben konnte oder wollte, welchen Tätigkeiten sie in dieser Zeit nachgegangen war, konnte eine Lückenepisode generiert werden. Ausserdem bestand die Möglichkeit, Beginn- oder Enddaten von bereits erfassten Episoden zu korrigieren und neue Episoden zu erfassen. Falls neue Episoden erfasst wurden, leitete das CATI-Programm zu den entsprechenden Fragemodulen.

4.6 Interview-Personal

Ein CATI-Instrument, welches nicht nur ein Ablaufschema mit standardisierten Fragen enthält, sondern zusätzlich eine flexible Unterstützung durch die Interviewer/innen verlangt, stellt erhöhte Anforderungen an das Interviewpersonal. Für die TREE-Erhebung wurden Interviewer/innen eingesetzt, die bereits Erfahrung in anderen wissenschaftlichen Studien gesammelt hatten und sie wurden intensiv geschult. Dabei wurde Gewicht darauf gelegt, dass die Befragter/innen die

Abbildung 7 Screenshot; Prüfmodul für Lücken

The screenshot displays a software interface for recording life events. At the top, a legend titled "Episodentypen" lists activity types: Ausbildung, Schule, Betrieb, Praktikum, Erwerbstätigkeit, and Andere. Below this is a timeline from 2008 to 2010, with months labeled. A grey bar labeled "Journalist" spans from April 2008 to October 2009. A grey bar labeled "Werftbeschneide" spans from November 2009 to December 2010. A grey bar labeled "2013 Minier" is shown in the 2008 section. Below the timeline is a text box with the following content:

Zwischen dem 01.02.2009 und dem 30.10.2009 haben wir eine Phase, wo Sie mir keine Tätigkeit angegeben haben. Was haben Sie denn in dieser Zeit gemacht?

BEFR: Vorlesen!

Eine Tätigkeit ausgeübt, über die wir bereits gesprochen haben (Daten anpassen)

oder eine Tätigkeit ausgeübt, über die wir noch nicht gesprochen haben (neue Episode erfassen)

Verweigerung / Angabe nicht möglich

ZURÜCK WEITER

Hinweise zum Screenshot: Der Bereich für die Zeit, in welcher noch keine Tätigkeit erfasst wurde (Februar 2009 bis Oktober 2009), wird markiert.

„Geschichte“ der TREE-Studie kannten, ein Aspekt, welcher bei Längsschnittstudien immer relevant ist. Bisherige Resultate der TREE-Studie wurden präsentiert, so dass sie sich ein Bild über mögliche Auswertungsprojekte machen konnten. Das CATI-Personal wurde ausserdem mit den Eigenschaften des Schweizerischen Bildungssystems und mit den verwendeten Nomenklaturen vertraut gemacht. Der grösste Teil der Schulung bestand im interaktiven Üben bei der Eingabe von vorgegebenen und frei erfundenen Lebensverläufen.

Ein standardisierter Ablauf und standardisierte Frageformulierungen wurden eingesetzt, wann immer es möglich war. Weil die Interviewer/innen trotzdem die Freiheit hatten, zusätzliche Hilfestellungen zu bieten, Rückfragen zu stellen und die Frageformulierungen zu ergänzen, musste ihnen auch die entsprechende Verantwortung bewusst gemacht werden. Innerhalb einer Episode hatten die Interviewer/innen immer die Möglichkeit, einen „zurück“-Knopf zu verwenden, um unkorrekte Angaben zu korrigieren. Zum Abschluss der Erfassung einer Episode bestand noch die Möglichkeit, einen Kommentar einzugeben. Hier konnten aussergewöhnliche Umstände beschrieben werden, die sich kaum durch standardisierte Interviewfragen abdecken lassen. Die Anwesenheit der Projektleitungen – sowohl von Seiten des TREE-Teams als auch des Erhebungsinstituts – im Telefonlabor zu Beginn der Feldphase ermöglichte einen stetigen Austausch über Unklarheiten, die während des Feldes auftauchten. Während der gesamten Feldphase sorgten Supervisor/innen durch stichprobenartiges Zuhören bei Interviews und Unterstützung bei Unklarheiten für eine hohe Qualität der Interviews.

5 Nützlichkeit des Erhebungsinstruments der TREE-Befragung 2010

In der achten TREE-Erhebungswelle wurden 3.112 CATI-Interviews geführt. Die durchschnittliche Interviewdauer betrug 22 Minuten. Insgesamt wurden 9.704 neue Episoden erfasst. Tabelle 1 bietet eine Übersicht über die Häufigkeiten der verschiedenen Episodentypen. Im Durchschnitt wurden pro Interview 3,1 neue Episoden erfasst³². Dazu kommen Episoden, welche bereits bei der letzten Befragung erhoben worden waren und zu denen die aktuellen Angaben ergänzt wurden. Pro Interview wurden somit 3,3 Episoden besprochen³³.

