

Mai 2001

48

# NACHRICHTEN

GESIS

ZUMA

## Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)* ([www.wgl.de](http://www.wgl.de)).

**Vorsitzender des Trägervereins ZUMA e.V.:** Prof. Dr. Max Kaase

**Direktor:** Prof. Dr. Peter Ph. Mohler

### Hausanschrift

B 2, 1  
68 159 Mannheim

### Postanschrift

Postfach 12 21 55  
68 072 Mannheim

**Telefon** 0621/1246 - 0

**Fax** 0621/1246 - 100

**Redaktion:** 0621/1246 - 268

**E-Mail** [luettinger@zuma-mannheim.de](mailto:luettinger@zuma-mannheim.de)

**Internet** [www.gesis.org/zuma](http://www.gesis.org/zuma) ZUMA  
[www.gesis.org/](http://www.gesis.org/) GESIS

**Redaktion:** Dr. Paul Lüttinger

**ISSN 0941-1670** 25. Jahrgang

Die ZUMA-Nachrichten informieren über Forschungsarbeiten und Dienstleistungen des Instituts. Beiträge externer Autoren werden dann aufgenommen, wenn sie aus Kooperationsprojekten mit ZUMA resultieren oder von besonderem wissenschaftlichen Interesse für die Arbeiten im Hause und die Empirische Sozialforschung sind. Alle Beiträge unterliegen einem internen förmlichen Begutachtungsverfahren durch jeweils zwei Gutachter. Der Herausgeber entscheidet auf der Grundlage der Gutachten über die Annahme eines Beitrages. Hinweise zur Manuskriptgestaltung sind auf der letzten Seite des Heftes.

Namentlich gekennzeichnete Beiträge geben die Meinung der Autoren wieder. Der Nachdruck von Beiträgen ist nach Absprache möglich. Die ZUMA-Nachrichten erscheinen im Mai und November eines Jahres. Sie werden Interessenten auf Anforderung kostenlos zugesandt. Die ZUMA-Nachrichten finden Sie auch im WWW: [http://www.gesis.org/publikationen/zeitschriften/zuma\\_nachrichten/](http://www.gesis.org/publikationen/zeitschriften/zuma_nachrichten/).

**Druck:** Verlag Pfälzische Post GmbH, Neustadt/Weinstraße.

Gedruckt auf chlorfrei gebleichtem Papier.

© ZUMA

In eigener Sache .....	5
------------------------	---

## **FORSCHUNGSBERICHTE**

Participation in Web Surveys - A Typology. <i>Michael Bosnjak, Tracy L. Tuten &amp; Wolfgang Bandilla</i> .....	7
--	---

Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. <i>Andreas Diekmann &amp; Ben Jann</i> .....	18
--	----

Die statistische Aufdeckung kausalstrukturell unterschiedener Gruppen am Beispiel politischen Teilnahmeverhaltens. <i>Frank Faulbaum, Petra Stein &amp; Kai Kelleter</i> .....	28
--	----

Why it is Easy to Write Bad Questions. <i>Floyd Jackson Fowler Jr</i> .....	49
--	----

Stichprobenrealisierung bei einer bundesweiten telefonischen Befragung alter Menschen. <i>Olaf v.d. Knesebeck, Volker Hüfken &amp; Peter Dübbert</i> .....	67
--	----

Die Berechnung der Varianz von Populationsschätzern im Scientific Use File des Mikrozensus ab 1996. <i>Ulrich Rendtel &amp; Bernhard Schimpl-Neimanns</i> .....	85
---	----

## **MITTEILUNGEN**

Kompatibilität von ISCO-68, ISCO-88 und KldB-92. <i>Alofns Geis &amp; Jürgen H.P. Hoffmeyer-Zlotnik</i> .....	117
--	-----

Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. <i>Rainer Frietsch &amp; Heike Wirth</i> .....	139
--	-----

Daten des ALLBUS 2000 verfügbar.....	164
--------------------------------------	-----

Forschungspreis der Popper-Nährpflichtstiftung für ZUMA-Dissertation.....	170
---	-----

Aktivitäten im Forschungsbereich Ernährungsverhalten. <i>Erika Claupein, Kristina Nebel &amp; Ulrich Oltersdorf</i> .....	173
--	-----

Verbraucherpaneldaten für die empirische Sozial- und Wirtschaftsforschung.....	179
--	-----

Zur Erinnerung an Aage B. Sørensen.....	187
---	-----

## BUCHBESPRECHUNGEN

J. MERZ/M. EHLING (EDS.): Time Use – Research, Data and Policy .....	191
A. GODENZI/R. STADLER: Statistik. Interaktive Lernsoftware zur beschreibenden und schließenden Statistik .....	195
J. H.Z. HOFFMEYER-ZLOTNIK: Regionalisierung sozialwissenschaftlicher Umfragedaten. Siedlungsstruktur und Wohnquartier .....	198

## PUBLIKATIONEN

ZUMA-Arbeitsberichte .....	201
ZUMA-Methodenberichte .....	203
HOW-TO-Reihe .....	204
Buchhinweise .....	205

## ZUMA-WORKSHOPS, ZWEITES HALBJAHR 2001

Simulation for the Social Sciences, 27.-31. August.....	210
Grundlagen der Fragebogenkonstruktion, 25.-26. September .....	211
Eine Einführung in das Arbeiten mit Amos, 23.-24. Oktober.....	212
Using (Multidimensional) Rasch Models, 25. Oktober .....	212
Armut- und Reichtumsberichterstattung, 8.-9. November.....	213
Einführung in die computerunterstützte Inhaltsanalyse mit TEXTPACK 20.-21. November .....	214
Angewandte Multiattributive Discret-Choice-Analyse, 29.-30. November .....	215
Translation Techniques, 6.-7. Dezember.....	215
Pretesting, 11. Dezember.....	216

## IN EIGENER SACHE

In der Umfrageforschung unterscheidet man beim Antwortverhalten von Befragten zwischen Personen, die an der Befragung teilnehmen und alle Fragen vollständig beantworten, Personen, die zwar an der Befragung teilnehmen, aber einzelne Fragen unbeantwortet lassen (Item nonresponse) und Personen, die das Interview vollständig verweigern (Unit nonresponse). Bei Internetbasierten Befragungen können Daten über den Befragungsprozess miterhoben werden, die einen detaillierteren Einblick in das Geschehen bei der Beantwortung von Fragen erlauben. MICHAEL BOSNJAK und WOLFGANG BANDILLA unterscheiden in ihrem Beitrag sieben Nicht-Antwortmuster und diskutieren deren theoretische und praktische Implikationen.

ANDREAS DIEKMANN und BEN JANN gehen der Frage nach, in welchem Ausmaß Geschenke bei einer postalischen Befragung zur Erhöhung der Ausschöpfungsquote beitragen.

Ein Verfahren zur Aufdeckung unbekannter Gruppierungen unter Einbeziehung gruppenspezifischer Kausalstrukturen steht im Mittelpunkt des Beitrages von FRANK FAULBAUM, PETRA STEIN und KAI KELLETER. Mit dem beschriebenen Verfahren kann insbesondere überprüft werden, ob eine Population hinsichtlich der Konstruktvalidität theoretisch interessanter Variablen homogen ist.

FLOYD JACKSON FOWLER JR. diskutiert das Problem der Formulierung von Fragen. Er unterscheidet acht Eigenschaften, die Fragen haben sollten, und diskutiert die Implikationen, die sich daraus für die Formulierung von Fragen ergeben.

Die Stichprobenrealisierung bei einer bundesweiten telefonischen Befragung von in Privathaushalten lebenden alten Menschen beschreiben OLAF V.D. KNESEBECK, VOLKER HÜFKEN und PETER DÜBBERT. Besonderes Augenmerk richten die Autoren darauf, welche Auswirkungen Proxy-Interviews (d.h. Interviews, bei der stellvertretend für den zu Befragenden eine andere Person Auskunft gibt) und Konvertierungsmaßnahmen auf die Zusammensetzung der Stichprobe haben.

Im Scientific Use File des Mikrozensus von 1996 (dabei handelt es sich um Daten, die in anonymisierter Form von den Statistischen Ämtern für die Forschung zur Verfügung gestellt werden) wurden erstmals Informationen über die Stichprobe bereitgestellt, die eine Berechnung der Varianz von Populationsschätzern ermöglichen.

ULRICH RENDTEL und BERNHARD SCHIMPL-NEIMANNs entwickeln ein methodisches Konzept zur Berechnung der Varianz und wenden dies auf die Schätzung von Randverteilungen, Verhältnis- und Mittelwerten an. Sie vergleichen Varianzschätzungen auf der Basis des Scientific Use Files mit den Ergebnissen des Statistischen Bundesamtes.

Max Kaase

Vorsitzender des ZUMA e.V.

# **PARTICIPATION IN WEB SURVEYS A TYPOLOGY<sup>1</sup>**

*MICHAEL BOSNJAK, TRACY L. TUTEN & WOLFGANG BANDILLA*

While traditional survey literature has addressed three possible (non-)response patterns (unit nonresponse, item nonresponse, and complete response), Web surveys can capture data about a respondent's answering process, enabling researchers to attain more fine-grained information about individual reactions to such surveys. Based on this information, at least seven (non-)response patterns can be distinguished. This paper describes these seven patterns in a typology of (non-)response. Finally, theoretical and practical implications are discussed.

Während man in der Umfrageforschung im Wesentlichen von drei (Nicht-)Antwortmustern bei Befragungen ausgeht (Unit nonresponse, Item nonresponse und vollständige Befragungsteilnahme), können bei Web-basierten Befragungen (Web surveys) Daten über den Befragungsprozess miterhoben werden, die einen erweiterten Einblick in das tatsächliche Geschehen bei der Beantwortung von Fragen geben. Aufbauend auf diesen Bearbeitungsprozessdaten lassen sich mindestens sieben (Nicht-)Antwortmuster voneinander unterscheiden, die im Rahmen dieses Beitrages beschrieben sowie hinsichtlich ihrer theoretischen und praktischen Implikationen diskutiert werden.

## **Introduction**

Surveys are generally characterized by the fact that data may be missing for some units of a sample, either partially, or for all variables. This problem of missing data is generally known as 'Nonresponse', whereby one usually differentiates between unit and item nonresponse (Groves/Couper 1998). Unit nonresponse refers to the complete loss of

---

<sup>1</sup> An earlier version of this paper has been presented at the Fifth International Conference on Social Science Methodology in Cologne, Germany, October 4<sup>th</sup>, 2000. We thank Don A. Dillman and the two anonymous reviewers for valuable comments on an earlier draft of this article. The support of the Center for Survey Research and Methodology (ZUMA) in Mannheim, Germany, during the development and completion of this research and manuscript is gratefully acknowledged.

a survey unit, while item nonresponse refers to missing responses to individual questions. Past work has assumed the existence of three possible responses to requests for survey participation: complete response, item nonresponse, and unit nonresponse. This is the case for all modes of survey data collection, whether it be in-person (e.g., Groves/Cialdini/Couper 1992), by telephone (e.g., Frey 1976; Oksenberg/Coleman/Cannell 1986), mail (e.g., Armstrong/Overton 1977; Yu/Cooper 1983), or via the Web (e.g., Tuten/Urban/Bosnjak 2001).

With the exception of Web-based surveys, this has been necessarily so, since the process by which a sample member views and answers a self-administered questionnaire has been, for the most part, a black box. However, in Web surveys, the response process can be traced automatically. Such 'para' or 'meta-data' about the answering process can provide insight into the sequencing and completeness of responses. Such data suggest an alternative way of conceptualizing patterns of reacting to a survey which encompasses at least seven possible response types to requests for survey participation. We introduce this typology of response behaviors to more comprehensively describe the potential variations in participation possible in most Web-based surveys. We begin with a brief review of the literature on response behaviors in Web surveys, followed by a description of the response typology. Subsequently, we provide an illustration of the typology proposed. Finally, theoretical as well as practical implications are briefly discussed.

## **Response Behaviors in Web Surveys**

Findings related to nonresponse in Web surveys are relatively limited. Literature primarily reveals work concerned with explaining (1) volitionally-controlled drop-out behaviors (in this case, the response process is prematurely terminated) and (2) item nonresponse.

Based on a summary of nine Web surveys, Knapp and Heidingsfelder (2001) showed that increased drop-out rates can be expected when using open-ended questions or questions arranged in tables. Dillman et al. (1998) recommended avoiding graphically-complex or 'fancy' design options. They compared fancy versus plain designs and found higher quit rates when fancy designs were used. This is likely due to the corresponding increase in download time for pages with complex designs.

Dillman (2000) warned of commonly-used techniques in Web surveys which may alienate respondents who are uncomfortable with the Web. The use of pull-down



menus, unclear instructions on how to fill out the questionnaire, and the absence of navigational aids may encourage novice Web-users to break off the survey process.

Frick, Bächtiger and Reips (2001) conducted an experiment on the effect of incentives on response. They concluded that the chance to win prizes in a lottery resulted in lower drop-out rates than in those conditions where no prize draw entry was offered as an incentive. Of particular interest in this context are the opposing findings of an experimental study by Tuten, Bosnjak and Bandilla (2000) which found that the share of unit nonresponders is significantly higher when offering the chance to win a prize, than in cases where altruistic motives for participation are addressed (contribution to scientific research).

Frick, Bächtiger and Reips (2001) also investigated the effect of the order of topics on the amount of drop-outs in a Web survey. In one condition, personal details were requested at the beginning of the investigation (socio-demographic data and e-mail address). In the other condition, these items were positioned at the end of the questionnaire. Surprisingly, drop-outs were significantly lower in the first condition (10.3% versus 17.5%). In other words, when personal data were requested at the beginning, fewer drop-outs occurred.

Bowker and Dillman (2000) conducted an experiment on the effect of the physical layout of answers on the rate of item nonresponse. The results showed a significantly lower item nonresponse rate where response alternatives were right aligned compared to a frequently applied left-aligned format. As a possible explanation for this difference, fewer 'back and forth' eye movements as well as a decreased number of hand movements associated with completing the questionnaire could have resulted in a lower number of items being missed by some respondents.

A conceptually different approach for the explanation of item nonresponse in Web surveys has been proposed by Bosnjak (2001). In his view, nonresponse in Web surveys is to some extent explainable due to the fact that respondents are not willing to respond to all of the questions presented. As an explanatory approach for this 'noncompliant' behavior, he uses Ajzen's theory of planned behavior (Ajzen 1985, Ajzen 1991). The results reported indicate that the theory of planned behavior represents a suitable starting point for predicting the extent of missing data, respectively item nonresponse in open-ended questions, where significant predictions of a medium effect-size could have been demonstrated. In the case of closed-ended questions, however, the theory used seems to have little or no predictive power.

In summary, methods research with regard to Web surveys has focused primarily on detecting some determinants of drop-out and item nonresponse behavior. Certainly Web survey methodology is still in an infancy stage. However, the additional information provided when using the Web to collect data (e.g., automatically-generated log files, visitor tracing programs etc.) can provide a valuable insight into understanding nonresponse and response behaviors. It is no longer necessary to view responses to survey requests within the confines of three conceptually distinct (non)response types. The typology of response behaviors proposed hereafter serves as a descriptive model for operationalizing specific behaviors. This, thereby, ensures increased comparability of individual findings.

## **A Typology of Nonresponse Patterns in Web Surveys**

In traditional mail surveys<sup>2</sup>, the response process basically remains a mystery. We do not know whether a potential respondent received the questionnaire at all, read it, and began answering it. Such information can hardly be reconstructed afterwards without the aid of a follow-up study. Given this lack of information about the participation process, a survey researcher loses valuable information. If an individual does not return the questionnaire, was it a genuine refusal (i.e., volitionally-controlled) or was it something else (e.g. a technical artifact) to blame? In both cases, the questionnaire is simply categorized as one with unit nonresponse. If a questionnaire is returned incomplete, we do not know whether the participant chose not to answer the remaining questions purposefully, or if he or she merely dropped-out of the process. In either case, the questionnaire is categorized as one with item nonresponse.

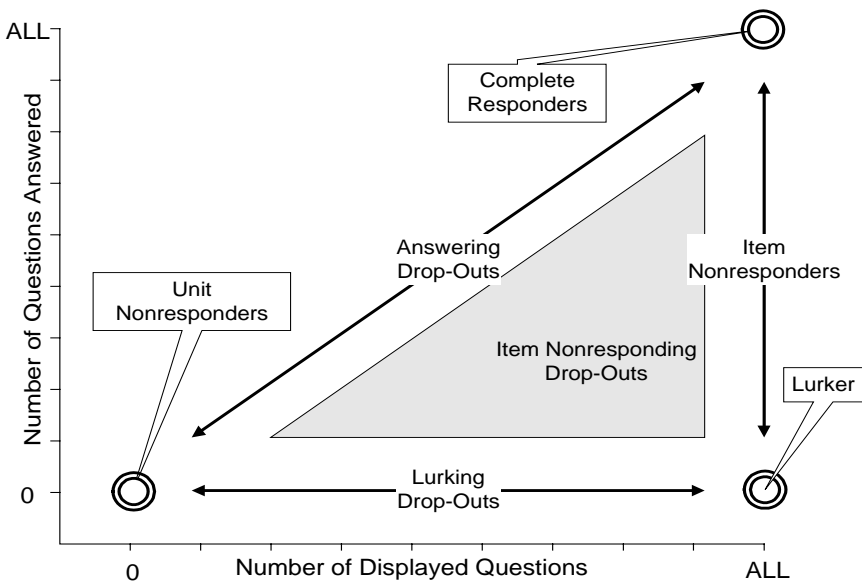
One of the substantial advantages of Web surveys, in comparison to mail surveys, is that they can supply para-data, or meta-data, in addition to responses to the substantive questions. There are several methods possible to trace the response process including the use of cgi scripts, java applets, and log files. Regardless of the specific approach used, the data allow the reconstruction of the response process (Batinic/ Bosnjak 1997). In order to log these individual response patterns completely, the following three conditions must be fulfilled: (1) each question must be displayed separately

---

**2** Due to the commonly accepted notion (see e.g., Bandilla/Bosnjak/Schneid/Stiegler 1999) mail surveys as well as Web surveys belong to the so-called 'self-administered surveys'. Therefore, we are focusing on the possibilities to trace response patterns within this class of surveys and will not elaborate on the possibilities offered by interviewer-administered surveys, like e.g. computer-assisted telephone interviews (CATT) or computer-assisted personal interviews (CAPI).

(screen-by-screen design), (2) the participants are not forced to provide an answer before being allowed to move on (non-restricted design), and (3) each page of the questionnaire must be downloaded separately from the server, and should not be allowed to reside in the Web browser's cache (cache passing pages)<sup>3</sup>. If these conditions are fulfilled, the data set containing information on the user's activities can be used to analyze the completeness and the sequence in which the questions have been processed by the respondents. Figure 1 illustrates the different response patterns.

**Figure 1: (Non-)Response Types in Web Surveys**



In Figure 1, the number of separately displayed questions (abscissa) is set in relation to the number of questions actually answered (ordinate). This graphical representation of observable response patterns allows for a differentiation between the fol-

<sup>3</sup> Various technical implementation methods are available, such as script-based downloading of pages, or integrating specific META tags. The precise technical procedures will not be elaborated upon here.

lowing seven processing types: 1) Complete responders, 2) Unit nonresponders, 3) Answering drop-outs, 4) Lurkers, 5) Lurking drop-outs, 6) Item nonresponders, and 7) Item non-responding drop-outs. Each pattern is described below.

Complete responders are those respondents who view all questions and answer all questions. Unit nonresponders are those individuals who do not participate in the survey. There are two possible variations to the unit nonresponder. Such an individual could be technically-hindered from participation, or he or she may purposefully withdraw after the welcome screen is displayed, but prior to viewing any questions. Answering drop-outs consist of individuals who provide answers to those questions displayed, but quit prior to completing the survey. Lurkers view all of the questions in the survey, but do not answer any of the questions. Lurking drop-outs share some characteristics with 'answering drop-outs' and 'lurkers'. Such participants view some of the questions without answering, but also quit the survey prior to reaching the end. Item nonresponders view the entire questionnaire, but only answer some of the questions. Item non-responding drop-outs represent a mixture of 'answering drop-outs' and 'item nonresponders'. Individuals displaying this response behavior view some of the questions, answer some but not all of the questions viewed, and also quit prior to the end of the survey. In our opinion, this typology of response patterns is a more accurate depiction of actual events in Web surveys than the relatively basic categorization of complete participation, unit non-response, or item nonresponse.

Using the traditional categorization of possible response behaviors, some behaviors would be mistakenly categorized. Specifically, lurkers and lurking drop-outs would be classified as unit nonresponders. Answering drop-outs and item non-responding drop-outs would be classified the same as item nonresponders. Only 'complete responders' remain unaffected by the classification system used.

The variations among the seven (non)response types represent significant differences, particularly when one seeks to understand and possibly change response behaviors: unit nonresponders are commonly thought of as people who refused to answer or are hindered from answering. Lurkers and lurking drop-outs, however, are able to respond and are interested enough in the topic to peruse the questions. Yet, they refuse to answer. Lurkers show enough interest to view all questions. Lurking drop-outs either experience technical difficulties in continuing to view the survey or lose interest during the survey, and so do not view all of the questions. Item nonresponders are commonly thought of as people who were not comfortable answering certain questions but otherwise completed the survey. They may have felt a question

was too personal. In other words, we do not tend to assume that item nonresponders lack motivation to respond, but rather that the question(s) influenced their response, or lack thereof. Answering drop-outs, however, begin the survey process much like a complete responder but they drop out prior to completion. These participants may drop-out due to technical difficulties or because they purposefully decide to drop-out. Item non-responding drop-outs begin the survey process like item nonresponders but also quit prior to the end of the survey. This responder type may be more similar to a unit nonresponder than to an item nonresponder.

In all forms of nonresponse and drop-out is always the possibility of both volitional and non-volitional behaviors. With volitionally-controlled, or intentional non-response types, the (potential) respondent decides for him or herself to which extent he or she will or will not participate in a survey. Non-volitional nonresponse is caused by technical artifacts, or other external obstacles. For example, an internet session may be disconnected, or a specific net-use pattern may prevent recognition of a request for participation (e.g., Lukawetz 2001). In principle, these two classes of causes must be taken into consideration as an explanation in all drop-out types, as well as for unit nonresponse.

In complete responders, lurkers and item nonresponders, one can assume that all actions are volitionally-controlled due to the evidence that the participants view all questions in the survey. The only possible exception here are the lurkers. While lurkers are not hindered by technical artifacts, it is also possible that lurkers are not always part of the desired sample. Specifically, one must note that it is not necessarily only humans who access, download, and fail to answer the appropriate questionnaire pages, but possibly robots, worms, or wanderers. These non-human, more or less 'intelligent' (software-based) agents are e.g. used by search engines to find and compile information. Technical procedures for excluding the majority of these robots may exist, yet complete exclusion is not possible due to the diversity of different systems. Dillman (2000) suggested restricting access to people with a valid PIN number in order to reduce coverage error. This technique may be useful for excluding robots.

## **An Illustration**

A Web-based survey was conducted on the topic of "the roles of men and women in family and work life." The survey questions were arranged according to the design guidelines described above for the identification of different response patterns: (1) each question was displayed separately, (2) participants were not forced to pro-

vide answers before allowed to move on, and (3) each page of the Web questionnaire was protected from being cached. Because our goal was to investigate response patterns, no incentive for participation was offered.

Web users were 'invited' to the survey through advertising placed on search engines and Web catalogs (e.g., Yahoo, Altavista, etc.). In total, 1469 people participated in the study. Of those answering demographic questions, 35.4% were male and 64.6% female. The mean age in this group was 27.6 years (SD = 8.4 years) and most of the participants were employed (46.5%) or students (34.8%). It is important to note, though, that not all participants are represented in the demographic descriptions. For instance, lurkers viewed the questions, but did not answer them.

Participants were classified into the appropriate segments of the seven-fold typology by analyzing data from both the automatically-generated log file and data set. Specifically, we tracked the questions viewed and answered for each participant. As anticipated, seven specific response types were identifiable.

In this study, 25.3% of the participants were complete responders, 10.2% were unit nonresponders, and 4.3% were answering drop-outs. 6.9% of the respondents were lurkers while 13.3% were lurking drop-outs. 36% of the participants were item nonresponders and 4% were item nonresponding drop-outs (see right column in Table 1).

**Table 1: Comparison of results for the conventional three-fold and the suggested seven-fold typology for the Web survey conducted (n= 1469)**

Conventional Three-Fold Typology		Suggested Seven-Fold Typology	
<i>(Non)Response Type</i>	%	<i>(Non)Response Type</i>	%
Unit Nonresponse	30.4	Unit Nonresponse	10.2
Item Nonresponse	44.3	Item Nonresponse	36.0
Complete Response	25.3	Complete Response	25.3
		Lurking Drop-Out	13.3
		Lurker	6.9
		Answering Drop-Out	4.3
		Item Nonresponding Drop-Out	4.0

Using the traditional categories of complete response, unit nonresponse, and item nonresponse (see left column in Table 1), the study described above would have reported unit nonresponse at 30.4% (10.2% unit nonresponders as defined above plus 6.9% lurkers plus 13.3% lurking drop-outs) with a response rate somewhere between 25.3% (according to our definition) and 69.6% (depending upon the degree of unanswered questions in each case). As discussed previously, if using only three response types, lurkers and lurking drop-outs are grouped with unit non-responders. While unit non-responders and lurking drop-outs may have experienced technical difficulties which prevented further participation, it is likely that the three groups differ significantly from each other. If one seeks to minimize nonresponse by encouraging those individuals who are likely to refuse to respond, these differences must be better understood. For instance, given that lurkers do not experience technical problems and willingly choose to view the entire survey, perhaps it is not lack of interest nor motivation, which prevents response but some other attitude.

Similarly, using only item nonresponse, unit nonresponse, and complete response as categories, item nonresponse would have been estimated at 44.3% of returned surveys. Using the seven-fold response typology, we see that 8.3% of the participants answered some questions but dropped out prior to completing the survey. This is an important distinction. The 36% who finished the survey but left missing answers to some questions maintained enough involvement in the survey to complete the activity and did not experience problems completing the survey. However, the answering drop-outs and item nonresponding drop-outs either chose to quit or possibly experienced some problem which interrupted the session. If the drop-out was volitionally controlled, we must learn what variables may have affected that decision. This is especially important for answering drop-outs, as this segment represents individuals who answered all questions up until the decision to quit. Answering drop-outs may be easily converted into complete responders if we develop an understanding of the reasons behind the mechanisms and/or choice to end participation.

## **Discussion and Further Suggestions**

In conclusion, this paper identifies seven distinct response patterns in Web surveys. The patterns are based upon the questions viewed and answered in a Web survey. In our opinion, the typology suggested here is both of practical and theoretical relevance, as it provides a detailed insight into the individual response patterns in Web surveys, and illuminates the previous 'black box' model of response patterns.

Concerning the explanation of these response patterns, we assume at least two broad complementary approaches to understand and explain nonresponse in Web surveys. The first approach, mostly found in the sociological literature on nonresponse in Web surveys (e.g., Bowker/Dillman, 2000; Dillman, 2000), is primarily focused on design-specific causes of nonresponse error. The second approach, which is in our view complementary to the first, refers closely to the psychological notion of nonresponse as some sort of 'non-compliance' (see, e.g., Rogelberg/Luong, 1998, for an overview). A combination of these two approaches may give us a comprehensive understanding of the processes underlying the nonresponse phenomenon in general and especially for every single response pattern in Web surveys presented above.

From a practical standpoint, the typology suggested may be used to provide indications of questionnaire quality during the pre-test stage. Changes can then be made, as appropriate, based on the distribution of response types reflected in the pre-test.

## References

- Ajzen, I., 1985: From intentions to actions: A theory of planned behavior. pp. 11-39 in: J. Kuhl/J. Beckmann (Eds.), *Action control*. Berlin: Springer Verlag.
- Ajzen, I., 1991: The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Armstrong, J. S./Overton, T.S., 1977: Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research*, 16 (August), 396-402.
- Bandilla, B./Bosnjak, M./Schneid, M./Stiegler, A. 1999: Interaktive Medien als Instrument und Gegenstand der empirischen Sozialforschung. [Interactive media as an instrument and a subject-matter for the empirical social sciences] pp. 127-148 in: M. Bergahus (Ed.), *Interaktive Medien [Interactive media]*. Opladen: Westdeutscher Verlag.
- Batinic, B./Bosnjak, M., 1997: Fragebogenuntersuchungen im Internet. [Questionnaire studies on the Internet] pp. 221-244 in: B. Batinic (Ed.), *Internet für Psychologen [Internet for Psychologists]*. Göttingen: Hogrefe.
- Bosnjak, M., 2001: Participation in Non-Restricted Web Surveys: A Typology and Explanatory Model for Item Non-Response. pp. 193-208 in: U.-D. Reips/M. Bosnjak (Eds.), *Dimensions of Internet Science*. Lengerich: Pabst Science Publishers.
- Bowker, D./Dillman, D.A., 2000, May: An Experimental Evaluation of Left and Right Oriented Screens for Web questionnaires. Paper presented at the 55<sup>th</sup> American Association for Public Opinion Research Annual Conference, Portland, OR.



- Dillman, D. A., 2000: *Mail and Internet Surveys – The Tailored Design Method*. New York, NY: Wiley.
- Dillman, D. A./Totoro, R.D./Conrad, J./Bowker, D., 1998. Influence of plain versus fancy design on response rates for web surveys. Paper presented at annual meeting of the American Statistical Association, Dallas, TX.
- Frey, J., 1976: *Survey Research by Telephone*. Thousand Oaks, CA: Sage Publications.
- Frick, A./Bächtiger, M.T./Reips, U.-D., 2001: Financial Incentives, Personal Information and Drop-Out Rate in Online Studies. pp. 209-220 in: U.-D. Reips/M. Bosnjak (Eds.), *Dimensions of Internet Science*. Lengerich: Pabst Science Publishers.
- Groves, R. M./Cialdini, R.B./Couper, M.P., 1992: Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, 56, 475-495.
- Groves, R.M./Couper, M.P., 1998: *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Knapp, F./Heidingsfelder, M., 2001: Drop-out analysis: the effect of research design. pp. 221-230 in: U.-D. Reips/M. Bosnjak (Eds.), *Dimensions of Internet Science*. Lengerich: Pabst Science Publishers.
- Lukawetz, G., 2001: Empirically Quantifying Unit-Nonresponse-Errors in Online Surveys and Suggestions for Computational Correction-Methods. In press in B. Batinic/U.-D. Reips/M. Bosnjak (Eds.), *Online Social Sciences*. Seattle, WA: Hogrefe & Huber.
- Oksenberg, L./Coleman, L./Cannell, C.F., 1986: Interviewers' Voices and refusal rates in telephone surveys. *Public Opinion Quarterly*, 50 (1), 97-111.
- Rogelberg, S.G./Luong, A., 1998: Nonresponse to Mailed Surveys: A Review and Guide. *Current Directions in Psychological Science*, 7(2), 60-65.
- Tuten, T.L./Bosnjak, M./Bandilla, W., 2000: Banner-Advertised Web-Surveys. *Marketing Research*, 11(4), 17-21.
- Tuten, T./Urban, D.J./Bosnjak, M., 2001. Internet Surveys and Data Quality – A Review. In press in: B. Batinic/U.-D. Reips/M. Bosnjak (Eds.), *Online Social Sciences*. Seattle, WA: Hogrefe & Huber.
- Yu, J./Cooper, H., 1983: A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research*, 20 (February), 36-44.

# ANREIZFORMEN UND AUSSCHÖPFUNGSQUOTEN BEI POSTALISCHEN BEFRAGUNGEN. EINE PRÜFUNG DER REZIPROZITÄTSHYPOTHESE\*

ANDREAS DIEKMANN UND BEN JANN<sup>+)</sup>

Mittels eines Methodenexperiments mit zwei randomisierten Versuchsgruppen (N jeweils 200) und einer Kontrollgruppe (N=200) wird untersucht, ob und in welchem Ausmaß Geschenke bei einer postalischen Befragung zur Erhöhung der Ausschöpfungsquote beitragen. Es zeigt sich, dass ein versprochenes Geschenk (Telefonkarte im Wert von 10 Schweizer Franken) die Ausschöpfungsquote nicht erhöht, während ein dem Fragebogen beigelegtes Geschenk zu einem Anstieg der Quote um zirka 10 Prozentpunkte führt. Die Befunde stehen in Einklang mit der Reziprozitätshypothese, derzufolge Vorleistungen von vielen Personen auch dann honoriert werden, wenn die reziproke Handlung nicht dem unmittelbaren Eigeninteresse eines Akteurs entspricht.

The paper presents findings from a study on the effect of promising a reward versus providing a reward on response rates to a mail survey. Two experimental groups and a control group were involved (N=200 in each). The incentive used in each case was a phone card to the value of 10 Swiss Francs (about 6 US \$). Subjects in the first experimental group were promised a phone card if they returned the completed questionnaire, those in the second group received a phone card with the first mailing. The control group were not given or offered an incentive. Subjects were assigned randomly to an experimental group or the control group. The promise of an incentive did

---

<sup>+)</sup>  Institut für Soziologie der Universität Bern

<sup>\*)</sup>  Diese Studie wurde im Rahmen des Schwerpunktprogrammes „Zukunft Schweiz“ vom Schweizerischen Nationalfonds (SNF) gefördert. Unser Dank gilt Markus Bühlmann, Andrea Hungerbühler, Philipp Kayser und Alex Keller für ihrer wertvolle Hilfe bei der Durchführung des Experiments. Für Anregungen und Hinweise bedanken wir uns bei Fritz Strack und den beiden anonymen Gutachtern bzw. Gutachterinnen dieses Artikels.

not improve response, while there was a significantly higher rate of response (about 10 percent points) among the group who received a telephone card with the first mailing. The findings support the hypothesis of reciprocity which assumes that people will behave in a reciprocal fashion, even if this is not in their narrow material self-interest.

## 1. Einleitung

Bei einer schriftlichen Befragung kann bekanntlich die Ausschöpfung durch ein Bündel geeigneter Maßnahmen wesentlich erhöht werden. Nicht selten erreicht man Ausschöpfungsquoten, die denjenigen telefonischer oder von Face-to-Face-Befragungen entsprechen oder diese noch übertreffen (Dillman 1978; Hippler/Seidel 1985). Eine dieser Maßnahmen ist ein kleines Geschenk, das die Zielpersonen ermuntern soll, den Fragebogen auszufüllen und zu retournieren. Bei der postalischen Befragung eignen sich zum Beispiel Telefonkarten, Veranstaltungsbillets oder Geschenkgutscheine, die beim Versand des Fragebogens beigelegt werden können. Bei zwei landesweiten telefonischen Befragungen in der Schweiz haben wir im Rahmen des International Social Survey Programme (ISSP) jeweils eine schriftliche Nachbefragung durchgeführt. Beim Versand des Fragebogens wurde als Geschenk eine Telefonkarte mit Projektaufdruck im Wert von 10 CHF mitversandt. Nun handelte es sich bei den schriftlichen Nachbefragungen um selektive Stichproben derjenigen Personen, die bereits an den telefonischen Interviews teilgenommen hatten, so dass ohnehin relativ hohe Ausschöpfungsquoten zu erwarten waren. Mit Ausschöpfungsquoten von 88 Prozent beim „Umweltsurvey 1994“ (Diekmann/Franzen 1995) und 84 Prozent beim „Arbeitsmarktsurvey 1998“ (Diekmann et al. 1999) lagen diese aber noch über den Erwartungen, so dass die Vermutung begründet erscheint, dass der Versand der Telefonkarte an diesem Ergebnis einen nicht unerheblichen Anteil hatte. Ob und in welchem Ausmaß dies der Fall war, lässt sich allerdings nur auf der Basis eines Kontrollgruppendesigns beantworten. Bei den beiden Nachbefragungen wollten wir die Ausschöpfungsquoten nicht „verderben“, so dass auf eine Methodenstudie mit Kontrollgruppendesign verzichtet wurde. Eine solche Studie bot sich aber bei einer kürzlich durchgeführten Panelbefragung an. Im Frühjahr 2000 haben wir eine Teilstichprobe aus den Befragten des Arbeitsmarktsurveys nach dem Zufallsprinzip gezogen und diesen Personen erneut einen schriftlich zu beantwortenden Fragebogen zukommen lassen. Anders als bei vorliegenden Arbeiten zu diesem Thema im deutschen Sprachraum ist das Ziel unserer Studie, nicht nur die Wirkung eines beigelegten, sondern auch eines versprochenen Geschenks auf die Ausschöpfungsquote zu untersuchen.

## 2. Hypothesen und Design der Untersuchung

Aus zahlreichen Studien ist bekannt, dass „incentives“ (Geschenke wie Briefmarken, Kugelschreiber oder auch eine monetäre Belohnung) die Ausschöpfungsquoten postalischer Befragungen erhöhen, wobei allerdings je nach Art des Geschenks (nicht-monetär versus monetär, Höhe eines Geldbetrags) und der Methode der Übergabe (mitgeschickt versus „versprochen“) erhebliche Variationen bestehen. Neuere Literaturüberblicke findet man bei Singer (1998), Porst (1999) sowie in der Metaanalyse von Church (1993). Letzterer analysierte die Ergebnisse von 38 Studien mit Kontroll- oder Vergleichsgruppe und insgesamt 74 „Incentive-Gruppen“. Es zeigte sich, dass versprochene incentives allenfalls eine geringe, nicht-signifikante Erhöhung gegenüber der Kontrollgruppe, deren Mitglieder keinen incentive erhielten, aufweisen. In 55 Fällen wurde ein Geschenk oder ein Geldbetrag (arithmetischer Mittelwert zirka 1,40 US \$ bezogen auf das Jahr 1989) vorab mit dem Fragebogen verschickt. Nicht-monetäre Anreize erhöhten die Rücklaufquote (Response Rate) um 7,9 Prozentpunkte, bei monetären Anreizen waren es 19,1 Prozentpunkte. In einer kürzlich publizierten Studie berichten Singer et al. (2000) von der Wirkung eines mit dem Anschreiben verschickten Geldgeschenks von 5 US \$ auf die Bereitschaft zur Teilnahme an Telefoninterviews. Die Response Rate konnte durch diese Maßnahme gegenüber der Kontrollgruppe (Anschreiben ohne Geldgeschenk) um 10,7 Prozentpunkte erhöht werden. Versprochene Geldgeschenke hatten dagegen keinen signifikanten Einfluss auf die Teilnahmebereitschaft.

Nach unseren Recherchen wurden im Zeitraum der vergangenen zehn Jahre drei im deutschsprachigen Raum durchgeführte Studien mit Kontrollgruppendesign publiziert. Harkness et al. (1998) berichten von zwei Experimenten im Rahmen der ISSP-Befragung in Deutschland. Als Incentive dienten Briefmarken im Wert von 4 DM, die dem Fragebogen beigelegt wurden. Im zweiten Experiment wurde eine Lotterie in Aussicht gestellt mit drei Hauptgewinnen von je 1000 DM. Im ersten Experiment erhöhte sich die Rücklaufquote um fünf, im zweiten Experiment um rund drei Prozentpunkte. Arzheimer und Klein (1998) erzielten eine Steigerung der Rücklaufquote von etwa 10 Prozentpunkten (von zirka 50 auf über 60 Prozent) durch Beilage einer Telefonkarte im Wert von 6 DM. Porst (1999) verwendete als Anreiz Briefmarken im Wert der Frankatur von vier Standardbriefen. In den zwei Varianten mit unterschiedlichen Einstiegsfragen zeigte sich in der ersten Feldphase (bei geringer Fallzahl und insgesamt niedriger Rücklaufquote) ein nicht-signifikanter Anstieg von 4,6 auf 17,9 Prozent beziehungsweise von 12,0 auf 17,4 Prozent.