Die folgenden Abschnitte geben Auskunft darüber, wie oft die erarbeiteten Werkzeuge zur Anwendung kamen und somit zu einer besseren Datenqualität beigetragen haben. Dabei gilt es zu bedenken, dass nicht nur die Anzahl der Korrekturen etwas über die Nützlichkeit der Instrumente aussagt, sondern auch die Anzahl der Bestätigungen von erfassten Angaben³⁴.

5.1 Episodenanschluss mit RDI

Tabelle 2 bietet eine Übersicht der Anzahl der Anwendungen von RDI beim Episodenanschluss. Aus den Rohdaten der Erhebung ist zu erkennen, wie oft neu erfasste Erwerbstätigkeiten oder Ausbildungen automatisch mit bereits bestehenden Episoden verbunden wurden, weil alle relevanten Variablen übereinstimmen³⁵. Ein automatischer Episodenanschluss wurde bei 208 Erwerbsepisoden und 134 Ausbildungsepisoden vom System vorgenommen.

32 Wenn man zusätzlich die Betriebsepisoden und die Schulepisoden berücksichtigt, welche zusammen mit den Ausbildungen erfasst wurden (Subepisoden), sind es sogar 3,9 neue Episoden pro Interview (518 Betriebsepisoden; 1.874 Schulepisoden).

33 Bei den Episoden, die bereits vorhanden waren, für welche aber die aktuellen Angaben ergänzt wurden, handelt es sich um 476 Erwerbsepisoden und 146 Ausbildungsepisoden.

34 Bei der Ausarbeitung des Erhebungsinstruments stand das Ziel im Vordergrund, konsistente Daten zu generieren und die Befragungssituation für die Befragten und die Interviewerinnen möglichst angenehm zu gestalten. Eine quantitative Evaluation der Elemente des Tools war ursprünglich nicht vorgesehen. Aus diesem Grund sind gewisse Informationen nicht gespeichert worden und es kann nicht zu allen Vorgängen eine exakte quantitative Auswertung gemacht werden. Beispielsweise ist bekannt, in welchen Interviews Datumsangaben korrigiert wurden. Wenn in einem Interview jedoch mehrere Datumsangaben korrigiert wurden, ist deren Anzahl nicht ersichtlich.

35 Bei Erwerbstätigkeiten: Berufsbezeichnung, Beginndatum und Betrieb; Bei Ausbildungen: Bezeichnung der Ausbildung (inkl. Fachrichtung) und Beginndatum.

Tabelle 1 Anzahl erfasste Episoden

Anzahl neu erfasste Episoden (Hauptepisoden)	9.704
davon Erwerbsepisoden	3.618
davon Ausbildungsepisoden	1.693
davon Praktikumsepisoden	475
davon Episoden des Typs „Anderes“ (NEET)	3.918
Durchschnittliche Anzahl <i>neu erfasste</i> Episoden pro Interview	3,1
Durchschnittliche Anzahl <i>besprochene</i> Episoden pro Interview	3,3

Tabelle 2 Anzahl Anwendungen von RDI für den Episodenanschluss

	Automatischer Episoden- anschluss	Nachfrage; kein Anschluss	Nachfrage; Anschluss, neue Daten korrekt	Nachfrage; Anschluss, alte Daten korrekt	Anwendung von RDI für Episo- denanschluss insgesamt
Erwerb	208 35%	116 20%	170 29%	98 17%	592 100%
Ausbildung	134 54%	102 41%	4 2%	8 3%	248 100%

Bei neu erfassten Erwerbstätigkeiten wurde 384 Mal nachgefragt, ob es sich bei der genannten Tätigkeit um die gleiche handelt, welche bereits in einer früheren Welle genannt wurde. In der Mehrheit der Fälle (268) handelte es sich tatsächlich um die gleiche Erwerbstätigkeit, wobei häufiger die aktuell gemachten Angaben für korrekt befunden wurden (170). Hierbei dürfte es sich oft um Fälle handeln, in denen die Betriebsangaben in der aktuellen Erhebung aufgrund verfeinerter Suchoptionen genauer erfasst und somit verbessert werden konnten. Insgesamt kam das Modul zum Episodenanschluss anhand RDI für Erwerbstätigkeiten 592 Mal zur Anwendung und bei vier von fünf Episoden ergab dies auch einen Anschluss. Die Anzahl der Episoden, welche potentiell für eine Anwendung von Episodenanschluss anhand RDI in Frage kommen, kann nur grob abgeschätzt werden. Demnach kam RDI in rund 67% der Erwerbsepisoden mit Beginndatum vor der letzten Erhebung zur Anwendung³⁶. Dies ist ein beachtlicher Prozentsatz und es lässt sich festhalten,