Es ist keineswegs selbstverständlich, dass beigelegte Geschenke die Ausschöpfungsquote erhöhen. Wenn das Ausfüllen und der Rückversand zumindest einem Teil der Zielpersonen Mühe und Aufwand bereiten, ist unter der Annahme strikter Rationalität prognostizierbar, dass diese Personen den Fragebogen nicht zurückschicken werden. Ein beigelegtes Geschenk ändert die Prognose nicht. Denn eine rationale Person behält das Geschenk und entsorgt den Fragebogen. Anders verhält es sich bei versprochenen Geschenken, das heißt bei einer bedingten Belohnung für die Teilnahme an der Befragung. Übersteigt die Belohnung wenigstens bei einem Teil der Befragten den Aufwand (abzüglich des Nutzens aus dem Interesse an der Befragung), dann werden diese Personen jetzt den Fragebogen ausfüllen und zurücksenden. Das versprochene Geschenk ist also ein zusätzlicher Anreiz, der zumindest bei einem Teil der Befragten das Resultat des Nutzenkalküls in den positiven Bereich verschieben kann. Die *Hypothese strikter Rationalität* prognostiziert, dass versprochene, nicht aber beigelegte Geschenke die Ausschöpfungsquote erhöhen werden.

Hierzu gegenläufige Effekte sind unter der Annahme der *Reziprozitätshypothese* zu erwarten. Der Fragebogen wird nicht deshalb um so eher ausgefüllt, weil man eine Belohnung erwartet; vielmehr wird die Vorleistung durch ein Geschenk als Verpflichtung empfunden, ebenfalls eine Leistung zu erbringen. Wie Gouldner (1960) in seinem klassischen Artikel bemerkt, haben Reziprozitätsnormen universellen Charakter. In neuen Modellen der Spieltheorie und modifizierten Rational-Choice-Modellen nehmen Reziprozitätsnormen einen zentralen Platz ein (Bolton/Ockenfels 2000; Fehr/Gächter 2000). Zahlreiche Experimente zeigen, dass Versuchspersonen eine starke Tendenz zur Reziprozität auch in nicht wiederkehrenden Situationen aufweisen und diese auch dann zeigen, wenn eigentlich „lohnendere“ Alternativen zur Verfügung stehen (Cialdini 1984). Ein ebenso einfaches wie eindrucksvolles Experiment stammt von Regan (1971; siehe auch Cialdini 1984). Den Versuchspersonen, die glaubten, zusammen mit einer anderen Person an einem Experiment zur Bewertung von Kunst teilzunehmen, wurde jeweils von der anderen Versuchsperson ein Gefallen erwiesen. In einer Pause nach dem Kunst-Rating brachte diese „Versuchsperson“, in Wirklichkeit ein Mitarbeiter des Experimentators, der wartenden Versuchsperson unaufgefordert eine Flasche Cola mit. Etwas später bat der Mitarbeiter die Versuchsperson, ihm einige Tombolalose abzukaufen. Probanden, denen ein Gefallen erwiesen wurde, kauften doppelt so viele Lose als die Versuchspersonen der Kontrollgruppe, denen keine solche Aufmerksamkeit entgegengebracht wurde.

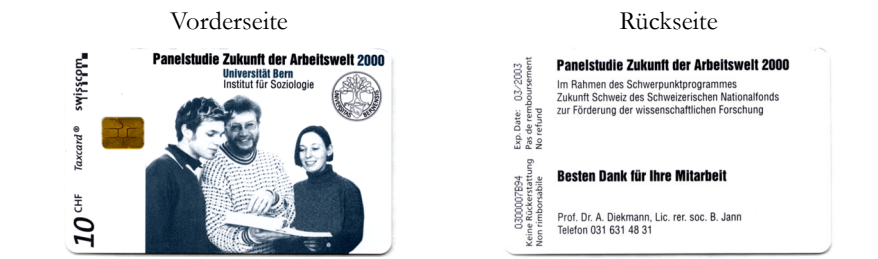
Konkret im Zusammenhang mit der Problematik der Rücklaufquoten haben Arzheimer und Klein (1998) sowie Singer (1998) auf die Bedeutung von Reziprozitätsnormen hingewiesen. Gemäß der Reziprozitätshypothese ist zu erwarten, dass bei-

gelegte Geschenke die Ausschöpfungs- und Rücklaufquote in stärkerem Maße erhöhen als versprochene Geschenke.

**Tabelle 1: Design des Telefonexperiments**

Versuchsbedingung	Zuordnung zu den Versuchsgruppen	Fallzahl N
Versuchsgruppe 1: Telefonkarte versprochen	per Random	200
Versuchsgruppe 2: Telefonkarte beigelegt	per Random	200
Kontrollgruppe: kein Geschenk	per Random	200

Telefonkarte mit Projektaufdruck:



Unser Design zur Prüfung der Hypothese sieht zwei Versuchsgruppen und eine Kontrollgruppe vor. In der Versuchsgruppe 1 wurde den Zielpersonen als Geschenk eine Telefonkarte mit Projektaufdruck im Wert von 10 CHF nach Retournierung des Fragebogens versprochen. In Versuchsgruppe 2 wurde die Telefonkarte beigelegt. Den Personen in der Kontrollgruppe wurde weder ein Geschenk versprochen noch beigelegt. Alle Personen in den drei Gruppen erhielten den Fragebogen und ein freigestempeltes Rückkuvert. Die drei Stichproben mit je 200 Personen wurden nach einem Zufallsverfahren aus der Stichprobe der Teilnehmerinnen und Teilnehmer an der schriftlichen Nachbefragung des „Arbeitsmarktsurveys 1998“ ausgewählt. Die Zielpersonen wurden damit den drei Gruppen per Random zugeordnet<sup>1</sup>. Tabelle 1 gibt einen Überblick zu dem gewählten Design.

<sup>1</sup> Für die Ziehung der Stichprobe wurden nur Erwerbspersonen im weiteren Sinne berücksichtigt (Personen, die zum Zeitpunkt der 1998er Befragung erwerbstätig, arbeitslos beziehungsweise stellensuchend, in Weiterbildung oder Umschulung, im Militär beziehungsweise

Die erste Feldphase erstreckte sich auf einen Zeitraum von 42 Tagen. Nach dieser Zeit erhielten Zielpersonen in der Stichprobe, die keinen Fragebogen retourniert hatten, einen Mahnbrief plus Fragebogen.

### 3. Ergebnisse

Von den 600 versandten Fragebogen sind 24 Ausfälle zu subtrahieren, so dass wir eine Nettostichprobe von  $N=576$  erhalten. Die wenigen Ausfälle sind über die drei experimentellen Gruppen fast gleich verteilt. Wir berichten im folgenden die auf die Nettostichprobe bezogenen Ausschöpfungsquoten in Prozent.<sup>2</sup> Betrachten wir zunächst die erste Feldphase ohne Mahnung. Die höchste Ausschöpfungsquote (72,4 Prozent) erhalten wir für die Gruppe „Geschenk beigelegt“. In der Gruppe „Geschenk versprochen“ beträgt die Ausschöpfungsquote dagegen nur 58,1 Prozent. Die Differenz ist für  $\alpha=0,05$  signifikant (zweiseitiger Test,  $p=0,003$ ). Sogar in der Kontrollgruppe ist die Ausschöpfungsquote mit 62,7 Prozent noch höher als bei den Personen, denen das Geschenk versprochen wurde, wobei die Differenz für  $\alpha=0,05$  aber nicht signifikant ist. Im Vergleich mit der Kontrollgruppe führt die beigelegte Telefonkarte zu einem knapp signifikanten Anstieg der Ausschöpfungsquote von 9,7 Prozentpunkten (Abbildung 1 und Tabelle 2).

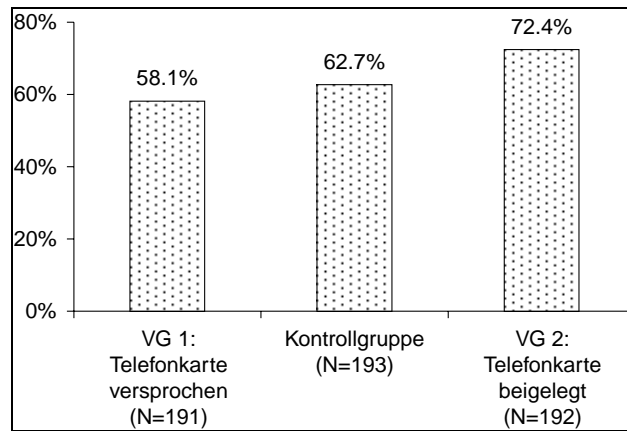
Die Mahnaktion erhöht die Ausschöpfungsquoten noch einmal um 12 bis 16 Prozentpunkte. Auch nach der Mahnung verbleibt zwischen den Gruppen „Geschenk beigelegt“ und „Geschenk versprochen“ noch eine Differenz von mehr als 10 Prozentpunkten. Ein ähnliches Bild erhält man, wenn die Rücklaufquoten, die sich auf die Bruttostichprobe beziehen, zugrunde gelegt werden (Tabelle 2 und Abbildung 2).

---

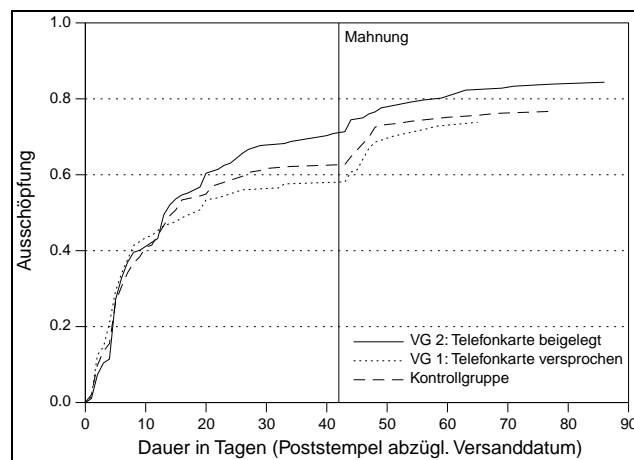
Zivildienst oder im Schwangerschaftsurlaub waren). Einschränkend wurden zudem nur Personen mit Wohnsitz in der Deutschschweiz und Alter zwischen 18 und 60 Jahren (zum Zeitpunkt der Erstbefragung) berücksichtigt. Zusätzlich wurde die Stichprobe noch um 300 Vollzeit erwerbstätige Zielpersonen aufgestockt, die keine Telefonkarte erhielten. Im Rahmen unseres Experiments spielt diese Gruppe keine Rolle.

**2** Die Ausschöpfungsquote ist der Quotient: (Anzahl zurückgesandter und gültiger Fragebogen)/(Anzahl verschickter Fragebogen abzüglich neutraler Ausfälle). Als Ausfallgründe zählen wir: Adresse ungültig und nicht ermittelbar, abgereist, unbekannt, im Ausland, gestorben. Wir folgen hier dem Schema von Hippler/Seidel 1985. Ob diese Ausfallgründe als „stichprobenneutral“ gelten können, kann kontrovers beurteilt werden. Für unser Experiment ist diese Zuordnung schon deshalb geeignet, weil damit Personen ausgeschlossen wurden, die keine Möglichkeit zur Beantwortung und Rücksendung hatten. Für Hinweise zur Präzisierung bedanken wir uns bei den Gutachtern bzw. Gutachterinnen.

**Abbildung 1: Ausschöpfungsquoten in Prozent nach Versuchsbedingungen (vor Mahnung)**



**Abbildung 2: Entwicklung der Ausschöpfungsquoten über die Dauer der Feldphase nach Versuchsbedingungen (N=576, Log-Rank Test:  $p=0,145$ )**





**Tabelle 2: Rücklaufstatistik vor und nach Mahnung nach Versuchsgruppen**

Versuchsbedingung	Teilnahme	keine Teilnahme	Ausfälle <sup>1</sup>	Ausschöpfung <sup>2</sup>	Rücklauf <sup>3</sup>
vor Mahnung <sup>4</sup>					
VG 1 (versprochen)	111	80	9	58,1%	55,5%
VG 2 (beigelegt)	139	53	8	72,4%	69,5%
Kontrollgruppe	121	72	7	62,7%	60,5%
nach Mahnung <sup>5</sup>					
VG 1 (versprochen)	141	50	9	73,8%	70,5%
VG 2 (beigelegt)	162	30	8	84,4%	81,0%
Kontrollgruppe	148	45	7	76,7%	74,0%
nur gemahnte					
VG 1 (versprochen)	30	50	-	37,5%	37,5%
VG 2 (beigelegt)	23	30	-	43,4%	43,4%
Kontrollgruppe	27	45	-	37,5%	37,5%

<sup>1</sup> Ausfallgründe: Adresse ungültig und nicht ermittelbar, abgereist, unbekannt, im Ausland, gestorben

<sup>2</sup> Ausschöpfungsquote = (zurückgesandte und gültige Fragebogen)/(verschickte Fragebogen – Ausfälle) = (zurückgesandte und gültige Fragebogen)/(Nettostichprobe)

<sup>3</sup> Rücklaufquote = (zurückgesandte und gültige Fragebogen)/(verschickte Fragebogen) = (zurückgesandte und gültige Fragebogen)/(Bruttostichprobe)

<sup>4</sup> Signifikanztests für Ausschöpfungsquoten (zweiseitig):  $p = 0,003$  für VG 2 versus VG1;  $p = 0,042$  für VG2 versus Kontrollgruppe;  $p = 0,359$  für VG1 versus Kontrollgruppe.

<sup>5</sup> Signifikanztests für Ausschöpfungsquoten (zweiseitig):  $p = 0,011$  für VG2 versus VG1;  $p = 0,057$  für VG2 versus Kontrollgruppe;  $p = 0,516$  für VG1 versus Kontrollgruppe.

Unsere Befunde stehen damit im Einklang mit vorliegenden Untersuchungen aus dem angelsächsischen Sprachraum (Church 1993). Sie sind für die empirische Sozialforschung von praktischem Interesse und dienen als weiterer Beleg für die Reziprozitätstheorie. Von praktischem Nutzen ist die Erkenntnis, dass versprochene Geschenke, geknüpft an die Bedingung der Rücksendung des ausgefüllten Fragebogens, nicht zur Steigerung der Ausschöpfungsquote beitragen können.<sup>3</sup> Will man

**3** Voraussetzung für die Gültigkeit unserer Ergebnisse ist allerdings, dass unser Versprechen als glaubwürdig angesehen wurde. Bei der Hauptbefragung haben wir versprochen, interessierte Befragte über die Ergebnisse zu informieren. Der ursprünglich für das Frühjahr 2000 vorgesehene Versand eines Informationsblattes wurde von uns auf später verschoben, um die Ergebnisse des Methodenexperiments nicht zu beeinflussen. Der Nachteil war, dass Befragte aufgrund des Ausbleibens des Informationsblattes Zweifel an der Glaubwürdigkeit unseres Versprechens haben konnten (wobei zu bemerken ist, dass den Befragten bei der

Geschenke zur Erhöhung der Ausschöpfung einsetzen, so empfiehlt es sich, diese direkt dem Fragebogen beizulegen. Ähnlich wie in der Untersuchung von Arzheimer und Klein (1998) konnte die Ausschöpfungsquote mittels der beigelegten Telefonkarte um rund zehn Prozentpunkte erhöht werden.

Von Interesse ist der Befund aber auch für die Reziprozitätstheorie, die in verschiedenen sozialen Kontexten – etwa bezüglich des Zusammenhangs von Lohn und Leistung in der Arbeitsmarktforschung – diskutiert wird (Überblick bei Fehr und Gächter 2000). Eine Vorleistung wird von vielen Personen eben auch dann honoriert, wenn die Reaktion nicht im unmittelbaren Eigeninteresse eines Akteurs liegt. Zu Gunsten der Reziprozitätshypothese liefert unser Telefonkartenexperiment einen klaren Beleg.

### **Korrespondenzadresse**

*Prof. Dr. Andreas Diekmann  
Unitobler, Lerchenweg 36  
CH-3000 Bern 9  
E-Mail: diekmann@soz.unibe.ch*

### **Literatur**

- Arzheimer, K./Klein, M., 1998: Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. *ZA-Informationen* 43: 6-31.
- Bolton, G.E./Ockenfels, A., 2000: ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90(1): 166-193.
- Cialdini, R. B., 1984: *Influence. The Psychology of Persuasion*. New York: William Morrow.
- Church, A.H., 1993: Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly* 57: 62-79.

---

Hauptbefragung keine Angabe zum Termin der Versendung des Informationsblattes gemacht wurden; es beklagte sich denn auch nur ein einziger Befragter, bisher noch keinen Ergebnisbericht erhalten zu haben). Es ist allerdings unwahrscheinlich, dass dieses Problem die Ergebnisse wesentlich beeinflusst haben sollte. Selbstverständlich wurden die versprochenen Telefonkarten nach dem Experiment den berechtigten Personen, die den Fragebogen retourniert hatten, zugeschickt.

- Diekmann, A./Franzen, A., 1995: Der Schweizer Umweltsurvey 1994. Codebuch. Universität Bern.
- Diekmann, A./Engelhardt, H./Jann, B./Armingeon, K./Geissbühler, S., 1999: Der Schweizer Arbeitsmarktsurvey 1998. Codebuch. Universität Bern.
- Dillman, D.A., 1978: Mail and Telephone Surveys. The Total Design Method. New York: Wiley.
- Fehr, E./Gächter, S., 2000: Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives* 14: 159-181.
- Gouldner, A., 1960: The Norm of Reciprocity. A Preliminary Statement. *American Sociological Review* 25: 161-178.
- Harkness, J./Mohler, P./Schneid, M./Christoph, B., 1998: Incentives in Two German Mail Surveys 1996/97 & 1997. S. 201-218 in: Koch, A./Porst R. (Eds.), *Nonresponse in Survey Research*. ZUMA, Mannheim: ZUMA-Nachrichten Spezial 4.
- Hippler, H.J./Seidel, K., 1985: Schriftliche Befragung bei allgemeinen Bevölkerungstichproben – Untersuchungen zur Dillmanschen „Total Design Method“. *ZUMA-Nachrichten* 16: 39-56.
- Porst, R., 1999: Thematik oder Incentives? Zur Erhöhung der Rücklaufquoten bei postalischen Befragungen. *ZUMA-Nachrichten* 45: 72-87.
- Regan, D. T., 1971: Effects of Favor and Liking on Compliance. *Journal of Experimental Social Psychology* 7: 627-639.
- Singer, E., 1998: Incentives for Survey Participation: Research on Intended and Unintended Consequences. Second Annual Wildenmann Lecture. *ZUMA-Nachrichten* 42: 7-29.
- Singer, E./van Hoewyk, J./Maher, M. P., 2000: Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly* 64: 171-188.

# DIE STATISTISCHE AUFDECKUNG KAUSALSTRUKTURELL UNTERSCHIEDENER GRUPPEN AM BEISPIEL POLITISCHEN TEILNAHMEVERHALTENS

*FRANK FAULBAUM, PETRA STEIN & KAI KELLETER*

Der vorliegende Beitrag beschäftigt sich mit der spezifischen Anwendung eines Verfahrens zur Aufdeckung unbekannter Gruppierungen unter Einbeziehung möglicherweise unterschiedlicher gruppenspezifischer Kausalstrukturen. Das Verfahren basiert auf einem Mischverteilungsansatz, bei dem auf ausgewählte exogene Variablen bedingte Komponenten einer multivariaten Normalverteilung bestimmt werden, deren Mittelwerts- und Kovarianzstrukturen wiederum wie bei der Analyse von Strukturgleichungsmodellen parametrisiert werden können. Es ist besonders geeignet zu überprüfen, ob eine Population hinsichtlich der Konstruktvalidität theoretisch interessanter Variablen homogen ist. Die Ergebnisse für das hier ausgewählte Beispiel relevanter Variablen des politischen Teilnahmeverhaltens zeigen, wie sich Gruppierungen finden lassen, die sich durch unterschiedliche Mess- und Strukturmodelle unterscheiden, so dass nicht davon ausgegangen werden kann, dass für die Population dasselbe gemeinsame Modell gilt.

This paper presents a specific application of a statistical method for finding unknown groups in survey data by including group-specific differences in causal structure. The procedure consists in the identification of the normal components of a multivariate normal mixture distribution conditional on selected exogenous variables. The mean and covariance structures of the components can be parameterized, as in structural equation modeling. The procedure is especially useful for testing whether a population can be considered homogenous with respect to the construct validity of substantially interesting variables. For the example of important variables of political participation discussed here, our findings point to different groups with differing measurement and structural models. This indicates that no one common model can be assumed to hold for the entire population.

## 1. Problemstellung

Der vorliegende Aufsatz stellt anhand einer politikwissenschaftlichen Anwendung ein Verfahren vor, das im Unterschied zur klassischen Clusteranalyse unbekannte Gruppierungen in Daten nicht auf der Basis von Distanzmaßen, sondern auf der Basis komplexer Strukturunterschiede zwischen den Gruppen aufzufinden versucht. Die dabei mit ins Kalkül gezogenen Strukturinformationen können in vollständigen Strukturgleichungsmodellen mit latenten Variablen, in Messmodellen oder nur in Regressionsmodellen mit ausschließlich beobachteten Variablen bestehen. Analyseverfahren dieser Art sind etwa geeignet, um Gruppen aufzudecken, die durch unterschiedliche Messmodelle oder Strukturmodelle charakterisierbar sind. Entfällt z.B. in einer Gruppe für eine latente Variable ein Indikator, so ändert sich sofort nicht nur die Semantik des Messmodells, sondern auch die des Strukturmodells; d.h., das Verfahren ist besonders geeignet, um Gruppen mit unterschiedlicher Konstruktvalidität zu entdecken.

Der verwendete statistische Ansatz ist an anderer Stelle bereits ausführlicher dargestellt worden (vgl. Arminger/Stein 1997; Faulbaum/Stein 2000; Stein 1997, 2000). Er beruht auf einer Weiterentwicklung von Mischverteilungsansätzen. In multivariaten Mischverteilungsansätzen werden unbekannte Gruppen in Umfragedaten als Komponenten (finite Mischungen) multivariater Normalverteilungen aufgefasst. Die angesprochenen Weiterentwicklungen betreffen die Einbeziehung exogener Variablen und der parametrischen Struktur von Kovarianzmatrizen, d.h. die Berücksichtigung unterschiedlicher kausaler Strukturen mit oder ohne unbeobachtete Variablen in den einzelnen Gruppen. Hier sind vor allem die Arbeiten von Stein (1997, 2000), Arminger und Stein (1997) sowie Arminger, Stein und Wittenberg (1999) zu nennen, die auf der Basis von Vorarbeiten von Yung (1994; vgl. auch Yung 1997) sowie DeSarbo und Cron (1988) entsprechende Erweiterungen des Mischverteilungsansatzes vorgenommen haben. Angewendet wurden diese Ansätze bisher auf Lebensstilvariablen (vgl. Stein 1997), auf Variablen der Lebenszufriedenheit (vgl. Stein 2000) und auf Einstellungen zur Integration von Ausländern (vgl. Faulbaum/Stein 2000). Ein alternatives Vorgehen ohne Einbeziehung exogener Variablen, unbekannte Gruppen mit unterschiedlichen Kausalstrukturen zu finden, stellen Jedidi, Jagpal und DeSarbo (1998) vor. Das vollständige zugrundegelegte Mischverteilungsmodell hat folgende mathematische Form:

$$f(\mathbf{y}_i | \mathbf{x}_i) = \sum_{k=1}^K \pi_k \phi(\mathbf{y}_i; \gamma_k + \Pi_k \mathbf{x}_i, \Sigma_k)$$

In dieser Formel bezeichnen  $K$  die Anzahl der Komponenten (Gruppen),  $\pi_k$  die mischenden Wahrscheinlichkeiten,  $\phi(\mathbf{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  die auf die Komponenten  $k$  bedingten multivariaten Normalverteilungen des Vektors  $\mathbf{y}_i$  von Zufallsvariablen mit Mittelwertsvektoren  $\boldsymbol{\mu}_k$  und Kovarianzmatrizen  $\boldsymbol{\Sigma}_k$  als Verteilungsparametern.  $\boldsymbol{\gamma}_k$  bezeichnet den Vektor der Regressionskonstanten und  $\boldsymbol{\Pi}_k$  die Matrix der Regressionskoeffizienten, wobei für den Mittelwertsvektor  $\boldsymbol{\mu}_{ik}$  der bedingten Verteilung von  $\mathbf{y}_i$  gegeben Regressoren  $\mathbf{x}_i$  und die  $k$ -te Komponente gilt, dass  $\boldsymbol{\mu}_{ik} = \boldsymbol{\gamma}_k + \boldsymbol{\Pi}_k \mathbf{x}_i$ .  $\boldsymbol{\Sigma}_k$  ist auf die exogenen Variablen  $\mathbf{x}_i$  bedingte Kovarianzmatrix.

$\boldsymbol{\mu}_{ik}$  und  $\boldsymbol{\Sigma}_k$  können im Rahmen einer Spezifikation von Strukturgleichungsmodellen mit oder ohne latente Variablen weiter parametrisiert werden. Das Modell hat dann die Form:

$$f(\mathbf{y}_i | \mathbf{x}_i) = \pi_1 \phi(\mathbf{y}_i; \boldsymbol{\mu}_{i1}(\boldsymbol{\vartheta}), \boldsymbol{\Sigma}_1(\boldsymbol{\vartheta})) + \pi_2 \phi(\mathbf{y}_i; \boldsymbol{\mu}_{i2}(\boldsymbol{\vartheta}), \boldsymbol{\Sigma}_2(\boldsymbol{\vartheta})) \\ \dots + \dots \pi_K \phi(\mathbf{y}_i; \boldsymbol{\mu}_{iK}(\boldsymbol{\vartheta}), \boldsymbol{\Sigma}_K(\boldsymbol{\vartheta})).$$

Im Vektor  $\boldsymbol{\vartheta}$  werden die Parameter der Strukturgleichungsmodelle (z.B. Ladungen, Strukturkoeffizienten, Fehlervarianzen, etc.) gesammelt.

Für die Parameterschätzung bieten sich verschiedene ein- oder zweistufige Schätzstrategien an, die in Arminger/Stein/Wittenberg (1999) genauer beschrieben sind (für eine zusammenfassende Darstellung vgl. auch Faulbaum/Stein 2000). Alle Verfahren verwenden die ein oder andere Form des EM-Algorithmus. Im vorliegenden Fall wird das zweistufige Verfahren mit anschließender Optimierung durch einen einstufigen Algorithmus zugrundegelegt (vgl. Faulbaum/Stein 2000). Als EDV-Programm wurde MECOSA 3 (vgl. Arminger/Wittenberg/Schepers 1996) verwendet.

Von den vom Schätzalgorithmus gelieferten Ergebnissen werden hier nur die Anzahl der Gruppen (Komponenten) berichtet, nicht aber die geschätzten mischenden Wahrscheinlichkeiten, d.h., die Wahrscheinlichkeiten, dass ein Element der Population zu der betreffenden Gruppe gehört sowie die Parameter der Strukturgleichungsmodelle, also die gruppenspezifischen Ladungen und Fehlervarianzen der Messmodelle sowie die Einflussparameter der exogenen auf die endogenen sowie der endogenen auf die endogenen Variablen.

Die hier geschilderte Anwendung bezieht sich auf den Datensatz des ALLBUS 1998 und einige der dort erhobenen Variablen, die für das politische Teilnahmeverhalten mitbestimmend sein könnten. Dazu gehören die „Political Efficacy“, das politische Interesse, normative Vorstellungen zur politischen Partizipation sowie der Zufriedenheit von Regierung und Demokratie in der BRD. Unter Political Efficacy

wird das Gefühl verstanden, durch das eigene Verhalten Einfluss auf die Politik nehmen zu können (vgl. z.B. Campbell et al. 1954), wobei zwischen den Dimensionen interner und externer Political Efficacy unterschieden werden kann (vgl. Lane 1959, Balch 1974, Vetter 1997). Im zuerst genannten Fall geht es um die eigenen Mittel bzw. die eigenen Kompetenzen, auf die Politik Einfluss zu nehmen, im zweiten Fall um die Überzeugung, dass die Politiker auf das eigene Verhalten reagieren (Responsivität der Politiker und Volksvertreter). Ziel der Analyse war es herauszufinden, ob sich Gruppen finden lassen, die sich in der Bedeutung der Political Efficacy unterscheiden bzw. die durch unterschiedliche Messmodelle der Political Efficacy sowie durch unterschiedliche Beziehungen der Political Efficacy zu den oben aufgeführten weiteren Variablen charakterisiert werden können. Im Mittelpunkt stand also die Frage, ob sich Gruppierungen in der Gesellschaft finden lassen, die sich durch bestimmte Intensitätsmuster von Efficacy-Variablen, Partizipationsnormen, politischem Interesse und Zufriedenheit mit Regierung und Demokratie auszeichnen, ob es latente Merkmale gibt, die diese Muster erklären können und ob sich die Gruppen durch Einflussstrukturen zwischen diesen Variablen unterscheiden.

Die Ergebnisdarstellung erfolgt in der Form, dass zunächst über Anzahl und Umfang der aufgefundenen Gruppen berichtet wird, wobei die Identifikation der Anzahl auf der Basis eines Ad-hoc Tests auf die Anzahl der Komponenten (Gruppen) erfolgte. Anschließend werden die Gruppen durch die un konditionalen Mittelwerte der Variablen beschrieben. Schließlich wird über die unterschiedlichen kausalen Strukturen in den einzelnen Gruppen berichtet. Zum Verständnis der Darstellung ist es wichtig sich vor Augen zu halten, dass die Gruppen *unter der Bedingung* ausgewählter exogener Variablen wie Geschlecht, Alter, etc., gesucht wurden. Sowohl die eingeführten exogenen Variablen als auch die endogenen Variablen gehen in die Spezifikation der kausalen Abhängigkeiten in den einzelnen Gruppen ein. Bei der Interpretation der Ergebnisse muss die politische Situation vor der Bundestagswahl 1998 mit in das Kalkül einbezogen werden.

## 2. Ausgewählte Modellvariablen

In die Analyse wurden die folgenden endogenen und exogenen Variablen des ALLBUS 1998 einbezogen, wobei in Klammern die in den Tabellen und Abbildungen aus Platzgründen verwendeten mnemotechnischen Abkürzungen angegeben sind:

## Endogene Variablen:

- **V150 (KUEMM):** Die Politiker kümmern sich nicht viel darum, was Leute wie ich denken (1: Stimme voll und ganz zu, 4: Stimme überhaupt nicht zu).
- **V153 (KOMPLEX):** Die ganze Politik ist so kompliziert, dass jemand wie ich gar nicht versteht, was vorgeht(1: Stimme voll und ganz zu , 4: Stimme überhaupt nicht zu).
- **V154 (INTER):** Die Politiker bemühen sich im Allgemeinen darum, die Interessen der Bevölkerung zu vertreten (1: Stimme voll und ganz zu, 4: Stimme überhaupt nicht zu).
- **V157 (ÜBERLASS):** Politik sollte man den gewählten Volksvertretern überlassen (1: Stimme voll und ganz zu; 4:Stimme überhaupt nicht zu).
- **V160 (VOLKSAB):** Volksabstimmungen sind ein notwendiger Bestandteil der Demokratie (1: Stimme voll und ganz zu, 4: Stimme überhaupt nicht zu).
- **V161 (DEMOGES):** Demokratie sollte nicht auf den politischen Bereich beschränkt bleiben, sondern in allen gesellschaftlichen Bereichen verwirklicht werden (1: Stimme voll und ganz zu, 4: Stimme überhaupt nicht zu).
- **V162 (ZUFDEMO):** Wie zufrieden sind Sie – alles in allem – mit der Demokratie, so wie sie in Deutschland besteht? (1: Sehr zufrieden, 6: Sehr unzufrieden).
- **V163 (REGLEIST):** Wie zufrieden sind Sie – insgesamt betrachtet – mit den gegenwärtigen Leistungen der Bundesregierung? (1: Sehr zufrieden, 6: Sehr unzufrieden).
- **V169 (POLINT):** Wie stark interessieren Sie sich für Politik: sehr stark, stark, mittel, wenig oder überhaupt nicht? (1:Sehr stark, 5:Überhaupt nicht).

## Exogene Variablen:

- **V3 (GEBIET):** Erhebungsgebiet (0: Neue Bundesländer, 1: Alte Bundesländer).
- **V37 (ALTER):** Alter in Jahren, geteilt durch 10.
- **V 141 (GESCHLECHT):** Geschlecht des Befragten (0: Männlich, 1: Weiblich).
- **V142 (BILDUNG):** Ausbildungsabschluss (0: kein Abitur, 1: Abitur).

Die Variablen V150 (KUEMM) und V153 (KOMPLEX) werden traditionell als Indikatoren für Externe bzw. Interne Efficacy angesehen (zum Überblick über Messmodelle der Political Efficacy vgl. Vetter 1997). V154 (INTER) würde man wohl eher der Externen Efficacy zuordnen. V157 (UEBERLASS), V160 (VOLKSAB) und V161 (DEMOGES) können als Partizipationsnormen aufgefasst werden.



### 3. Diskussion der Ergebnisse

#### 3.1 Anzahl und Größe der Gruppen

Im Regelfall ist die Anzahl der Komponenten einer Mischverteilung nicht a priori durch theoretische Vorannahmen festgelegt, sondern wird in der Analyse von Mischverteilungen geschätzt. Als Test auf die Anzahl der Komponenten eines Mischverteilungsmodells würde sich eigentlich ein Likelihood-Ratio-Test anbieten, wobei die Likelihood-Ratio-Statistik  $\lambda$  die Nullhypothese  $H_0: K = G$  Gruppen gegen die Alternativhypothese  $H_1: K = G + 1$  Gruppen prüft.

Wie mehrere Autoren festgestellt haben (vgl. MacLachlan/Basford 1988, Kap. 1.10), ist die Anwendung der Likelihood-Ratio-Statistik bei Mischverteilungen problematisch. Die Gründe hierfür sind zusammenfassend in Stein (1997) dargestellt. Deshalb ist man auf Ad-hoc-Methoden oder auf das parametrische Bootstrapverfahren angewiesen, um die Anzahl der Komponenten zu testen. Als Beispiel für ein Ad-hoc Verfahren ist das von Wolfe (1971) zu nennen. Wolfe fand in einer Simulationsstudie mit zwei Komponenten, dass die Verteilung der Likelihood-Ratio-Statistik  $\lambda$  unter der Nullhypothese  $G = 1$  Komponenten durch eine  $\chi^2_d$  - Verteilung mit  $d = 2(d_1 - d_0)$  Freiheitsgraden approximiert werden kann.

Die Nullhypothese wird auf dem Signifikanzniveau  $\alpha$  zurückgewiesen, wenn

$$\lambda > \chi^2_{(G+1)(d_1-d_0)}.$$

Statistisch korrekt, aber sehr aufwendig ist die Anwendung eines parametrischen Bootstrap-Verfahrens zur Schätzung der Verteilung von  $\lambda$ . Anwendungen des parametrischen Bootstraps finden sich in Stein (1997) sowie Arminger, Stein und Wittenberg (1999).

**Tabelle 1: Ergebnisse des Tests auf Anzahl der Komponenten**

Anzahl der Komponenten	Likelihood-Funktion für $K$	Likelihood-Ratio Teststatistik	df	$\pi$	$\alpha$
1	-31738.146	1042.472	180	0.318	0.000
2	-31216.910	531.389	270	0.448	0.000
3	-30951.216	197.054	360	0.234	1.000

Tabelle 1 zeigt die Ergebnisse eines Ad-hoc Tests auf die Anzahl der Komponenten. Jede Zeile der Tabelle enthält die Likelihood-Funktion und die Likelihood-Ratio-Teststatistik für ein spezifisches  $K$ . Der Likelihood-Ratio-Test testet die Nullhypothese  $H_0 : K = G$  Gruppen gegen die Alternativhypothese  $H_1 : K = G + 1$  Gruppen.

Die Tabelle zeigt, dass für alle  $K$  außer  $K = 3$  die Nullhypothese zurückgewiesen werden muss, da alle Überschreitungswahrscheinlichkeiten  $\alpha$  kleiner als das Signifikanzniveau 0.05 sind. Da für  $K = 3$  gilt, dass  $\alpha = 1.000$ , kann die Nullhypothese  $K = 3$  nicht zurückgewiesen werden. Dies bedeutet, dass die Population in bezug auf die betrachteten Variablen heterogen ist und mutmaßlich aus drei verschiedenen Gruppen besteht.

Wie die Tabelle weiter zeigt, ergeben sich für die geschätzten mischenden Wahrscheinlichkeiten  $\pi$  die Werte  $\pi_1 = 0.318$ ,  $\pi_2 = 0.448$  und  $\pi_3 = 0.234$ . Dies bedeutet bei insgesamt in die Analyse einbezogenen  $N=2.667$  Befragten, dass die erste Gruppe aus 848 Befragten, die zweite Gruppe aus 1.195 Befragten und die dritte Gruppe aus 624 Befragten besteht.

### 3.2 Beschreibung der Komponenten durch unbedingte Mittelwerte

Einen ersten Eindruck der Komponenten erhält man, wenn man sich die unbedingten Mittelwerte der Variablen näher ansieht. Tabelle 2 gibt für jede Gruppe einen Überblick über die Mittelwerte aller Variablen.

**Tabelle 2: Unbedingte Mittelwerte**

Variablen	Gruppe 1	Gruppe 2	Gruppe 3
<b>Exogene Variablen:</b>			
GEBIET (Ost/West)	0.651	0.684	0.721
BILDUNG	0.234	0.209	0.207
ALTER	4.907	4.795	4.740
GESCHLECHT	0.526	0.536	0.484
<b>Endogene Variablen:</b>			
KUEMM (V150)	1.775	2.021	2.072
KOMPLEX (V153)	2.782	2.679	2.610
INTER (V154)	2.873	2.730	2.916
UEBERLASS (V157)	2.893	2.664	2.569
VOLKSAB (V160)	0.966	1.647	2.151
DEMOGES (V161)	0.976	1.825	2.249
ZUFDEMO (V162)	3.234	2.622	3.671
REGLEIST (V163)	4.270	3.957	4.291
POLINT (V169)	2.732	2.987	2.994

Gruppe 1 zeichnet sich gegenüber den anderen beiden Gruppen durch eine besonders niedrige mittlere Einschätzung der Responsivität der Politiker (Variable KUEMM) bei gleichzeitig hohen Partizipationsnormen bzw. starke Forderung nach „mehr Demokratie“ aus. Entsprechend sind die mittleren Zufriedenheitswerte mit der Demokratie in der BRD und den Leistungen der Regierung am geringsten. Sie umfasst die geringste Anzahl von Westdeutschen.

Gruppe 3 weist ähnlich geringe Zufriedenheiten mit der Demokratie in der Bundesrepublik und den Leistungen der Bundesregierung bei eher niedrigen Partizipationsnormen aus. Die Gruppe sieht die Verantwortung für die Politik eher bei den Volksvertretern und steht einer Ausweitung der demokratischen Rechte eher skeptisch gegenüber. Es liegt die Vermutung nahe, dass die Unzufriedenheiten dieser Gruppe mit der Demokratie in der BRD eher auf Vorbehalte gegenüber der bestehenden politischen Ordnung zurückgehen. Angesichts der politischen Situation in der Bundesrepublik vor der Bundestagswahl 1998 könnte die Unzufriedenheit auch als Unzufriedenheit mit der Art und Weise gewertet werden, wie die Bundesregierung die ihr obliegende politische Verantwortung wahrgenommen hat. Diese Gruppe hat den geringsten Umfang und enthält die meisten Männer und die meisten Befragten aus den alten Bundesländern.

Gruppe 2 besteht aus der höchsten Anzahl von Befragten und weist die höchsten mittleren Zufriedenheitswerte auf. Sie stellen stärkere demokratische Forderungen als Gruppe 3. Ihr politisches Interesse ist größer als das von Gruppe 1 und ähnlich hoch wie das von Gruppe 3.

### **3.3 Ergebnisse der Modellanpassung**

#### **3.3.1 Anpassungsbeurteilung**

Im Folgenden werden die Ergebnisse der simultanen Anpassung der kausalen Modelle in den einzelnen Gruppen dargestellt (vgl. Tabelle 3). Die Modellanpassung basiert auf einer Minimum-Distanz-Schätzung, bei der die Strukturparameter der Kovarianzstrukturmodelle (z.B. Ladungen der Indikatoren auf den latenten Variablen, die Fehlervarianzen, die Strukturkoeffizienten, etc.) in den Gruppen so geschätzt werden, dass die Distanz zwischen den Parameterschätzungen der Regression der exogenen Variablen auf die endogenen Variablen und den durch die Strukturparameter vorhergesagten Regressionsparameter ein Minimum ergeben.

**Tabelle 3: Anpassungsindizes des Gesamtmodells**

$\chi^2_M$	NFI	NNFI	CFI	RMSEA
896.000 (df = 88)	0.998	0.995	0.998	0.520

In der Tabelle bezeichnen  $\chi^2_M$  den  $\chi^2$ -Wert des betrachteten Modells, NFI den Normed Fit Index, NNFI den Nonnormed Fit Index, CFI den Comparative Fit Index und RMSEA den Root Mean Square Error of Approximation.

Ein Vergleich von  $\chi^2_M$  mit dem  $1 - \alpha$  - Quantil einer zentralen  $\chi^2$  - Verteilung mit df = 88 Freiheitsgraden ergibt, dass die Überschreitungswahrscheinlichkeit kleiner als 0.5 ist, was zu einer Zurückweisung des Modells führen müsste.

Auf der anderen Seite wäre ein anderes Ergebnis aufgrund des großen Stichprobenumfangs auch nicht zu erwarten gewesen. Die Höhe des RMSEA deutet auf eine weitere Verbesserungsfähigkeit des Gesamtmodells hin. Angesichts der Größe von NFI, NNFI und CFI nahe 1.000 und angesichts des Sachverhalts, dass die residualen Kovarianzen der hier nicht abgebildeten residualen Kovarianzmatrix ausreichend klein sind, kann die Modellanpassung als akzeptabel eingestuft werden, zumal davon auszugehen ist, dass weitere Modellmodifikationen die Stichprobenabhängigkeit der Ergebnisse verstärkt hätten.

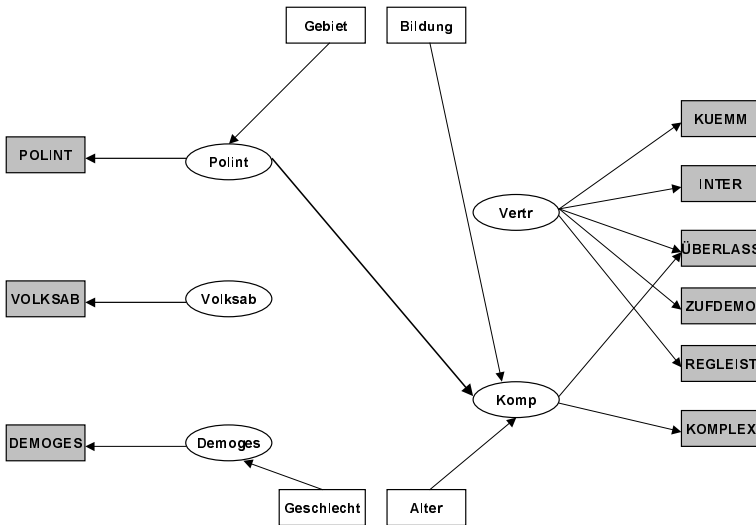
Bei der Anwendung von Anpassungsindizes zur Beurteilung der Anpassungsgüte der Kovarianzstrukturmodelle in den Gruppen muss die besondere Form der Minimum-Distanz-Schätzung für das oben beschriebene konditionale Mischverteilungsmodell berücksichtigt werden. Inkrementelle Anpassungsindizes müssen an die konditionale Modellspezifikation angepasst werden (vgl. Stein 1997). Bei den inkrementellen Anpassungsindizes NFI (Normed Fit Index), NNFI (Nonormed Fit Index) und CFI (Comparative Fit Index), die im Rahmen der traditionellen Analyse von Kovarianzstrukturen verwendet werden, wird die Anpassung des analysierten Modells mit der des Unabhängigkeitsmodell verglichen, bei dem nur die Diagonalelemente der Kovarianzmatrix der beobachteten Variablen als von Null verschieden angenommen werden. Implizit werden dabei die Mittelwerte freigesetzt, da nur die Matrix der zweiten Momente betrachtet wird. Im Fall einer Mischverteilung werden die Regressionskoeffizienten freigesetzt, die bedingten Kovarianzmatrizen  $\Sigma_k$  werden als Diagonalmatrizen angenommen. Schließlich wird für alle Komponenten angenommen, dass die Regressionskoeffizienten gleich sind. Dies entspricht der

Hypothese einer Homogenität der Wirkungen der abhängigen auf die unabhängigen Variablen.

### 3.3.2 Ergebnisse für Gruppe 1

Im Folgenden sollen die einzelnen Komponenten hinsichtlich ihrer auf die exogenen Variablen bedingten kausalen Modelle charakterisiert werden. Dabei werden zunächst die Ergebnisse für die erste Gruppe vorgestellt. In Abbildung 1 ist das Kausalmodell für diese Gruppe graphisch dargestellt. Wie aus der Abbildung, in die nur die statistisch signifikanten Einflussbeziehungen aufgenommen worden sind, ersichtlich, enthält dieses Modell insgesamt fünf endogene Faktorvariablen, von denen zwei latente Variablen mit jeweils fünf bzw. zwei Indikatoren darstellen.

Abbildung 1: Kausalmodell für Gruppe 1



Um eine Interpretation der latenten Variablen vornehmen zu können, ist eine Inspektion der Ladungen der Indikatoren unerlässlich. Tabelle 4 enthält die Ladungen der Indikatoren, die Varianzen der Messfehler (in der Spalte „Fehler“) und die addi-

tiven Konstanten. Die Kovarianzen zwischen unterschiedlichen Indikatoren wurden auf Null gesetzt, so dass über diese nicht berichtet werden muss. Wie Tabelle 4 noch einmal zeigt, besteht das Messmodell von Gruppe 1 aus fünf Faktorvariablen, wobei die Variablen **Volksab**, **Demoges** und **Polint** identisch sind mit den entsprechenden beobachteten Indikatoren VOLKSAB (V160), DEMOGES (V161) und POLINT (V169).

**Tabelle 4: Messmodell für Gruppe 1**

Items	Konstante	Ladungen					Fehler
		Vertr	Komp	Volkab	Demoges	Point	
KUEMM	1.650* (18.383)	1.000	–	–	–	–	0.479 (13.598)
KOMPLEX	2.781* (30.785)	–	1.000	–	–	–	0.748 (13.102)
INTER	3.085 (26.858)	-1.373* (-5.796)	–	–	–	–	0.344 (8.119)
UEBERLASS	3.186* (28.045)	-0.526* (-2.787)	1.2484* (3.757)	–	–	–	0.617 (10.157)
VOLKSAB	0.894* (73.283)	–	–	1.000	–	–	0.000 (0.000)
DEMOGES	0.982* (77.283)	–	–	–	1.000	–	0.000 (0.000)
ZUFDEMO	3.537* (28.124)	-1.478* (-5.984)	–	–	–	–	0.000 (0.000)
REGLEIST	4.449* (39.592)	-1.184* (-5.407)	–	–	–	–	0.000 (0.000)
POLINT	2.754* (43.591)	–	–	–	–	1.000	0.000 (0.000)

\*signifikant auf dem 5%-Niveau

Beim ersten Faktor „Vertrauen“ (**Vertr**) handelt es sich um eine latente Variable, welche die Variablen der Zufriedenheit mit Demokratie und Regierungsleistungen sowie die Variablen INTER (V154), ÜBERLASS (V157) schwach negativ, die Variable KUEMM (V150) aber positiv beeinflusst. Die Ausprägungen der Individuen auf diesem Faktor steigen mit wachsender Zufriedenheit mit Regierung und Demokratie, mit zunehmender Intensität der Überzeugung, dass Politiker die Interessen der Bevölkerung vertreten, dass man die Politik den Politikern überlassen sollte und dass Politiker sich darum kümmern, was die Leute denken, steigen. Dieser Faktor

scheint wohl am ehesten eine Intensitätsdimension des allgemeinen Vertrauens in die Politik und die Politiker der zum Zeitpunkt der Umfrage regierenden Koalition und die durch sie in der Bundesrepublik praktizierten Demokratie zu repräsentieren. Der Aspekt des Vertrauens manifestiert sich im Kontext der übrigen Indikatoren vor allem im Grad der Zustimmung zur Aussage, dass man die Politik den gewählten Volksvertretern überlassen sollte. Dieser Indikator fehlt in den übrigen Gruppen. Eine Erklärung hierfür wäre, dass die Variable V157 in einer Gruppe besonders akzentuiert wird, die verstärkt für die Ausdehnung demokratischer Rechte eintritt. In einer nur geringen Intensität der Zustimmung zu diesem Item drücken sich Zweifel an der universellen Kompetenz der Volksvertreter und eine Betonung eigener politischer Handlungskompetenzen aus. Durch die Einbettung in die Menge der übrigen Indikatoren bekommt eine Abnahme des Vertrauens in die Politiker die zusätzliche Konnotation einer Abnahme des Vertrauens in deren Handlungskompetenz.

Der zweite als **Komp** bezeichnete Faktor ist positiv korreliert mit den Indikatoren KOMPLEX und ÜBERLASS; d.h. die Ausprägungen auf dieser latenten Variablen steigen mit abnehmender Zustimmung, dass die Politik zu kompliziert ist und dass man die Politik den Politikern überlassen sollte. Er repräsentiert wohl am ehesten den Glauben an die eigene Kompetenz bei starker Betonung der kognitiven Kompetenz. Insofern stellt er eine Komponente der Internen Efficacy dar.

Tabelle 5 gibt einen Überblick über die Einflussstärken der exogenen Variablen. Sie zeigt einen positiven Einfluss des Erhebungsgebiets auf den ersten Faktor. Unter Zugrundelegung der obigen Interpretation dieses Faktors als Faktor des Vertrauens in die Politik würde dies bedeuten, dass dieses Vertrauen 1998 im Westen stärker ausgeprägt war als im Osten. Wie zu erwarten hat die Bildung einen positiven Einfluss auf den Faktor **Komp** sowie einen negativen Einfluss auf das politische Interesse. Das Geschlecht hat in Gruppe 1 einen positiven Einfluss auf das politische Interesse in dem Sinne, dass das politische Interesse der Frauen leicht geringer ist als das der Männer.

Die Einflüsse der Faktorvariablen ergeben sich aus Abbildung 1. Als einziger signifikanter Einfluss ergab sich ein negativer Einfluss des politischen Interesses auf den Kompetenzfaktor ( $\beta_{32} = -0.109; t = -3.265$ ).

Dies würde bedeuten, dass mit zunehmendem politischen Interesse der Glaube an die eigene politische Kompetenz steigt.

**Tabelle 5: Wirkungen der exogenen Variablen in der Gruppe 1**

	Exogene Variablen			
	<i>GEBIET</i>	<i>BILDUNG</i>	<i>ALTER</i>	<i>GESCHLECHT</i>
<b>Vertr</b>	0.181* (3.364)	0.095 (1.081)	0.005 (0.350)	0.000 (0.000)
<b>Komp</b>	0.000 (0.000)	0.309* (3.670)	-0.035* (-2.233)	0.000 (0.000)
<b>Volksab</b>	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<b>Demoges</b>	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<b>Polint</b>	0.000 (0.000)	-0.441* (-4.341)	0.000 (0.000)	0.215* (2.769)

\*signifikant auf dem 5%-Niveau

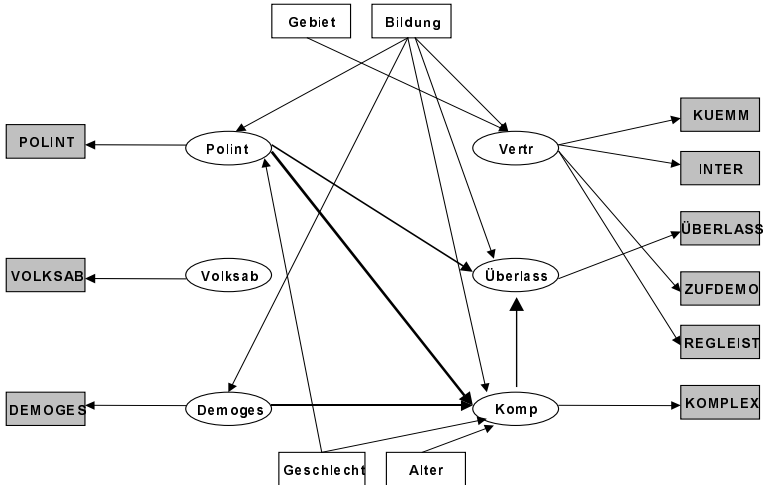
### 3.3.3 Ergebnisse für Gruppe 2

In Abbildung 2 ist das Kausalmodell für Gruppe 2 dargestellt. Wie aus ihr und den Ladungen des Messmodells in Tabelle 6 zu ersehen ist, ergab sich für diese Gruppe nur noch *ein* Faktor mit multiplen Indikatoren. Er besitzt bis auf den Indikator ÜBERLASS (V157) die gleichen Indikatoren wie in Gruppe 1, deren Ladungen auch das gleiche Vorzeichen aufweisen. Der Vertrauensfaktor repräsentiert damit nur noch das Vertrauen in die Responsivität der Politiker, aber nicht mehr das Vertrauen in deren Handlungskompetenz. Die Variable ÜBERLASS repräsentiert nunmehr eine eigenständige, vom Faktor Vertr unabhängige Faktorvariable Überlass.

Der Kompetenzfaktor fällt nunmehr mit der beobachteten Variablen KOMPLEX zusammen, repräsentiert also die rein kognitive, informative Kompetenz. Wir sehen bereits an dieser Stelle, wie heterogen sich die Population hinsichtlich der Messmodelle erweist.



**Abbildung 2: Kausalmodell für Gruppe 2**



**Tabelle 6: Messmodell für Gruppe 2**

Items	Konstante	Ladungen						Fehler
		Vertr	Komp	Volksab	Demoges	Polint	Überlass	
KUEMM	1.714* (30.410)	1.000	-	-	-	-	-	0.390 (10.873)
KOMPLEX	2.941* (29.904)	-	1.000	-	-	-	-	0.000 (0.000)
INTER	3.087* (61.027)	-0.936* (-9.002)	-	-	-	-	-	0.324* (11.282)
ÜBERLASS	2.805* (44.826)	-	-	-	-	-	1.000	-37275* (342.111)
VOLKSAB	1.629* (60.783)	-	-	1.000	-	-	-	0.000 (0.000)
DEMOGES	1.872* (70.734)	-	-	-	1.000	-	-	0.000 (0.000)
ZUFDEMO	2.884* (55.625)	-0.921* (-8.030)	-	-	-	-	-	0.000 (0.000)
REGLEIST	4.262* (66.601)	-0.653* (-5.432)	-	-	-	-	-	0.000 (0.000)
POLINT	2.839* (44.069)	-	-	-	-	1.000	-	0.000 (0.000)

\*signifikant auf dem 5%-Niveau

Tabelle 7 zeigt die Höhe der in Abbildung 2 dargestellten Einflüsse der exogenen Variablen auf die endogenen Faktorvariablen. Wie in Gruppe 1 ergibt sich ein Einfluss des Erhebungsgebiets auf **Vertr**. Die Variable Bildung hat positive Einflüsse auf **Vertr**, **Komp** und auf **Überlass**; d.h. mit zunehmender Bildung steigt in dieser Gruppe der Glaube an die Responsivität der Politiker und die Überzeugung, politische Vorgänge zu verstehen bei gleichzeitig steigender Skepsis, den Politikern das Feld der Politik zu überlassen. Negative Einflüsse von Bildung ergeben sich für **Demoges** und **Polint**; d.h. mit steigender Bildung steigen der Wunsch nach Ausdehnung der Demokratie auf weitere Bereiche außerhalb der Politik sowie das politische Interesse. Das Alter hat einen leicht negativen Einfluss auf **Komp**; d.h. mit zunehmendem Alter wird nach Selbsteinschätzung die Politik für den Einzelnen undurchschaubarer. Die Variable Geschlecht hat ebenfalls einen negativen Einfluss auf **Komp** sowie einen positiven Einfluss auf **Polint**; d.h. Frauen bekennen sich eher zu einem mangelnden Verständnis politischer Vorgänge als Männer und sind nach Selbsteinschätzung politisch weniger interessiert.

**Tabelle 7: Wirkungen der exogenen Variablen in der Gruppe 2**

	Exogene Variablen			
	<i>GEBIET</i>	<i>BILDUNG</i>	<i>ALTER</i>	<i>GESCHLECHT</i>
<b>Vertr</b>	0.365* (6.515)	0.144* (2.285)	0.000 (0.000)	0.000 (0.000)
<b>Komp</b>	0.000 (0.000)	0.346* (4.091)	-0.039* (-2.2871)	-0.149* (-2.298)
<b>Volksab</b>	0.000 (0.000)	-0.036 (-0.704)	0.000 (0.000)	0.000 (0.000)
<b>Demoges</b>	0.000 (0.000)	-0.218* (-4.394)	0.000 (0.000)	0.000 (0.000)
<b>Polint</b>	-0.101 (-1.614)	-0.519* (-6.500)	0.000 (0.000)	0.485* (8.205)
<b>Überlass</b>	-0.097 (-1.357)	0.201* (2.525)	0.000 (0.000)	0.000 (0.000)

\* signifikant auf dem 5%-Niveau

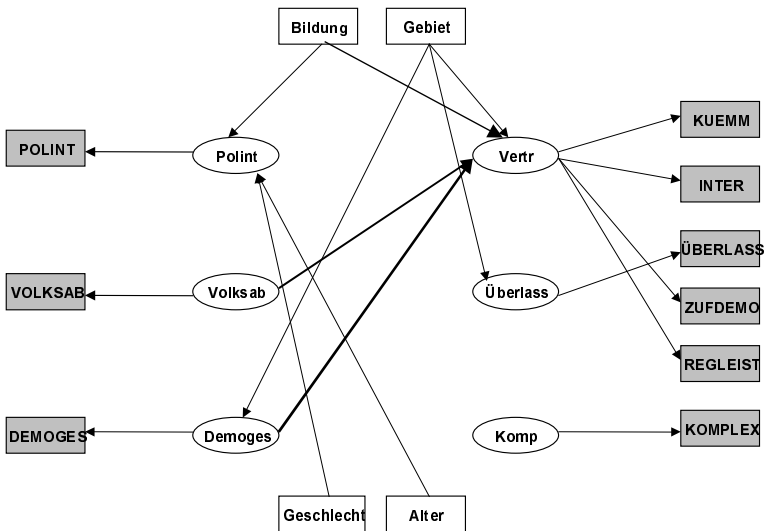
Die Struktur der Abhängigkeiten zwischen den endogenen Variablen geht aus Abbildung 2 hervor. Die Höhe der Einflüsse der endogenen Variablen untereinander sind für diese Gruppe ebenfalls nicht in einer separaten Tabelle dargestellt. Die mit der Variablen KOMPLEX (V153) identische Faktorvariable **Komp** hat einen signifikant po-

sitiven Effekt auf die Variable **Überlass** (V157), ( $\beta = 0.113; t = 2.921$ ). D.h.: Je stärker das Vertrauen in das Verständnis politischer Vorgänge, desto geringer die Bereitschaft, die Politik den gewählten Volksvertretern zu überlassen. Die Variable DEMOGES (V161) hat einen signifikant negativen Einfluss auf **Komp**; d.h. je stärker die Meinung vertreten wird, die Demokratie sollte auf weitere Bereiche ausgedehnt werden, desto größer das Vertrauen an den eigenen politischen Durchblick. Die Variable politisches Interesse hat einen negativen Einfluss auf **Komp** und **Überlass**; d.h. mit steigendem politischem Interesse steigen der Glaube an die Verstehbarkeit politischer Vorgänge und die Skepsis, die Verantwortung für Politik den gewählten Volksvertretern zu überlassen.

### 3.3.4 Ergebnisse für Gruppe 3

Abbildung 3 zeigt das Kausalmodell für Gruppe 3. Wie zu erkennen ist, ist die Indikatorenstruktur der Messmodelle identisch mit der für Gruppe 2.

**Abbildung 3: Kausalmodell für Gruppe 3**



Die Ladungen sind in Tabelle 8 wiedergegeben. Wie zu ersehen ist, weisen die Indikatoren für den ersten Faktor die gleichen Vorzeichen wie in Gruppe 2 auf. Auch die Größenverhältnisse sind innerhalb der beiden Gruppen identisch, so dass sich eine ähnliche Interpretation der Messmodell ergibt.

Die Einflussstärken der in Abbildung 3 dargestellten Wirkungen der exogenen Variablen auf die Faktorvariablen finden sich Tabelle 9. Wie man bereits der Abbildung 2 entnehmen kann, ergaben sich häufiger signifikante Einflüssen des Erhebungsbereichs. Letzteres wirkt sich nicht nur wie in Gruppe 2 auf **Vertr** aus, sondern auch auf **Demoges** (V161) und **Überlass** (V157), wobei der Einfluss auf **Überlass** negativ ist. Wir haben in dieser Gruppe also die Situation, dass in den alten Bundesländern einer Ausweitung der Demokratie auf weitere Bereiche in geringerem Ausmaß zugestimmt wird als in den neuen Bundesländern und dass in den alten Bundesländern die Skepsis gegenüber den Volksvertretern größer ist.

**Tabelle 8: Messmodell für Gruppe 3**

Items	Konstante	Ladungen						Fehler
		Vertr	Komp	Volksab	Demoges	Polint	Überlass	
KUEMM	1.854* (24.540)	1.000	-	-	-	-	-	0.636* (8.480)
KOMPLEX	2.600* (32.938)	-	1.000	-	-	-	-	0.000 (0.000)
INTER	3.125* (45.256)	-1.193* (-5.864)	-	-	-	-	-	0.559* (11.005)
ÜBERLASS	2.790* (27.424)	-	-	-	-	-	1.000	0.000 (0.000)
VOLKSAB	2.087* (21.284)	-	-	1.000	-	-	-	0.000 (0.000)
DEMOGES	2.023* (19.801)	-	-	-	1.000	-	-	0.000 (0.000)
ZUFDEMO	4.387* (30.085)	-3.342* (-5.986)	-	-	-	-	-	0.000 (0.000)
REGLEIST	4.931* (35.863)	-3.014* (-5.986)	-	-	-	-	-	0.000 (0.000)
POLINT	3.645* (24.640)	-	-	-	-	1.000	-	0.000 (0.000)

\* signifikant auf dem 5%-Niveau

Die Bildung wirkt sich dagegen in dieser Gruppe nur auf **Vertr** und auf **Polint** aus. Die Wirkung der Variablen **ALTER** wirkt sich im Unterschied zu den übrigen auf das politische Interesse in der Richtung aus, dass das politische Interesse mit zunehmendem Alter stärker wird. Das Geschlecht hat wie in Gruppe 1 nur eine Wirkung auf politisches Interesse.

**Tabelle 9: Wirkungen der exogenen Variablen in der Gruppe 3**

	<b>Exogene Variablen</b>			
	<i>GEBIET</i>	<i>BILDUNG</i>	<i>ALTER</i>	<i>GESCHLECHT</i>
<b>Vertr</b>	0.222* (4.101)	0.133* (2.512)	0.000 (0.000)	0.000 (0.000)
<b>Komp</b>	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<b>Volksab</b>	0.183* (1.774)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<b>Demoges</b>	0.273* (2.546)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<b>Polint</b>	0.000 (0.000)	-0.795* (-5.412)	-0.128* (-4.872)	0.444* (4.393)
<b>Überlass</b>	-0.295* (-2.533)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

Was die Beziehungen der endogenen Variablen untereinander angeht, so sind die signifikanten, in Abbildung 3 dargestellten Einflüsse der Faktoren **Volksab** (V160) ( $\beta_{31} = 0.09; t = 3.591$ ) und **Demoges** (V161) ( $\beta_{41} = 0.080; t = 3.111$ ) auf **Vertr** zu schwach, um in eine inhaltliche Interpretation einbezogen zu werden. Alle anderen Einflussgrößen liegen unterhalb der Signifikanzgrenze.

#### 4. Zusammenfassung und Interpretation

Ziel der hier vorgestellten Analyse war es, die Anwendung eines Mischverteilungsansatzes zur Aufdeckung unbekannter Gruppierungen mit noch unbekanntem gruppenspezifischen kausalen Strukturen, insbesondere gruppenspezifischen Messmodellen in den Daten des ALLBUS 1998 vorzustellen. Die für die Analyse ausgewählten Variablen standen im Zusammenhang mit dem politischen Teilnahmeverhalten und seinen einstellungsbezogenen Voraussetzungen. Einbezogen wurden auch die Zufriedenheiten mit Regierungsleistungen und Demokratie. Das Verfahren ist in der Lage, mit den Gruppen zugleich auch deren Kausalstruktur aufzudecken und damit herauszufinden, ob eine Population hinsichtlich der Konstruktvaliditäten von Variablen homogen ist.

Die Ergebnisse der Analyse zeigen eindrucksvoll, wie heterogen die Population hinsichtlich der Messmodelle und der Abhängigkeitsstrukturen einzuschätzen ist. Auf-

gefunden werden konnten drei Gruppen, wobei bereits vor einer Beurteilung der Kausalstruktur vorgenommene Inspektion der Mittelwerte der in die statistische Analyse einbezogenen Variablen deutlichen Gruppenunterschiede offenbarte.

Die erste Gruppe zeichnete sich durch eine eher geringe mittlere Einschätzung der Responsivität der Politiker, repräsentiert durch die Variable KUEMM (V150) aus. Die mittleren Zufriedenheiten mit Regierungsleistungen und praktizierter Demokratie sind in dieser Gruppe ebenfalls am geringsten. Zugleich werden die Forderungen nach einer Ausdehnung der Demokratie und Volksabstimmung in dieser Gruppe stärker vertreten. Diese Gruppe fordert also mehr Teilnahme, wobei offen bleiben muss, ob die geringe Einschätzung der Responsivität hierfür ursächlich ist. Eine Charakterisierung der Messmodelle in dieser Gruppe ergab zwei latente Variablen, die als Vertrauen in die Politiker und die Politik sowie als Selbstzuschreibung von politischer Eigenkompetenz durch die Befragten interpretiert wurden. Es ist für diese Gruppe kennzeichnend, dass in den Konnotationen der latenten Variablen die Handlungskompetenz immer mit akzentuiert wird. Der Grad der Selbstzuschreibung politischer Kompetenz hängt dabei signifikant vom politische Interesse ab.

Diese Struktur findet sich in den übrigen Gruppen nicht wieder. Die Variable V157 („Politik sollte man den gewählten Volksvertretern überlassen“) findet sich nunmehr als eigenständige Variable wieder und verliert ihre Funktion als Indikator für irgendein Konstrukt. Dadurch verliert der Kompetenzfaktor den Aspekt der Handlungskompetenz und wird zur rein kognitiven Kompetenz. Allerdings ergab sich für Gruppe 2, dass mit zunehmender Selbstzuschreibung kognitiver Kompetenz der Grad von Zuschreibung von Handlungskompetenz an die Politiker abnimmt; d.h. in Gruppe 2 ergibt sich ein einseitig gerichteter Zusammenhang zwischen beiden Variablen.

Wie die Betrachtung der Mittelwerte der analysierten Variablen ergab, weist Gruppe 3 wie Gruppe 1 Zufriedenheiten mit der Demokratie und den Regierungsleistungen auf, die über der in Gruppe 2 liegen. Im Unterschied zu Gruppe 1 ist Gruppe 3 aber eher gegen Ausweitungen demokratischer Rechte. Das Messmodell sieht ähnlich aus wie in Gruppe 2. Im Unterschied zu den beiden übrigen Gruppen ergibt sich aber kein Zusammenhang mehr zwischen der Selbstzuschreibung eigener kognitiver Kompetenz und der Zuschreibung von Handlungskompetenz an die Politiker. Offensichtlich hat hier die Selbstzuschreibung keine Referenz mehr auf den Mangel an Politikerkompetenz. Die Vermutung liegt nahe, dass es sich bei dieser Gruppe um die politisch eher Passiven und Verzagten handelt.

Zusammenfassend ergibt sich, dass die unbeobachtete Heterogenität bei teilnahmebezogenen Variablen eine Rolle spielt und nicht davon ausgegangen werden kann,

dass in allen Gruppierungen die gleiche Abhängigkeits- und Indikatorenstruktur vorliegt. Leider müssen die Interpretationen der Faktoren noch als vorläufig gelten. Erst die Einbeziehung weiterer Indikatoren und eine Charakterisierung durch weitere Merkmale würde hier mehr Interpretationssicherheit bringen.

## Korrespondenzadresse

*Prof. Dr. Frank Faulbaum*  
*Gerhard-Mercator-Universität Duisburg*  
*FB 1 Soziologie*  
*Lotharstraße 65*  
*D-47048 Duisburg*  
*Tel.: 0203/379-2532 (-2804 Sekretariat)*  
*Fax: 0203/379-4350*  
*E-Mail: [faulbaum@uni-duisburg.de](mailto:faulbaum@uni-duisburg.de)*

## Literatur

- Arminger, G./Stein, P., 1997: Finite mixtures of covariance structure models with regressors. *Sociological Methods & Research* 26, 148-182.
- Arminger, G./Stein, P./Wittenberg, J., 1999: Mixtures of conditional mean- and covariance structure models. *Psychometrika* 64, 475-494.
- Balch, G.I., 1974: Multiple indicators in survey research: The concept of "Sense of Political Efficacy". *Political Methodology* 1, 1-43.
- Bollen, K. A., 1989: *Structural equations with latent variables*. New York: Wiley.
- Campbell, A./Gurin, G./Miller, W.F., 1954: *The voter decides*. New York: Wiley.
- DeSarbo, W. S./Cron, W. L., 1988: A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification* 5, 249-282.
- Faulbaum, F./Stein, P., 2000: Wie homogen sind Einstellungen gegenüber Ausländern? Zur Aufdeckung und Modellierung unbeobachteter Heterogenität in Umfragedaten. S. 486-518 in: Alba, R./Schmidt, P./Wasmer, M. (Hrsg.), *Deutsche und Ausländer: Freunde, Fremde oder Feinde?* Wiesbaden: Westdeutscher Verlag.
- Jedidi, K./Jagpal, H.S./DeSarbo, W.S., 1997: Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science* 16, 1997, 39-59.
- Lane, R.E., 1959: *Political life. Why people get involved in politics*. Glencoe: Free Press.

- Loehlin, J.C., 1998: *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Mahwah: Lawrence Erlbaum Associates.
- McLachlan, G. J./Basford, K.E., 1988: *Mixture Models*. New York: Marcel Dekker.
- Stein, P., 1997: *Konstruktion und sozialwissenschaftliche Anwendung finiter Mischungen von Kovarianzstrukturmodellen*. Köln: Eul-Verlag.
- Stein, P., 2000: Modelle zur Aufdeckung unbeobachteter Heterogenität bei der Erklärung von Lebenszufriedenheit. *Zeitschrift für Soziologie* 29, 138-159.
- Vetter, A., 1997: *Political Efficacy: Alte und neue Meßmodelle im Vergleich*. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49, 53-73.
- Wolfe, J.H., 1971: *A Monte Carlo Study of the sampling distribution of the likelihood ratio or mixtures of multinormal distributions*. Technical Bulletin STB 72-2. San Diego: U.S. Naval Personnel and Training Research Laboratory.
- Yung, Y.F., 1994: *Finite mixtures in confirmatory factor-analytic models*. Dissertation, University of California at Los Angeles, Department of Psychology.
- Yung, Y.F., 1997: Finite mixtures in confirmatory factor-analysis models. *Psychometrika* 62, 297-330.



---

# WHY IT IS EASY TO WRITE BAD QUESTIONS<sup>1</sup>

FLOYD JACKSON FOWLER JR.

There are at least 8 standards that survey questions should meet. One reason it is easy to write poor questions is that researchers focus on some standards but ignore others. A more complex problem is posed by the fact that designing a question that is good according to one standard can make it a poor question when judged against another standard. Third, typical pretest and question evaluation procedures do not provide information about all 8 standards. It is easy to ignore a standard if there is no information about it. Finally, researchers often are committed to questions that have been used previously, even when there is evidence that they are flawed. This paper presents the 8 standards, the ways in which they can be in conflict, the challenges of evaluating questions, and the implications for standard survey practice.

## 1. Introduction

“Total survey design” refers to the idea that in designing or evaluating a survey project, it is important to consider all aspects of the data collection protocol, sampling, data collection procedures, interviewing, non-response, and question design, not just a few. The quality of survey data will be no better than the worst aspect of the data collection protocol, considering all aspects of it (Groves, 1989; Fowler, 1993).

In the area of survey question design and evaluation, a comparable construct, “total survey question design”, may be appropriate. One of the important reasons that designing good questions is difficult is that there are numerous standards that survey questions should meet, and they are not necessarily related to one another. In a manner parallel to the concept of total survey design, if a question proves to be in-

---

<sup>1</sup> This manuscript was prepared while the author was a guest professor at the Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, Germany. ZUMA’s hospitality and support is warmly acknowledged.

adequate with respect to any one of these design standards, it is likely to be a poor survey question.

The purpose of this paper is to lay out 8 standards that survey questions should meet, to briefly describe why each is important, and then to discuss the implications of applying these 8 standards for the design of good survey questions.

## **2. The 8 Standards**

The following are 8 standards that survey questions need to meet.

### **Content standards**

Researchers have to decide what questions to ask, what content to ask about. Although that might seem an obvious standard, one that would be hard to fail to meet, in fact methodologists often find that researchers have not thought through what they really should ask about.

Decisions about question content emerge from an integration of two kinds of information. First, researchers need a set of analytic objectives, analytic questions that they want to answer based on the survey data they collect. Second, they need to find out what respondents have to say, what they are able and willing to tell them, that can provide the information they need. Often, a result of careful pre-survey evaluation of questions is that researchers become more informed about what questions they should ask in order to gather the information they need.

The content standard has two components:

- a) Whether or not the answers to the questions asked will meet the analytic objectives;
- b) Whether or not the questions are the right ones to get respondents to provide the information they have that will enable the researchers to address the analytic questions.

### **Cognitive standards**

The official beginning of the cognitive aspects of survey methodology (CASM) movement in the United States is usually identified as 1983. In that year, an advanced seminar on CASM was convened, bringing together statisticians, survey researchers, and researchers from the cognitive sciences to explore their mutual interests. A similar event was held at ZUMA in 1984. The results of these two conferences

were published by Jabine/Tanur/Tourangeau (1984) and by Hippler/Schwarz/Sudman (1987).

Since those conferences, and the related publications, there has been growing agreement that questions should meet cognitive standards before they are used in surveys (see Sirken, et al. 1999). Cognitive standards include:

- a) That questions should be consistently understood by all respondents.
- b) That respondents have access to, that is they know or can remember, the information required to answer the questions.
- c) That the answers given to the questions accurately reflect the reality that respondents are being asked to describe.

### **Interpersonal Standards**

A survey instrument is not only a set of questions for respondents to answer. When it is used by an interviewer, it becomes a script for (at least one side of) an interaction. Thus, the fact that the questions will be used in an interpersonal context, in a particular way to collect data, produces another set of issues that need to be addressed in designing a survey instrument.

- a) The questions need to be designed so that they can be asked exactly as written. When the questions are read orally, they should be clear enough that the respondent is ready to answer the question. Questions that require extensive probing, or that lead to respondents to frequently ask for clarification, have been shown to have adverse effects on data quality (Fowler/Mangione 1990; Fowler/Cannell 1996).
- b) The interview schedule also must produce an interaction that seems reasonable and appropriate to respondents. Instruments that seem repetitive, overly detailed, or irrelevant to the purposes of the survey from the respondents perspective may have an adverse effective on respondent motivation.
- c) The content of the survey instrument also must appear appropriate to the respondent. If questions call for answers that are seen by respondents to be highly personal or potentially embarrassing, special efforts are likely to be needed in order to obtain accurate reporting. Researchers need to assess the appropriateness of the data collection context, with particular attention to the kind of relationship the interviewer and respondent may have developed, when thinking about what questions to ask and how to ask them.

## **Psychometric standards**

Survey questions must meet standards with respect to the information that they provide. The three main standards include:

- a) How answers are distributed is one key to how much information is provided by a question. If there is little variation in the answers, there may be little value to the question. Also, if questions apply to only a subset of a sample, their value is related to the number of people to whom they apply and who will actually answer them.
- b) Measures of validity, how well the answers to a question measure what the researcher intends to measure are, of course, central to any assessment of the quality of a survey question (Cronbach/Meehl 1955).
- c) Reliability, the extent to which answers are consistent over time (in the absence of change) is another aspect of psychometric analysis. Reliability does not ensure validity, of course, but unreliability places a limit on how valid measures can be.

## **Usability**

Survey instruments usually are in some sort of paper form or on a computer. They will be used either by interviewers or by respondents themselves. Survey instruments need to be as easy as possible to use so that whoever is using them, interviewer or respondent, can focus on the question-and-answer process and not face challenges in figuring out which questions are to be answered or how to answer them.

## **Multi-mode capability**

Increasingly, it is desirable to design survey questions that can be asked in more than one mode of administration. Survey instruments can be interviewer administered or self administered; interviews are done in person and by telephone. It is highly desirable to be able to compare answers that are collected by different modes. A particular incentive for comparability across modes is that multi-mode designs, whereby more than one mode of data collection is used to maximize the rate of response, are becoming increasingly used (see Dillman 2000). However, to use such designs, it is essential to be able to assume that the answers to questions are not affected by the mode of data collection itself.

To meet this standard, researchers need to think about how questions would be administered both with and without an interviewer. Questions should be designed so that they form a script for an interviewer, as well as being appropriate for self-administration. The response alternatives need to be simple enough that they can be administered on the telephone, as well as being appropriate in self administration.

### **Multi-language capability**

It may be obvious that considering how well questions can be translated into other languages should be a concern in the design of survey questions. In the past, at least in the United States, researchers often designed questions in English, then worried about ease of translation later. We know that if researchers think about how easy or hard it is to translate a particular question at the time of its initial design, the likelihood that questions can be comparable across languages is greatly increased.

### **Cost effective use of survey time**

Finally, there is always pressure on survey researchers to use survey time well. It is very common for researchers to want to ask more questions than budgets will permit or than respondents will put up with. Thus, in addition to all the previous standards, researchers also must attend to setting priorities and to using the survey time to ask the questions that will yield the most valuable information for the research purposes; in short, they must try to achieve their research objectives in the most parsimonious way, consistent with data quality.

Table 1 provides a summary of the 8 standards.

**Table 1: Summary of Standards**

<ol style="list-style-type: none"><li>1. Right content?</li><li>2. Cognitive Standards.</li><li>3. Interpersonal Standards.</li><li>4. Psychometric Standards.</li><li>5. Usability Standards.</li><li>6. Multi-mode Capability.</li><li>7. Multi-language Capability.</li><li>8. Cost-effective use of survey time.</li></ol>
--

### 3. Why it is easy to write bad questions

There are four main reasons that researchers write bad questions:

1. There are many standards; researchers often fail to attend to all of them, either through inattention or lack of appreciation of their importance.
2. The standards sometimes conflict with one another; making a question better in one respect can make it worse in some other respect.
3. Presurvey evaluation procedures often are inadequate to provide information about whether or not standards have been met.
4. Researchers do not like to change questions that have been used before.

#### Many Standards

It is a common experience for those testing questions to learn that designers of questions have not even thought through what they want to measure and why (standard 1), much less begun to think through carefully the next and more complex step of which questions to ask to achieve their poorly specified objectives. Many questions are given birth without any real thought about standards at all.

Fundamental to the problem of question design is that many of these standards have not been widely appreciated as being important. For many years, “usability”, whether or not interviewers could administer the survey instrument, was the main focus of presurvey testing. In some disciplines, studies of validity and reliability are routinely done. In other disciplines, researchers are satisfied with “face validity”, the appearance of validity, assuming questions measure what they appear to measure, without any empirical evidence.

New research has documented the importance of standards that were not previously appreciated. It may seem strange to say that it is new to demand that questions be consistently understood and ask for answers that respondents are able to provide. However, it is only in the past decade that cognitive evaluation of questions began to be done anywhere, and even today it is routine only in a small, though growing, number of centers (see Willis/DeMaio/Harris-Kojetin 1999).

It was over 30 years ago that Charles Cannell began documenting the effects of the way interviewers relate to respondents on data quality. He observed interviews and demonstrated that interviewers often asked questions in their own words, rather than the words that were provided (Cannell/Marquis/Laurent 1977). He showed he could manipulate interviewer behavior and improve reporting (Cannell/Oksen-

berg/Converse 1977). More recent studies show how standardized interviewing breaks down when questions are poorly designed and demonstrate that questions that consistently require interviewer probing and clarification are associated with more interviewer-related error (Fowler/Mangione 1990; Fowler 1991; Fowler/Cannell 1996), and how socially sensitive material is reported more favorably when an interviewer collects data than when questions are self administered (Dillman/Tarnai 1991; Aquilino 1994; Tourangeau/Smith 1998; Turner/Forsyth/O'Reilly, et al. 1998). Despite all the evidence, attention to how question design will affect the interviewer-respondent relationship is among the most neglected of the standards for questions.