36 Rund 880 Erwerbstätigkeiten, welche in t8 genannt wurden, haben ein Beginndatum, das vor April 2007 liegt. Der grösste Anteil der Interviews in der 7. Erhebungswelle fand im Februar oder März statt. Einige Befragte hatten in der 7. Welle nicht an der Befragung teilgenommen. Zudem ist es gut möglich, dass gewisse (Neben-)Tätigkeiten

dass sich der Einsatz von RDI für den Episodenanschluss von Erwerbstätigkeiten bewährt hat. Auch wenn eine hohe Anzahl Erwerbsepisoden automatisch verbunden wurden, gab es bei Erwerbstätigkeiten etwas mehr Episoden, die während des Interviews verbunden wurden, nachdem kleine Differenzen in den Angaben zwischen der befragten Person und der Interviewer/in geklärt waren.

Anders sieht es für die Ausbildungsepisoden aus³⁷. Ein automatischer Zusammenschluss kam für 134 Episoden zur Anwendung. Bei 114 Episoden wurde zwar die Frage gestellt, ob es sich um die gleiche Episode handle, die in einer früheren Welle genannt wurde, jedoch wurden die Episoden nur in wenigen Fällen auch verbunden (12 Fälle). Eine nachträgliche Betrachtung dieser Interviewdaten hat gezeigt, dass in der Mehrheit der Fälle (rund 85%) zwei Ausbildungsepisoden stehen gelassen wurden. Dabei handelt es sich grösstenteils um ein Bachelor- und ein Masterstudium des gleichen Fachs. Auch bei Personen, die aktuell ein Doktorat absolvieren, generierte das CATI-Instrument die Frage nach einer Verbindung mit dem Studium, jedoch wurde während des Interviews korrekterweise ein Enddatum für das Studium gesetzt und das Doktorat als neue Episode erfasst. Insofern ist festzuhalten, dass sich das Prüfmodul trotz der hohen Anzahl von Fällen, in denen es bei der Anschlussprüfung keine Episodenverbindung gab, bewährt hat. Es wurden im Nachhinein nur wenige Interviews identifiziert, in denen aus Sicht des Datenaufbereitungsteams ein Episodenanschluss plausibler gewesen wäre.

Während der Erhebung hat sich gezeigt, dass die Befragten kaum verärgert auf die Konfrontation mit Unstimmigkeiten zwischen den Angaben der verschiedenen Erhebungswellen reagierten.

5.2 Nützlichkeit der anderen Plausibilisierungsinstrumente

Prüfmodul für Lücken

Die Anwendung des Plausibilisierungsinstruments für Lücken gibt nicht nur Auskunft darüber, wie oft die Datenqualität dadurch verbessert werden konnte, sondern sagt auch etwas über die Antwortqualität insgesamt aus. In den 3.112 Inter-

Fortsetzung Fußnote 36

in der 7. Erhebungswelle nicht genannt wurden, obwohl sie damals schon ausgeübt wurden. Dadurch lässt sich nicht genau sagen, für wie viele Episoden der Anschluss an bestehende Daten potenziell möglich war.

37 Für Ausbildungsepisoden kann keine Aussage darüber gemacht werden, wie viele potentiell für einen Episodenanschluss in Frage kommen. Rund 1860 Ausbildungsepisoden, die in der achten Erhebungswelle besprochen wurden, haben ein Beginndatum vor April 2007. Allerdings kann nicht eruiert werden, für wie viele dieser Episoden in der achten Erhebungswelle lediglich das Enddatum erfasst wurde.

Tabelle 3 Generierte Episoden nach der Plausibilisierung von Lücken

Episodentyp	Anzahl	Dauer (Monate) der eingefügten Episode ³⁹			
		Mittelwert	Median	Maximum	Minimum
Erwerb	61	12,3	6,5	41	0,5
Ausbildung	29	8,3	5,4	37	0,5
Praktikum	13	7,2	5,5	18	2,5
Anderes:					
Militär	18	6,7	6,0	14	1,5
Zivildienst	4	6,2	6,5	9	2,5
Arbeitslosigkeit (nicht RAV gemeldet)	3	5,2	5,5	7	3,5
Arbeitslosigkeit (RAV gemeldet)	5	5,7	3,5	14	1,5
Ferien	54	5,4	3,5	39	1,5
Suche Ausbildungsplatz	1	2,5	2,5	2,5	2,5
Mutterschaft/Vaterschaft	2	18,0	18,0	21	15,0
ehrenamtliche Arbeit	6	8,2	7,5	14	2,5
Arbeit im Familienbetrieb	2	5,0	5,0	7	3,5
Krankheit	4	8,0	8,0	14	2,5
Sprachaufenthalt	17	5,3	3,5	11	1,5
Anderes	12	8,7	8,0	20	2,5
Lücke (Angabe n. mögl./ Verweigerung)	123	6,6	4,5	38	2,5
Total	354	7,6	4,5	41	0,5