Three other standards have emerged as important as the world has changed. Until recently, designers of survey questions thought their questions would be used in a single language. In the United States that is no longer the case. Hardly any surveys of importance in the U.S. are done only in English, even for studies restricted to the U.S. Researchers in other parts of the world have been concerned about multi-lingual use far longer than those in the U.S. Moreover, the desire for cross-national studies is growing daily.

There are fundamental difficulties in achieving comparable measurement across languages and cultures. However, the chances of success are better if question designers think about multi-language use before, not after, their questions are designed. For too long, researchers designed their questions in the primary language, then turned to the problem of translation. Some words and concepts translate more easily than others. To the extent that question designers take that issue into account when designing their instruments, the comparability of questions across languages will be better.

A similar issue arises with respect to mode of data collection. Increasingly, researchers want to collect data using more than one mode of data collection: with and without interviewers, with and without computer assistance. There are some features of question design that facilitate the comparable collection of data across modes. Interviewers need scripted questions that can be read as worded. Response alternatives that are long and wordy may be all right in a self-administered form, but they are hard to use on the telephone. Complex skipping instructions are easy for interviewers, especially with computer assistance. However, they pose a challenge for self administration. If a researcher designs an instrument with only one mode of data collection in mind, the chances are good that it will not adapt easily to other modes. Designing for multi-mode data collection from the start will greatly increase the chances that an instrument can produce comparable data across modes.

Finally, the introduction of computers into all phases of data collection adds a new and challenging dimension to the concept of usability. A solid premise of survey instrument design is that the tools of data collection should be as easy to use as possible, so the participants can concentrate on the question-and-answer process. Initially, when questions were put on computers, designers more or less mirrored the designs they had used on paper. However, recent studies of the interaction between interviewers and computers has shown that there is much room for improvement in the way computer-assisted data collection tools are designed (Couper/Hansen/Sadowsky 1997). Interviewers are found to waste considerable time trying to navigate, have great difficulty making corrections, cannot take advantage of many of the tools and aids that computers have to offer, and spend more time looking at the computer than looking at respondents. There are similar issues with computer-based surveys designed for respondents to use themselves (see, for example, Dillman 2000). The point is that the concept of usability has risen to a new level of complexity with the introduction of computers, and how easily interviewers and respondents can use them is a critical part of the evaluation of how well a survey instrument is designed.

### **Conflicting Standards**

A second set of challenges for designing good questions stems from the fact that sometimes improving a question from the perspective of one standard can make it worse with respect to another standard. Two places where standards are particularly likely to conflict include:

- a) Providing detailed definitions and explanations to make questions clearer can also make questions complicated and harder to administer, particularly for interviewer administration (see Conrad/Schober 2000).
- b) Having more response categories, and labeling the response alternatives with adjectives, have both been shown to improve the psychometric performance of questions. However, having numerous response alternatives makes a question harder for an interviewer to administer, particularly on the telephone. Also, using numerous labels on alternatives makes it less likely that a question can be easily translated into other languages. The following pairs of questions provide examples of the conflicts that are inherent in question design.



- 
- Example 1a: In the past 7 days, how many times did you exercise?
- Example 1b: Exercise can sports, running, swimming include, and cycling, as well as working out at a club or gym. It can also include walking. In the past 7 days, not counting exercise you got while working or doing chores at home, how many times did you exercise—none, 1 or 2, 3 to 5, or 6 or more?
- Comment: Example 1a leaves the definition of “exercise” up to the respondent. Also, by asking for a specific number of times, it may pose a difficult task for respondents to be as precise as the question demands.
- Example 1b provides information about what is meant by exercise, and it allows respondents to answer in categories. However, it is a much wordier question, which may prove to be confusing to respondents and hard for an interviewer to administer. Also, the addition of labeled categories of times may prove to be harder for respondents to cope with than giving a precise number, particularly in a telephone interview, in which respondents must remember the categories.
- Example 2a: How would you rate the way the Chancellor is doing his job - excellent, good, not so good, not good at all?
- Example 2b: Using a rating scale from 0 to 10, where 0 is as bad as possible and 10 is as good as possible--- What number from 0 to 10 would you use to describe the way the Chancellor is doing his job?
- Comment: Example 1a is a classic 4-category rating, with each category labeled. The numerical rating in 2b only has labels at the ends. Also, explaining the 0 to 10 task will take more time and be more complex than asking question 2a. However, using numbers has three advantages. First, psychometricians will like the fact that it provides 11 categories, which is likely to provide more information than 4 categories. Second, in a telephone interview, it is easier to remember all the possible answers, 0 to 10, for the numerical task than for the 4 labeled answers. Finally, the problem of translation across languages is easier for numbers than for words.
- It is important to note that standards do not always conflict. Sometimes, applying several standards leads consistently to the same conclusion about which question is best.

Example 3a: Tell me what you think about the following statement: The news reported in newspapers and on television is not to be trusted—do you strongly agree, generally agree, neither agree nor disagree, generally disagree, or strongly disagree?

Example 3b: How much do you trust the news reported in the newspapers and on television—very much, some, a little, or not at all?

Comment: Example 3a is one of the most popular question forms in survey research. Question 3b is an alternative way to accomplish a similar research objective, to measure trust in the media. Putting aside for the moment whether or not trust in newspapers and magazines is a good concept about which to ask and whether or not newspapers and television should be combined in one question or separated, if one applies the standards outlined in this paper, question 3b is definitely a better question than question 3a.

1. Applying cognitive standards, the agree-disagree format has been consistently found to be confusing to respondents. It is particularly confusing when a respondent has to disagree with a negative statement in order to express a positive opinion. In this case, respondents have to disagree that media are not to be trusted to say that they can be trusted.
2. The response task is complicated in the agree-disagree form, with difficult categories to remember, making it difficult to administer by interviewers, particularly on the phone.
3. Interpersonal forces have been shown to produce acquiescence bias in some respondents, especially those with less education. That means that they consistently will be more likely to “agree” when a question is in this form than to give a comparable answer to a question in some other form. A significant number of respondents will agree with this question and with its opposite phrased positively (e.g. news in newspapers and on television is to be trusted) (see Schuman/Presser 1981; Dillman/Tanai 1991; Converse/Presser 1986).
4. Psychometrically, agree-disagree questions have disadvantages as well:
  - a) The agree-disagree questions are almost always analyzed as two-category variables. In contrast, question 3b distributes answers across four categories. Thus, there usually is more information provided by the four-category scale.

- b) There is frequent debate about where to place those who neither agree nor disagree: are they in the middle or should they be treated as missing data? It is not at all clear that those in the middle category can be put in order with the rest of the respondents.

One advantage of the agree-disagree question form is that one can ask almost any opinion or attitude question in that form. Thus, respondents (and interviewers) can enjoy a consistency of response tasks that is harder to provide with more direct questions. However, in virtually all other respects, alternatives to the agree-disagree question form will better meet the standards for question design.

In conclusion, there is the potential for real conflict among standards, which accounts for some of the shortcomings in the way questions are designed. Nonetheless, if researchers are mindful about the various standards, they usually can find a way to design a question that will do a reasonable job of meeting them all.

### **Inadequate Presurvey Testing**

The third reason that poor questions are used in surveys is that question testing protocols prior to surveys often are not adequate to evaluate all aspects of the question that are important. Historically, pretesting consisted mainly of a small field test, in which senior interviewers conducted some test interviews and reported back on their experiences. In the last decade or so, question testing has been improved.

The core steps for presurvey question evaluation include:

- a) Systematic review of questions (Lessler/Forsyth 1996; Fowler 1995)
- b) Cognitive testing (Lessler/Tourangeau 1989; Willis/DeMaio/Harris-Kojetin 1999)
- c) Behavior coding of field pretests (Fowler/Cannell 1991; Oksenberg/Cannell/Kalton 1991).

These steps are becoming increasingly common, particularly at centers that place great emphasis on methodological excellence. Each of the steps has the potential to identify question problems before a survey is done. Other presurvey evaluation steps that are done less frequently include:

- a) focus group discussions prior to drafting the survey instrument to learn what respondents have to say on the survey topic and to explore vocabulary issues
- b) an evaluation of the literacy level of questions by a specialist in reading

- c) tests of the usability of survey instruments, in paper and pencil or in computer form
- d) debriefing of respondents after field pretest interviews.

**Table 2: Value of Various Evaluation Options for Providing Information About How Well Questions Meet Various Standards**

**Standards for Survey Instruments**

<b>Evaluation Options</b>	Content	Cognitive	Inter-personal	Psychometric	Usability	Multi-mode Capability	Multi-Language
<b>Pre-Field Testing</b>							
Expert Review	YES	H	----	----	H	H	H
Focus Groups	YES	H	----	----	----	----	----
Cognitive Ints	H	YES	----	H	----	----	H
Usability Tests	----	----	----	----	YES	----	----
<b>Field Testing</b>							
Interviewer debriefing	----	H	H	----	YES	----	----
Respondent debriefing	----	H	H	H	H	----	----
Behavior coding	----	H	YES	----	----	----	----
Analysis of pilot data	----	----	----	YES	----	----	----
Reinterviews	----	H	----	YES	----	----	----
Split Form (Split Ballot) tests	----	----	----	YES	----	YES	YES

YES = provides information; H = helpful, but not sufficient.

Table 2 is an attempt to relate different testing steps to the 8 standards outlined at the beginning of this paper. The cells are labeled “yes” to indicate that a test can provide information about how well a question meets a certain standard, “H” if it can “help” to evaluate it, and blank if it makes no contribution. Although the coding admittedly is arbitrary and subject to debate, two points stand out from the table: First, the most common testing strategies—expert review, focus groups, cognitive testing, and behavior coding of field interviews -- address the first three standards plus usability. However, they do very little to help with the other four. Particularly lacking are measures of data quality, validity and reliability, and comparability of data across modes and across languages. Second, there are three underutilized kinds of evaluation steps that would add considerably to knowledge about key elements of survey instruments: analysis of pilot or pretest data, reinterviews, and split-sample testing.

Pretest samples are often a little too small for meaningful analysis. Samples of 25 to 35 are large enough for usability testing and to get meaningful behavior coding measures for questions that are asked of most respondents. Probably having pretest samples of size 50 would be better if some psychometric evaluation is planned. By simply tabulating marginal distributions and cross tabulating variables that may potentially be redundant, questions can be identified that are yielding little information because there is little variation in answers, they apply to too small a percentage of the sample to be statistically useful, or they overlap with other questions. Correlational analyses to assess construct validity can be conducted to identify items that can be dropped from multi-item indices.

Reinterviews after pretests offer two potential contributions to question evaluation that are hard to achieve in any other way. First, they provide the opportunity to assess the reliability of answers. By asking the same question twice, at different points in time, a measure of the stability of answers can be obtained. Items measuring something thought to be stable that show low reliability are obvious candidates for revision or dropping.

An even more interesting use for reinterviews has become possible with the advent of computer-assisted interviewing. A reinterviewer now can ask a set of questions that have been previously asked, blind to the original answers. Then, either after a given answer has been given or at the end of the reinterview, the computer can check reinterview answers against original answers, notify the interviewer of discrepancies, and enable interviewers to ask probe questions to understand the reason for the differences. Such a process can be a valuable adjunct to cognitive testing,

particularly because it can reveal confusion and response inconsistency under realistic data collection conditions.

Split-ballot tests are among the most useful question evaluation activities. Their disadvantage, of course, lies in the fact that they entail additional data collection at the presurvey stage. However, the great strength of such designs is that they permit addressing a key question that cannot be addressed in any other way: whether two versions of the same question produce the same or different estimates. This, of course, is crucial to evaluating the comparability of data collected by phone and self administration, as well as assessing the comparability of questions to be asked in different languages. In addition, when changes are made to improve questions, based on findings from cognitive testing or field pretests, there is always a need to know how the changes will affect the results. If a change is designed to clear up some ambiguity found in a question, one would like to know if the resulting estimate is affected in the expected way. If a question was found awkward to administer, we should find out if the interviewer-respondent interaction is improved. In some cases, an improved question should produce a different result. In other cases, if a question has been streamlined to improve the ease of administration, the hope may be that the resulting estimate will be essentially the same as from its predecessor. In all these cases, it is only by comparing results from comparable samples that one will know the answer. Small split-ballot field tests, with samples of 150 to 200 interviews randomized to alternative forms of a survey instrument can provide information on these and many more topics in a cost-effective manner.

In conclusion, one important way to get researchers to tend to various standards is routinely to have test procedures to evaluate how well questions are designed. Recent developments in testing, such as cognitive testing and behavior coding of pretests, have definitely improved the evaluation of some aspects of questions. However, current protocols are generally inadequate to evaluate all the aspects of questions that should be evaluated. Any important aspect of questions that is not routinely evaluated is likely to be a continuing source of problem questions.

### **Resistance to New Questions**

There are some reasons that have merit for wanting to repeat previously used questions. Using previously established measures permits replication of previous research findings, as well as the possibility of measuring change over time. Reviewers, for grant proposals and for articles submitted to journals, often are reassured when researchers propose to use, or have used, questions that have been used previously.

The question researchers and methodologists must face is whether these reasons are compelling enough to continue to ask a question that has demonstrable flaws.

Every discipline that uses survey research has a past filled with questions that were designed and used before we knew as much as we do today about question design. The repeated use of these questions constitutes an important source of poor quality survey research.

There may be compelling reasons in some circumstances for repeating questions that are less than perfect. However, there is no reason that all key questions in a survey instrument should not be evaluated. If a question to which a researcher is committed is found to be problematic in some way, the researcher then is in a position to make an informed choice among three reasonable options:

1. Revise the question to make it better
2. Use the flawed question, but be aware of its problems when analyzing the results and inform data users of the problems that were identified
3. Use both the original and a revised, improved question, so that results from the original question can be compared with previous studies, the impact of the revisions can be evaluated, and an even more informed choice, with improved question options, can be made by future researchers studying the same topic area.

#### **4. Conclusion**

Knowledge about the design of survey questions has progressed considerably over the past decade or so, as has the commitment to better question evaluation. However, better survey instrument design provides one of the most important opportunities to improve the quality of survey research practice. There are several needed steps.

1. We need continued research that links characteristics of questions to the quality of the resulting data. More empirically based generalizations about question design and survey error will accomplish two important goals: to teach us details about how to design better questions and to heighten general awareness of the significance of poor question design for the quality of survey estimates.
2. We need a broader appreciation of the complexity of designing good survey questions. While the 8 standards outlined in this paper might be subject to some debate, there is no doubt that good questions must meet numerous,

largely unrelated standards. To design good questions is not something that can be done by the amateur, by committee, or without serious evaluation activities and the input of experts.

3. More and better presurvey tests need to be used routinely. Even the testing that goes on at our best research institutions is not enough to provide the information researchers need to know about their questions. The next addition to routine testing protocols, beyond systematic expert review, cognitive testing, and behavior coding of field pretests, should be small split-ballot pilot tests of revised or alternative forms of questions. Such tests would provide information on whether the question revisions emerging from the early testing had the expected effect on results, and they would permit psychometric evaluation of questions prior to surveys. However, there are other kinds of evaluations mentioned above that are valuable. The important point is not to specify exactly which protocol should be used but to emphasize that attention should be given to more of the important aspects of questions than is the case with current protocols.
4. Finally, researchers, reviewers at funding agencies, and those who review journal articles need to be educated so they understand that new questions, properly evaluated, are likely to be better than old questions. A question that was used in the past, usually designed without the benefit of current testing wisdom, should not be an object of reverence, but rather a subject for evaluation and, usually, improvement. Researchers say that a certain question was used previously and it “worked”, when there was no real evaluation of the response process or the quality the answers—in short, no real evidence that it “worked”. The standard for whether or not a question is a good one for a particular purpose should rest not on its pedigree but on the evidence gleaned from an appropriate evaluation protocol

In conclusion, good researchers write bad questions: 1) because they do not appreciate all the standards that apply to survey questions; 2) because sometimes addressing multiple standards can be in conflict; 3) the evaluation protocols that would flag all the potential weaknesses of their questions are not routinely in place; and 4) because they overvalue previously used questions. The challenges for dealing with all these issues are obviously daunting. However, the real key to change is better question evaluation. When that is in place, better questions will follow.



## Contact

*Floyd Jackson Fowler Jr.*  
*Center For Survey Research*  
*University of Massachusetts Boston*  
*100 Morrissey Boulevard*  
*Boston, MA 02125-3393*  
*U. S. A.*  
*Phone (617) 287-7200; Fax: (617) 287-7210*  
*csr@umb.edu*

## References

- Aquilino, W. S. 1994: Effects of interview mode on measuring depression in younger adults. *Journal of Official Statistics*, 14(1), 15-30.
- Cannell, C. F./Marquis, K./Laurent, A. 1977: A summary of studies. *Vital & Health Statistics, Series 2*, 69. Washington, DC: Government Printing Office.
- Cannell, C. F./Oksenberg, L./Converse, J. 1977: Experiments in interviewing techniques: Field experiments in health reporting, 1971-1977. Hyattsville, MD: National Center for Health Services Research.
- Conrad, F. G./Schober, M. F. 2000: Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64, (1), 1-28.
- Converse, J./Presser, S. 1986: *Survey questions*. Beverly Hills, CA: Sage.
- Couper, M. P./Hansen, S. E./Sadowsky, S. A. 1997: Evaluating interviewer use of CAPI technology. pp. 267-286 in: Lyberg, L. E./Beimer, P./Collins, M./et al. (Eds). *Survey measurement and process quality*. New York: John Wiley and Sons.
- Cronbach, L./Meehl, P. 1955: Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dillman, D. A. 2000: *Mail and Internet surveys: The tailored design method*. New York: John Wiley.
- Dillman, D. A./Tarnai, J. 1991: Mode effects of cognitively designed recall questions: A comparison of answers to telephone and mail surveys. pp. 287-310 in: Beimer, P. N./Corres, R. M./Lyberg, L. E./Mathiewetz, N. A./Sudman, S. (Eds.), *Measurement errors in surveys*. New York: John Wiley.
- Fowler, F. J./Mangione, T. W. 1990: *Standardized survey interviewing*. Thousand Oaks, CA: Sage.
- Fowler, F. J. 1995: *Improving survey questions*. Thousand Oaks, CA: Sage.

- Fowler, F. J. 1991: Reducing interviewer related error through interviewer training, supervision, and other means. pp. 259-278 in: Beimer, P. N./Groves, R. M./Lyberg, L. E./Mathiewetz, N. A./Sudman, S. (Eds.), *Measurement errors in surveys*. New York: John Wiley.
- Fowler, F. J./Cannell, C. F. 1996: Using behavioral coding to identify cognitive problems with survey questions. pp. 15-36 in: Schwarz, N. A./Sudman, S. (Eds.), *Answering questions*. San Francisco: Jossey-Bass.
- Groves, R. M. 1989: *Survey errors and survey costs*. New York: John Wiley.
- Hippler, H. J./Schwarz, N./Sudman, S. (Eds.), 1987: *Social information processing and survey methodology*. New York: Springer-Verlag.
- Jabine, T. B./Straf, M. L./Tanur, J. M./Tourangeau, R. (Eds.). 1984: *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Lessler, J. T./Forsyth, B. H. 1996: A coding system for appraising questionnaires. pp. 259-292 in: Schwartz, N. A./Sudman, S. (Eds.). *Answering questions*. San Francisco: Jossey-Bass.
- Lessler, J. T./Tourangeau, R. 1989, May: *Questionnaire design in the cognitive research laboratory*. Vital Health & Statistics, Series 6, 1. Washington, DC: Government Printing Office.
- Oksenberg, L./Cannell, C. F./Kalton, G. 1991: New strategies of pretesting survey questions. *Journal of Official Statistics*, 7(3), 349-366.
- Schuman, H./Presser, S. 1981: *Questions and answers in attitude surveys*. New York: Academic Press.
- Sirken, M. G./et al. (Eds.). 1999: *Cognition and survey research*. New York: John Wiley.
- Tourangeau, R./Smith, T. W. 1998: Collecting sensitive data with different modes of data collection. pp. 431-453 in: Couper, M. P., et al. (Eds.), *Computer assisted survey information collection*. New York: John Wiley.
- Turner, C. F./Forsyth, B. H./O'Reilly, J. M./et al. 1998: Automated self-interviewing and the survey measurement of sensitive behaviors. pp. 455-473 in: Couper, M. P., et al. (Eds.), *Computer assisted survey information collection*. New York: John Wiley.
- Willis, F. B./DeMaio, T./Harris-Kojetin, B. 1999: Is the bandwagon headed to the methodological promised land? Evaluating the validity of cognitive interviewing techniques. pp. 133-154 in: Sirken, M. G., et al. (Eds.), *Cognition in survey research*. New York: John Wiley and Sons.

# STICHPROBENREALISIERUNG BEI EINER BUNDESWEITEN TELEFONISCHEN BEFRAGUNG ALTER MENSCHEN

*OLAF V.D. KNESEBECK, VOLKER HÜFKEN UND PETER DÜBBERT<sup>1</sup>*

Im vorliegenden Beitrag wird die Stichprobenrealisierung bei einer bundesweiten telefonischen Befragung von in Privathaushalten lebenden alten Menschen (60 Jahre und älter) beschrieben. Dabei steht die Frage nach den Auswirkungen von Proxy-Interviews und Konvertierungsmaßnahmen auf die Zusammensetzung der Stichprobe im Zentrum der Analysen. Ein Vergleich zwischen Personengruppen mit unterschiedlich ausgeprägter Befragbarkeit und Kooperationsbereitschaft zeigt, dass durch die ausschöpfungssteigernden Maßnahmen mehr Hochaltrige, Verheiratete, alte Menschen mit geringer Bildung und geringem Einkommen sowie alte Menschen mit schlechterem Gesundheitszustand in die Untersuchung einbezogen werden konnten. Zudem ergeben sich zum Teil deutliche Unterschiede in der Ausprägung multivariater Zusammenhänge zwischen kooperationsbereiten Befragungsteilnehmern und temporären Verweigerern. Die Ergebnisse verdeutlichen die Notwendigkeit ausschöpfungssteigernder Maßnahmen zur Reduzierung von durch Ausfällen bedingten Verzerrungen insbesondere bei telefonischen Befragungen alter Menschen.

The article describes a survey of older people ( $\geq 60$  years) living in private accommodation in Germany. The aim was to investigate the impact of proxy-interviews and successful conversion on the structure of the realised sample. A comparison of groups shows that conversion techniques lead to higher participation of the oldest old, of married people, and of people with either low socio-economic status or poor health. In addition, considerable differences were found in multivariate relationships between co-operative respondents on the one hand and those initially unwilling to participate on the other. Our findings point to the importance of employing response enhancement measures to reduce non-response, in particular in telephone surveys of older and elderly people.

---

<sup>1</sup> Olaf v.d. Knesebeck, Peter Dübbert Institut für Medizinische Soziologie, Universität Düsseldorf; Volker Hüfken (Sozialwissenschaftliches Institut, Universität Düsseldorf).

## 1. Einleitung

Ältere Menschen und ihre Lebenssituation werden in den letzten Jahren immer häufiger zum Gegenstand der Forschung gemacht. Dies betrifft in zunehmendem Maße auch sozialwissenschaftlich orientierte Untersuchungen, in denen Daten durch Umfragen erhoben werden (z.B. Kohli/Künemund 2000). Betrachtet man die deutschsprachige Literatur, so wird deutlich, dass Telefonumfragen in diesem Bereich die Ausnahme sind (Knesebeck/Lüschen 1998). Dies verwundert umso mehr, als die wenigen methodischen Untersuchungen, die sich mit Befragungen von in Privathaushalten lebenden alten Menschen beschäftigen, darauf hindeuten, dass keine der drei Befragungsarten (persönlich, postalisch, telefonisch) aufgrund ihrer Überlegenheit zum Standard bei der Befragung alter Menschen erhoben werden kann (Herzog/Rogers 1988a; Tennstedt/Dettling/McKinlay 1992; Kühn/Porst 1999). Da Telefoninterviews bei allgemeinen Bevölkerungsbefragungen in Deutschland in den letzten Jahren an Bedeutung gewonnen haben, stellt sich die Frage, welche spezifischen Aspekte bei der Organisation und Durchführung telefonischer Befragungen mit alten Menschen berücksichtigt werden müssen. Zu diesem Thema liegen bislang kaum Erfahrungen oder Untersuchungen vor. Vor diesem Hintergrund steht die Beschreibung des erhebungsmethodischen Vorgehens bei einer bundesweiten telefonischen Befragung alter Menschen im Zentrum des vorliegenden Beitrages. Dabei werden vor allem populationsspezifische Ausfallursachen und ausschöpfungssteigernde Maßnahmen sowie deren Auswirkungen auf die Zusammensetzung der Stichprobe dargestellt und analysiert.

Als ein Indikator für die Beurteilung der methodisch-wissenschaftlichen Qualität von Befragungen auf der Basis von Zufallsstichproben werden im allgemeinen Ausschöpfungsquoten bzw. das Ausmaß der Befragungsausfälle (Unit-Nonresponse) herangezogen. Bei telefonischen Befragungen der Allgemeinbevölkerung in Deutschland werden in der Regel Ausschöpfungsquoten von 40 bis 60 Prozent erreicht (Kviz 1977; Porst 1993; Schnell 1997; Schulte 1997). Die hauptsächlichen Ursachen für systematische Ausfälle bei telefonischen Befragungen bestehen darin, dass Zielpersonen nicht erreicht werden oder nicht kooperationsbereit sind. Aufgrund der mit dieser Erhebungsmethode verbundenen direkten, schnellen und preisgünstigen Art der Kontaktaufnahme werden im Allgemeinen mehrere Kontaktversuche zu unterschiedlichen Zeiten unternommen, um das Problem der Erreichbarkeit zu reduzieren und die Ausschöpfung der Stichprobe zu erhöhen. Zur Reduzierung der Ausfälle, die auf mangelnde Kooperationsbereitschaft zurückzuführen sind, können - neben Maßnahmen wie z.B. schriftlichen Vorabinformationen - nicht kooperationsbereite Personen nochmals kontaktiert werden, mit dem Ziel, diese doch noch

zu einem Interview zu bewegen. Einschlägige Untersuchungen zeigen, dass es möglich ist, auf diese Weise 30 bis 50 Prozent der potentiellen Verweigerer zu konvertieren. Überdies weisen die wenigen Analysen, die sich mit den Merkmalen konvertierter Verweigerer beschäftigen, darauf hin, dass eine temporäre Verweigerung kein zufälliges Ereignis darstellt, sondern durch bestimmte Personenmerkmale beeinflusst wird (Porst/Schneid 1989; Knesebeck/Lüschen 2000; Reuband/Blasius 2000).

Generell wird davon ausgegangen, dass bei der Befragung älterer Menschen mit höheren Ausfällen zu rechnen ist (Herzog/Rogers 1988b; Tennstedt/Dettling/McKinlay, 1992; Schnell 1997; Kühn/Porst 1999). Diese werden weniger auf Probleme der Erreichbarkeit, als auf die zum Teil eingeschränkte Befragbarkeit und auf die mangelnde Kooperationsbereitschaft älterer Menschen zurückgeführt (Bungard 1979; Nuthmann/Wahl 1999; Koch 1997; Knesebeck/Lüschen 1998; Kühn/Porst 1999). Das Problem eingeschränkter Befragbarkeit ergibt sich aus einer erhöhten Wahrscheinlichkeit von Erkrankungen, Hör- oder Verständnisproblemen bei alten Menschen. Eine Möglichkeit, diesem Problem entgegenzuwirken, besteht darin, stellvertretend eine Person zu befragen, die mit der Situation des alten Menschen vertraut ist (Mosely/Wolinsky 1986; Moore 1988). Je enger die Beziehung zur Person ist, über die Aussagen getroffen werden, und je mehr es sich um direkt beobachtete Sachverhalte handelt, desto genauer sind Proxy-Angaben. Einstellungen werden dabei stärker verzerrt erfasst als Verhaltensweisen, aber auch diese sind von Messfehlern nicht frei (Mingay et al. 1994).

Im Hinblick auf einen altersbedingten Rückgang der Teilnahmebereitschaft liegen recht widersprüchliche Ergebnisse vor. Einige Untersuchungen zeigen, dass die Teilnahmewilligkeit bei älteren Personen reduziert ist, was auf unterschiedliche Faktoren zurückgeführt wird. Dazu zählen ein generell reduziertes Interesse an den meisten Untersuchungsthemen, erhöhte Kriminalitätsfurcht und die im Allgemeinen geringere Bildung alter Menschen, die mit unklaren Konsequenzbefürchtungen hinsichtlich der Teilnahme und der Angst einhergeht, der Interviewsituation möglicherweise nicht gewachsen zu sein (Schnell 1997; Kühn/Porst 1999; Schräpler 2000). Für den letztgenannten Erklärungsfaktor sprechen die Ergebnisse von Reuband und Blasius (2000) aus einer telefonischen Befragung, die zeigen konnten, dass die kognitive Kompetenz, dem Interview zu folgen und die gestellten Fragen zu verstehen, bei konvertierten Verweigerern geringer ausgeprägt ist als bei kooperationsbereiten Befragungsteilnehmern.

Andererseits deuten Ergebnisse aus anderen Untersuchungen darauf hin, dass sich eine lineare Zunahme von Verweigerungen mit zunehmendem Alter nicht feststellen

lässt (Herzog/Rogers 1988b). Zudem hat sich wiederholt gezeigt, dass die durch Verweigerungen bedingten Ausfälle durch entsprechenden Aufwand bei der Planung und Durchführung der Befragung (z.B. sorgfältige Gestaltung der Erhebungsinstrumente, Vorkontaktierungen oder Konvertierungsbemühungen) reduziert werden können (Schnell 1997; Knesebeck/Lüschen 1998).

Im vorliegenden Beitrag sollen Besonderheiten bei der Organisation und Durchführung von bundesweiten telefonischen Befragungen bei in Privathaushalten lebenden alten Menschen sowie die damit verbundenen Möglichkeiten und Probleme aufgezeigt und diskutiert werden. Dabei wird die Frage verfolgt, inwieweit das Verständnis der im Interview gestellten Fragen mit soziodemographischen und sozialstrukturellen Merkmalen variiert. Ein weiterer Schwerpunkt liegt in der Thematisierung populationsspezifischer Ausfallursachen und ausschöpfungssteigernder Maßnahmen sowie deren Auswirkungen auf die Zusammensetzung der Stichprobe.

## **2. Beschreibung des erhebungsmethodischen Vorgehens**

Grundlage der Analysen ist eine in der Zeit zwischen April und Juli 2000 mit Hilfe von CATI durchgeführte telefonische Befragung zum Thema „sozialer Status und Gesundheit bei alten Menschen“.<sup>2</sup> Grundgesamtheit der Untersuchung waren alle Personen, die zum Zeitpunkt der Befragung 60 Jahre oder älter waren und in einem Privathaushalt in Deutschland lebten. Die Stichprobe basiert auf einem modifizierten Random Digit Dialing-Design, das von Gabler und Häder (1999) entwickelt wurde. Bei der Stichprobenziehung, die von ZUMA durchgeführt wurde, sind sowohl eingetragene als auch nicht eingetragene Telefonnummern einbezogen worden.

Der Anteil der nicht in das Telefonbuch eingetragenen Haushalte innerhalb der Population der älteren Menschen (60 Jahre und älter) ist bundesweit noch vergleichsweise gering (Kühn/Porst 1999). Vor diesem Hintergrund wurden nur solche Telefonnummern in die Untersuchung aufgenommen, die laut Ergebnisdatei (Gabler/Häder 1999) in ein Telefonbuch eingetragen waren.

---

**2** Die Befragung ist Teil einer deutsch-amerikanischen Vergleichsuntersuchung, die von der Stiftung Deutsch-Amerikanisches Akademisches Konzil im Rahmen des TransCoop-Programms gefördert wird. Durchführende Institutionen sind auf deutscher Seite von der Universität Düsseldorf das Institut für Medizinische Soziologie in Kooperation mit dem Sozialwissenschaftlichen Institut und auf amerikanischer Seite das Institute of Social Research am Department of Sociology der Universität Alabama/Birmingham.

Insgesamt wurden 4800 Telefonnummern aus der von ZUMA zur Verfügung gestellten Stichprobe ausgewählt. Diese Telefonnummern wurden mit den Telefonbuchangaben von November 1999 verglichen und auf die Verfügbarkeit einer Adresse hin überprüft. Dabei wurden 294 Nummern aussortiert, die eindeutig nicht zu einem Privathaushalt, sondern z.B. zu einer Firma oder Behörde gehörten. 282 Telefonnummern waren im Telefonbuch von November 1999 nicht mehr verzeichnet, in 393 Fällen war die im Telefonbuch enthaltene Adresse unvollständig. Die übrigen 3831 ermittelten Adressen wurden mit Hilfe von Post-Direkt auf Zustellbarkeit überprüft. Im Zuge dieser Überprüfung erwiesen sich 248 Adressen als falsch, während 3583 Adressen verifiziert werden konnten. Diese Haushalte wurden angeschrieben und damit auf den Anruf vorbereitet. Das Anschreiben enthielt eine Beschreibung des Untersuchungszieles und des Auswahlverfahrens, einen Hinweis auf die vertrauliche Behandlung der Daten und die Ankündigung des Anrufes durch einen Interviewer. Die übrigen 923 Telefonnummern, die nicht gelistet waren oder bei denen die Adresse nicht vollständig bzw. richtig ermittelt werden konnte, wurden ohne vorheriges Anschreiben telefonisch kontaktiert (vgl. Tabelle 1).

**Tabelle 1: Ergebnis der Adressermittlung und -verifizierung**

	N	%
Bei Adressermittlung als Geschäftsanschluss identifiziert (nicht kontaktiert)	294	6.1
Adresse ermittelt und verifiziert (brieflicher Vorkontakt)	3583	74.7
Nummer nicht gelistet, Adresse unvollständig oder falsch (kein brieflicher Vorkontakt)	923	19.2
Gezogene Telefonnummern	4800	100.0

Bei der telefonischen Kontaktierung der 4506 Nummern wurden bis zu 18 Versuche zu möglichst unterschiedlichen Zeiten unternommen, wobei zunächst durch ein Screening festgestellt wurde, ob und wie viele Personen im betreffenden Haushalt leben, die 60 Jahre oder älter sind. Lebte keine solche Person in diesem Haushalt, so wurde dieser aus der Untersuchung ausgeschlossen. Lebte dort mehr als eine solche Person, so wurde auch auf der Haushaltsebene mittels der Geburtstagsmethode (Salmon/Nichols 1983) ein Wahrscheinlichkeitsauswahlverfahren angewandt, indem nach der Person im Haushalt gefragt wurde, die als letztes Geburtstag hatte.

Als Erhebungsinstrument wurde für das Telefoninterview ein standardisierter Fragebogen konzipiert, der in unterschiedliche Fragenkomplexe untergliedert war und insge-

samt etwa 220 Fragen umfasste. Die Befragungszeit lag im Durchschnitt bei etwa 43 Minuten (Minimum etwa 20 Minuten; Maximum etwa 105 Minuten). Im Fragebogen wurden verschiedene Indikatoren für den Gesundheitszustand, allgemeine und gesundheitsbezogene Einstellungen, Informationen zu sozialen Beziehungen, kritische Lebensereignisse, berufsbezogene Merkmale, Informationen zur gesundheitlichen Versorgung sowie verschiedene Indikatoren für den sozialen Status erhoben.

In der Konzeption der Untersuchung wurde auf der Basis vorangegangener Erfahrungen bei einer regionalen telefonischen Befragung alter Menschen (Knesebeck/Lüschen 1998) davon ausgegangen, dass nicht mehr alle ausgewählten alten Personen aufgrund von Erkrankungen, Hör- oder Verständnisproblemen zu einem telefonischen Interview in der Lage sind oder sich zu diesem nicht mehr in der Lage fühlen. Um jedoch auch über diese Personen einige Informationen zu bekommen und somit Ausfällen aufgrund von Nichtbefragbarkeit entgegenzuwirken, wurde neben dem „normalen“ Fragebogen, eine etwa 40 Fragen umfassende Kurzfassung entwickelt, die von einer, mit der Situation des alten Menschen vertrauten Person, stellvertretend beantwortet wurde (Proxy-Interview).

### **3. Ergebnisse**

#### **3.1 Stichproben- und Ausfallbeschreibung**

Tabelle 2 zeigt die Stichproben- und Ausfallbeschreibung der bundesweiten telefonischen Altenbefragung unterteilt nach stichprobenneutralen Ausfällen, Ausfällen vor der Personenauswahl, durchgeführten Interviews und Ausfällen nach der Personenauswahl.

Der hohe Anteil stichprobenneutraler Ausfälle (insgesamt 69.7%) ergibt sich vor allem durch die Haushalte, in denen keine Person lebt, die 60 Jahre oder älter ist (53.5%). Bei der Beurteilung dieses Anteils gilt es allerdings zu berücksichtigen, dass die im Rahmen des Screenings gegebene Auskunft, dass im Haushalt keine Person lebt, die 60 Jahre oder älter ist, auch eine 'versteckte Verweigerung' sein kann. Der Tabelle ist weiterhin zu entnehmen, dass zusätzlich zu den 294 bereits vor der telefonischen Kontaktierung ausgeschlossenen Geschäftsanschlüssen weitere 130 Telefonnummern (insgesamt 424=8.9%) als nicht zu einem Privathaushalt gehörig identifiziert wurden. Insgesamt konnten 821 Interviews realisiert werden, von denen 139 anhand des Kurzfragebogens geführt wurden. Wie bereits angesprochen, wurden die Fragen bei der Kurzversion in der Regel von einer Referenzperson beantwortet;



in einigen Fällen waren die alten Menschen auch nur zu der 40-minütigen Befragung nicht mehr in der Lage, beantworteten den Kurzfragebogen jedoch selbst.

**Tabelle 2: Stichproben- und Ausfallbeschreibung (Altenerhebung BRD 2000)**

	N	%	%(N)
<b>Bruttstichprobe</b>	<b>4800</b>	<b>100.0</b>	
<b>Stichprobenneutrale Ausfälle</b>			
Niemand im HH ist 60 Jahre oder älter	2568	53.5	
Kein HH: Betrieb, Behörde usw.	424	8.9	
Ungültige Nummer	285	5.9	
Sonstiges: Z.B. Fax, Modem	68	1.4	
<b>Bereinigte Bruttostichprobe</b>			<b>100.0 (1319)*</b>
<b>Ausfälle vor der Auswahl</b>			
Kein Kontakt nach 18 Versuchen	163	3.4	5.8 (76)
Sprachprobleme	59	1.2	2.1 (27)
Ablehnung: Keine Information über HH	31	0.7	1.1 (14)
<b>Ablehnungen und Abbrüche</b>			
Ablehnung: HH mit Untersuchungseinheit	279	5.8	21.1 (279)
Abbruch des Interviews	9	0.2	0.7 (9)
<b>Sonstige Ausfälle nach der Auswahl</b>			
Krankheit ohne Proxy	93	1.9	7.0 (93)
<b>Abgeschlossene Interviews</b>			
Interview (Langfassung)	682	14.2	51.7 (682)
Kurzinterview (Proxy)	139	2.9	10.5 (139)

\* Die Berechnung der bereinigten Bruttostichprobe stellt bei der Berücksichtigung der Ausfälle vor der Auswahl den hohen Anteil von Haushalten in Rechnung, in denen keine Person lebt, die 60 Jahre oder älter ist.

Es ergibt sich eine *Antwortrate von 62.2 Prozent*, was für telefonische Befragungen in Relation zu den üblicherweise erzielten Werten und angesichts einer Population, die in der Literatur als schwer befragbar gilt, ein relativ gutes Ergebnis darstellt. Auch die ermittelte *Verweigerungsrate von 22.9 Prozent* liegt im Rahmen der in der Literatur berichteten Werte für Telefonumfragen von 20- bis 30 Prozent (Frey/Kunz/Lüschen 1990; Schulte 1997; Schnell 1997). Da bislang kaum Erfahrungen mit bundesweiten telefonischen Befragungen von Altenpopulationen vorliegen, ist eine endgültige Beurteilung der Antwort- wie auch der Verweigerungsrate allerdings nur eingeschränkt möglich.

Vergleicht man die Stichproben- und Ausfallbeschreibung der Teilstichprobe, die brieflich vorkontaktiert wurde (N=3583) mit der ohne brieflichen Vorkontakt (N=923), so fällt auf, dass der Anteil stichprobenneutraler Ausfälle in der Teilstichprobe ohne brieflichen Kontakt höher ist (77.3% gegenüber 65.3%). Dies ist vor al-

lem auf eine deutlich höhere Anzahl ungültiger Nummern zurückzuführen. Diese hohe Anzahl ist dadurch zu erklären, dass rund ein Drittel der Telefonnummern, die nicht brieflich kontaktiert wurden, im aktuellsten Telefonbuch nicht mehr verzeichnet waren und somit eine hohe Wahrscheinlichkeit nicht geschalteter Anschlüsse gegeben war. Die Antwortrate ist in der Teilstichprobe mit brieflichem Kontakt mit 63.8 Prozent höher als in der anderen Teilstichprobe (54.6%), während sich die Verweigerungsraten nur geringfügig unterscheiden.

### **3.2 Interviewerangaben zum Fragenverständnis**

Am Ende der Langinterviews wurden die Interviewer um eine Einschätzung gebeten, wie gut die einzelnen Befragten generell die Fragen verstanden haben. Wie einleitend angesprochen, ist in der Angst, dem Interview nicht folgen und die gestellten Fragen nicht verstehen zu können, ein potentieller Erklärungsfaktor für fehlende Teilnahmebereitschaft zu sehen. Der Einschätzung der Interviewer zufolge war das Verständnis der Fragen in knapp 80 Prozent der Fälle gut oder sehr gut, in 16 Prozent zufriedenstellend und in 4 Prozent schlecht oder sehr schlecht. Trotz dieser insgesamt positiven Einschätzung lassen sich verschiedene Faktoren identifizieren, die Zusammenhänge mit dieser Variable aufweisen. So wurde das Fragenverständnis bei den männlichen Befragten etwas besser eingeschätzt als bei den weiblichen, was vor allem in dem höheren Anteil der als sehr gut eingeschätzten Befragten zum Ausdruck kommt (Tabelle 3). Ein deutlicherer Zusammenhang ergibt sich erwartungsgemäß mit dem Alter: Hier nimmt der Anteil der als sehr gut eingeschätzten Befragten mit zunehmendem Alter linear ab.

Darüber hinaus zeigt sich ein Zusammenhang gleicher Stärke zwischen dem Ausmaß des Fragenverständnisses und der Bildung: Bei rund 41 Prozent der Haupt- und Volksschulabsolventen wurde das Verständnis als sehr gut eingeschätzt, 35 Prozent weniger als bei den Befragten mit Abitur oder Fachhochschulreife. Auch bestehen Unterschiede hinsichtlich des Verständnisses in Abhängigkeit von der subjektiven Gesundheit der Befragten. Die in den dokumentierten Kreuztabellen gefundenen Zusammenhänge bleiben im Übrigen auch in multivariaten Analysen bei wechselseitiger Kontrolle der einzelnen Effekte im Wesentlichen erhalten.

Grundsätzlich kommen zur Interpretation dieser Ergebnisse zwei Möglichkeiten in Betracht: Zum einen ist denkbar, dass die Befragtenangaben zu Bildung, Alter und Gesundheit Einfluss auf die Interviewereinschätzung haben. Die andere Möglichkeit besteht darin, dass Befragte mit geringerer Bildung, höherem Alter und schlechterer Gesundheit tatsächlich mehr Schwierigkeiten hatten, die Fragen zu verstehen. Für die letztgenannte Interpretationsmöglichkeit spricht, dass die Interviewer

darauf hingewiesen wurden, dem Rating ihren Gesamteindruck und nicht einzelne Variablen zugrunde zu legen. Die Interviewer wurden darüber hinaus gebeten, die Entwicklung des Verständnisses im Verlauf der Befragung zu beurteilen. Demzufolge ließ sich in 88.4 Prozent der Fälle ein gleichmäßiges Verständnis konstatieren, während bei 4.1 Prozent das Verständnis anfangs schlecht, aber später besser und bei 7.5 Prozent anfangs gut und später schlechter ausgeprägt war, wobei sich das Verständnis bei den ältesten Befragten (76 Jahre und mehr) im Befragungsverlauf häufiger verschlechtert hat als bei den jüngeren. Diese insgesamt positive Einschätzung kann als Hinweis darauf gesehen werden, dass die Länge der Befragung nur in wenigen Fällen zu Schwierigkeiten bei den Befragten geführt hat

**Tabelle 3: Interviewerrating zum Fragenverständnis (Zeilenprozentwerte)**

Variable	sehr gut	gut	zufriedenstellend	schlecht/ sehr schlecht	N (100.0%)
<b>Geschlecht<sup>1</sup></b>					
männlich	57.3	22.5	17.0	3.2	253
weiblich	46.5	32.4	16.6	4.5	374
<b>Alter<sup>2</sup></b>					
60-65 Jahre	69.0	20.9	8.0	2.1	187
66-70 Jahre	53.3	28.7	15.3	2.7	150
71-75 Jahre	47.4	32.8	17.2	2.6	116
76 Jahre und mehr	31.8	33.5	27.2	7.5	173
<b>Bildung<sup>3</sup></b>					
Haupt-/Volksschule	41.2	31.1	22.3	5.4	354
Mittlere Reife	56.4	29.1	11.5	3.0	165
Fachhochschulreife/ Abitur	76.2	17.1	6.7	0.0	105
<b>subjektive Gesundheit<sup>4</sup></b>					
sehr gut	69.2	21.2	9.6	0.0	52
gut	56.5	28.2	13.0	2.3	216
zufriedenstellend	46.7	28.0	21.3	4.0	225
nicht so gut	44.0	30.0	18.0	8.0	100
schlecht	35.3	38.2	17.6	8.8	32

1) Cramer's V=.12, p<.05; 2 Cramer's V=.18, p<.001; 3 Cramer's V=.20, p<.001; 4 Cramer's V=.12, p<.01

### **3.3 Auswirkungen von ausschöpfungssteigernden Maßnahmen auf die Zusammensetzung der Stichprobe**

Zu den wesentlichen Ausfallursachen in telefonischen Befragungen mit alten Menschen gehören mangelnde Teilnahmebereitschaft und Nichtbefragbarkeit aufgrund von Erkrankungen, Hör- oder Verständnisproblemen. Um dem Problem der Nichtbefragbarkeit entgegenzuwirken, wurde die bereits angesprochene Fragebogenkurzfassung für Proxy-Interviews entwickelt. Zur Reduzierung der durch mangelnde Kooperationsbereitschaft bedingten Ausfälle wurden verschiedene Maßnahmen zur Konvertierung eingesetzt.<sup>3</sup>

Haushalte, die keine Informationen über ihre Haushaltsstruktur gegeben hatten, wurden nochmals telefonisch kontaktiert. Durch diese Maßnahme konnten etwa 20 zusätzliche Interviews realisiert und ca. 50 Ausfälle als stichprobenneutral identifiziert werden.

Ebenso wurden Zielpersonen, die zu einem Interview nicht bereit waren, von denen die Adresse jedoch nicht bekannt war, nochmals telefonisch kontaktiert und um ein Interview gebeten.

Zielpersonen, die zu einem Interview nicht bereit waren, und von denen die Adresse bekannt war, wurden in einem Brief nochmals um ihre Teilnahme gebeten. Als zusätzlicher Anreiz war diesem Schreiben ein Marken-Kugelschreiber beigelegt. Etwa 50 Prozent der angeschriebenen Personen wurden daraufhin telefonisch kontaktiert und um ein Interview gebeten. Bei den übrigen 50 Prozent wurde der Fragebogen zur schriftlichen Beantwortung dem Anschreiben beigelegt.

Wie Tabelle 4 zu entnehmen ist, konnte die Antwortrate durch die Proxy-Interviews bzw. durch den Einsatz des Kurzfragebogens vor der Konvertierungsphase um 8.7 Prozent erhöht werden. Durch die Konvertierungsbemühungen konnten insgesamt zusätzliche 114 Interviews (davon 24 anhand des Kurzfragebogens) realisiert werden, was einer Erhöhung der Antwortrate um weitere 8.6 Prozent entspricht. Die Konvertierungsrate durch das nochmalige Anschreiben und das Präsent betrug etwa 26 Prozent, wobei die Rate bei der telefonischen Nachbefragung mit 30 Prozent etwas höher war als die der postalischen Nachbefragung (23%).

---

**3** Die im folgenden beschriebenen Konvertierungsmaßnahmen kamen nur bei solchen nicht kooperationsbereiten Personen zur Anwendung, bei denen der Grund für eine Verweigerung erkennbar situationsspezifischen Umständen und nicht einer expliziten, generell ablehnenden Haltung zuzuschreiben war. Auch wurde auf die Maßnahmen bei solchen Personen verzichtet, bei denen eine nochmalige Kontaktierung aus moralischen Gründen nicht in Frage kam (z.B. bei einem Todesfall in der Familie).

**Tabelle 4: Ausschöpfungssteigerung durch Kurzinterviews und Konvertierung**

	N	Antwortrate
Langinterviews vor Konvertierung	592	44.9%
Kurzinterviews vor Konvertierung	115	8.7%
Lang- und Kurzinterviews durch Konvertierung	114	8.6%
Gesamt	821	62.2%

Im Folgenden wird der Frage nachgegangen, welche Auswirkungen die eingesetzten ausschöpfungssteigernden Maßnahmen auf die Zusammensetzung der Stichprobe haben, indem bereitwillige Befragungsteilnehmer, Teilnehmer der Kurzbefragung und konvertierte Verweigerer im Hinblick auf soziodemographische, sozioökonomische und gesundheitsbezogene Merkmale verglichen werden. Die Analysen sollen Aufschluss über die Notwendigkeit ausschöpfungssteigernder Maßnahmen zur Verbesserung der Stichprobenqualität und zur Reduzierung von durch Ausfällen bedingten Verzerrungen geben.

Tabelle 5 zeigt, dass bei den Kurzinterviews Frauen leicht und Hochaltrige deutlich überrepräsentiert sind. Die durch einen Vergleich mit dem Mikrozensus (Statistisches Bundesamt 1998) festgestellte Überrepräsentierung der 75-jährigen und älteren in der Gesamtstichprobe ist demnach auf die Kurzinterviews mit Referenzpersonen zurückzuführen, d.h. ohne besondere Behandlung der Nicht-Befragbaren ergibt sich hinsichtlich der Altersverteilung in der Stichprobe ein Ergebnis, das eher der Verteilung in der Grundgesamtheit entspricht (zu den mit einem Vergleich von Aggregatstatistik und Stichprobenergebnissen verbundenen Problemen vgl. Schnell 1997). In Bezug auf den Familienstand lässt sich konstatieren, dass Verheiratete vergleichsweise häufig temporär verweigern und der Anteil von Verwitweten bei den Kurzinterviews höher ist als in den beiden anderen Subgruppen.

Hinsichtlich der sozioökonomischen Merkmale Einkommen und Bildung zeigt sich folgendes Bild: Durch die Konvertierung wurden mehr alte Menschen mit geringem Einkommen in die Untersuchung aufgenommen; überdies ist der Anteil der höchsten Einkommensgruppe unter den konvertierten Verweigerern vergleichsweise gering. Auffällig ist zudem der hohe Anteil fehlender Werte bei den Kurzinterviews, der vor allem darauf zurückzuführen ist, dass den Referenzpersonen das Einkommen der Zielpersonen nicht bekannt war. Bezogen auf die Bildung weisen die Ergebnisse darauf hin, dass im Zuge der ausschöpfungssteigernden Maßnahmen der Anteil der unteren Bildungsgruppen (Haupt- und Volksschulabsolventen) in der Untersuchung erhöht wurde. Vor dem Hintergrund des Vergleiches mit den Daten des Mikrozensus ist da-

von auszugehen, dass dies zu einer Verbesserung der Qualität der Stichprobe beigetragen hat.

Erwartungsgemäß sind die alten Menschen, die durch Proxy-Interviews in die Untersuchung einbezogen wurden, durch einen vergleichsweise schlechten Gesundheitszustand gekennzeichnet, der auch in einer verstärkten Einschränkung bei alltäglichen Aufgaben zum Ausdruck kommt.

So sind fast 37 Prozent der durch ein Kurzinterview Befragten durch ihren Gesundheitszustand bei der Erfüllung ihrer alltäglichen Aufgaben sehr stark eingeschränkt. Ein ähnliches Bild - wenn auch deutlich schwächer ausgeprägt - zeigt sich bei der Verteilung von chronischen Erkrankungen. Eine hier nicht dokumentierte Analyse ergab darüber hinaus, dass die Fähigkeit, die Fragen zu verstehen und dem Interview zu folgen bei temporären Verweigerern nicht geringer ausgeprägt ist als bei kooperationsbereiten Befragungsteilnehmern.

### **3.4 Auswirkungen auf Variablenbeziehungen**

Um ein weitergehendes Verständnis der Auswirkungen von Konvertierungsbemühungen auf die Zusammensetzung der Stichprobe zu erlangen, wird im Folgenden der Frage nachgegangen, ob bei konvertierten Verweigerern anders ausgeprägte Zusammenhänge zwischen verschiedenen Variablen festzustellen sind als bei Kooperationsbereiten. Dies soll am Beispiel des für die medizinsoziologische Forschung zentralen Zusammenhanges zwischen sozialer Ungleichheit und Gesundheit erfolgen: Sind die Zusammenhänge zwischen sozialer Ungleichheit, psychosozialen Merkmalen und subjektiver Gesundheit bei konvertierten Verweigerern anders ausgeprägt als bei Kooperationsbereiten? Zur Untersuchung dieser Frage werden mit der subjektiven Gesundheit als abhängiger Variable stratifizierte multiple Regressionsanalysen durchgeführt. Da in diese Analysen Variablen eingehen, die im Rahmen der Kurzinterviews nicht erhoben wurden, beziehen sich die folgenden Aussagen auf die 682 Langinterviews. In die Analysen wird als Indikator für die subjektiv wahrgenommene emotionale Unterstützung die Existenz einer vertrauensvollen Beziehung bzw. einer Vertrauensperson und als Indikator für das soziale Netzwerk die Häufigkeit von Kontakten mit Freunden und Bekannten einbezogen. Zusätzlich werden körperliche und psychische Arbeitsbelastungen berücksichtigt, die jeweils durch Einzelitemmessung erhoben wurden.

**Tabelle 5: Auswirkungen ausschöpfungssteigernder Maßnahmen auf die Zusammensetzung der Stichprobe (Spaltenprozentwerte)**

Variable	Langinterviews (nicht verweigert) (N=592)	Kurzinterviews (nicht verweigert) (N=115)	konvertierte Verweigerer (N=114)	Gesamt (N=821)
<b>Geschlecht</b>				
männlich	41.4	34.5	40.9	40.3
weiblich	58.6	65.5	59.1	59.7
<b>Alter<sup>1</sup></b>				
60-65 Jahre	30.6	16.4	26.4	28.1
66-70 Jahre	22.6	9.1	30.0	21.9
71-75 Jahre	18.4	9.1	17.3	17.0
76 Jahre und mehr	28.2	65.5	26.4	33.0
<b>Familienstand<sup>2</sup></b>				
Verheiratet	54.0	47.0	61.1	54.0
Verwitwet	35.0	44.3	35.4	36.3
Geschieden	6.3	7.7	0.9	5.8
Ledig	4.7	1.0	2.7	3.8
<b>Region</b>				
Ostdeutschland	16.4	18.2	14.4	16.4
Westdeutschland	83.6	81.8	85.6	83.6
<b>Einkommen<sup>3</sup></b>				
unter 30.000 DM	21.1	20.0	34.5	22.7
30.000-40.000 DM	34.0	23.5	29.2	31.8
über 40.000 DM	29.8	21.0	12.4	26.1
weiß nicht/keine Angabe	15.2	35.5	23.9	19.4
<b>Bildung<sup>4</sup></b>				
kein Abschluss	1.7	2.9	0.9	1.7
Haupt-/Volksschule	53.6	77.7	73.1	59.4
Mittlere Reife	27.2	14.6	15.7	24.0
Fachhochschulreife	5.2	0.0	6.5	4.7
Abitur	12.2	4.8	3.7	10.1
<b>subjektive Gesundheit<sup>5</sup></b>				
sehr gut	8.1	2.7	4.5	6.8
gut	35.0	15.9	27.7	31.3
zufriedenstellend	35.3	30.6	45.5	36.0
nicht so gut	16.4	28.9	14.3	17.8
schlecht	5.2	21.9	8.0	8.0
<b>Einschränkung durch Gesundheit<sup>6</sup></b>				
Sehr	9.0	36.8	12.5	13.5
ziemlich	13.7	22.8	15.2	15.2
etwas	26.5	17.5	24.1	24.8
wenig	13.3	7.9	21.4	13.6
gar nicht	37.5	14.9	26.8	32.8
<b>chronische Erkrankung</b>				
ja	58.0	64.0	58.2	58.8
nein	42.0	36.0	41.8	41.2

<sup>1</sup> Cramer's V=.20, p<.001; <sup>2</sup> Cramer's V=.10, p<.01; <sup>3</sup> Cramer's V=.17, p<.001; <sup>4</sup> Cramer's V=.16, p<.001; <sup>5</sup> Cramer's V=.20, p<.001; <sup>6</sup> Cramer's V=.23, p<.001. Die statistischen Kennwerte der übrigen Tabellen sind auf dem 5%-Niveau nicht signifikant.

In Tabelle 6 zeigen sich deutliche Effektunterschiede bei den soziodemographischen Variablen Geschlecht, Alter und Region zwischen temporären Verweigerern und kooperationsbereiten Befragungsteilnehmern. Zudem wird die Selbsteinschätzung der Gesundheit bei Kooperationsbereiten stärker durch das Einkommen beeinflusst als bei konvertierten Verweigerern, während sich hinsichtlich körperlicher Arbeitsbelastungen bei den konvertierten Verweigerern stärkere negative Einflüsse auf die subjektive Gesundheit ergeben als bei kooperationsbereiten Befragungsteilnehmern. Auch bezüglich des Anteils erklärter Varianz zeigen sich leichte Unterschiede zwischen den beiden Subgruppen.

**Tabelle 6: Lineare Regression auf die subjektive Gesundheit mit soziodemographischen, sozialstrukturellen und psychosozialen Merkmalen getrennt für Kooperationsbereite und konvertierte Verweigerer (Beta und Signifikanz (\*<.05))**

VARIABLE	nicht verweigert/konvertiert (N=592)	verweigert/konvertiert (N=90)
WEIBLICH	-.11*	.08
ALTER	-.03	-.20
WESTDEUTSCHLAND	.03	-.10
EINKOMMEN	.20*	.10
BILDUNG	.09	.07
SOZIALE KONTAKTE	.14*	-.04
VERTRAUENSPERSON	.00	.06
KÖRPERLICHE ARBEITSBELASTUNGEN	-.07	-.28*
PSYCHISCHE ARBEITSBELASTUNGEN	-.22*	-.27*
R <sup>2</sup>	.17	.25

#### 4. Zusammenfassung und Diskussion

Alte Menschen gelten als schwer befragbar: Zu den zentralen Schwierigkeiten gehört die gegenüber allgemeinen Bevölkerungsumfragen erhöhte Anzahl systematischer Ausfälle. Diese sind vor allem durch mangelnde Kooperationsbereitschaft und eingeschränkte Befragbarkeit in Folge einer erhöhten Wahrscheinlichkeit von Erkrankungen, Hör- oder Verständnisproblemen bedingt. Im vorliegenden Beitrag wurden Maßnahmen zur Reduzierung solcher Ausfälle und deren Auswirkungen auf



die Stichprobenzusammensetzung am Beispiel einer bundesweiten telefonischen Befragung alter Menschen (60 Jahre und älter) dargestellt. Es lässt sich zunächst konstatieren, dass man durch gezielte Maßnahmen (Proxy-Interviews, Konvertierung) den Hauptursachen für systematische Ausfälle bei telefonischen Befragungen älterer Menschen entgegenwirken und relativ hohe Ausschöpfungsquoten (hier etwa 62 Prozent) erzielen kann. Zudem hat sich gezeigt, dass bei den „regulär“ Befragten - zumindest nach Einschätzung der Interviewer - die Fähigkeit, die Fragen zu verstehen und dem relativ langen Interview zu folgen, in den meisten Fällen gegeben ist (Lundberg/Thorslund 1996).

Darüber hinaus wurden beim Vergleich zwischen Personengruppen mit unterschiedlich ausgeprägter Befragbarkeit und Kooperationsbereitschaft Differenzen im Hinblick auf soziodemographische, sozioökonomische und gesundheitsbezogene Merkmale deutlich. Demzufolge konnten durch die ausschöpfungssteigernden Maßnahmen Teilnehmer mit anders ausgeprägten Merkmalen in die Untersuchung einbezogen werden (Hochaltrige, Verheiratete, alte Menschen mit geringer Bildung und geringem Einkommen sowie alte Menschen mit schlechterem Gesundheitszustand). Überdies haben sich zum Teil deutliche Unterschiede in der Ausprägung multivariater Zusammenhänge zwischen kooperationsbereiten Befragungsteilnehmern und temporären Verweigerern gezeigt. Ein Verzicht auf die Konvertierung von Verweigerern, deren Verhalten weniger auf grundsätzlichen als auf situationsspezifischen Erwägungen beruht, hätte demnach zu tendenziell anderen Aussagen über Richtung und Stärke des Einflusses von soziodemographischen, sozioökonomischen und psychosozialen Variablen auf die subjektive Gesundheit im Alter geführt. Differenzen zwischen Personengruppen mit unterschiedlicher Kooperationsbereitschaft drücken sich demzufolge nicht nur in unterschiedlichen Merkmalsausprägungen und Antwortprofilen, sondern auch in komplexen Variablenbeziehungen aus (Knesebeck/Lüschen 1999). Solche Differenzen lassen Rückschlüsse auf die Bedeutung ausschöpfungssteigernder Maßnahmen zur Verbesserung der Stichprobenqualität und zur Reduzierung von durch (vermeidbaren) Ausfällen bedingten Verzerrungen zu.

Wie in einer jüngst veröffentlichten Arbeit von Keeter et al. (2000) deutlich gemacht wurde, konnte das Verhältnis zwischen Ausschöpfung und Verzerrungen durch Ausfälle bislang nicht eindeutig geklärt werden: „High nonresponse rates can still yield low nonresponse errors (if the difference between respondents and nonrespondents is small), and low nonresponse rates can yield high nonresponse errors (if respondents and nonrespondents differ dramatically).“ (Keeter et al. 2000: 126) So ist es beispielsweise nicht von vornherein klar, inwieweit Verzerrungen reduziert werden, wenn man die Ausschöpfung von 40 auf 60 Prozent erhöht. Die Tatsache, dass sich die

Verteilung bestimmter Variablen - wie in der vorliegenden Untersuchung - mit zunehmender Ausschöpfung verändert, spricht allerdings dafür, dass Maßnahmen zur Ausschöpfungsoptimierung durch die Einbeziehung anderer Bevölkerungsgruppen dazu beitragen, dass ein weniger verzerrtes Abbild der Grundgesamtheit entsteht.

Im Hinblick auf die Stichprobenrealisierung bei telefonischen Befragungen alter Menschen stellt sich die Frage nach der Qualität der in der vorliegenden Untersuchung eingesetzten Screening-Prozedur. Es wurde bereits angedeutet, dass das zugrunde gelegte Verfahren Raum für 'versteckte Verweigerungen' lässt. Es ist nur schwer zu prüfen, ob die Frage „Lebt bei Ihnen im Haushalt eine Person, die 60 Jahre oder älter ist?“ wahrheitsgemäß beantwortet wird oder nicht. Eine Möglichkeit, Hinweise auf die Validität der Screening-Prozedur zu bekommen, besteht darin, den durch die amtliche Statistik dokumentierten Anteil der Privathaushalte, in denen mindestens eine Person lebt, die 60 Jahre oder älter ist, mit dem Anteil aus der vorliegenden Untersuchung zu vergleichen. Eine Auswertung des statistischen Bundesamtes mit dem Mikrozensus ergab für das Jahr 1999 einen Anteil von 36.6 Prozent. In der hier vorgestellten Untersuchung betrug dieser Anteil etwa 46.5 Prozent. Diese relativ hohe Differenz ist nicht zuletzt darauf zurückzuführen, dass in der Untersuchung ausschließlich im Telefonbuch verzeichnete Haushalte berücksichtigt wurden. Wie einleitend angesprochen, ist der Anteil nicht gelisteter Haushalte innerhalb der Population der älteren Menschen geringer als bei den jüngeren. Dies bedeutet umgekehrt, dass der Anteil alter Menschen bei den gelisteten Haushalten - und somit auch in der vorliegenden Untersuchung - höher ist als bei der Gesamtheit aller Privathaushalte. Darüber hinaus ist davon auszugehen, dass es sich bei den nicht erreichten Telefonnummern (n=163) vor allem um Haushalte mit jüngeren Personen handelt.

## **Korrespondenzadresse**

*Dr. Olaf v.d. Knesebeck  
Institut für Medizinische Soziologie  
Universität Düsseldorf  
Postfach 101007  
40001 Düsseldorf  
E-Mail: [knesebec@uni-duesseldorf.de](mailto:knesebec@uni-duesseldorf.de)*

## **Literatur**

- Bungard, W., 1979: Methodische Probleme bei der Befragung älterer Menschen. *Zeitschrift für experimentelle und angewandte Psychologie* 26, 2: 211-237.
- Frey, J.H./Kunz, G./Lüschen, G., 1990: *Telefonumfragen in der Sozialforschung*. Opladen: Westdeutscher Verlag.
- Gabler, S./Häder, S., 1999: Generierung von Telefonstichproben mit TelSuSa. *ZUMA-Nachrichten* 44: 138-148.
- Herzog, A.R./Rogers, W.L., 1988a: Interviewing Older Adults. *Public Opinion Quarterly* 52: 84-99.
- Herzog, A.R./Rogers, W.L., 1988b: Age and Response Rates to Interview Sample Surveys. *Journal of Gerontology: Social Sciences* 43: 200-205.
- Keeter, S./Miller, C./Kohout, A./Groves, R.M./Presser, S., 2000: Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly* 64: 125-148.
- Knesebeck, O.v.d./Lüschen, G., 1998: Telefonische Befragungen bei alten Menschen: Ausschöpfung, Kontakthäufigkeit und Verweigerungen. *ZA-Information* 43: 32-47.
- Knesebeck, O.v.d./Lüschen, G., 1999: Die Bedeutung ausschöpfungssteigernder Maßnahmen für die medizinsoziologische Forschung. *Sozial- und Präventivmedizin* 44: 233-241.
- Knesebeck, O.v.d./Lüschen, G., 2000: Zum Problem von Verweigerungen in telefonischen Befragungen alter Menschen. S. 123-138 in: Hüfken, V. (Hrsg.), *Methoden in Telefonumfragen*. Opladen: Westdeutscher Verlag.
- Koch, A., 1997: Teilnahmeverhalten beim ALLBUS. Soziodemographische Determinanten von Erreichbarkeit, Befragungsfähigkeit und Kooperationsbereitschaft. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 98-122.
- Kohli, M./Künemund, H. (Hrsg.), 2000: *Die zweite Lebenshälfte. Gesellschaftliche Lage und Partizipation im Spiegel des Alters-Survey*. Opladen: Leske+Budrich.
- Kühn, K./Porst, R., 1999: Befragung alter und sehr alter Menschen: Besonderheiten, Schwierigkeiten und methodische Konsequenzen. Ein Literaturbericht. Mannheim: ZUMA-Arbeitsbericht 99/03.
- Kviz, F.J., 1977: Toward a Standard Definition of Response Rate. *Public Opinion Quarterly* 41: 265-267.
- Lundberg, O./Thorslund, M., 1996: Fieldwork and Measurement Considerations in Surveys of the Oldest Old. *Social Indicators Research* 37: 165-187.

- Mingay, D.J. et al., 1994: Self and Proxy Reports of Everyday Events. S. 235-250 in: Schwarz, N./Sudman, S. (Hrsg.), *Autobiographical Memory and the Validity of Retrospective Reports*. Springer: New York.
- Moore, J.C., 1988: Self/Proxy Response Status and Survey Response Quality. *Journal of Official Statistics* 4: 155-172.
- Mosely, R.R./Wolinsky, F.D., 1986: The Use of Proxies in Health Surveys. *Medical Care* 24: 496-510.
- Nuthmann, R./Wahl, H.-W., 1999 Methodische Aspekte der Erhebungen der Berliner Altersstudie. S. 55-83 in: Mayer, K.U./Baltes, P.B. (Hrsg.), *Die Berliner Altersstudie*. Berlin: Akademie.
- Porst, R., 1993: Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Annäherungen aus der ZUMA-Perspektive. Mannheim: ZUMA-Arbeitsbericht 93/12.
- Porst, R./Schneid, M., 1989: Ausfälle bei der Panelbefragung. *Planung und Analyse* 1: 8-13.
- Reuband, K.-H./Blasius, J., 2000: Situative Bedingungen des Interviews, Kooperationsverhalten und Sozialprofil konvertierter Verweigerer. S. 139-167 in: Hüfken, V. (Hrsg.), *Methoden in Telefonumfragen*. Opladen: Westdeutscher Verlag.
- Salmon, C.T./Nichols, J.S., 1983: The Next-Birthday Method of Respondent Selection. *Public Opinion Quarterly* 47: 270-276.
- Schnell, R. 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.
- Schräpler, J.-P., 2000: Was kann man am Beispiel des SOEP bezüglich Nonresponse lernen? *ZUMA-Nachrichten* 46: 117-149.
- Schulte, W. 1997: Telefon- und Face-to-Face-Umfragen und ihre Stichproben. Allgemeine Bevölkerungsumfragen in Deutschland. S. 148-195 in: Gabler, S./Hoffmeyer-Zlotnik, J. (Hrsg.): *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Tennstedt, S.L./Dettling, U./McKinlay, J.B., 1992: Refusal Rates in a Longitudinal Study of Older People: Implications for Field Methods. *Journal of Gerontology: Social Sciences* 47: 313-318.

# **DIE BERECHNUNG DER VARIANZ VON POPULATIONSSCHÄTZERN IM SCIENTIFIC USE FILE DES MIKROZENSUS AB 1996\***

*ULRICH RENDTEL UND BERNHARD SCHIMPL-NEIMANNS*

Erstmals wurden im Scientific Use File des Mikrozensus 1996 Stichprobeninformationen bereitgestellt, die eine Berechnung der Varianz von Populationsschätzern ermöglichen. Nach der Darstellung der Ziehung des Mikrozensus und des Scientific Use Files wird in Anlehnung an Särndal et al. (1992) ein methodisches Konzept zur Berechnung der Varianz entwickelt und auf die Schätzung von Totals, Verhältnis- und Mittelwerten angewendet. Für ausgewählte Merkmale werden Varianzschätzungen auf der Basis des Scientific Use Files mit Ergebnissen des Statistischen Bundesamtes verglichen. Zur Behandlung von hochgerechneten, an die Bevölkerungsfortschreibung angepassten Mikrozensusergebnissen wird eine Regressionschätzung vorgestellt. Erhebliche Differenzen zwischen angepassten und nicht angepassten Schätzergebnissen werfen die Frage verzerrter Populationsschätzungen auf. Abschließend wird der Nutzen von Approximationen der Varianz über die Design-Effekte diskutiert. Es zeigt sich, dass dieses Hilfsmittel auch für das Scientific Use File zu brauchbaren Ergebnissen führt, in Einzelfällen aber mit erheblichen Über- bzw. Unterschätzungen der Varianz verbunden ist. Mit dem Vorliegen der Stichprobeninformationen ab dem Scientific Use File des Mikrozensus 1996 ist es jedoch nicht mehr nötig, diese grobe Varianzschätzung zu verwenden.

---

\* Diese Arbeit entstand als Resultat eines Gastaufenthalts des Erstautors im Oktober 1999 bei ZUMA. Für anregende Diskussionen danken wir Wolf Bihler (Statistisches Bundesamt), Ralf Münnich (Universität Tübingen) sowie Siegfried Gabler, Sabine Häder und Michael Wiedenbeck (ZUMA). Eine erste Fassung des Manuskripts wurde auf der 2. Nutzerkonferenz „Forschung mit dem Mikrozensus“ am 12.-13.10.2000 in Mannheim vorgetragen und liegt in einer kürzeren Version vor (Rendtel/Schimpl-Neimanns 2000). Wir danken Wolf Bihler, Ulrich Pötter und Götz Rohwer sowie zwei Gutachtern der ZUMA-Nachrichten für hilfreiche Kommentare zu einer früheren Fassung.

The Scientific Use File of the German Microcensus 1996 was the first to provide anonymized sample information necessary for calculating variances. The paper outlines the sampling design of the 1996 Microcensus and the selection procedure of the Scientific Use File. Following Särndal et al. (1992), a procedure is presented which uses the available design information to calculate the variance of parameter estimates for totals, ratios, and means. The variance estimates based on the Scientific Use File are compared with results from the Federal Statistical Office. A regression estimator (group mean model) is presented for the post-stratification of the Microcensus estimates. Large differences between the post-stratified data and the unadjusted data raise the question as to which of the two population estimates is biased. We then discuss the behaviour of the linear regression of the design effect, which is frequently used as a tool for variance estimation. In general terms this provides reasonable results, but in individual cases leads to either considerable over- or underestimation. The sample design information now available for the Microcensus means researchers no longer need to rely on such coarse approximations.

## 1. Einleitung

Mit einem Stichprobenumfang von einem Prozent der Personen und Haushalte werden im Mikrozensus jährlich vielfältige Informationen über die demographische, soziale und wirtschaftliche Struktur der Bevölkerung erhoben (vgl. Emmerling/Riede 1997; Lüttinger/Riede 1997). Der Forschung steht eine faktisch anonymisierte 70-Prozent-Substichprobe des Mikrozensus als Scientific Use File mit über 500.000 Personen in rund 230.000 Haushalten zur Verfügung.<sup>1</sup> Dieser große Stichprobenumfang erlaubt die differenzierte Analyse auch kleinerer Bevölkerungsgruppen. Neben der Stichprobengröße ist der sehr kleine Unit-Nonresponse hervorzuheben. Auf Grund der Auskunftspflicht liegt die Teilnahmequote der Haushalte bei rund 97 Prozent. In diesem Zusammenhang und unter der Annahme, dass mit Mikrozensusdaten Verteilungs- und Zusammenhangsmaße für die Population verlässlich geschätzt werden können, dient der Mikrozensus häufig als Referenzstatistik für die normalen Bevölkerungsumfragen, die vielfach Ausschöpfungsquoten unter 60 Prozent aufweisen.

Bei der Berechnung von Stichprobenfehlern ist zu berücksichtigen, dass der Mikrozensus keine uneingeschränkte Zufallsauswahl, sondern eine geschichtete Klumpenstichprobe ist. Weil die Scientific Use Files des Mikrozensus bis einschließlich 1995

---

1 Vgl. zu den Anonymisierungsmaßnahmen im Mikrozensus Müller et al. (1991).

keine Informationen zum Stichprobenplan enthielten, konnten die Nutzer die Varianzschätzungen nur unter der Annahme einer uneingeschränkten Zufallsauswahl durchführen. Da aber die Klumpung in der Regel mit einer Varianzvergrößerung verbunden ist, wird hierbei der Stichprobenfehler unterschätzt. Nur sehr eingeschränkt war mit Hilfe veröffentlichter Fehlerrechnungen des Statistischen Bundesamtes (1998a) eine Korrektur für Design-Effekte möglich.

Aus Sicht der Nutzer der Scientific Use Files war deshalb eine sachgerechte Lösung dieses Problems erforderlich. Um so wichtiger wurde dies, nachdem die Zahl der Mikrozensusnutzer in Folge eines vom Bundesministerium für Bildung und Forschung (BMBF) finanzierten Pilotprojekts (siehe Köhler et al. 2000) stark anstieg. Der im Vergleich zu früher vor allem in finanzieller Hinsicht wesentlich erleichterte Datenzugang hat maßgeblich dazu beigetragen, dass der Mikrozensus von der Forschung zunehmend für die Bearbeitung verschiedenster Fragestellungen herangezogen wird.<sup>2</sup>

Um den vielfältigen Forschungszwecken bei der Analyse von Scientific Use Files gerecht zu werden, reichen die nur für wenige Merkmale vom Statistischen Bundesamt veröffentlichten Design-Effekte nicht aus. Damit Varianzschätzungen auch für die von den Forschern selbst gebildeten Variablen durchgeführt werden können, sind Angaben zur Schichtung und Klumpung im Scientific Use File notwendig. Auf Nachfrage von Mikrozensusnutzern stellt nun das Statistische Bundesamt ab dem Mikrozensus 1996 Informationen über den Stichprobenplan in anonymisierter Form zur Verfügung. Damit ist in der Forschungspraxis erstmals eine adäquate Berechnung der Stichprobenfehler möglich geworden. In diesem Beitrag wird diskutiert, wie man selbst Varianzschätzungen unter Berücksichtigung des Stichprobendesigns mit den faktisch anonymisierten Mikrozensusdaten vornehmen kann und welche Unterschiede ggf. zu den Fehlerrechnungen der statistischen Ämter bestehen.

Dieser Beitrag konzentriert sich auf den Stichprobenfehler, der wie bei allen Stichproben für die Beurteilung der Qualität der Schätzungen von Bedeutung ist. Grundsätzlich sind aber auch systematische Fehler zu beachten, die bei der Berechnung des Stichprobenfehlers unberücksichtigt bleiben. Systematische Fehler entstehen durch Untererfassung, Befragungsausfälle, falsche Angaben und Kodierfehler etc. Systematische Fehlerquellen werden in diesem Beitrag im Zusammenhang mit dem

---

<sup>2</sup> Siehe u.a. die auf zwei Nutzerkonferenzen zum Mikrozensus vorgestellten Arbeiten in Lüttinger (1999) sowie im World Wide Web (URL [http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Veranstaltungen/Veranstaltungen\\_index.htm](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Veranstaltungen/Veranstaltungen_index.htm)).

Unit-Nonresponse und der Anpassung an die Bevölkerungsfortschreibung angesprochen.

Im Folgenden wird zunächst das Erhebungsdesign des Mikrozensus skizziert. Daran schließt sich die Darstellung der Ziehung des Scientific Use Files an. Danach wird gezeigt, wie die Varianz von Totals sowie von Verhältnis- und Mittelwerten bestimmt werden kann. Weiterhin wird die Anpassung der Mikrozensusergebnisse an die Bevölkerungsfortschreibung mit Hilfe des von Särndal et al. (1992) entwickelten Regressionsansatzes behandelt. Anschließend wird die bisher verwendete Methode der linearen Approximation der Design-Effekte untersucht, die auch für die Scientific Use Files vor dem Erhebungszeitpunkt 1996 noch relevant ist. In den abschließenden Bemerkungen fassen wir die wichtigsten Ergebnisse zusammen und nennen Möglichkeiten, wie die Datenbereitstellung verbessert werden kann.

## 2. Das Erhebungsdesign des Mikrozensus

Das Scientific Use File, das im Folgenden mit dem Kürzel FAMZ für faktisch anonymisierter Mikrozensus bezeichnet wird, ist eine 70-prozentige Substichprobe des Mikrozensus (MZ). Um die Varianz von Populationsschätzungen auf Basis des FAMZ bestimmen zu können, ist es deshalb notwendig, zunächst das Erhebungsdesign des MZ darzustellen (siehe Krug et al. 1999: 304ff.; Meyer 1994; Statistisches Bundesamt 1999).<sup>3</sup> Im MZ wurde 1990 ein neuer Stichprobenplan eingeführt, der im Vergleich zum Stichprobenplan für den MZ 1972-1989 eine tiefere regionale Schichtung und eine Verringerung der Klumpengröße aufweist. Zunächst wurden Primäreinheiten (Primary Sampling Units = PSU's) gebildet, die als Auswahlbezirke bezeichnet werden. Sie bestehen durchschnittlich aus neun benachbarten Wohnungen, die in einer Gebäudegruppe oder innerhalb eines größeren Gebäudes liegen.