views wurde 386 Mal eine Lücke von mindestens drei Monaten festgestellt, für welche keine Tätigkeit genannt wurde. Diese relativ geringe Anzahl weist auf eine gute Erinnerungs- und Antwortqualität der Befragten hin. In mindestens 32 Fällen handelte es sich um einen Irrtum bei einem Beginn- oder Enddatum, und die Lücke konnte durch eine Korrektur von Datumsangaben beseitigt werden³⁸. 354 neue Episoden wurden während des Prüfmoduls für Lücken erhoben. Tabelle 3 zeigt die Verteilung und Dauer der generierten Episoden nach der Plausibilisierung von Lücken.

In 35% der Fälle (123) wollten oder konnten die befragten Personen nicht angeben, was ihre Tätigkeiten in dieser Zeitperiode waren. Da der Mittelwert mit 6,6 Monaten etwas unter demjenigen für das Total der nachträglich eingefügten Episoden liegt, gehen wir davon aus, dass es sich in der Mehrheit der Fälle um Erinnerungsschwierigkeiten handelt. Weil aber auch für längere Phasen keine Angaben

38 Aus den Rohdaten ist ersichtlich, ob in einem Interview eine Datumsangabe angepasst wurde, jedoch nicht, ob im gleichen Interview mehrere Datumsangaben geändert wurden.

gemacht wurden, dürfte es sich bei einem Teil der Lückenepisoden um Verweigerungen der Angaben zur ausgeübten Tätigkeit handeln.

Bei den meisten Episoden, die erst im Prüfmodul für Lücken erfasst wurden, handelt es sich um Episoden des Typs „Anderes“ (128 Fälle, 36%) – insbesondere um Ferien und Sprachaufenthalte. Weil nach diesen Episodentypen während des regulären CATI-Durchlaufs nicht aktiv gefragt wurde, liegt es auf der Hand, dass sie lediglich im Prüfmodul genannt wurden.

Erfreulich ist, dass nur eine geringe Anzahl an Phasen von Arbeitslosigkeit in diesem Modul genannt wurden. Das weist darauf hin, dass die Befragten Arbeitslosigkeitsepisoden, nach denen im CATI-Fragebogen gefragt wurde, in der Regel nicht verschwiegen haben. Auch aus dem Vergleich der Arbeitslosenquote im TREE-Datensatz mit den offiziellen Statistiken, welche vom Staatssekretariat für Wirtschaft (Seco) veröffentlicht werden, lässt sich ableiten, dass Arbeitslosenphasen durchaus berichtet wurden⁴⁰.

61 Erwerbstätigkeiten – die Hälfte davon mit einer Dauer von mehr als 6,5 Monaten – wurden erst im Prüfmodul für Lücken genannt, was erstaunlich ist. Es handelt sich um Erwerbstätigkeiten mit einer relativ hohen Anzahl Wochenarbeitsstunden (Mittelwert 29,2). Die Betrachtung der ausgeübten Berufe zeigt, dass es sich in einigen wenigen Fällen um Saisonstellen oder Nebenerwerbe handeln könnte (Barman, Cateringangestellte, Skilehrerin etc.), jedoch sind es in den meisten Fällen feste Berufstätigkeiten (Krankenpfleger, Hochbauzeichnerin, Wirtschaftsprüfer etc.). Es sind demnach nicht nur unbedeutende Nebenjobs mit kleinem Arbeitspensum, die im Prüfmodul für Lücken erfasst wurden, sondern auch reguläre Erwerbstätigkeiten mit einem relativ hohen Arbeitspensum und verhältnismässig langer Dauer. Es ist anzunehmen, dass es sich nicht bei all diesen Episoden um Erinne-

39 Berechnung mit der Annahme, dass Episoden, die im gleichen Monat starten und enden, 0,5 Monate gedauert haben, Angaben gerundet, wenn die Episode aktuell war, wurde das Interviewdatum als Enddatum verwendet.

40 Offizielle Arbeitslosenquote (Seco) für 26-Jährige: 2008: 3,3%, 2009: 5,1%, 2010: 4,6 %. Arbeitslosenquote im TREE-Datensatz 2010: 4,2%. Für die Jahre 2008 und 2009 kann keine durchschnittliche Arbeitslosenquote berechnet werden. Rund 10 % aller Befragten haben im Verlauf der letzten drei Jahre mindestens eine Phase von Arbeitslosigkeit (welche bei der Regionalen Arbeitsvermittlungsstelle gemeldet wurde) durchlebt. Zu bedenken gilt es, dass ein Vergleich der beiden Quoten nur bedingt möglich ist, weil sich die TREE-Population von der Population der Wohnbevölkerung des entsprechenden Jahrgangs unterscheidet. Die TREE-Population besteht aus den Personen, welche das letzte obligatorische Schuljahr in der Schweiz absolviert haben (und in der Schweiz geblieben sind). Personen dieses Alters, die nach dem Absolvieren der obligatorischen Schulzeit zugewandert sind, sind in der TREE-Population nicht enthalten.

rungsfehler handelt, sondern dass teilweise auch unkorrekte Angaben aufgrund von Missverständnissen während des regulären CATI-Durchlaufs geklärt wurden.

Bei den nachträglich erfassten Ausbildungen handelt es sich grösstenteils um Sprachkurse, Ausbildungen des Typs „andere Ausbildung“ und Vorkurse. Allerdings wurden auch einige Ausbildungen auf Tertiärstufe erst nachträglich genannt, wobei es sich bei diesen um Ausbildungen handelt, bei welchen der Anteil an Selbststudium hoch ist (Vorbereitung auf das Anwaltspatent und Vorbereitung auf Meisterprüfung).

Insgesamt erfüllte das Prüfmodul für Lücken verschiedene Zwecke: Einerseits konnte die Anzahl Phasen identifiziert werden, zu denen die Befragten keine Angabe machen wollten oder konnten. Andererseits konnten Tätigkeiten erhoben werden, welche während des regulären CATI-Ablaufs vergessen wurden. Drittens bestand im Prüfmodul für Lücken auch die Möglichkeit, Angaben zu korrigieren und fehlerhafte Daten aufgrund von Missverständnissen während des regulären CATI-Durchlaufs zu verbessern⁴¹.

Löschen von Episoden, Korrekturen vom Datum und von geleisteten Arbeitsstunden

Während des Interviews war es möglich, in der aktuellen Befragung erfasste Episoden wieder zu löschen. Denkbar war, dass ein Episodenanschluss mit RDI nicht glückte und dass Episoden generiert wurden, die parallel zueinander verliefen und den gleichen Inhalt hatten. Auch wenn aufgrund von Missverständnissen zu viele Episoden erfasst wurden und die aktuell geleisteten Arbeitsstunden hoch waren, war eine Korrektur durch Löschen von Episoden möglich. Das Löschen von Episoden wurde also entweder durch die Plausibilisierung der aktuell geleisteten Arbeitsstunden oder durch das Korrekturmodul für Überschneidungen eingeleitet. In 96 Interviews wurde mindestens eine Episode gelöscht⁴².

Wenn diese Löschungen nicht bereits während des Interviews stattgefunden hätten, wären die zu löschenden Episoden bei der Datenkorrektur höchstwahrscheinlich schwierig zu identifizieren gewesen. Insofern dürfte sich die Daten-

41 Weil Lücken auch dann geschlossen werden, wenn es sich um Tätigkeiten handelt, welche nur wenig Zeit in Anspruch nehmen (etwa ein Nebenerwerb), ist nicht auszuschliessen, dass im Lückenmodul beispielsweise Nebenerwerbe mit einem geringen Arbeitspensum berichtet wurden, statt der eher negativ besetzten parallelen Phase von Arbeitslosigkeit. Der Vergleich mit den offiziellen Daten zur Arbeitslosigkeit deutet allerdings nicht auf diesbezügliche Probleme hin. Um zu verhindern, dass Arbeitslosigkeitsepisoden nicht berichtet werden, dürfte der konsequent modulare CATI-Ablauf wichtiger gewesen sein als das Prüfmodul für Lücken.

42 Bzw. in den Daten als „gelöscht“ markiert.

qualität durch das Löschen von Episoden während des Interviews erhöht haben. Die relativ geringe Anzahl an Löschungen weist jedoch erneut auf eine generell gute Antwortqualität hin. Die Korrektur von Datumsangaben kam etwas häufiger zum Zug. In 111 Interviews wurden bei mindestens einer Episode Datumsangaben angepasst. Stundenangaben wurden in 32 Interviews bei mindestens einer Episode angepasst.