Die Bildung dieser PSU's erfolgte innerhalb von 4 Gebäudeklassen (sogenannte fachliche Schichtung) auf Basis von Ergebnissen der Volkszählung 1987 bzw. des Zentralen Einwohnerregisters in den neuen Bundesländern ab 1991. Der MZ 1996 enthält über 40.000 PSU's. Da alle Haushalte eines ausgewählten PSU befragt werden, ist der MZ eine Klumpenstichprobe (Cluster Sample). Tendenziell vergrößert

---

<sup>3</sup> Siehe hierzu ergänzend auch die Darstellungen der statistischen Landesämter: Frank/Kafurke 1990; Reinders 1993; Schmidt 1990; Werner 1994.



die Klumpung die Varianz der Populationschätzer im Vergleich zu einer uneingeschränkten Zufallsauswahl.<sup>4</sup>

Das zweite Element des MZ ist die regionale Schichtung, die den umgekehrten Effekt hat, nämlich die Varianz zu verringern. Großstädte über 200 Tsd. Einwohner und sonstige Regionen über 250 Tsd. Einwohner bilden eigene Schichten. Innerhalb der Schichten werden darüber hinaus Schichtuntergruppen identifiziert, die aus Regionaleinheiten mit mindestens 100 Tsd. Einwohnern bestehen.

Vor der eigentlichen Auswahl wurden die PSU's innerhalb der fachlichen und regionalen Schichten nach den Merkmalen Bundesland, Regierungsbezirk, Kreis, Gemeindegroßenklasse und Gemeinde angeordnet. Innerhalb der Gemeinden wurden die PSU-Nummern, welche die regionale Anordnung innerhalb der Schichtuntergruppen widerspiegeln, zu einer Anordnung der PSU's der Auswahlgesamtheit verwendet. Jeweils 100 aufeinander folgende PSU's wurden zu einer Zone zusammengefasst.

Der Kern des Ziehungsverfahrens besteht in einer zufälligen Zuordnung von jeweils 100 PSU's einer Zone zu den Zahlen 0 bis 99. Für jede Zone wurde eine neue zufällige Permutation der Zahlen 0 bis 99 verwendet. Die PSU's zu einer Zahl zwischen 0 bis 99 bildeten dann jeweils eine MZ-Stichprobe.

Nach dem Stichprobendesign des MZ wird eine Zerlegung der 1-Prozent-Stichproben in 0,25-Prozent-Stichproben benötigt, da im Rahmen des Rotationsverfahrens jährlich ein Viertel des Bestandes ausgetauscht wird.<sup>5</sup> Hierzu wurden jeweils vier aufeinanderfolgende Zonen einer Zufallszahl von 1 bis 4 zugeordnet. Dabei wurde für jeden Block von vier aufeinanderfolgenden Zonen eine Permutation der Zahlen 1 bis 4 verwendet. PSU's mit gleicher Nummer zählten zum gleichen Rotationsviertel, so dass nach diesem Verfahren die Auswahlgesamtheit der PSU's in Rotationsviertel zerlegt wurde. Aus dieser Auswahlgesamtheit von 100 1-Prozent-Stichproben der PSU's wurden 20 Vorratsstichproben über eine Ziehung aus der Urne ausgewählt.

Interpretiert man dieses Verfahren innerhalb des Schemas der klassischen Stichprobentheorie, so entsprechen die Zonen Schichten. Für die Berechnung der Gesamt-

---

**4** Der Stichprobenfehler ist in der Regel um so größer, je homogener die Klumpen hinsichtlich der interessierenden Merkmale, je größer die Klumpen, und je unterschiedlicher die Klumpengrößen sind.

**5** Der MZ ist als rotierende Panelstichprobe angelegt, bei der ein Auswahlbezirk bzw. die darin wohnenden Haushalte vier Jahre lang befragt werden, wobei jährlich ein Viertel der PSU's ausgetauscht wird. Wegziehende Personen bzw. Haushalte werden allerdings nicht weiterbefragt, sondern durch die nachziehenden Haushalte ersetzt (Prinzip der Flächenstichprobe).

varianz würde die Varianz der PSU-Totals innerhalb einer Zone benötigt. Da aber aus jeder Zone jeweils nur eine Primäreinheit gezogen wurde, kann die Varianz der PSU-Totals innerhalb einer Zone nicht berechnet werden. Dieser Ansatz führt deshalb nicht weiter. Eine Möglichkeit mit dieser Schwierigkeit umzugehen, besteht darin, die Schichten größer zu definieren und die Zonenstruktur des Ziehungsverfahrens zu ignorieren; das heißt, für diese größeren Schichten wird eine uneingeschränkte Zufallsauswahl angenommen. Diese Strategie wird auch in den Varianzberechnungen des Statistischen Bundesamtes angewendet. Innerhalb jeder Gebäudeklasse bilden die insgesamt über 200 Regionalschichten jeweils eine Schicht bei der Varianzberechnung. Die Annahme größerer Schichten und das Ignorieren der kleinräumigen Struktur bei dem Ziehungsverfahren führt in der Tendenz zu einer Überschätzung der Varianz. Vor dem Hintergrund, dass auf Grund fehlender Detailangaben zum Stichprobendesign im FAMZ die Varianz vermutlich nur annähernd berechnet werden kann, ist es vielleicht überraschend, dass schon im Fall des MZ - so, wie er in den statistischen Ämtern vorliegt - die Stichprobenvarianz ebenfalls nur näherungsweise bestimmt werden kann.

Die Benutzung eines vereinfachten Berechnungsverfahrens wird auch durch einen zweiten Sachverhalt nahegelegt. In den Jahren nach der Ziehung der MZ-Stichproben sind Neubauten entstanden bzw. ganze Flächen neu bebaut worden. Die Neubauten werden bei der Aktualisierung der Stichproben in einer zusätzlichen „Neubauschicht“ berücksichtigt und unterliegen einem gesonderten Ziehungsverfahren.

Bei der Aktualisierung der Stichprobe werden zunächst auf Kreis- bzw. Gemeindeebene Primäreinheiten auf Basis der Bautätigkeitsstatistik gebildet. Diese PSU's werden nach Gebäudegrößenklassen geschichtet und innerhalb jeder Schicht nach der Reihenfolge ihrer Bildung durchnummeriert. Im Unterschied zu dem oben beschriebenen Verfahren erfolgt die Ziehung der Neubau-PSU's aber durch systematisches Ziehen mit festem Intervall bei zufälligem Startpunkt.

Für das systematische Ziehungsverfahren ist bekannt, dass keine erwartungstreuen Varianzschätzer existieren (vgl. Wolter 1985: 248ff.). Man muss sich auch in diesem Fall mit einer Vereinfachung begnügen. Erfahrungen aus Simulationsstudien zeigen, dass die Approximation über die Varianz einer in den Schichten uneingeschränkten Zufallsstichprobe in der Regel zu einer Überschätzung der Varianz führt (vgl. Wolter 1985: 282).

Da alle Stichproben nach dem selben Verfahren ermittelt wurden, ergibt sich die Möglichkeit, die Variation der Schätzergebnisse über die Rotationsviertel für die Varianzschätzung zu benutzen. Allerdings ist im FAMZ die Zugehörigkeit der Haus-

halte zu den Rotationsgruppen bisher nicht dokumentiert. Bei Kenntnis der Rotationsgruppenzugehörigkeit könnte man die Varianz beispielsweise über das Jackknife-Verfahren analysieren (vgl. Wolter 1985: 153ff) und hierbei Effekte der Schichtung durch Anordnung sowie des systematischen Ziehens bei der Neubauschicht berücksichtigen.

### **3. Die Ziehung des Scientific Use Files aus dem MZ 1996**

Aus Datenschutzgründen konnte die Regionalschicht als regionales Schichtungsmerkmal nicht in den FAMZ aufgenommen werden.<sup>6</sup> Darüber hinaus wurde im Vergleich zum MZ ein anderes Ziehungsverfahren verwendet, das aus folgenden Schritten besteht.

Zunächst wurden die Haushalte des MZ nach Bundesland, Regierungsbezirk, Gemeindegrößenklasse, Zahl der Personen im Privathaushalt, Auswahlbezirksnummer und Haushaltsnummer angeordnet. Diese Anordnung unterscheidet sich von der Schichtung im MZ hauptsächlich hinsichtlich der Zahl der Personen im Privathaushalt in der Sortierfolge. In dieser Anordnung wurden die Haushalte neu durchnummeriert. In die Auswahl von 70 Prozent aller Haushalte wurden alle Haushalte – einschließlich aller Personen in diesen Haushalten – übernommen, deren letzte Platzziffer der Haushaltsnummer von 2, 5 und 9 verschieden war. Im Anschluss an die Stichprobenziehung wurden die Haushalte umsortiert und erhielten eine neue, fortlaufende Nummerierung, so dass die alten Haushaltsendziffern nicht mehr rekonstruierbar sind.

Dieses vom MZ abweichende Verfahren erschwert die Berechnung der Stichprobenvarianz. Insbesondere ist die fehlende Kenntnis über die Anzahl der Haushalte pro PSU im MZ beziehungsweise der Verlust der Information über die letzte Platzziffer beim Auswahlverfahren insofern bedauerlich, als man aus der Varianz der Schätzergebnisse über die sieben verbliebenen Platzziffern Rückschlüsse auf die zusätzliche Varianzkomponente gewinnen könnte, die durch die 70-Prozent-Auswahl verursacht wird.

---

<sup>6</sup> Zur Zeit stehen der Forschung Scientific Use Files des Mikrozensus nur als sogenanntes Grundfile zur Verfügung, das durch tiefe fachliche Gliederungen und sehr grobe regionale Gliederungen charakterisiert ist. Die Bereitstellung eines faktisch anonymisierten Mikrozensus-Regionalfile durch die statistischen Ämter ist in Vorbereitung.

#### 4. Die Berechnung der Varianz im Scientific Use File

Obwohl das Scientific Use File somit im strengen Sinne über ein 2-phasiges Ziehungsverfahren auf Haushaltsebene gewonnen wurde, ist das Verfahren in guter Näherung durch ein zweistufiges Auswahlverfahren beschreibbar.<sup>7</sup> Hierbei entspricht die erste Stufe der geschichteten Auswahl der PSU's im Mikrozensus. Die zweite Stufe kann als eine einfache 70-Prozent-Auswahl von Haushalten aus einer PSU betrachtet werden.

Wie im letzten Abschnitt festgehalten, ist die Annahme fester Stichprobenumfänge für die zweite Auswahlstufe durch das systematische Ziehungsverfahren nicht immer gegeben und bildet eine Quelle für die Variation der Schätzergebnisse, die mangels entsprechender Information ignoriert werden muss. Weiterhin sind aufgrund des systematischen Ziehungsverfahrens die Ziehungen der Haushalte zwischen den einzelnen PSU's nicht voneinander unabhängig. Es kann jedoch vermutet werden, dass die aus beiden Sachverhalten resultierende Unterschätzung der Varianz eine untergeordnete Bedeutung hat (vgl. hierzu auch Särndal et al. 1992: 349ff).

Zusammenfassend ist festzuhalten, dass der FAMZ nicht alle Schichtungsmerkmale der ersten Auswahlstufe enthält, sondern nur die Merkmale Bundesland, Gemeindegroßenklasse sowie die Gebäudeschicht. Grundsätzlich führt die Nichtberücksichtigung von Schichtungsmerkmalen zu einer Überschätzung der Varianz, so dass man bei der Bildung von Konfidenzintervallen konservativ bleibt.

Die im Folgenden benutzte Notation lehnt sich an das Lehrbuch von Särndal et al. (1992) an. Es bezeichne  $h \in \{1, \dots, H\}$  den Schichtindex. Die PSU's werden mit  $i \in \{1, \dots, N_h\}$  indiziert, wobei  $N_h$  die Anzahl der PSU's in der  $h$ -ten Schicht ist. Haushalte werden mit  $k \in \{1, \dots, N_i\}$  indiziert, wobei  $N_i$  die Anzahl der Haushalte im  $i$ -ten PSU ist. Der Merkmalswert  $y_{h,i,k}$  bezieht sich somit auf Haushalt  $k$  in PSU  $i$  in Schicht  $h$ ;  $y_k$  verweist auf den Merkmalswert von Haushalt  $k$ , unabhängig von dessen PSU-Nummer und der Schichtzugehörigkeit.

---

<sup>7</sup> Mehrphasige unterscheiden sich von mehrstufigen Stichproben darin, dass bei Phasenauswahlen die Auswahlseinheiten (z.B. 1. Phase: Auswahl der MZ-Haushalte, 2. Phase: 70-Prozent-Auswahl der FAMZ-Haushalte) gleich sind, während mehrstufigen Auswahlen unterschiedliche Auswahlseinheiten zu Grunde liegen (z.B. 1. Stufe: PSU's, 2. Stufe: Haushalte). Siehe zum Vergleich von Varianzschätzungen bei 2-phasigen und 2-stufigen Stichproben Särndal et al. 1992: 349-350).

Für ein 2-stufiges geschichtetes Ziehungsverfahren müssen unterschiedliche *Grundgesamtheiten* berücksichtigt werden. Es sei  $U_I$  die Menge aller Primäreinheiten im Erhebungsgebiet (Anzahl =  $N_I$ ).  $U_{I,h}$  sei die Menge der Primäreinheiten in Schicht  $h$  (Anzahl =  $N_{I,h}$ ). Die Menge aller Haushalte im  $i$ -ten PSU sei  $U_i$  (Anzahl =  $N_i$ ). Schließlich bezeichne  $U$  die Menge aller Haushalte im Erhebungsgebiet (Anzahl =  $N$ ).

Den Grundgesamtheiten entsprechen die unterschiedlichen *Stichproben*:  $s_I$  ist die Stichprobe der PSU's vom Umfang  $n_I$ . Diese Stichprobe verteilt sich auf die  $H$  Schichten als  $s_{I,h}$  mit dem Umfang  $n_{I,h}$ .

Die Stichprobe der Haushalte aus PSU  $i$  ist  $s_i$  und besitzt den Umfang  $n_i$ . Schließlich bezeichnet  $s$  die Stichprobe aller Haushalte mit dem Umfang  $n$ .

Weiterhin bezeichnet  $\pi_{I,i}$  die *Ziehungswahrscheinlichkeit* von PSU  $i$  und  $\pi_{k|i}$  die bedingte *Ziehungswahrscheinlichkeit* von Haushalt  $k$  aus PSU  $i$ , wenn PSU  $i$  gezogen wurde. Im Falle des MZ gilt  $\pi_{I,i} = 0,01$ . Für den FAMZ nehmen wir  $\pi_{k|i} = 0,7$  an.

## 5. Schätzung von Gesamtwerten qualitativer Merkmale (Totals)

Bevor wir zeigen, wie sich diese Stichprobeninformationen für die Varianzschätzung verwenden lassen, soll ein einführendes Beispiel aus der Situation vor der Weitergabe von Angaben zur Schichtung und Klumpung im FAMZ die Ausgangsproblematik und den möglichen Informationsgewinn verdeutlichen.

Nehmen wir an, ein Forscher sei im Rahmen einer Analyse von Lebenslagen allein lebender Personen an der Zahl weiblicher 1-Personenhaushalte in der Grundgesamtheit interessiert. Der Einfachheit halber wird bei der Hochrechnung die Anpassung der Ergebnisse an die Bevölkerungsfortschreibung nicht berücksichtigt (siehe Abschnitt 8). Die Auswertung des FAMZ 1996 ergibt, dass unter den insgesamt 229.221 Privathaushalten 49.071 weibliche 1-Personenhaushalte sind. Nach Designgewichtung der Fallzahlen mit dem Kehrwert der *Ziehungswahrscheinlichkeiten* ( $1/(0,01 \cdot 0,7)$ ) ergibt sich hochgerechnet ein Gesamtwert (Total) von 7.010,1 Tsd. Haushalten. Um die stichprobenbedingte Unsicherheit dieser Schätzung zu berücksichtigen, musste ein Nutzer bislang von einer uneingeschränkten Zufallsauswahl ausgehen. Unter Annahme einer Binomialverteilung berechnet man den relativen Standardfehler und erhält einen Wert von 0,4 Prozent (siehe Statistisches Bundesamt 1998a: 17; siehe auch Gleichung (28) im Abschnitt 9). Um die Schichtung und

Klumpung des MZ näherungsweise zu berücksichtigen, ist dieser Wert mit dem Design-Effekt zu multiplizieren. Verwendet man die Näherungsfunktion des Statistischen Bundesamtes (1998a: 17), beträgt der Design-Effekt für 1-Personenhaushalte 1,36. Für das interessierende Merkmal erhält man schließlich den Wert 0,54 ( $=0,4 \cdot 1,36$ ) als Schätzung des relativen Standardfehlers. Ein 95-Prozent-Konfidenzintervall für den geschätzten Populationswert reicht somit von circa 6.936 bis 7.084 Tsd. Haushalten. Würde man, wie vom Statistischen Bundesamt empfohlen, den für den MZ berechneten design-basierten Standardfehler an den geringeren Stichprobenumfang des FAMZ anpassen und entsprechend mit 1,2 multiplizieren, verbreiterte sich das Konfidenzintervall auf circa 6.921 bis 7.099 Tsd. Haushalte.

Wie können nun die ab dem FAMZ 1996 vorliegenden Informationen zum Stichprobendesign für die direkte Varianzschätzung verwendet werden? Der Gesamtwert (Total) eines Merkmals  $y$  ist gegeben durch:

$$(1) \quad t = \sum_{k \in U} y_k = \sum_{i \in U_i} t_i \quad \text{wobei } t_i = \sum_{k \in U_i} y_k$$

$$(2) \quad = \sum_{h=1}^H t_h \quad \text{wobei } t_h = \sum_{i \in U_{i,h}} t_i$$

Hierbei ist (1) die Darstellung über die Totals der Primäreinheiten und (2) die Darstellung über die Schichttotals.

Der  $\pi$ -Schätzer von  $t$  basiert auf dem Kehrwert der Ziehungswahrscheinlichkeiten. Man erhält für das PSU-Total die Schätzung:

$$(3) \quad \hat{t}_i = \sum_{k \in s_i} \frac{y_k}{\pi_{k|i}} = \frac{1}{0,7} \sum_{k \in s_i} y_k$$

Das Schichttotal  $t_h$  lässt sich schätzen durch:

$$(4) \quad \hat{t}_h = \sum_{i \in s_{i,h}} \frac{\hat{t}_i}{\pi_{I,i}} = 100 \sum_{i \in s_{i,h}} \hat{t}_i$$

Man erhält damit als Schätzung für das Gesamt-Total:

$$(5) \quad \hat{t} = \sum_{h=1}^H \hat{t}_h = \frac{100}{0,7} \sum_{k \in s} y_k$$

Wir wollen im folgenden  $V(\hat{t})$ , die Varianz von  $\hat{t}$ , bestimmen. Bei Unabhängigkeit der Ziehung zwischen den Schichten erhält man:

$$(6) \quad V(\hat{t}) = \sum_{h=1}^H V_h(\hat{t}_h)$$

$$(7) \quad \hat{V}(\hat{t}) = \sum_{h=1}^H \hat{V}_h(\hat{t}_h)$$

Das heißt, es genügt, die Varianz des geschätzten Totals  $\hat{t}_h$  für die einzelnen Schichten zu bestimmen. Bei einem 2-stufigen Ziehungsverfahren, mit dem auf jeder Stufe eine einfache (simple: SI) Stichprobe gezogen wird (Typ: (SI, SI) in der Notation von Särndal et al. 1992: 142), erhält man:

$$(8) \quad V_{SI,SI}(\hat{t}_h) = N_{I,h}^2 \frac{1-f_h}{n_{I,h}} S_{U_{I,h}}^2 + \frac{N_{I,h}}{n_{I,h}} \sum_{i \in U_{I,h}} N_i^2 \frac{1-f_i}{n_i} S_{U_i}^2$$

wobei  $f_h = \frac{n_{I,h}}{N_{I,h}} = 0,01$ ,  $f_i = \frac{n_i}{N_i} = 0,7$  und:

$$(9) \quad S_{U_{I,h}}^2 = \frac{1}{N_{I,h} - 1} \sum_{i \in U_{I,h}} (t_i - \bar{t}_{U_{I,h}})^2$$

= Varianz der PSU - Werte in der Schicht  $h$   
= "Varianz between PSU"

$$(10) \quad S_{U_i}^2 = \frac{1}{N_i - 1} \sum_{k \in U_i} (y_k - \bar{y}_{U_i})^2$$

= Varianz der y - Werte im  $i$ -ten PSU  
= "Varianz within PSU"

Für die gegebenen Werte vereinfacht sich  $V_{SI,SI}$  zu:

$$(11) \quad V_{SI,SI} = 100^2 \times 0,99 \times n_{I,h} \times \text{"Varianz between PSU"} \\ + 100 \sum_{i \in U_{I,h}} \frac{0,3}{0,7^2} n_i \times \text{"Varianz within PSU"}$$

Indem in Gleichung (8) bzw. (11) die Between- und Within-Varianzen durch ihre Stichprobenpendants ersetzt werden, ergibt sich die Varianzschätzung  $\hat{V}_{SI,SI}$ :

$$(12) \quad S_{S_{I,h}}^2 = \frac{1}{n_{I,h} - 1} \sum_{i \in S_{I,h}} \left( \hat{t}_i - \hat{\bar{t}}_{S_{I,h}} \right)^2 \\ = \text{Varianz der PSU - Werte in Schicht } h$$

$$(13) \quad S_{S_i}^2 = \frac{1}{n_i - 1} \sum_{k \in S_i} \left( y_k - \bar{y}_{S_i} \right)^2 \\ = \text{Varianz der } y \text{ - Werte im } i \text{-ten PSU}$$

Man beachte allerdings, dass die beiden Varianzterme in Gleichung (11) durch die Stichprobenpendants nicht erwartungstreu geschätzt werden (vgl. Särndal et al. 1992: 142). Die Between-Komponente wird im Mittel durch die Varianz der geschätzten PSU-Totals überschätzt. Lediglich die Summe der beiden Stichproben-terme ist eine erwartungstreue Schätzung für  $V_{SI,SI}$ .

Die Berechnung des zweiten Teils von  $\hat{V}_{SI,SI}$  verlangt die Ermittlung der Standardabweichung der y-Werte innerhalb von circa 40.000 PSU's. Dies ist zwar rechenaufwendig, jedoch kein prinzipielles Problem. Särndal et al. (1992: 139ff.) schlagen als Näherung für  $\hat{V}_{SI,SI}$  die Berechnung lediglich des ersten Teils über die Between-Varianzen vor. Dies ist gerechtfertigt, wenn die Auswahlwahrscheinlichkeit der Primäreinheiten klein ist. Wie die numerischen Ergebnisse in Tabelle 1 belegen, ist im FAMZ die Within-Varianz eine zu vernachlässigende Größe im Vergleich zur Between-Varianz. Für die Schätzung der Varianz können deshalb Standardprozeduren verwendet werden, die mittlerweile in Statistikpaketen wie STATA und SAS implementiert sind. In diesen Prozeduren wird lediglich die erste Komponente der Between-PSU Varianz innerhalb der Schichten berechnet.



Die Tabelle 1 zeigt für einige ausgewählte Merkmale die Schätzung von Totals auf Basis des FAMZ und den Vergleich der Schätzgenauigkeit mit den Angaben für den MZ (Statistisches Bundesamt 1998b). Weiterhin werden die relativen Standardfehler miteinander verglichen sowie die in Abschnitt 9 behandelte Näherungsfunktion auf Basis der Design-Effekte. Für das eingangs geschilderte Beispiel weiblicher 1-Personenhaushalte finden wir bei Anwendung des direkten Schätzverfahrens mit 0,47 Prozent im Vergleich zu 0,54 Prozent bei der bisher notwendigen Vorgehensweise einen deutlich geringeren relativen Standardfehler. Damit erhalten wir mit einem 95-Prozent-Konfidenzintervall von circa 6.946 bis 7.075 Tsd. Haushalten eine präzisere Schätzung.

Betrachten wir die geschätzten Totals in Tabelle 1, fällt auf, dass die Werte für den FAMZ stets kleiner sind als für den MZ. Bei den meisten Merkmalen liegen die MZ-Werte weit außerhalb des konventionellen 95-Prozent-Konfidenzintervalls der FAMZ-Werte. Diese systematische Abweichung ist aber kein Hinweis auf eine verzerrte Substichprobenziehung, sondern darauf zurückzuführen, dass das Statistische Bundesamt für seine Berechnungen eine Kompensationsgewichtung für die circa 2,5 Prozent Nonresponse-Fälle auf Haushaltsebene benutzt. Die für die einzelnen Haushalte unterschiedlichen Gewichte liegen jedoch im FAMZ nicht vor und können folglich auch nicht für die Schätzung von Totals benutzt werden.<sup>8</sup>

Der Vergleich der Standardfehler zeigt, dass die FAMZ-Werte stets über den MZ-Werten liegen. Dies war auch angesichts des um den Faktor 0,7 niedrigeren Stichprobenumfangs zu erwarten. Die Reduktion des Stichprobenumfangs würde bei einer uneingeschränkten Zufallsauswahl zur Vergrößerung des Standardfehlers um rund 20 Prozent führen. Empirisch ergibt sich jedoch nur eine Vergrößerung um durchschnittlich 9 Prozent.<sup>9</sup> Der Grund für die Halbierung des Fallzahleffekts liegt in der Verringerung des Klumpeneffekts, der sich aus der Auswahl von Haushalten unter approximativer Erhaltung der Schichtung bei der Substichprobenziehung ergibt.

---

<sup>8</sup> Bezogen auf das Standardauswertungsprogramm von über 450 Merkmalen betragen die FAMZ-Totals durchschnittlich 97,7 Prozent der entsprechenden MZ-Totals.

<sup>9</sup> Das Statistische Bundesamt hat auf einem Workshop zur Nutzung des FAMZ 96 für die Berechnung des Stichprobenfehlers empfohlen, den Design-Effekt mit dem Faktor 1,2 zu multiplizieren. Statt des Faktors  $(1/0,7)^2 = 1,195$  beträgt der Faktor für den FAMZ bezogen auf die über 450 Merkmale des Standardprogramms durchschnittlich 1,09.

**Tabelle 1: Schätzung von Gesamtwerten (Totals) für ausgewählte Merkmale**

Merkmal	Total (in 1000)		Standardfehler (in 1000)			relativer Standardfehler (in Prozent)			Design- Effekt		
	FAMZ	MZ	FAMZ			MZ	Näherung			FAMZ	MZ
			Insges.	Betw.	Within		FAMZ	MZ	Näherung		
Ausländische Erwerbspersonen <sup>1)</sup>	2.283,7	2.330,6	29,4	29,4	1,2	27,3	1,29	1,17	0,92	1,66	1,83
Sofort verfügbare Erwerbslose <sup>1)</sup>	2.976,7	3.034,2	24,1	24,0	1,2	21,4	0,81	0,71	0,81	1,20	1,26
Überwiegender Lebensunt. durch Erwerbstätigkeit <sup>1)</sup>	29.607,9	30.285	81,4	81,4	3,2	77,7	0,28	0,26	0,30	1,63	1,85
Weibliche Erwerbst., monatl. Nettoeink. <600 DM <sup>1)</sup>	1.545,4	1.585,4	15,9	15,8	0,8	13,9	1,03	0,88	1,11	1,08	1,12
Privathaushalte mit 1 Person, weiblich <sup>2)</sup>	7.010,1	7.259,6	33,1	33,1	1,5	32,0	0,47	0,44	0,54	1,19	1,35

**Quelle:** FAMZ: Mikrozensus 1996 (faktisch anonymisierte 70%-Substichprobe); MZ: Frei hochgerechnete Werte nach Kompensation der bekannten Ausfälle (Fehlerrechnungen zum Mikrozensus 1996; Statistisches Bundesamt 1998b).

**Näherung:** Aus den Besetzungszahlen der Tabellenfelder geschätzter relativer Standardfehler unter Verwendung der veröffentlichten Zuschlagsfaktoren zum Mikrozensus 1990 (Statistisches Bundesamt 1998a: 17); Näherungsfunktion; s. Abschnitt 9.

**Subpopulationen:** 1) Bevölkerung am Hauptwohnsitz; 2) Privathaushalte.

Die Betrachtung des relativen Standardfehlers (=Standardfehler/Total) belegt die hohe Präzision von Ergebnissen auf Basis des FAMZ, die ihresgleichen bei den üblichen sozialwissenschaftlichen Erhebungen sucht. Selbst bei relativ schwach besetzten Merkmalen, wie zum Beispiel erwerbstätigen Frauen mit einem Nettoeinkommen unter 600 DM, liegt der relative Standardfehler bei 1 Prozent.<sup>10</sup> Wie aber der Vergleich der FAMZ-Totals mit den durch das Kompensationsgewicht hochgerechneten MZ-Werten gezeigt hat, liegen bereits bei kleinen Stichprobenausfällen die MZ-Schätzwerte weit außerhalb des Konfidenzintervalls, das man für den

<sup>10</sup> Nicht zu verwechseln mit einem Prozentpunkt.

FAMZ erhält. Dieser Sachverhalt erklärt sich aus dem Fehlen von Nonresponse-Informationen im FAMZ, das einen Bias in zweifacher Hinsicht bewirkt.

Zum einen werden die Populationstotals systematisch unterschätzt. Diese Unterschätzung ließe sich beheben, wenn die Kompensationsgewichtung im FAMZ erhalten wäre.<sup>11</sup> Die Kompensationsgewichte geben die geschätzten Ausfallraten innerhalb der 401 regionalen Untergruppen für 19 verschiedene Merkmalskombinationen an.<sup>12</sup> Die Kompensationsgewichte könnten für eine Korrektur der Inklusionswahrscheinlichkeiten  $\pi_{k|i}$  auf der zweiten Auswahlstufe verwendet werden. Ersatzweise kann man die durchschnittliche Ausfallrate von rund 2,5 Prozent verwenden. Multipliziert man die FAMZ-Schätzwerte in Tabelle 1 mit dem Faktor 1,025 so überdecken alle Konfidenzintervalle die entsprechenden Werte aus dem MZ.

Zum anderen bewirkt das Ausblenden des Nonresponse jedoch auch eine gewisse Unterschätzung des Stichprobenfehlers, da die Streuung der Stichprobenvarianz aufgrund des Nonresponse ignoriert wird. Hinsichtlich der geringen Ausfallrate dürfte dieser Effekt aber zu vernachlässigen sein.

Auch wenn die genannten Probleme die Varianzschätzung beeinträchtigen, kann zusammenfassend festgehalten werden, dass die direkte Varianzberechnung im Vergleich zur bisher notwendigen Vorgehensweise für die Nutzer des FAMZ erhebliche Vorteile einer präziseren Varianzschätzung unter Berücksichtigung des Stichprobendesigns besitzt.

## 6. Schätzung eines Verhältniswerts

Für die im Mikrozensus überwiegenden qualitativen Merkmale werden neben Gesamtwerten häufig auch Verhältnis- oder Anteilswerte betrachtet. Da bei dieser Verhältnissschätzung sowohl für das Zähler- als auch das Nennermerkmal Stichprobenfehler zu berücksichtigen sind, ist die Berechnung nicht ganz so einfach wie bei Totals.

---

<sup>11</sup> Das Grundfile enthält als Gewichtungsvariable lediglich die Gewichtung nach Anpassung an die Bevölkerungsfortschreibung, die sowohl die unterschiedlichen Ausfallraten als auch die Soll-Ist Anpassung nach der Bevölkerungsfortschreibung berücksichtigt; vgl. hierzu nachfolgenden Abschnitt 8.

<sup>12</sup> Auf Haushaltsebene (d.h. für Haushaltsbezugspersonen) werden die Merkmalsgruppen über die Haushaltsgröße (1,2,3≤), die Nationalität (Deutsche/Ausländer), den Wohnsitz (Haupt-/Nebenwohnsitz; nur für Deutsche) sowie für 1-Personenhaushalte das Geschlecht und das Alter (≤60, >60) gebildet.

Es sei  $t_y$  ein Total bezüglich eines Merkmals  $y$  und  $t_z$  ein Total bezüglich eines Merkmals  $z$ . Die zu schätzende Größe wird durch  $\hat{R} = \hat{t}_y / \hat{t}_z$  dargestellt. Zur Berechnung der Varianz von  $\hat{R}$  wird eine Taylorentwicklung der Funktion  $f(t_y, t_z) = t_y / t_z$  benutzt:

$$\begin{aligned}
 f(\hat{t}_y, \hat{t}_z) &\approx f(t_y, t_z) + \frac{\partial f}{\partial t_y} \Big|_{t_y} (\hat{t}_y - t_y) + \frac{\partial f}{\partial t_z} \Big|_{t_z} (\hat{t}_z - t_z) \\
 &= f(t_y, t_z) + \frac{1}{t_z} (\hat{t}_y - t_y) - \frac{t_y}{t_z^2} (\hat{t}_z - t_z) \\
 &= \text{Konstante} + \frac{1}{t_z} (\hat{t}_y - R \hat{t}_z) \\
 (14) \quad &= \text{Konstante} + \frac{1}{t_z} \sum_{k \in s} \frac{u_k}{\pi_k} \quad \text{wobei } u_k = (y_k - R z_k)
 \end{aligned}$$

Damit ist bis auf die Konstante  $1/t_z$  lediglich die Varianz des  $\pi$ -Schätzers für das U-Total zu bestimmen. Wegen der linearen Approximation der Schätzfunktion wird allerdings nur eine Näherung der Varianz von  $\hat{R}$  bestimmt, die für große Beobachtungsumfänge mit  $V(\hat{R})$  übereinstimmt (Bezeichnung: Asymptotische Varianz  $AV(\hat{R})$ ):

$$(15) \quad AV(\hat{R}) = \frac{1}{t_z^2} V(\hat{t}_u)$$

Zur Berechnung von  $V(\hat{t}_u)$  kann man wieder Gleichung (8)-(11) verwenden. Man hat dabei lediglich  $y_k$  durch  $u_k$  zu ersetzen.  $S_{U_i}^2$  misst die Varianz der Hilfsgrößen  $u_k$  innerhalb eines PSU. Die Hilfsgröße  $u_k$  kann als Abweichung der  $y_k/z_k$  Werte von dem Populationswert  $R$  interpretiert werden.

Eine asymptotisch erwartungstreue Schätzung von  $AV(\hat{R})$  erhält man, indem die unbekanntenen Populationswerte  $t_z$  und  $R$  durch die Schätzwerte  $\hat{t}_z$  und  $\hat{R}$  ersetzt und die Gleichungen (12) und (13) benutzt werden. Hierbei ist  $\hat{t}_i$  gegeben durch:

$$(16) \quad \hat{t}_i = \sum_{k \in s_i} \frac{u_k}{\pi_{k|i}} = \frac{1}{0,7} \sum_{k \in s_i} (y_k - \hat{R}z_k)$$

Der Mittelwert der unter (16) geschätzten PSU Werte über alle PSU's der Schicht  $h$  entspricht dem Ausdruck  $\hat{t}_{s_i, h}$  in Gleichung (12).

Die Schätzung der Varianz von  $\hat{R}$  bedarf damit keiner gesonderten Programmierung. Man hat lediglich statt des Merkmals  $y_k$  das Merkmal  $u_k = y_k - \hat{R}z_k$  zu verwenden und darf am Schluss die Division durch  $\hat{t}_z^2$  nicht vergessen.

## 7. Die Varianz von Populationsmitteln

Um auch für quantitative Variablen, wie beispielsweise Alter oder Zahl der Kinder im Haushalt etc., die Varianz von Populationsmitteln zu bestimmen, kann wie folgt vorgegangen werden. Der Populationsmittelwert zu einem Merkmal  $y$  ist gegeben durch:

$$\bar{y}_U = \frac{1}{N} \sum_{k \in U} Y_k = \frac{t_y}{N}$$

Falls der Populationswert  $N$  bekannt ist, lässt sich  $\bar{y}_U$  schätzen durch:

$$\hat{y}_{U, \pi} = \frac{\hat{t}_y}{N}$$

Eine geringere Varianz hat jedoch häufig das gewichtete Stichprobenmittel  $\hat{y}_s$  (vgl. Särndal et al. 1992: 182):

$$\hat{y}_s = \frac{\sum_{k \in s} Y_k / \pi_k}{\sum_{k \in s} 1 / \pi_k}$$

Dieser Schätzer wird von den meisten Programmpaketen benutzt, sobald eine Gewichtungvariable verwendet wird. In dieser Form ist  $\hat{y}_s$  ein Spezialfall von  $\hat{R}$  mit  $z_k = 1$ , so dass die Varianzschätzung für Verhältniswerte  $\hat{R}$  benutzt werden kann.

Man erhält mit  $\hat{N} = \sum_{k \in s} 1 / \pi_k$  für das Hilfsmerkmal  $u_k$  den folgenden Wert:

$$(17) \quad u_k = y_k - \frac{\hat{t}_y}{\hat{N}} \cdot 1 = y_k - \hat{y}_s$$

Das Hilfsmerkmal  $u_k$  misst also die Abweichung der  $y$ -Werte vom gewichteten Stichprobenmittel. Der Faktor  $1/\hat{t}_z^2$  ist durch  $1/\hat{N}^2$  gegeben.

Die Tabelle 2 zeigt für die Merkmale von Tabelle 1 den geschätzten Anteil (in Prozent) der Merkmalsträger an der Bevölkerung am Hauptwohnsitz beziehungsweise den Privathaushalten insgesamt. Die letzte Spalte in Tabelle 2 zeigt den Wert für den Standardfehler, wenn man die Insgesamt-Werte als bekannte Größe benutzt, indem man beispielsweise die im letzten Abschnitt geschätzten Standardfehler der Totals auf Anteilswerte umrechnet. Man erhält das auf den ersten Blick verblüffende Ergebnis, dass diese Schätzung in allen Fällen ungenauer ist als die Schätzung mit variablem Nenner. Dieser Effekt ist um so größer, je häufiger das Merkmal in der Population vertreten ist, das heißt je stärker Zähler- und Nennermerkmal kovariieren. Beispielsweise wird der Anteil der Personen mit überwiegendem Lebensunterhalt aus Erwerbstätigkeit um rund 30 Prozent genauer geschätzt, wenn der Nenner ebenfalls geschätzt wird. Für das Beispielmerkmal weibliche 1-Personenhaushalte wirkt sich dieser Effekt praktisch nicht aus. Sowohl im Fall eines geschätzten wie auch bekannten Nennermerkmals beträgt das 95-Prozent-Konfidenzintervall 21,2 bis 21,6 Prozent. Man beachte auch hier die extreme Genauigkeit; selbst bei relativ seltenen Merkmalen.

**Tabelle 2: Schätzung von Anteilswerten für ausgewählte Merkmale**

Merkmal Y	Anteil R (in %)	Standardfehler (in %)	
		N gesch.	N bek.
Ausländische Erwerbspersonen <sup>1)</sup>	3,2	0,040	0,041
Sofort verfügbare Erwerbslose <sup>1)</sup>	4,1	0,033	0,034
Überwiegender Lebensunterhalt durch Erwerbstätigkeit <sup>1)</sup>	41,3	0,081	0,114
Weibliche Erwerbstätige mit einem monatlichen Nettoeinkommen unter 600 DM <sup>1)</sup>	2,2	0,021	0,022
Privathaushalte mit 1 Person, weiblich <sup>2)</sup>	21,4	0,093	0,101

**Quelle:** Mikrozensus 1996 (faktisch anonymisierte 70%-Substichprobe).

N. gesch./bek.: Standardfehler bei Verwendung der Fallzahl im Nenner (Z) als zu schätzende bzw. bekannte Größe.

Z-Merkmal: 1) Bevölkerung am Hauptwohnsitz; 2) Privathaushalte.