Situationen, die durch Nachfragen bestätigt wurden

Nicht nur die Anzahl der Korrekturen, sondern auch die Anzahl bestätigter Situationen sagt etwas über die Nützlichkeit der Korrekturmodule im CATI-Instrument aus. Wenn Konstellationen, welche aus Sicht des Forschenden unplausibel erscheinen, explizit angesprochen und bestätigt werden, müssen diese bei der nachträglichen Datenaufbereitung nicht mehr in Frage gestellt werden. In 77 Interviews wurde die sehr hohe aktuell geleistete Anzahl an Wochenarbeitsstunden durch Nachfragen bestätigt. In fast 600 Fällen wurde eine Überschneidung von zwei Episoden bestätigt.

6 Fazit

Das Erhebungsinstrument für die achte TREE-Erhebungswelle hat sich insgesamt als nützlich erwiesen. Ein Episodenanschluss an die Daten aus früheren Erhebungswellen konnte mit RDI insbesondere für Erwerbsepisoden erfolgreich umgesetzt werden. Erinnerungshilfen, die durch die Kombination eines modularen Vorgehens und einer flexiblen Unterstützung der Interviewer/innen geleistet wurden, scheinen erfolgreich gewesen zu sein. Dies lässt sich aus der relativ geringen Anzahl von Lücken und Korrekturen ablesen. Auch die beachtliche Anzahl von bestätigten Situationen weist darauf hin, dass der grundsätzliche Aufbau des CATI-Instruments eine Erfassung von konsistenten Angaben erleichtert. Durch den Einsatz der Plausibilisierungswerkzeuge konnte die Datenqualität verbessert werden. Für die nachträgliche Datenedition ist es hilfreich zu wissen, dass gewisse Plausibilisierungen nicht nötig sind, weil sie bereits während der Erhebung thematisiert wurden. Abschliessend ist zu vermerken, dass die Interviewer/innen das Instrument als benutzerfreundlich und hilfreich beurteilt haben.

Literatur

- Barsalou, L., 1988: The Content and Organization of Autobiographical Memories. S. 193-243 in: U. Neisser und E. Winograd (Hg.): *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*. Cambridge: Cambridge University Press.
- Belli, R. F., E. H. Lee, F. P. Stafford und C-H. Chou, 2004: Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics* 20 (2): 185-218.
- Bluck, S. und T. Habermas, 2000: The life story schema. *Motivation & Emotion* 25 (2): 121-147.
- Buchholz, S., D. Hofacker, M. Mills, H-P. Blossfeld, K. Kurz und H. Hofmeister, 2009: Life Courses in the Globalization Process: The Development of Social Inequalities in Modern Societies. *European Sociological Review* 25: 53-71.
- Conway, M., 1996: Autobiographical Knowledge and Autobiographical Memories. S. 67-93 in: D. Rubin (Hg.): *Remembering our Past: Studies in Autobiographical Memory*. Cambridge/ England: Cambridge University Press.
- Conway, M. und C. Pleydell-Pearce, 2000: The Construction of Autobiographical Memories in the Self Memory System. *Psychological Review* 107: 261-288.
- Drasch, K. und B. Matthes, 2009: Improving Retrospective Life Course Data by Combining Modularized Self-reports and Event History Calendars. Experiences from a Large Scale Survey. IAB Discussion Paper 21/2009. Nürnberg: IAB. <http://doku.iab.de/discussionpapers/2009/dp2109.pdf> (05.08.2011).
- Glaser, T. und W. van der Vaart, 2009: Applications of calendar instruments in social surveys: A review. *Quality and Quantity* 43 (3): 333-349.
- Dürnberger, A., K. Drasch, und B. Matthes, 2011: Kontextgestützte Abfrage in Retrospektiv-erhebungen. *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 5 (1): 3-35.
- Hale, A. und S. Michaud, 1995: *Dependent Interviewing: Impact on Recall and on Labour Market Transitions*. SLID Research Paper Series. Ottawa, Canada: Statistics Canada.
- Huttenlocher, J., L. V. Hedges und N. M. Bradburn, 1990: Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology, Learning, Memory and Cognition* 16: 196-213.
- Huttenlocher, J., L. V. Hedges und V. Prohaska, 1988: Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review* 95: 471-484.
- Jäckle, A. und P. Lynn, 2007: Dependent Interviewing and Seam Effects in Work History Data. *Journal of Official Statistics* 23 (4): 529-51.
- Keller, A., S. Hupka-Brunner und T. Meyer, 2010: *Nachobligatorische Ausbildungsverläufe in der Schweiz: Die ersten sieben Jahre. Ergebnisübersicht des Jugendlängsschnitts TREE, Update 2010*. Basel: TREE.
- Larsen, S. F. und C. P. Thompson, 1995: Reconstructive memory in the dating of personal and public events. *Memory and Cognition* 23: 780-790.
- Loftus, E. F. und W. Marburger, 1983: Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory and Cognition* 11: 114-120.
- Lynn, P. und E. Sala, 2006: Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing. *International Journal of Public Opinion Research* 18 (4): 500-509.