## 8. Die Varianz von Populationsschätzern nach der Anpassung an die Bevölkerungsfortschreibung

Wir sind bisher bei der Hochrechnung ausschließlich von den Stichprobenwerten ausgegangen. Von den statistischen Ämtern werden jedoch die Mikrozensusergebnisse anhand von Eckzahlen der laufenden Bevölkerungsfortschreibung hochgerechnet. Mit der Bereitstellung dieser Hochrechnungsfaktoren werden zwei Ziele verfolgt: Zum einen soll die Varianz der Populationsschätzungen verringert werden. Zum anderen soll ein möglicher Bias aufgrund der Stichprobenausfälle von rund 2,5 Prozent kompensiert werden. Zu diesem Zweck wird ein Hochrechnungsfaktor auf der Basis von regionalen Eckzahlen zu sechs Anpassungsklassen<sup>13</sup> gebildet, der im Wesentlichen das Verhältnis der Sollzahlen aus der Bevölkerungsfortschreibung zu den Ist-Zahlen der im Mikrozensus Befragten abbildet (vgl. Heidenreich 1994). Die Nutzer der faktisch anonymisierten Mikrozensusdaten können für diese sogenannte gebundene Hochrechnung auf die im Datensatz enthaltenen Hochrechnungsfaktoren für Personen und Haushalte zurückgreifen. Bei deren Verwendung sollten sich die FAMZ-Ergebnisse nicht wesentlich von den veröffentlichten Gesamtwerten der amtlichen Statistik unterscheiden. Im Gegensatz zu der Veröffentlichung angepasster MZ-Ergebnisse basieren die Fehlerrechnungen des Statistischen Bundesamtes nur auf den Stichprobenwerten, korrigiert um den Unit-Nonresponse auf Haushaltsebene. Die Anpassung der Fallzahlen an die Bevölkerungsfortschreibung bleibt bei der Berechnung des Stichprobenfehlers außer Acht. Es ist deshalb zu fragen, in welcher Weise die Anpassung an die Bevölkerungsfortschreibung („Gewichtung“) bei der Varianzschätzung berücksichtigt werden kann.

Formal kann die Verwendung von Gewichten, die aus der Anpassung der MZ-Fallzahlen an die Bevölkerungsfortschreibung resultieren, als Regressionsschätzung interpretiert werden. Der hier benutzte Regressionsschätzer basiert auf dem Group Mean Modell, das sich wie folgt darstellen lässt.

---

<sup>13</sup> Die Anpassung der MZ-Ergebnisse an die Bevölkerungsfortschreibung nach den Klassen Geschlecht in Kombination mit Staatsangehörigkeit (Deutsche/Ausländer) erfolgt regional auf der Ebene von Anpassungsschichten; das sind regionale Einheiten mit wenigstens 500.000 Einwohnern. Bei der Anpassung der Ergebnisse für Soldaten (einschl. Bundesgrenzschutz und Bereitschaftspolizei) und Wehrpflichtige werden Bestandsmeldungen aus dem Verteidigungs- bzw. Innenministerium auf Regierungsbezirks- bzw. Länderebene herangezogen.

Die Grundgesamtheit  $U$  lässt sich in  $G$  disjunkte Teilmengen  $U_g$ ,  $g \in \{1, \dots, G\}$  zerlegen. Für die Elemente innerhalb jeder Gruppe  $g$  gilt:

$$(18) \quad E_{\xi}(y_k) = \beta_g \quad k \in U_g$$

$$(19) \quad V_{\xi}(y_k) = \sigma_g^2 \quad k \in U_g$$

Hierbei bezeichnet  $E_{\xi}(\cdot)$  die Erwartungswertbildung bezüglich einer Verteilung  $\xi$  über den Merkmalen und  $V_{\xi}(\cdot)$  die entsprechende Varianz. Die Erwartungswertbildung bezieht sich hier also nicht auf den Wert eines Populationsschätzers bezüglich des Erhebungsdesigns, sondern auf die Realisierung der Merkmalswerte bei einer gegebenen Stichprobe.<sup>14</sup>

Die OLS-Schätzung von  $\beta_g$  auf Basis der Daten für die Grundgesamtheit ist:

$$(20) \quad B_g = \frac{1}{N_g} \sum_{k \in U_g} y_k = \bar{y}_{U_g}$$

wobei:  $N_g =$  Anzahl Elemente von  $U_g$ . Die Gruppendifinition liefert eine Zerlegung der Stichprobe:  $s_g = s \cap U_g$ . Als Schätzer für  $B_g$  benutzt man:

$$(21) \quad \hat{B}_g = \frac{\sum_{k \in s_g} y_k / \pi_k}{\sum_{k \in s_g} 1 / \pi_k} = \frac{1}{\hat{N}_g} \sum_{k \in s_g} \frac{y_k}{\pi_k}$$

Hierbei ist  $\hat{N}_g$  der geschätzte Wert für den Umfang der Gruppe  $g$ . Im Fall des FAMZ sind die Zichungswahrscheinlichkeiten aller Einheiten gleich. Folglich gilt:

$$(22) \quad \hat{B}_g = \bar{y}_{s_g} \quad g = (1, \dots, G)$$

---

<sup>14</sup> Die Verwendung des statistischen Modells wird nur für die Herleitung des Populationsschätzers benötigt. Die Eigenschaften des resultierenden Schätzers werden jedoch wieder unter dem klassischen Design-basierten Ansatz ermittelt. Särndal et al. (1992: 238) sprechen daher von einem ‚model assisted‘ Ansatz. Der Vorteil der hier gewählten Darstellung liegt in allgemein hergeleiteten Formeln für die Schätzung der Varianz von Regressionsschätzern.



Es bezeichne  $\hat{y}_k$  den durch das Regressionsmodell geschätzten Wert von  $y_k$ . Im vorliegenden Fall erhält man:

$$(23) \quad \hat{y}_k = \hat{B}_g = \bar{y}_{s_g} \quad k \in s_g$$

Der Regressionsschätzer  $\hat{t}_{reg}$  hat für das Group Mean Modell die folgende Gestalt:

$$(24) \quad \hat{t}_{reg} = \sum_{k \in U} \hat{y}_k = \sum_{g=1}^G \sum_{k \in U_g} \hat{B}_g = \sum_{g=1}^G N_g \hat{B}_g = \sum_{g=1}^G \sum_{k \in s_g} \frac{N_g}{\hat{N}_g} \cdot \frac{y_k}{\pi_k} = \sum_{g=1}^G \sum_{k \in s_g} w_g \frac{y_k}{\pi_k}$$

Der Faktor  $w_g = N_g / \hat{N}_g$  beschreibt das Verhältnis von  $N_g =$  Umfang von Gruppe  $g$  in der Grundgesamtheit (= Soll - Vorgabe) zu  $\hat{N}_g =$  geschätzter Umfang von Gruppe  $g$  (= Ist - Wert).

Bei der Anpassung an die Bevölkerungsfortschreibung wird für  $N_g$  jeweils der Wert für bestimmte Gruppen gemäß dieser Fortschreibung gewählt. Im FAMZ ist eine Gewichtungvariable  $w_k$  enthalten, die für jede Person  $k \in s_g$  den jeweiligen Wert  $N_g / \hat{N}_g$  („Soll durch Ist“) annimmt.<sup>15</sup> Der Regressionsschätzer lässt sich damit als ein „gewichtetes Mittel“ darstellen:

$$(25) \quad \hat{t}_{reg} = \sum_{k \in s} w_k \frac{y_k}{\pi_k}$$

Eine wesentliche Eigenschaft des Regressionsschätzers liegt allgemein darin, dass er für Merkmale, die in die Soll/Ist-Anpassung eingingen, unter jeder Stichprobe wieder die Soll-Werte liefert (vgl. Särndal et al. 1992: 324). Folglich hat  $\hat{t}_{reg}$  für diese Merkmale die Varianz Null. Allerdings ist der FAMZ nur eine 70-Prozent-Substichprobe aus dem MZ und wurde nicht extra an die Bevölkerungsfortschreibung angepasst. Daher wird die Varianzschätzung für den Regressionsschätzer auch im Fall der Anpassungsmerkmale positive Werte liefern.

---

<sup>15</sup> Darüber hinaus liegt ein Personengewicht für die Hochrechnung der Unterstichprobe vor, in der Merkmale der Arbeitskräftestichprobe der EU sowie Ergänzungs- und Zusatzprogramme erhoben werden. Für Haushalte bzw. Familien wurde das Haushaltsgewicht als arithmetisches Mittel der Personenfaktoren im Haushalt gebildet.

Für die Herleitung von  $V(\hat{t}_{reg})$  wird wieder eine Taylorentwicklung von  $\hat{t}$  benutzt. Der lineare Teil der Taylorentwicklung ist durch die folgende Hilfsgröße  $u_k$  gegeben (vgl. Särndal et al. 1992: 331):

$$(26) \quad u_k = \frac{N_g}{\hat{N}_g} (y_k - \hat{B}_g) = w_k (y_k - \bar{y}_{s_g}) \quad k \in s_g$$

Dieser lineare Anteil ist also die mit  $w_k$  gewichtete Abweichung des Merkmalswerts  $y_k$  von dem jeweiligen Gruppenmittelwert  $\bar{y}_{s_g}$  in der Stichprobe. Als asymptotische Varianz  $V(\hat{t}_{reg})$  wird die Varianz dieses Hilfsmerkmals  $u$  verwendet. Bei der praktischen Berechnung hat man also lediglich in Gleichung (8) - (13)  $y_k$  durch  $u_k$  zu ersetzen.<sup>16</sup>

In Tabelle 3 wird für die Merkmale von Tabelle 1 der Einfluss der Anpassung an die Bevölkerungsfortschreibung auf die Schätzung von Totals und ihrer Varianz dargestellt. Es zeigt sich, dass der relative Standardfehler durch die Anpassung etwas verringert wird. Für Merkmale, die eng mit den Anpassungsklassen zusammenhängen, fällt die Reduktion stärker aus. Beispielsweise reduziert sich der relative Standardfehler für das Merkmal „1-Personenhaushalte, weiblich“ von 0,47 auf 0,27 Prozent; ähnlich stark auch beim Merkmal „Ausländische Erwerbsperson“. Jedoch besteht ein bemerkenswerter Unterschied zwischen den Schätzwerten in der Größenordnung von teilweise mehreren Millionen, bei weiblichen 1-Personenhaushalten um fast 900 Tsd. Haushalte. Dies zeigt die Problematik bei der Verwendung der Hochrechnungsfaktoren auf Basis der Bevölkerungsfortschreibung: Sie führen insgesamt betrachtet praktisch kaum zu einer Verringerung der Varianz, sondern verdecken vielmehr einen Bias. Entweder liefert der MZ und damit auch der FAMZ verzerrte Populationsschätzer oder aber die Bevölkerungsfortschreibung produziert ihrerseits verfälschte Schätzungen.

---

<sup>16</sup> Für Haushaltsauswertungen können die Anpassungsklassen nach den Eigenschaften der Haushaltsbezugsperson gebildet werden.

**Tabelle 3: Schätzung von Bevölkerungstotalen und relativer Standardfehler (in Prozent) für ausgewählte Merkmale mit und ohne Anpassung an die Bevölkerungsfortschreibung**

Merkmal Y	Total (in 1000)		relativer Standardfehler (in %)	
	mit Anp.	ohne Anp.	mit Anp.	ohne Anp.
Ausländische Erwerbspersonen <sup>1)</sup>	3.609,7	2.283,7	0,72	1,29
Sofort verfügbare Erwerbslose <sup>1)</sup>	3.490,3	2.976,7	0,79	0,81
Überwiegender Lebensunterhalt durch Erwerbstätigkeit <sup>1)</sup>	33.806,1	29.607,9	0,20	0,28
Weibliche Erwerbstätige mit einem monatlichen Nettoeinkommen unter 600 DM <sup>1)</sup>	1.740,7	1.545,4	1,00	1,03
Privathaushalte mit 1 Person, weiblich <sup>2)</sup>	7.896,6	7.010,1	0,27	0,47

**Quelle:** Mikrozensus 1996 (faktisch anonymisierte 70%-Substichprobe)

Subpopulation — verwendeter Hochrechnungsfaktor:

- 1) Bevölkerung am Hauptwohnsitz — Personen-Hochrechnungsfaktor
- 2) Privathaushalte — Haushalts-/Familienhochrechnungsfaktor.

Als potentielle Quelle dieses Bias kommen beim Mikrozensus Probleme hinsichtlich der Realisierung der Stichprobe in Frage, wie beispielsweise Aktualität des Stichprobenplans, Erreichbarkeit der Haushalte und Unit-Nonresponse usw. In diesem Zusammenhang darf vermutet werden, dass die Untererfassung der Soldaten und Wehrpflichtigen, die sich in einem Soll/Ist-Verhältnis von ca. 1,6 widerspiegelt, durch Nichterreichbarkeit verursacht wird.<sup>17</sup> Aber auch die Bevölkerungsfortschreibung ist nicht fehlerfrei. Insbesondere bei den Ausländern wird vermutet, dass Wegzüge nur unzureichend erfasst sind und die Bevölkerungsfortschreibung deshalb zu hohe Ausländerzahlen ausweist. Mit der Gewichtung werden die Mikrozensusergebnisse auf Ergebnisse der Bevölkerungsfortschreibung adjustiert, was zur Übertragung von Fehlern der Bevölkerungsfortschreibung auf den Mikrozensus führen kann. Die Anpassung führt beim FAMZ zu einer Hochgewichtung der Deutschen um einen Faktor von circa 1,1 und der Ausländer sogar um über 1,5.

<sup>17</sup> Gemäß dem Stichprobenplan umfassen die Auswahlbezirke (PSU's) keine Kasernen. Soldaten und Wehrpflichtige sind somit nicht in den Kasernen, sondern nur an ihrem Haupt- bzw. Nebenwohnsitz erreichbar.

Eine „Korrektur“ in dieser Größe lässt sich einerseits nicht aus den Ergebnissen der Feldarbeit herleiten (siehe Heidenreich 1994: 116). Andererseits zeigen Vergleiche der Ergebnisse der Bevölkerungsfortschreibung und der Volkszählung 1987 wesentlich kleinere Abweichungen (Heidenreich 1989: 328; Jäger 1992: 105f.; vgl. auch Krug et al. 1999: 234). Die beträchtlichen Differenzen zwischen MZ und Bevölkerungsfortschreibung können somit keinesfalls überwiegend auf Fehler der Bevölkerungsfortschreibung zurückgeführt werden. Zur Klärung dieser Frage fehlen jedoch systematische Untersuchungen über die Fehlerquellen und zu Abweichungen zwischen Mikrozensus und anderen Bevölkerungsstatistiken.

## 9. Design-Effekte

Wie am Anfang des vierten Abschnitts beispielhaft dargestellt, waren bisher die Nutzer des FAMZ bei der Schätzung der Varianz von Totals auf die Verwendung von Ergebnissen der Fehlerrechnung des Statistischen Bundesamtes (1998a) angewiesen, in denen Design-Effekte („Zuschlagsfaktoren“) berichtet werden. Ein Design-Effekt beschreibt das Verhältnis des design-basierten Standardfehlers des MZ zum Standardfehler einer Stichprobenziehung gleichen Umfangs, die jedoch ohne Klumpung und Schichtung durchgeführt worden wäre – also unter der Annahme einer uneingeschränkten Zufallsauswahl. Kern dieses Ansatzes ist eine lineare Regression der Design-Effekte auf den geschätzten Populationsanteil der durch das Tabellenfeld definierten Merkmalsträger.

Durch die Substichprobenziehung verringert sich im FAMZ die Zahl der Haushalte und Personen pro Auswahlbezirk im Vergleich zum MZ, und damit auch der Klumpeneffekt. Es ist deshalb davon auszugehen, dass die für den MZ veröffentlichten Design-Effekte nicht einfach auf den FAMZ übertragbar sind und ihre Verwendung zu einer Überschätzung der Varianz im FAMZ führt. Wir wollen daher prüfen, ob die Ziehung des FAMZ zu anderen Design-Effekten führt. Außerdem soll untersucht werden, wie gut die lineare Approximation der Design-Effekte ist.

Es bezeichne  $U_d$  eine Teilmenge (Domain) der Grundgesamtheit, die das Feld einer Tabelle charakterisiert.  $N_d$  sei die Anzahl von  $U_d$  in der Grundgesamtheit und  $n_d$  in der Stichprobe.

Mit Hilfe der Indikatorfunktion

$$y_k = \begin{cases} 1 & \text{falls } k \in U_d \\ 0 & \text{sonst} \end{cases}$$

kann man die Schätzung von Domains auf die Schätzung von Totals zurückführen, da in diesem Falle  $N_d = \sum_{k \in U} y_k = t_y$  gilt. Für die Schätzung des Populationsanteils  $P_d = N_d/N$  durch  $\hat{P}_d = \hat{t}_y/N$  erhält man im Fall einer uneingeschränkten Zufallsauswahl (vgl. Särndal et al. 1992: 70):

$$(27) \quad \hat{V}_{SI}(\hat{P}_d) = \frac{1-f}{n-1} p_d(1-p_d)$$

wobei  $p_d = n_d/n$  das Stichprobenpendant zu  $P_d$  ist.

Der relative Standardfehler ist der Variationskoeffizient von  $\hat{P}_d$ . Für die uneingeschränkte Zufallsauswahl ist er gegeben durch:

$$(28) \quad cv_{SI} = \frac{\sqrt{\hat{V}_{SI}(\hat{P}_d)}}{\hat{P}_d} = \sqrt{\frac{1-f}{n-1} \frac{1-p_d}{p_d}}$$

Der Design-Effekt  $k(\hat{p}_d)$  ist dann definiert durch:<sup>18</sup>

$$(29) \quad k(\hat{p}_d) = \frac{\sqrt{\hat{V}(\hat{p}_d)}}{\sqrt{\hat{V}_{SI}(\hat{p}_d)}}$$

Dieses Verhältnis kann für jede Subpopulation anders ausfallen. Das Statistische Bundesamt benutzt für drei Merkmalsgruppen (Bevölkerung und Erwerbstätige (B/E), Ausländer und Erwerbstätige in der Landwirtschaft (A/L) und Haushalte (H)) eine unterschiedliche einfache lineare Regression von  $k(\hat{p}_d)$  auf  $\hat{p}_d$ :

$$(30) \quad k(\hat{p}_d) \approx a + b\hat{p}_d$$

---

<sup>18</sup> Häufig wird der Design-Effekt als das Verhältnis der Varianzen definiert; so zum Beispiel bei Särndal et al. (1992: 54).

Für gegebene Werte von  $a$  und  $b$  erhält man unter Verwendung von (29) und (30) die folgende Näherung für den relativen Standardfehler in Abhängigkeit von  $p_d$ :

$$(31) \quad cv = k(p_d)cv_{SI} = (a + bp_d) \sqrt{\frac{1-f}{n-1} \frac{1-p_d}{p_d}}$$

Tabelle 4 vergleicht zusammenfassend die auf Basis des FAMZ jeweils für die einzelnen Gruppen ermittelten Regressionskoeffizienten mit den für den MZ 1990 veröffentlichten Werten und den Berechnungen für den MZ 1996. Insgesamt verlaufen die Geraden für den FAMZ 1996 flacher als für den MZ 1996. Bezogen auf die über 450 Merkmale des Standardauswertungsprogramms beträgt die Reduktion etwa 10 Prozent und ist für die drei Merkmalsgruppen B/E, A/L und H in etwa gleich. Diese Verringerung des Design-Effekts hängt mit der Reduktion des Klumpeneffekts bei der Substichprobenziehung zusammen. Lässt man einmal außer acht, dass die in der Näherungsfunktion verwendeten Koeffizienten für den MZ 1990 veraltet sind und sich nur auf das frühere Bundesgebiet beziehen, kann durch Vergleich der MZ90- und FAMZ-Werte in Tabelle 4 festgestellt werden, dass der relative Standardfehler des FAMZ für die hier ausgewählten Merkmale in brauchbarer Näherung wiedergegeben wird.

**Tabelle 4: Vergleich der Koeffizienten einer einfachen linearen Regression des Design-Effekts auf den geschätzten Bevölkerungsanteil**

Gruppe	Datenbasis	Konstante	Steigung
Bevölkerung und Erwerbstätige (B/E)	FAMZ	1.009	1.84
	MZ96	1.042	2.44
	MZ90	1.136	1.61
Ausländer und Erwerbstätige in der Landwirtschaft (A/L)	FAMZ	1.088	21.69
	MZ96	1.162	25.47
	MZ90	1.169	25.04
Haushalte (H)	FAMZ	0.988	1.01
	MZ96	1.009	1.60
	MZ90	1.119	1.14

Datenbasis:

FAMZ: Mikrozensus 1996 (faktisch anonymisierte 70%-Substichprobe).

MZ96: Eigene Berechnungen auf Basis unveröffentlichter Fehlerrechnungen zum Mikrozensus 1996 (Statistisches Bundesamt 1998b).

MZ90: Fehlerrechnung zum Mikrozensus 1990 (Statistisches Bundesamt 1998a: 22).

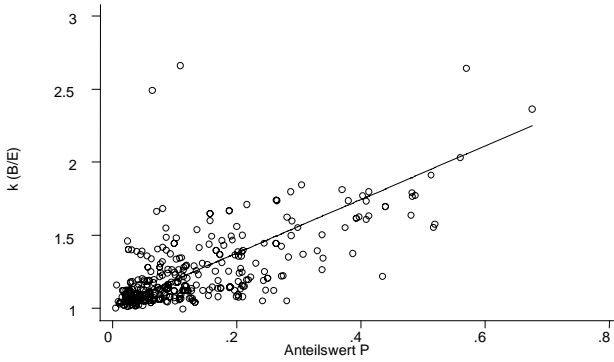
Es bleibt noch zu überprüfen, wie gut die lineare Beziehung in Gleichung (30) zu den berechneten Design-Effekten passt.<sup>19</sup> Die Abbildungen 1 bis 3 zeigen die Regressionen der Design-Effekte für die Merkmale des MZ-Standardauswertungsprogramms für die Gruppen Bevölkerung und Erwerbstätige (B/E), Ausländer und Erwerbstätige in der Landwirtschaft (A/L) sowie Haushalte (H). Für das Beispiel weibliche 1-Personenhaushalte wurde empirisch ein Design-Effekt von 1,19 ermittelt (siehe Tabelle 1). Verwendet man die lineare Approximation, erhält man mit einem Wert von 1,20 eine sehr gute Schätzung. Hingegen streuen die empirisch ermittelten Design-Effekte für die Merkmalsgruppe Bevölkerung und Erwerbstätige (B/E) wesentlich stärker. Für einen Anteilswert von rund 20 Prozent liegt das Minimum der Design-Effekte bei 1,08 und das Maximum bei 1,56; durch die Regression wird für diesen Anteil ein Design-Effekt von rund 1,38 geschätzt. Die auf Basis direkter Varianzschätzungen ermittelten Konfidenzintervalle können sich folglich um mehrere Zehntausend Personen von dem auf der linearen Approximation basierenden Konfidenzintervall unterscheiden und zu verschiedenen Aussagen führen.

Insgesamt zeigt sich, dass die lineare Approximation zwar ein brauchbares Modell zur Beschreibung des Design-Effekts liefert, aber bei einzelnen Merkmalen doch beträchtliche Abweichungen des jeweiligen Design-Effekts von der Regressionsgeraden vorliegen. In Einzelfällen führt also die Verwendung der Design-Effekte zu erheblichen Über- bzw. Unterschätzungen der Varianz, wobei über die Richtung des Fehlers keine Aussage gemacht werden kann. Mit dem Vorliegen der Information über den Auswahlbezirk und die Gebäudeschicht ist es jedoch nicht mehr nötig, sich auf diese Varianzabschätzung zu verlassen. Nutzer der Mikrozensus 1991-1995, in denen diese Stichprobeninformationen nicht enthalten sind, können ersatzweise die hier berichteten Design-Effekte für den FAMZ 1996 bei Varianzschätzungen verwenden.

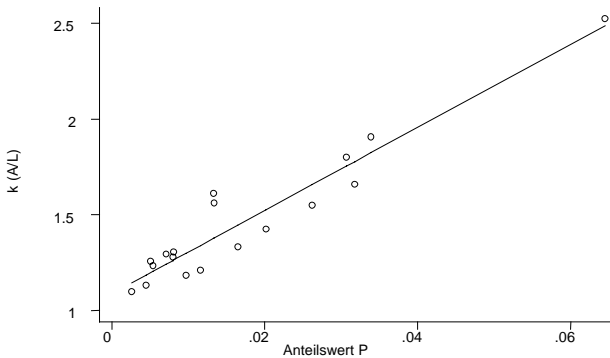
---

<sup>19</sup> Für den MZ sind bisher nur für 1990 die entsprechenden Regressionskoeffizienten veröffentlicht worden. Eine Überprüfung der Angemessenheit des einfachen Regressionsmodells ist bis auf den Hinweis, daß die Abweichungen der berechneten von den geschätzten Design-Effekten im Mittel kleiner als 15-20 Prozent sind (Statistisches Bundesamt 1998a: 17), bisher nicht dokumentiert worden.

**Abbildung 1: Regression des Design-Effekts ( $k$ ) für die Merkmalsgruppe Bevölkerung und Erwerbstätige (B/E) auf den Anteilswert für 346 Merkmale des FAMZ**

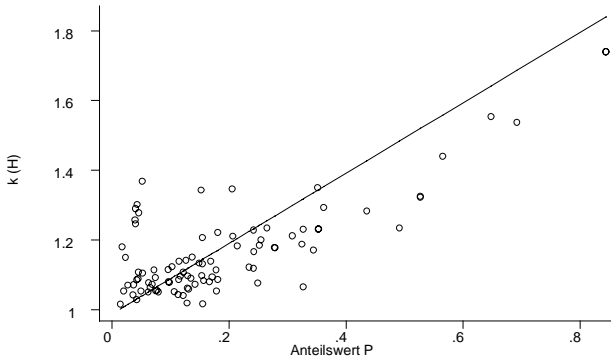


**Abbildung 2: Regression des Design-Effekts ( $k$ ) für die Merkmalsgruppe Ausländer und Erwerbstätige in der Landwirtschaft (A/L) auf den Anteilswert für 18 Merkmale des FAMZ**





**Abbildung 3: Regression des Design-Effekts ( $k$ ) für die Merkmalsgruppe Haushalte (H) auf den Anteilswert für 94 Merkmale des FAMZ**



## 10. Abschließende Bemerkungen

Im vorliegenden Beitrag wurde gezeigt, dass die ab dem Scientific Use File des Mikrozensus 1996 mögliche Varianzschätzung gegenüber dem bisher verwendeten Näherungsverfahren zu niedrigeren Standardfehlern führt. Dieses Ergebnis hängt hauptsächlich mit der Reduzierung des Klumpeneffekts infolge der Substichprobenziehung zusammen. Die wichtigsten Voraussetzungen für die verbesserte Varianzschätzung sind mit der Weitergabe anonymisierter Informationen über die Zugehörigkeit zum Auswahlbezirk und zur Gebäudeschicht sowie einigen groben Regionalmerkmalen erfüllt. Dies eröffnet den Nutzern die Möglichkeit, eine wesentliche Eigenschaft dieses Datensatzes, nämlich die hohe Präzision der Schätzergebnisse bei zugleich sehr geringem Nonresponse, effizient auszuschöpfen. Neben Methoden für die Schätzung von Totals, Anteils- und Mittelwerten wurde eine Regressionsschätzung für Totals entwickelt. Damit können die an die Bevölkerungsfortschreibung angepassten Mikrozensusergebnisse statistisch angemessen behandelt werden. Es hat sich gezeigt, dass die Varianzschätzungen für das Scientific Use File des Mikrozensus ab 1996, die nur unter Berücksichtigung der ersten Auswahlstufe der PSU's durchgeführt werden, eine sehr gute Näherung der Gesamtvarianz erge-

ben. Für diese vereinfachte Schätzung können deshalb auch Standardprozeduren der Statistikpakete STATA und SAS verwendet werden.<sup>20</sup>

Die Analysen haben gleichzeitig die Grenzen der klassischen Stichprobentheorie deutlich gemacht, die sich ausschließlich auf den unter der Annahme einer idealen Umsetzung des Stichprobenplans berechenbaren Stichprobenfehler konzentriert und die in der Praxis auftretenden systematischen Fehler – Befragungsausfälle, Antwort- und Verkodungsfehler etc. – außer Acht lässt. Insbesondere die Differenzen der Mikrozensusergebnisse vor und nach der Anpassung an die Bevölkerungsfortschreibung weisen darauf hin, dass für Aussagen zur Qualität statistischer Daten ein großer Bedarf an systematischen Untersuchungen besteht. Vor dem Hintergrund, dass die nächste Volkszählung nicht mehr in der Form einer primärstatistischen Totalerhebung durchgeführt wird, sondern statt dessen verschiedene Datenquellen – unter anderem Ergebnisse des Mikrozensus – in einem registergestützten System zusammengeführt werden, sind Fragen der Datenqualität von zentraler Bedeutung. Hier ist aber nicht nur die amtliche Statistik gefordert.<sup>21</sup>

Damit auch von Seiten der akademischen Forschung Methodenstudien durchgeführt werden können, ist es wünschenswert, weitere feldbedingte Merkmale im Scientific Use File zur Verfügung zu stellen. In bezug auf Varianzschätzungen ist die Zugehörigkeit zu den Rotationsgruppen im Mikrozensus von besonderem Interesse. Bei Kenntnis der Rotationsgruppen kann man die Varianz der Schätzungen durch die Varianz der Schätzergebnisse über die Rotationsgruppen schätzen. Eine Identifikation der Zugehörigkeit zur Rotationsgruppe ist außerdem für den Vergleich von Kennwerten auf der Basis einzelner Querschnittsangaben (zum Beispiel Veränderungen 1996-1997) für Varianzschätzungen nötig. In diesem Zusammenhang sei angemerkt, dass die ab dem Mikrozensus 1996 mögliche Zusammenführung der einzelnen Mikrozensusserhebungen zu einem rotierenden Panel für die Forschung von allergrößtem Interesse ist. Es wäre wünschenswert, wenn auch die Scientific Use Files zu einem Rotationspanel zusammengeführt werden können.

---

**20** Eine Dokumentation der hier verwendeten SAS-, SPSS- und STATA-Programme, die auch die zweite Auswahlstufe berücksichtigen, ist als ZUMA-Methodenbericht in Vorbereitung. Die Programme werden in Kürze auf den WWW-Seiten der Abteilung Mikrodaten bei ZUMA ([www.gesis.org/dauerbeobachtung/mikrodaten](http://www.gesis.org/dauerbeobachtung/mikrodaten)) zur Verfügung gestellt.

**21** In anderen Ländern wird das Thema Datenqualität schon seit längerer Zeit verstärkt untersucht. Beispielsweise hat das statistische Amt im Vereinigten Königreich hierzu eine Methods and Quality Division eingerichtet und arbeitet auf diesem Gebiet eng mit Statistikern der Universität Southampton zusammen (Holt/Jones 1998).

Hinsichtlich der Varianzschätzung beziehen sich weitere Verbesserungsmöglichkeiten des Datensatzes auf die Bereitstellung des Kompensationsgewichts und von Informationen über die Ziehung der 70-Prozent-Substichprobe aus dem Mikrozensus. Bei Kenntnis der Zahl der Haushalte vor der Substichprobenziehung ist es möglich, die durch diese Auswahlstufe bedingte Varianzkomponente zu schätzen. Da die Merkmale Kompensationsgewicht, Rotationsgruppe und Zahl der MZ-Haushalte pro PSU in bezug auf den Datenschutz kein grundsätzliches Problem darstellen dürften, würde die Weitergabe der Variablen neue methodische und inhaltliche Analysen des Scientific Use File des Mikrozensus bieten.

## **Korrespondenzadresse**

*Prof. Dr. Ulrich Rendtel  
Johann Wolfgang Goethe Universität  
Institut für Statistik und Mathematik, Fachbereich Wirtschaftswissenschaften (02)  
Mertonstraße 17, R138B, 60054 Frankfurt am Main  
E-Mail: rendtel@em.uni-frankfurt.de*

## **Literatur**

- Emmerling, D./Riede, T., 1997: 40 Jahre Mikrozensus. *Wirtschaft und Statistik* (3): 160-174.
- Frank, E./Kafurke, A., 1990: Die Mikrozensusstichprobe ab 1990 auf neuer Auswahlgrundlage. *Baden-Württemberg in Wort und Zahl* (4): 154-164.
- Heidenreich, H.-J., 1989: Erwerbstätigkeit im April 1988. *Wirtschaft und Statistik* (6): 327-339.
- Heidenreich, H.-J., 1994: Hochrechnung des Mikrozensus ab 1990. S. 112-123 in: S. Gabler/J.H.P. Hoffmeyer-Zlotnik/D. Krebs (Hrsg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Holt, T./Jones, T., 1998: Quality work and conflicting quality objectives. Paper for the 84<sup>th</sup> DGINS conference in Stockholm 28-29 May 1998 <URL: [http://www.statistics.gov.uk/methods\\_quality/downloads/qualitywork.doc](http://www.statistics.gov.uk/methods_quality/downloads/qualitywork.doc); letzter Zugriff: 03.11.00>.
- Jäger, M., 1992: Im Westen was Neues? - Im Osten was Besseres? Möglichkeiten der Nutzung von Daten der Einwohnermelderegister für statistische Zwecke. S. 103-124 in: *Statistisches Bundesamt (Hrsg.): Volkszählung 2000 - oder was sonst?* Band 21 der Schriftenreihe Forum der Bundesstatistik. Stuttgart: Metzler-Poeschel.

Köhler, S./Schimpl-Neimanns, B./Schwarz, N., 2000: Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten. *Wirtschaft und Statistik* (1): 30-37.

Krug, W./Nourney, M./Schmidt, J., 1999: *Wirtschafts- und Sozialstatistik. Gewinnung von Daten*. München: Oldenbourg (5., völlig neubearb. Auflage).

Lüttinger, P., (Hrsg.), 1999: *Sozialstrukturanalysen mit dem Mikrozensus*. ZUMA Nachrichten Spezial, Band 6. Mannheim: ZUMA.

Lüttinger, P./Riede, T., 1997: Der Mikrozensus: amtliche Daten für die Sozialforschung. ZUMA-Nachrichten Nr. 41: 19-43.

Meyer, K., 1994: Zum Auswahlplan des Mikrozensus ab 1990. S. 106-111 in: S. Gabler/J.H.P. Hoffmeyer-Zlotnik/D. Krebs (Hrsg.), *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.

Müller, W./Blien, U./Knoche, P./Wirth, H. u.a., 1991: *Die faktische Anonymität von Mikrodaten*. Band 19 der Schriftenreihe Forum der Bundesstatistik, herausgegeben vom Statistischen Bundesamt. Stuttgart: Metzler-Poeschel.

Reinders, M., 1993: Fehlerrechnung zum Mikrozensus 1990. *Statistische Rundschau Nordrhein-Westfalen* (8): 398-404.

Rentel, U./Schimpl-Neimanns, B., 2000: *Varianzschätzungen für den faktisch anonymisierten Mikrozensus*. *Jahrbücher für Nationalökonomie und Statistik* 220(6): 759-776.

Särndal, C.-E./Swensson, B./Wretman, J., 1992: *Model Assisted Survey Sampling*. New York: Springer.

Schmidt, G., 1990: Der Auswahlplan des Mikrozensus ab 1990. *Bayern in Zahlen* (6): 217-221.

Statistisches Bundesamt, 1998a: *Fachserie 1, Bevölkerung und Erwerbstätigkeit. Reihe 4.1.1, Stand und Entwicklung der Erwerbstätigkeit 1996 (Ergebnisse des Mikrozensus)*. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt, 1998b: *Fehlerrechnung Mikrozensus 1996 (nach Kompensation der bekannten Ausfälle)*. Wiesbaden (unveröffentlichte Tabellen; StBA VIII C; September 1998).

Statistisches Bundesamt, 1999: Zum Auswahlplan des Mikrozensus ab 1990. S. E2 49-54 in: *Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957 (Loseblattsammlung)*. Wiesbaden.

Werner, J., 1994: *Regionalisierung des Mikrozensus. Baden-Württemberg in Wort und Zahl* (6): 278-285.

Wolter, K., 1985: *Introduction to Variance Estimation*. New York: Springer.

# KOMPATIBILITÄT VON ISCO-68, ISCO-88 UND KLDB-92

*ALFONS J. GEIS & JÜRGEN H.P. HOFFMEYER-ZLOTNIK*

Nachdem sich unser Beitrag in den letzten ZUMA-Nachrichten mit den Besonderheiten der Berufscodierung beschäftigt hat, soll hier die Vergleichbarkeit der verschiedenen Berufsklassifikationen behandelt werden.

Angenommen, es liegt eine Berufscodierung des Statistischen Bundesamtes vor und es sollen Prestigewerte generiert werden, dann wird eine ISCO-Codierung gebraucht, da nur diese einen Zugriff auf eine der Prestige-Skalen ermöglicht. Oder umgekehrt: Es liegt eine internationale Studie mit ISCO-Codes vor, die mit Daten der nationalen Arbeitsamtsstatistik verknüpft werden soll, dann muss man wissen, welche Kategorien sich jeweils entsprechen. Noch ein dritter Fall kommt häufig vor: Eine aktuelle Studie mit ISCO-88-Codierung soll mit Ergebnissen von 1985 verglichen werden, als noch ausschließlich nach der ISCO-68 codiert wurde. Jedesmal handelt es sich im Fragebogen um dieselbe Berufsangabe, die in den jeweiligen Systematiken zu verorten ist, und es ist schwer nachvollziehbar, warum dies nicht im Sinne einer Recodier-Routine oder mit geringem Aufwand möglich sein sollte. In der Praxis wird in der Regel neu codiert. Dieser Frage soll in diesem Beitrag nachgegangen werden. Dazu sind zunächst einige grundsätzliche Anmerkungen notwendig.

Der Sinn der Codierung nach Klassifikationen besteht darin, die Vielfalt der unterschiedlichen Aussagen systematisch zu reduzieren und in Kategorien zusammenzufassen, so dass sie übersichtlich, zählbar und beschreibbar werden. Deshalb erhält z.B. nicht jede der ca. 30.000 Berufsbenennungen, die das Statistische Bundesamt derzeit registriert hat, einen eigenen Code, da diese nicht handhabbar wären, sondern mehrere Berufe werden unter einem gemeinsamen Aspekt - z.B. Branche, Art der Tätigkeit oder Arbeitsmittel - in einer Kategorie zusammengefasst. Je nach Gesichtspunkt, Fragestellung oder Zweck der Zusammenfassung kann ein und derselbe Beruf zu jeweils einer anderen Gruppe gezählt werden. Daraus ergeben sich unterschiedliche Klassifikationen, die jeweils in sich stimmige Gliederungssysteme darstellen, ausgerichtet auf einen jeweiligen Zweck.

Die für den deutschsprachigen Raum wichtigsten Klassifikationen werden zunächst beschrieben. Dann wird untersucht, wie stabil eine bestimmte Kategorie der einen Klassifikation gemeinsam mit einer bestimmten Kategorie der anderen Klassifikation auftritt und ob es valide Umsteigeschlüssel zwischen den Klassifikationen geben kann. Welche Schlussfolgerungen sich für die sozialwissenschaftliche Praxis daraus ergeben, wird am Ende erörtert.

In der empirischen Sozialforschung in der Bundesrepublik Deutschland sind zur Klassifikation der Berufe derzeit drei unterschiedliche Klassifikationsschemata gebräuchlich. Alle drei Klassifikationsschemata ordnen die in Umfragen benannten Berufsbeschreibungen zufriedenstellend ein. Damit, so könnte man meinen, ist die Wahl des zu benutzenden Klassifikationsschemas eher Geschmacksache. Allerdings folgen alle drei Klassifikationsschemata jeweils einer eigenen Logik. Dieses verhindert eine unproblematische Übersetzung des einen in ein anderes. Hinzu kommen jeweils spezifische Vorteile und Merkmale des einen oder anderen Klassifikationsschemas.

Die drei Schemata sind:

- die „International Standard Classification of Occupations“ von 1968 (ISCO-68),
- die „International Standard Classification of Occupations“ von 1988 (ISCO-88),
- die „Klassifikation der Berufsbenennungen“ von 1992 (KldB-92).

Beide Instrumente der „International Standard Classification of Occupations“ sind international geltende Klassifikationen, die unter der Schirmherrschaft des Internationalen Arbeitsamtes in Genf (International Labour Organisation: ILO) jeweils von einer multinationalen Expertenkommission entwickelt wurden. Das dritte Instrument, die Klassifikation der Berufsbenennungen von 1992 ist ein nationales Instrument, entwickelt von der amtlichen Statistik in Deutschland für deren Belange.