- Mathiowetz, N. A. und K. A. McGonagle, 2000: An Assessment of the Current State of Dependent Interviewing in Household Surveys. *Journal of Official Statistics* 16 (4):401-18.
- Matthes, B., M. Reimer und R. Künster, 2007: Techniken und Werkzeuge zur Unterstützung der Erinnerungsarbeit bei der computergestützten Erhebung retrospektiver Längsschnittdaten. *Methoden – Daten – Analysen. Zeitschrift für empirische Sozialforschung* 1 (1): 69-92.
- OECD (Hg.) 2009: *Education at a Glance 2009*. Paris: OECD.
- Rubin, D.C., und A.Baddeley, 1989: Telescoping is not time compression: A model of the dating of autobiographical events. *Memory and Cognition* 77 (6): 653-661.
- TREE (Hg.) 2008: *Projekt-Dokumentation 2000-2008*. Bern/Basel: TREE.
- Reimer, M., 2001: Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Lebensverlaufsdaten: Kognitive und erhebungspragmatische Aspekte. Berlin: Max-Planck-Institut für Bildungsforschung.
- Reimer, M., 2005: *Autobiografisches Gedächtnis und retrospektive Datenerhebung: Die Rekonstruktion und Validität von Lebensverläufen (Studien und Berichte 70)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Reimer, M. und B. Matthes, 2006: Collecting Event Histories with TrueTales: Techniques to Improve Autobiographical Recall Problems in Standardized Interviews. *Quality & Quantity* 41: 711-735.
- Solga, H., 2001: Longitudinal surveys and the study of occupational mobility: panel and retrospective design in comparison. *Quality and Quantity* 35: 291-309.
- Uhrig, N. und E. Sala, 2011: When Change Matters: An Analysis of Survey Interaction in Dependent Interviewing on the British Household Panel Study. *Sociological Methods & Research* 40 (2): 333-366.
- Van der Vaart, W., 2004: The time-line as a device to enhance recall in standardized research interviews: A split ballot study. *Journal of Official Statistics* 20 (2): 301-317.
- Wagenaar, W. A., 1986: My memory: A study of autobiographical memory over six years. *Cognitive Psychology* 18: 225-252.

Anschrift des Autors

Melania Rudin
Büro für arbeits- und sozialpolitische Studien (BASS)
Konsumstrasse 20
3007 Bern
Schweiz
melania.rudin@buerobass.ch

Christoph Müller
M.I.S. Trend S.A.
Pont-Bessières 3
1005 Lausanne
Schweiz
c.mueller@mistrend.ch

Autorinnen und Autoren Jahrgang 7 (2013)

- Tobias Bachteler, Duisburg
- Anna Baumert, Landau
- Inna Becher, Nürnberg
- Constanze Beierlein, Mannheim
- Michael Blohm, Mannheim
- Heidi Bruderer Enzler, Zürich
- Claudia Diehl, Konstanz
- Patrick Fick, Konstanz
- David Glowsky, Berlin
- Tobias Gramlich, Duisburg
- Rolf Hackenbroch, Mannheim
- Julia Haubrich, Hagen
- Timo Heydasch, Hagen
- Katharina Hörstermann, Köln
- Julia Jerke, Leipzig
- Christoph J. Kemper, Mainz
- Mira Celine Klein, Mannheim
- Achim Koch, Mannheim
- Anastassiya Kovaleva, Bielefeld
- Ivar Krumpal, Leipzig
- Johannes Lutz, Potsdam
- Jutta Margraf-Stiksrud, Mannheim
- Martin Messingschlager, Bamberg
- Reto Meyer, Zürich
- Christoph Müller, Lausanne
- Marcel Preising, Bamberg
- Beatrice Rammstedt, Mannheim
- Jörg Reiher, Duisburg
- Karl-Heinz Renner, Hagen
- Melania Rudin, Basel
- Stefan Schipolowski, Berlin
- Alexander Schmidt-Catran, Köln
- Manfred Schmitt, Landau
- Rainer Schnell, Duisburg
- Ulrich Schroeders, Berlin
- Jürgen Schupp, Berlin
- Menno Smid, Bonn
- Mark Trappmann, Bamberg
- Oliver Wilhelm, Ulm

Rezensentinnen und Rezensenten Jahrgang 7 (2013)

- Volker Hüfken, Düsseldorf
- Ralf Schultheiß, Gießen
- Cornelia Züll, Mannheim

Gutachterinnen und Gutachter Jahrgang 7 (2013)

Wir danken folgenden Kolleginnen und Kollegen, die für den Jahrgang 7 (2013) der methoden, daten, analysen Manuskripte begutachtet haben:

- Mitja Back, Münster
- Annelies Blom, Mannheim
- Michael Bosnjak, Mannheim
- Michael Braun, Mannheim
- Evelyn Bytzek, Landau
- Claus-Christian Carbon, Bamberg
- Daniel Danner, Mannheim
- Stefanie Eifler, Eichstätt
- Tobias Enderle, Mannheim
- Siegfried Gabler, Mannheim
- Tanja Gerlach, Darmstadt
- Anja Göritz, Freiburg
- Andrea Hildebrandt, Berlin
- Dominique Joye, Lausanne
- Martin Kersting, Giessen
- Oliver Lipps, Lausanne
- Jochen Mayerl, Kaiserslautern
- Alexandra Nonnenmacher, Siegen
- Rainer Schnell, Duisburg
- Manfred Schmitt, Landau
- Stefan Schmukle, Leipzig
- Andreas Schwerdtfeger, Graz
- Martin Spieß, Hamburg
- Matthias Ziegler, Berlin

Information for Authors

Methods, data, analyses (mda) publishes research on all questions important to quantitative methods, with a special emphasis on survey methodology. In spite of this focus we welcome contributions on other methodological aspects.

Manuscripts that have already been published elsewhere or are simultaneously submitted to other journals will not be considered. As a rule we do not restrict authors' rights. All rights remain with the author, and articles in mda are published under the CC-BY open-access license.

Mda aims for a quick peer-review process. All papers submitted to mda will first be screened by the editors for general suitability and then double-blindly reviewed by at least two reviewers. The decision on publication is made by the editors based on the reviews. The editorial team will contact the authors by email with the result at the latest eight weeks after submission; if the reviews have not been received by then, we provide a status update with a new target date.

When preparing a paper for submission, please consider the following guidelines:

- Please submit your manuscript by e-mail to [mda\(at\)GESIS\(dot\)org](mailto:mda(at)GESIS(dot)org).
- The total length of the manuscript shall not exceed 10.000 words.
- Manuscripts should...
 - be written in English, using American English spelling. Please use correct grammar and punctuation. Non-native English speakers should consider a professional language editing prior to publication.
 - be typed in a 12 pt Roman font, double-spaced throughout.
 - start with a cover page containing the title of the paper and contact details / affiliations of the authors, but be anonymized for review otherwise.
- Please also send us an abstract of your paper (approx. 300 words), a brief biographical note (no longer than 250 words), and a list of 5-7 keywords for your paper.
- Acceptable formats for Graphics are
 - Tiff
 - Jpeg (uncompressed, high quality)
 - EMF/WMF
- Please ensure a resolution of at least 300 dpi and take care to send high-quality graphics. Line art images should have a resolution of 500-1000 dpi. Please note that we cannot print color images.
- The type area of our journal is 11.5 cm (width) x 18.5 cm (height). Please consider this when producing tables or graphics.
- Footnotes should be used sparingly.
- By submitting a paper to mda the authors agree to make data and program routines available for purposes of replication.

Please follow the APA guidelines when preparing in-text references and the list of references.

Entire Book:

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Journal Article (with DOI):

Klimoski, R., & Palmer, S. (1993). The ADA and the hiring process in organizations. *Consulting Psychology Journal: Practice and Research*, 45(2), 10-36. doi:10.1037/1061-4087.45.2.10

Journal Article (without DOI):

Abraham, K. G., Helms, S., & Presser, S. (2009). How social processes distort measurement: The impact of survey nonresponse on estimates of volunteer work in the United States. *American Journal of Sociology*, 114(4), 1129-1165.

Chapter in an Edited Book:

Dixon, J., & Tucker, C. (2010). Survey nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research*. Second Edition (pp. 593-630). Bingley: Emerald.

Internet Source (without DOI):

Lewis, O., & Redish, L. (2011). Native American tribes of Wisconsin. Retrieved April 19, 2012, from the Native Languages of the Americas website: www.native-languages.org/wisconsin.htm

For more information, please consult the *Publication Manual of the American Psychological Association* (Sixth ed.).