## **1. Aufbau der Klassifikationen**

Die Fragestellung dieses Artikels untersucht die Vergleichbarkeit der drei genannten Klassifikationen miteinander. Hierzu ist es notwendig, sich die jeweilige Struktur und den Aufbau dieser Klassifikationen anzusehen. Weitere Klassifikationen wie ISCO-58 oder KldB-75 sollen hier nicht weiter berücksichtigt werden. Soweit weitere Klassifikationen benutzt werden, betrifft dieses kleine Forschergruppen zu sehr spezifischen Forschungsfragen.

### 1.1 „International Standard Classification of Occupations“ von 1968: ISCO-68

Der Aufbau der Klassifikation der ISCO-68 hat folgende vier Ebenen, die, aufeinander aufbauend, sukzessive eine immer feinere Untergliederung bieten:

**Tabelle 1: Klassifikationsstruktur der ISCO-68**

Ebene	Einheit	deutsche Bezeichnung	Kategorienanzahl
1st level	minor groups	Berufshauptgruppe	8
2nd level	major groups	Berufsuntergruppe	83
3rd level	unit groups	Berufsgattung	284
4th level	occupational categories	Berufsfeld	1881

(ILO 1968: 1)

Die zweite und dritte Ebene liefern zusammenfassende Einheiten, die zur Darstellung berufsstatistischer Daten z.B. bei Arbeitskräfteerhebungen oder bei Zensen geeignet sind. Die vierte Ebene bietet eine Unterteilung in Berufe oder Tätigkeitskomplexe. Es wird ein Dezimalsystem verwendet, das von einer sektoralen Gliederung nach Beschäftigtensektoren in zwei Schritten zu einer fach- oder branchenbezogenen Gliederung übergeht. Diese wird schließlich in der vierten und fünften Stelle (bzw. auf der dritten und vierten Ebene) durch eine Klassifikation der Tätigkeit präzisiert.

Die vier Schritte über die vier Ebenen werden von den Autoren der ISCO an folgendem Beispiel verdeutlicht (ILO 1968: 1):

*for example 'typist' is found in:*

*major group 3 = Clerical and related worker;*

*minor group 3-2 = Stenographers, typists and card- and tape-punching machine operators;*

*unit group 3-21 = Stenographers, typists and teletypists;*

*occupational category 3-21.40 = Typist.*

Die „major groups“ orientieren sich an den Beschäftigungssektoren (Primärer Sektor: Code 6; Sekundärer Sektor: Code 7-9; Tertiärer Sektor: Code 2-4), ohne jedoch diese klar voneinander abzugrenzen, da der Bereich der wissenschaftlichen und leitenden Tätigkeiten (Code 0-2) nicht gemäß den Sektoren klassifiziert wird.

Die „minor groups“ stellen eine grobe Zusammenfassung von „occupations“ dar. „Since the whole range of vocational activities of civilian workers, whether in an industrialised or in a developing economy, is covered by the 83 'minor groups' of ISCO, the common occupational characteristics which link together the different

'unit groups' within a 'minor group' are frequently of a broad or general nature. In addition, the 'minor groups' have been designed to reflect, in large measure, the conventional broad groupings which are typical of classifications in current use for statistical purposes, including censuses of the population.“ (ILO 1968: 4).

„A 'unit group' in ISCO is a group of 'occupations' related to each other by similarity of the characteristics of the work they entail.“ (ILO 1968: 4). Damit weisen „occupations“, die in „unit groups“ zusammengefasst sind, eine gewisse Homogenität auf. Die Berufstätigkeiten von „unit groups“ sind miteinander verwandt. Die „unit groups“ gründen auf der beruflichen Tätigkeit: Es werden vergleichbare Tätigkeiten auf verwandten Gebieten zusammengefasst. Diese Ebene der Kategorisierung ist für den praktischen Gebrauch gedacht. Die relative Grobheit der Kategorisierung trägt nach Ansicht der Autoren der ISCO einer bei Großzählungen zu erwartenden Datenqualität Rechnung: Eine sehr detaillierte Abfrage auf den Ebenen der Tätigkeitsbeschreibung, der Qualifikation und der Verantwortlichkeiten ist nicht möglich.

Mit der vierten Ebene, der Ebene der „occupations“, ist „the smallest segment of work“ in der ISCO-Klassifikation erreicht. Mit dieser Klassifikation wird ein Typ von Arbeit identifiziert, nicht die Beschreibung eines individuellen Arbeitsplatzes oder einer spezifischen Position. „The definition of an 'occupation' covers various 'jobs' or 'positions' which are held by individual workers who perform one or another of the different possible combinations of the tasks described.“ (ILO 1968: 5). Die vierte Ebene zeigt das Grundprinzip der Klassifikation am deutlichsten: Im Zentrum steht eine Klassifikation nach der Art der verrichteten Tätigkeit, um die Ähnlichkeit verrichteter Arbeiten der Art nach zu bestimmen.

## **1.2 „International Standard Classification of Occupations“ von 1988: ISCO-88**

Die ISCO von 1968 war als Instrument für unterschiedliche Zwecke angelegt: Sie war nicht allein als Instrument für die amtliche Statistik konzipiert, sondern stellte zusätzlich ein Instrument dar, das für berufskundliche Fragestellungen, für Fragestellungen zur sozialen Mobilität und für die international vergleichende Forschung einsetzbar sein sollte. Daher wurde besonderer Wert auf eine detaillierte Beschreibung der untersten Gliederungseinheit, den Fünfsteller, gelegt, deren Grundlage die Tätigkeitsbeschreibung ist. Die ISCO von 1988 ist primär für Anwendungszwecke der amtlichen Statistik konzipiert. Nicht mehr die Differenzierung steht im Vordergrund, sondern eine strukturierte Reduktion. Damit erhält bei diesem Kategoriensystem die oberste Hierarchieebene, d.h. die erste Stelle, eine zentrale strukturierende Rolle.



Die ISCO von 1988 hat als neue Bestimmungsfaktoren zur Konkretisierung der Gleichheit bzw. der Verwandtschaft von beruflichen Tätigkeiten die Merkmale „skill level“ und „skill specialisation“ eingeführt:

„Skill level“ stellt hierbei im Sinne eines Anforderungsniveaus ein berufs- und arbeitsplatzbezogenes Merkmal dar. Das „skill level“-Konzept findet nur Anwendung auf die Bestimmung der „major groups“. Es unterscheidet in Anlehnung an die „International Standard Classification of Education (ISCED)“ nach vier relativ groben Kategorien von Bildungsqualifikationen bzw. von Anforderungen. Hierbei ist allein wichtig, dass ein bestimmtes Kenntnis- und Fertigungsniveau vorhanden ist. Das Zertifikat zählt.

„Skill specialisation“ ist ein berufsfachliches Kriterium. Es hebt einerseits auf die Art der ausgeübten Tätigkeit innerhalb einer gegebenen Qualifikationsebene ab und versucht andererseits Merkmale wie verarbeitetes Material, Berufsmilieu, Art der produzierten Güter und Dienstleistungen, Besonderheiten der Fertigungsverfahren etc. mit einzubeziehen.

**Tabelle 2: Klassifikationsstruktur der ISCO-88 in Verbindung mit dem „skill level“-Konzept**

Major groups (Hauptgruppe)	Sub-major groups (Untergruppe)	Minor groups (Gattung)	Unit groups (xxx)	skill level
1: Legislators, senior officials and managers	3	8	33	--
2: Professionals	4	18	55	4th
3: Technicians and associate professionals	4	21	73	3rd
4: Clerks	2	7	23	2nd
5: Service workers and shop and market sales workers	2	9	23	2nd
6: Skilled agricultural and fishery workers	2	6	17	2nd
7: Craft and related trades workers	4	16	70	2nd
8: Plant and machine operators and assemblers	3	20	70	2nd
9: Elementary occupations	3	10	25	1st
armed forces	1	1	1	--
Totals	28	116	390	

(ILO 1990: 3)

Im Gegensatz zur ISCO-68 geht die ISCO-88 nicht auf die Ebene der beruflichen Tätigkeit („occupations“) hinunter. Auch wenn in der Regel mehr als eine berufliche Tätigkeit („occupation“) einer Berufsgruppe („unit group“) zuzuordnen sind, so erscheint die Ebene der „unit groups“ den Autoren der ISCO-88 (ILO 1990: 4) die sinnvollere und aussagekräftigere Ebene, denn: auf unterschiedlichen nationalen Ebenen weichen die „occupations“ voneinander ab, „depend on the size of the economy and the level of economic development, the level and type of technology, work organisation and historical circumstances“ (ILO 1990: 4). Die ISCO-88 ist damit eher als ein Instrument für die amtliche Statistik zu sehen.

Wegen der grundlegenden Unterschiede in der Struktur lässt sich die Ebene der „major groups“ der ISCO-88 nicht mit denen der ISCO-68 vergleichen. Kontinuität zum Zwecke des Vergleichs wurde dennoch angestrebt: Ein Vergleich sei auf der Ebene der „unit groups“ gegeben, d.h. die 284 dreistelligen „unit groups“ der ISCO-68 sind mit den 390 vierstelligen „unit groups“ der ISCO-88 vergleichbar. Vergleichbarkeit bedeutet jedoch nicht, dass eine Übersetzbarkeit von einem Code der einen Klassifikationen in einen der anderen möglich sein muss. Dieses wird allein schon durch die unterschiedliche Anzahl der „unit groups“ in den beiden Klassifikationen deutlich. Die ISCO-88 ist auf dieser Ebene differenzierter als die ISCO-68.

### **1.3 Aufbau der „Klassifikation der Berufsbenennungen“ von 1992: KIdB-92**

Auf nationaler deutscher Ebene wurde 1992 die „Klassifikation der Berufsbenennungen“ des Statistischen Bundesamtes von 1975 überarbeitet und den Veränderungen auf dem Feld der beruflichen Tätigkeiten angepasst. Aus Gründen der Vergleichbarkeit wurde jedoch die Gliederungsstruktur von 1975 im Prinzip unverändert beibehalten. Damit reicht für eine nationale deutsche Klassifikation eine Betrachtung des Instrumentes von 1992, obwohl hin und wieder, vor allem in der Arbeitsmarktstatistik, bis heute die Berufsklassifizierung von 1975 benutzt wird. Diese 75er-Klassifikation war vor allem von den berufsstatistischen Belangen der Arbeitsverwaltung geprägt. Die Revision von 1992 setzte auf der Ebene der Berufsgruppen mit einer Umstrukturierung, bedingt durch Umsetzungen von Berufsordnungen oder Teilen derselben, ein, die etwa ein Drittel aller Berufsgruppen betraf. Die Änderungen auf der Ebene der Berufsordnungen waren bedingt entweder durch eine Zusammenfassung von Berufen, die im Verlauf des gesellschaftlichen Wandels ihre Eigenständigkeit verloren haben, mit „neuen“ Berufen, oder durch eine Aufgliederung von Berufen zwecks höherer Transparenz. So gesehen ist zwar die Struktur, nicht aber das Codesystem von 1992 mit dem von 1975 identisch.

**Tabelle 3: Struktur der KldB**

Ebene	Gruppe der Klassifikation	Stelle des Codes	Anzahl der Kategorien
1.	Berufsbereiche	1 - 2	6
2.	Berufsabschnitte	1 - 2	33
3.	Berufsgruppen	1 - 2	88
4.	Berufsordnungen	1 - 3	369
5.	Berufsklassen	1 - 4	2287
6.	Berufsbenennungen	1 - 6	29527

StaBA 1992: 13

Die „Berufsbereiche“ fassen jene Berufe zusammen, die „in der Arbeitsaufgabe und in den allgemeinen Arbeitsanforderungen“ Gemeinsamkeiten aufweisen. Unterschieden wird in „Urproduktions-, Fertigungs- und Dienstleistungsberufe“ (StaBA 1992: 13).

Die „Berufsabschnitte stellen Zusammenfassungen von bis zu 6 Berufsgruppen dar“. Sie umfassen Bereiche, die sich hinsichtlich der Tätigkeit oder des verarbeiteten Materials ähneln (StaBA 1992: 13). „Berufsgruppen“ fassen Berufe nach Aufgaben und Tätigkeitsmerkmalen zusammen, wobei das verarbeitete Material als Leitprinzip dient. Die „Berufsordnungen“ werden als das statistische Basissystem angesehen. Auf dieser Ebene werden nach Aufgabe und Tätigkeit gleichartige Berufe zusammengefasst, die einen abgrenzbaren, eigenständigen Charakter haben. Lücken in der Nummerierung markieren Einheiten aus der 1975-er Klassifikation, die in der Revision aufgehoben wurden. „Berufsklassen“ sind Berufe oder Berufsarten, die einen gemeinsamen Tätigkeitstyp aufweisen. Die Mehrzahl der Berufsklassen weist nur einen Beruf oder Spezialisierungsformen in einzelnen Berufen auf. Als 6. Ebene und unterster Baustein der Klassifikation sind „Berufsbenennungen“ möglich. Die Liste der Berufsbenennungen wird jährlich aufgrund der Meldungen der Landesämter im Statistischen Bundesamt ergänzt.

Als zentrales Abgrenzungsmerkmal für die Berufe dient die „ausgeübte Tätigkeit“, die durch Titel und Bezeichnung definiert wird (StaBA 1992: 13). Erforderliche Bildung und Ausbildung für die Berufsausübung sowie die Stellung im Beruf oder die Stellung im Betrieb finden bei der Abgrenzung der Klassifikationsmerkmale keine Berücksichtigung (StaBA 1992: 16).

Das Erfassen des Berufes für die Klassifikation geschieht über die Berufsbezeichnung, nicht wie bei den ISCO-Klassifikationen über eine Beschreibung. Die KldB

sieht hinter der Bezeichnung die Definition der Tätigkeit. Wichtiges Strukturelement ist der Wirtschaftszweig.

## **2. Möglichkeiten für eine Vergleichbarkeit**

Ersetzbar eines durch das andere sind die drei Klassifikationen nicht, da, wie im Folgenden zu zeigen sein wird, keine eins-zu-eins-Übertragung von Codes von einer Klassifikation in die andere existiert. Zwar haben ISCO-68 und ISCO-88 eine gemeinsame Logik, dennoch gibt es Probleme bei der direkten Übertragung von der einen Klassifikation in die andere. Während die ISCO-88 z.B. über das „skill level“-Konzept Tätigkeiten nach der Voraussetzung der Ausbildung in wissenschaftlich und nicht-wissenschaftlich unterscheidet, womit sich schon eine Abweichung auf der ersten Stelle des Codes ergibt, fassen die unit groups (dreistelliger Code) der ISCO-68 Berufsbilder nach den Tätigkeitsfeldern zusammen. Damit gibt es nicht nur ein Problem bei der automatischen Zuordnung der 284 unit group-Kategorien der ISCO-68 zu den 390 Berufsgattungen (vierstelliger Code) der ISCO-88, da eine Zuordnung Zusatzinformationen erfordert (z.B. zur Ausbildung), sondern auch die Zuordnung von ISCO-88 zu ISCO-68 bringt Schwierigkeiten mit sich, da auch hier nicht allein eine Verdichtung von 390 auf 284 Kategorien gilt, sondern die Definition von Tätigkeitsfeldern den Ausschlag gibt.

Im Gegensatz zu den ISCO-Klassifikationen folgt die KldB einer anderen Logik: Hier geschieht die Differenzierung vom Speziellen zum Generellen, indem zunächst die Berufsbezeichnungen erfasst und diese dann zu Gruppen zusammengefasst werden. Damit kann ein Code der KldB mehreren Codes in der ISCO-68 und der ISCO-88 zugeordnet werden.

## **3. Empirischer Vergleich**

Die bisherigen Überlegungen gehen von einer theoretisch konzipierten Systematik der Klassifikationen aus. Im Folgenden steht die Anwendung in empirischen Untersuchungen im Vordergrund.

### **3.1 Datenquelle**

Für folgende Betrachtung liegen fünf Repräsentativbefragungen aus den Jahren 1994 bis 1999 vor, in denen die offenen Berufsangaben coderbasiert nach allen drei Klassifikationen codiert wurden, nach der Klassifikation der Berufsbenennungen 1992

(KldB-92), der International Standard Classification of Occupations 1968 (ISCO-68) und der International Standard Classification of Occupations 1988 (ISCO-88):

**Tabelle 4: Zusammensetzung der Daten**

Projekt	Zeit	Codes
A - 1	1994	4672
B - 2	1997	3559
C - 5	1997	1386
D - 6	1997	340
E - 7	1999	2836
Summe		12793

Insgesamt liegen also dreimal 12.793 Codierungen vor. Eliminiert wurden alle missing-Fälle, die keine inhaltliche Codierung darstellen, z.B. „trifft nicht zu“ (Code 0000) und „keine Angabe“ (Code 0009).

**Tabelle 5: Aufbau des Datensatzes**

Projekt-Kennung

	ISCO-68	ISCO-88	KldB-92
2	211	1210	0110
2	611	6130	0110
2	611	6130	0110
2	611	6130	0110
2	611	6130	0110
6	611	6130	0110
2	611	6130	0111

### 3.2 Vorgehensweise

In der Gegenüberstellung der verschiedenen Klassifikationen wird geprüft, wie zweifelsfrei ein Code der einen Klassifikation eindeutig und ohne Ausnahme einem bestimmten Code der zweiten Klassifikation entspricht. Erster wird Referenz-Code

genannt (Ausgangscode, auf den Bezug genommen wird), zweiter wird Inferenz-Code genannt (Zielcode, auf den geschlossen werden soll).

Die Code-Zuordnungen von jeder Klassifikation zu jeder werden systematisch verglichen, um so ein Maß für die Kompatibilität zwischen den unterschiedlichen Klassifikationssystemen zu finden und damit einen validierten Umsteigeschlüssel, der für die coderbasierte und computerunterstützte Berufsvercodung verwendet werden kann.

Eine vollständige Kompatibilität ist dann gegeben, wenn eine Kategorie der einen Klassifikation eindeutig einer einzigen bestimmten Kategorie der anderen Klassifikation entspricht. Der Vergleich, der feststellt, wie eindeutig eine Kategorie innerhalb des Kategoriensystems „übersetzt“ werden kann, unabhängig von Häufigkeit ihrer Verwendung, wird in den Ergebnistabellen dichotom genannt.

Wenn auch die Häufigkeit berücksichtigt wird, handelt es sich um den empirischen Vergleich, der die Bedeutung einer Kategorie aufgrund ihres wiederholten Auftretens in der Untersuchung gewichtet. Für einen Forscher ist nicht nur wichtig, dass für eine bestimmte Kategorie eine fehlerhafte Umsetzung möglich ist, sondern vor allem auch, wie umfangreich der potentielle Fehler ist. Ein fehlerhafter Umstieg von einem zum anderen System, der nur einmal bei über 10.000 Codierungen vorkommt, kann vernachlässigt werden.

Je mehr eine Kategorie nicht ausschließlich einer einzigen Kategorie der anderen Klassifikation zugeordnet werden kann und je häufiger darüber hinaus diese Kategorie in der Erhebung verwendet wird, um so weniger sind die Klassifikationen kompatibel.

Alle Paarzuordnungen von jeder zu jeder Klassifikation wurden überprüft, insgesamt also sechsmal 12.793, das sind 76.758 Vergleiche. Dazu wurde ein kleines BASIC-Programm geschrieben, das folgende Daten ermittelte:

- Anzahl der Paare,
- Anzahl/Auflistung der unterschiedlichen Referenz-Codes und ihrer Häufigkeit,
- Anzahl/Auflistung der unterschiedlichen Paarbildungen jedes Referenz-Codes und ihrer Häufigkeit.

Die Datenausgabe sah folgendermaßen aus:

**Tabelle 6: Ausgabe-Daten**

Entsprechungen von KldB-92 mit ISCO-88 (Test)					
R-Code	Max-Kombi	I-Code	Kombi-Freq	N	VON RC
0110		1210	1		
0110		1221	2		
0110		1311	1		
0110		6100	1		
0110		6121	1		
0110	6	6130	22		28
0111	1	6130	1		1
.					
9971		0004	37		
9971	2	2000	2		39

Anzahl der verarbeiteten Zeilen / Code-Paare: 12504  
 Anzahl der unterschiedlichen Paare: 1488  
 Anzahl der unterschiedlichen Referenz-Codes: 927  
 Durchschnittliche Paare je Referenz-Code: 1.6

### 3.3 Auswertung

Für jeden der sechs Vergleiche wurden in einer Tabelle folgende Ergebnisse zusammengetragen, jeweils für die dichotome (jede Kategorie zählt nur einmal, unabhängig vom Umfang des tatsächlichen Auftretens) und die empirische Betrachtungsweise (Berücksichtigung der Häufigkeit der Verwendung) sowie mit unterschiedlicher Prozentuierung/Gesamtheiten (die Ziffern in Klammern sind Beispielwerte aus der Tabelle 4, Vergleich ISCO-68 mit ISCO-88, und zwar dichotom und empirisch):

- Referenz-Codes, die nur mit einem einzigen Code des anderen Kategorienschemas (Referenzcode) ein Paar bilden (101/2418), unterschieden nach einmaligem Auftreten (22/22) und mehrfachem (79/2396).
- Referenz-Codes, die mit mehr als einem Inferenz-Code gemeinsam auftreten (116/10.375).
- Anzahl der vorkommenden unterschiedlichen Referenz-Codes (217/12.793).
- Anzahl der theoretisch möglichen Referenz-Codes, Anzahl der Kategorien der Klassifikation (291/-).

Der Wert mit der größten Aussagekraft, der Anteil aller verwendeten Referenz-Codes, die mehr als einmal ausschließlich immer mit demselben Inferenz-Code zusammen auftreten, ist fett dargestellt (18,7).

**Tabelle 7: Vergleich ISCO-68 und ISCO-88**

i68-188		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	22	21,8	10,1	7,6	22	0,9	0,2
	mehrfach	79	78,2	36,4	27,1	2396	99,1	18,7
	insgesamt	101	100,0	46,5	34,7	2418	100,0	18,9
mit 2 und mehr Inferenz-Codes		116		53,5	39,9	10375		81,1
Referenzcodes		217		100,0	74,6	12793		100,0
ISCO-68 - Codes		291			100,0			

**Tabelle 8: Vergleich ISCO-68 und KIdB-92**

i68-KIdB-92		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	22	55,0	10,1	7,6	22	11,3	0,2
	mehrfach	18	45,0	8,3	6,2	172	88,7	1,3
	insgesamt	40	100,0	18,4	13,7	194	100,0	1,5
mit 2 und mehr Inferenz-Codes		177		81,6	60,8	12599		98,5
Referenzcodes		217		100,0	74,6	12793		100,0
ISCO-68 - Codes		291			100,0			



**Tabelle 9: Vergleich ISCO-88 und ISCO-68**

i88-i68		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	43	21,5	13,3	10,3	43	1,1	0,3
	mehrfach	157	78,5	48,6	37,6	3718	98,9	29,1
	insgesamt	200	100,0	61,9	48,0	3761	100,0	29,4
mit 2 und mehr Inferenz-Codes		123		38,1	29,5	9032		70,6
Referenzcodes		323		100,0	77,5	12793		100,0
ISCO-88 - Codes		417			100,0			

**Tabelle 10: Vergleich ISCO-88 und KIdB-92**

i88-KIdB-92		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	43	54,4	13,3	10,3	43	12,3	0,3
	mehrfach	36	45,6	11,1	8,6	308	87,7	2,4
	insgesamt	79	100,0	24,5	18,9	351	100,0	2,7
mit 2 und mehr Inferenz-Codes		244		75,5	58,5	12442		97,3
Referenzcodes		323		100,0	77,5	12793		100,0
ISCO-88 - Codes		417			100,0			

**Tabelle 11: Vergleich KldB-92 und ISCO-68**

KldB-92-i68		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	235	35,5	25,4	10,3	235	3,5	1,8
	mehrfach	427	64,5	46,1	18,7	6494	96,5	50,8
	insgesamt	662	100,0	71,4	28,9	6729	100,0	52,6
mit 2 und mehr Inferenz-Codes		265		28,6	11,6	6064		47,4
Referenzcodes		927		100,0	40,5	12793		100,0
KldB-92 - Codes		2287			100,0			

**Tabelle 12: Vergleich KldB-92 und ISCO-88**

KldB-92-i88		dichotom				empirisch		
		abs.	%	%	%	abs.	%	%
	Vorkommen							
Referenz-Codes mit nur einem Inferenz-Code	einmal	235	37,5	25,4	10,3	235	3,8	1,8
	mehrfach	391	62,5	42,2	17,1	5887	96,2	46,0
	insgesamt	626	100,0	67,5	27,4	6122	100,0	47,9
mit 2 und mehr Inferenz-Codes		301		32,5	13,2	6671		52,1
Referenzcodes		927		100,0	40,5	12793		100,0
KldB-92 - Codes		2287			100,0			

Die Kennwerte aus den sechs Vergleichen sind in folgender Tabelle zusammengefasst:

**Tabelle 13: Vergleichende Kennwerte**

	s92-i68	s92-i88	i88-i68	i68-i88	i88-s92	i68-s92
2 u.mehr eindeutige Zuordnungen %	50,8	46,0	29,1	18,7	2,4	1,3
Anzahl unterschiedlicher Paare	1396	1503	556	556	1503	1396
maximale Paarbildung	8	14	10	24	29	38
durchschnittliche Paare insgesamt	9,2	8,5	23,0	23,0	8,5	9,2
durchschnittliche Paare je R-Code	1,5	1,6	1,7	2,6	4,7	6,4
Varianz der Max-Paare	0,80	0,96	0,77	1,61	3,71	5,61

Die durchschnittliche Anzahl von Kategorien, denen der Referenz-Code empirisch zugeordnet wurde, ist eine Maßzahl für die systematische Kompatibilität.

Der Anteil der Codierungen, der mehr als einmal ausschließlich zugeordnet wird, ist eine Maßzahl für die Kompatibilität in der konkret vorliegenden Untersuchung. Dass ein Referenz-Code ausschließlich einem einzigen Inferenz-Code entspricht, aber nur ein einziges Mal vorkommt, betrifft zwar einen zu vernachlässigenden kleinen Anteil, kann aber dennoch nicht als Beleg gelten, dass der Referenz-Code in einer weiteren Erhebung nicht doch zusammen mit einem anderen Inferenz-Code auftritt.

Es ist selbstverständlich, dass ein sehr ausdifferenziertes Kategorienschema wie die KldB-92 mit über 2.000 Kategorien leichter in eine gröbere Klassifikation mit weniger als 300 Kategorien, wie die ISCO-68, zu überführen ist als umgekehrt. Dieses Gefälle zeigt sich deutlich an den Maßzahlen.

Wenn bei annähernd gleich differenzierten Klassifikationen wie den beiden ISCO mit 300 bis 400 Codes dennoch in nicht einmal 30% der Paare eine ausschließliche Zuordnung möglich ist, liegt es an der unterschiedlichen Systematik der Klassifikationen.

Die Anzahl der möglichen Paarbildungen mit demselben Referenz-Code zeigt das Ausmaß der Streuung und damit der Fehlerwahrscheinlichkeit.

Zu berücksichtigen ist, dass nicht alle Kategorien der Klassifikationen überprüft wurden, sondern nur die in den untersuchten Studien benutzten. Allerdings ist zu erwarten, dass Kategorien, die in fast 13.000 Nennungen aus verschiedenen Repräsentativ-Studien unterschiedlicher Jahre immer noch nicht zur Anwendung gekommen sind, auch sonst wenig vertreten sind, so dass sie das Gesamtbild kaum verändern werden.

**Tabelle 14: Umfang verwendeter Kategorien**

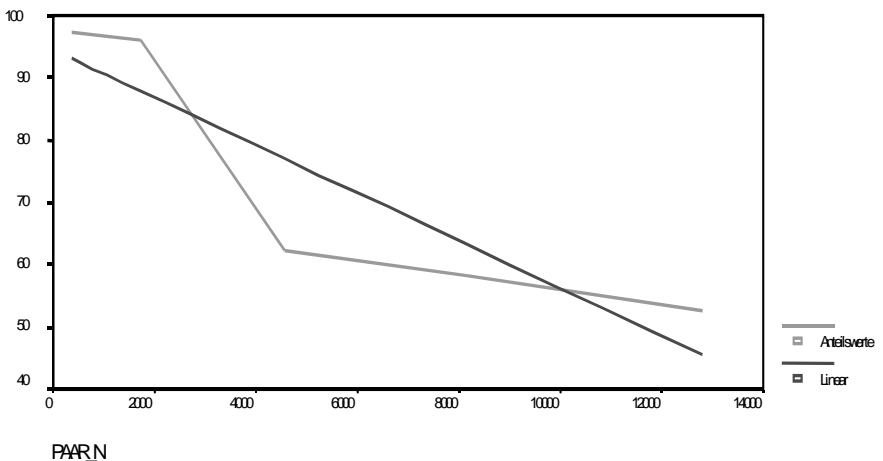
	100%	absolut	Anteil
<b>ISCO-68</b>	291	217	74,6
<b>ISCO-88</b>	417	323	77,5
<b>KldB-92</b>	2287	927	40,5

Die ILO-Dokumentation (1990) enthält einen Umsteigeschlüssel (Index of occupational titles according to ISCO-88 numerical order), in dem 5.114 Berufsbezeichnungen mit Codes nach ISCO-88 dem entsprechenden Code nach der ISCO-68 gegenübergestellt werden, sortiert nach der ISCO-88-Systematik. Von den 384 ISCO-88-Kategorien sind 217 (56,5%) nur einer einzigen ISCO-68-Kategorie (dreistellig) zugeordnet, 167 (43,5%) ISCO-88-Kategorien können in der ISCO-68 mit zwei und mehr verschiedenen Codes wiedergegeben werden. Zum Vergleich: Im verwendeten Datensatz wurden nur 324 ISCO-88-Kategorien tatsächlich verwendet, von denen bei dichotomer Betrachtung 200 (61,9%) ausschließlich und 123 (38,1%) mehrfach mit ISCO-68-Kategorien kombiniert auftreten.

Der Anteil der ausschließlichen Codezuordnungen mit mehr als einer Nennung nimmt mit zunehmender Fallzahl ab und dürfte sich bei einem Wert etwas über 40 Prozent einpendeln, wie die kumulative Auswertung (Tab. 15) zeigt.

**Tabelle 15: Kumulative Kennwerte**

	KUM1	KUM2	KUM3	KUM4	KUM5
Projekt-ID	6	6, 5	6, 5, 7	6, 5, 7, 2	6, 5, 7, 2, 1
Anzahl der Paare	340	1726	4562	8121	12793
Anzahl der R-Codes	179	287	567	745	927
eindeutig 1 x (abs.)	111	123	195	194	235
eindeutig 1 x %	32,6	7,1	4,3	2,4	1,8
2 u.mehr eindeutige Zuordnungen (abs.)	220	1535	2650	4549	6494
2 u.mehr eindeutige Zuordnungen %	64,7	88,9	58,1	56,0	50,8
Summe eindeutiger Paare %	97,4	96,1	62,4	58,4	52,6
Anzahl unterschiedlicher Paare	183	301	803	1084	1396
maximale Paarbildung	2	4	7	8	8
durchschnittliche Paare insgesamt	1,9	5,7	5,7	7,5	9,2
durchschnittliche Paare je R-Code	1,0	1,0	1,4	1,5	1,5

**Tabelle 16: Anteil der ausschließlichen Paarbildungen und Fallzahl**

### 3.4 Automatische Umsetzung von KldB-92 auf ISCO-88

Es liegt ein Umsteigeschlüssel des Statistischen Bundesamtes vor, der jedem Code der KldB-92 einen Code nach der ISCO-88 (COM) gegenüberstellt, manchmal mit Einschränkungen/Spezifizierungen bezüglich der Betriebsgröße und Selbständigkeit. Die Einträge mit Einschränkungen (2.234) wurden nicht ins Wörterbuch übernommen. Die 12.793 Fälle der Kumulativ-Datei wurden für die automatische Umsetzung aufbereitet und mit TEXTPACK und dem Umsteigeschlüssel als Wörterbuch codiert:

**Tabelle 17: Ergebnis der automatischen Umsetzung**

12.260	95,8%	Cui-Fälle: automatische Umsetzungen
533	4,2%	Lft-Fälle: es liegt kein passender ISCO-88-Code vor
12.793	100,0%	Summe

Das Ergebnis der automatischen Umsetzung wurde mit der Hand-Codierung verglichen: 9.327 (76,1%) aller Umsetzungen stimmen mit den ZUMA-Codes überein, 2.933 (23,9%) stimmen nicht überein.

Aufgegliedert nach Berufshauptgruppe, -gruppe und -untergruppe, ergibt sich für die Nicht-Übereinstimmungen folgende Verteilung:

**Tabelle 18: Aufgliederung der Nicht-Übereinstimmungen**

1.624	55,4	stimmen auf der 1. Stelle (Berufshauptgruppe) überein und weichen nur auf der Stelle 2 - 4 ab
1.162	39,6	stimmen auf der 1. und 2. Stelle (Berufsgruppe) überein und weichen nur auf der Stelle 3 - 4 ab
746	25,4	stimmen auf der 1. - 3. Stelle (Berufsuntergruppe) überein und weichen nur auf der Stelle 4 ab
2.933	100,0	stimmen auf der 1. - 4. Stelle (Berufsgruppe) nicht überein

Würde man die 4. Stelle des Codes, deren Zuordnung oft nicht zweifelsfrei möglich ist, unberücksichtigt lassen, so würden 10.073 (82,2%) der Umsetzungen übereinstimmen. In dem Vergleich der Code-Paare KldB-92 zu ISCO-88 gibt es 6.671 nicht ausschließliche Code-Zuordnungen; von diesen sind 2.312 (34,7%) nur bezüglich der 4. Code-Stelle der ISCO-88 unterschiedlich, während zwei Drittel bis zur Berufsuntergruppe übereinstimmen und nur in der Berufsgattung differieren.

Bereits einen ersten Hinweis auf die Art der Nichtübereinstimmung zeigen die entsprechenden ZUMA-Codes. So wurde 50 Mal der ZUMA-Code 0004 (das heißt: nicht codierbar), vom Diktionär aber ein vollwertiger ISCO-Code vergeben, d.h. die automatische Umsetzung erfolgt auch in den Fällen, wo die gegebenen Informationen zwar ausreichen, um einen KldB-Code zu vergeben, eine ISCO-Klassifizierung aber zusätzliche Angaben braucht, um wenigstens die Berufshauptgruppe (1. Stelle des Codes) bestimmen zu können. Ähnlich war für die manuelle Bearbeitung in 701 Fällen die Angabe zu allgemein und reichte nur bis zur Codierung der Berufsgruppe (in 512 Fällen sogar nur der Berufshauptgruppe), während die automatische Umsetzung offensichtlich eine unzulässige Spezifizierung unterstellte. Weitere Aufgliederungen zeigt die Tabelle 19.

**Tabelle 19: Nichtübereinstimmungen bestimmter manueller Codes**

Code	Inhalt	Anzahl
0004	Nicht codierbar, weil notwendige Informationen fehlen (ZUMA-Code)	50
2500	Erzieher ohne nähere Angaben (ZUMA-Code)	10
4300	Bürohilfskräfte	4
7510	Schlosser ohne nähere Angaben	8
7520	Elektroniker ohne nähere Angaben	1
7900	Meister, Vormänner in der Produktion	10
x000	Nur bis zur Berufshauptgruppe bestimmbar	560
xx00	Nur bis zur Berufsgruppe bestimmbar (inkl. 2500, 4300 und 7900)	749
	Summe der unzureichend bestimmbar Angaben ohne Überschneidungen	769
343x	Verwaltungsfachkräfte	179
344x	Zoll-, Steuer- und verwandte Fachkräfte der öffentlichen Verwaltung	94
	davon umgesetzt in 343x oder 344x	112

Weiter gibt es 112 Fälle mit den Codes 343x und 344x, in denen sich die Unterschiede innerhalb der beiden Untergruppen 343 und 344 befinden. Dabei handelt es sich um Kategorien, die in der Klassifikation nicht trennscharf definiert sind: Verwaltungsberufe, für die nicht immer klar ist, ob sie für den privaten oder öffentlichen Bereich gelten, und von denen einige mit „verwandt“ definiert werden, ohne dass ersichtlich ist, was die Verwandtschaft der Berufe ausmacht. Es sind also Kategorien, für die von vornherein ein Entscheidungsspielraum zwischen den beiden Berufsuntergruppen bleibt, so dass nicht zwischen „richtig“ und „falsch“ unterschieden werden kann.

Ließe man die 769 unzureichend bestimmbaren Angaben außer acht, verminderte sich der Anteil der automatischen Umsetzungen auf 89,8 % (10,2% Leftover-Fälle), die Übereinstimmung stieg aber auf 81,2% (9.327 von 11.491 Angaben). Unter Einbeziehung der 112 arbiträren Fälle ergibt sich eine Übereinstimmung von letztlich 82,1% (9.439 von 11.491).

Um die Abweichungen genauer beschreiben zu können, wurden die KldB-92-Codes in verschiedene Klassen der Eindeutigkeit/Ausschließlichkeit bezüglich der Zuordnung zu ISCO-88-Codes eingeteilt. Mit dieser Datei als „Wörterbuch“ konnten alle KldB-Codes mit Hilfe von TEXTPACK automatisch klassifiziert (codiert) werden. Für alle 12.793 Fälle und die 2.933 abweichenden ergibt sich folgende Verteilung:

**Tabelle 20: Eindeutigkeit der Referenz-Codes (KldB-92)**

Code	Beschreibung	alle Fälle		abweichende	
		N	%	N	%
1	Es gibt nur eine Kombination, die dreimal und mehrmals vorkommt.	5.695	44,5	353	12,0
2	Es gibt nur eine Kombination, die zweimal vorkommt.	192	1,5	44	1,5
3	Es gibt nur eine Kombination, die einmal vorkommt.	235	1,8	76	2,6
4	Es gibt 2 Kombinationen, die durchschnittlich dreimal und mehrmals vorkommen.	2.298	18,0	593	20,2
5	Es gibt 2 Kombinationen, die durchschnittlich ein- bis zweimal vorkommen.	191	1,5	116	4,0
6	Es gibt 3 Kombinationen, die durchschnittlich dreimal und mehrmals vorkommen.	2.134	16,7	512	17,5
7	Es gibt 3 Kombinationen, die durchschnittlich ein- bis zweimal vorkommen.	104	0,8	76	2,6
8	Es gibt 4 bis 7 Kombinationen.	1.524	11,9	833	28,4
9	Es gibt 7 und mehr Kombination (maximal 14).	420	3,3	330	11,2
	Summe	12.793	100,0	2.933	100,0

86,5% (2.536) der abweichenden Fälle gehören zu den Kategorien, die keine eindeutige Paarzuweisung von der KldB-92 zur ISCO-88 kennzeichnen, weil sie entweder mit unterschiedlichen ISCO-Codes oder nur ein einziges Mal auftreten.



## **4. Schlussfolgerungen für die empirische Praxis**

Alle drei Klassifikationen folgen einem jeweils unterschiedlichem Konzept. Damit ist ein Umsteigen von einer auf die andere Klassifikation weder theoretisch noch praktisch ohne Probleme möglich. Es gibt keine eins-zu-eins Zuweisungen und damit keinen Umstieg ohne zumindest teilweise neue Zuordnungsleistungen. Allerdings sollte dieser Befund nicht ganz entmutigen. Für den Umstieg von ISCO-68 zu ISCO-88 liegt ein Schlüssel des ILO vor, der für einen Teil der Codes eine automatische eins-zu-eins Zuweisung der Kategorien des einen Kategoriensystems zu den Kategorien des zweiten Kategoriensystems vornimmt. Der nicht unbeträchtliche Rest von Zuweisungen erfordert die kognitive Leistung einer Neuvercodung.

Der Umstieg von der KldB-92 auf die ISCO-88 ist mit einem befriedigenderem Ergebnis automatisch möglich. Da das Kategoriensystem der KldB-92 eine so umfangreiche Anzahl von Codes aufweist, ist hier eine Zuordnung zu dem kompakten Kategoriensystem der ISCO-88 mit relativ hoher Treffsicherheit möglich. Zumindest lässt sich der Aufwand einer neuen Codierung nach ISCO-88 erheblich verringern, wenn eine KldB-Codierung vorliegt. Allerdings ist auch hier eine zusätzliche Codierleistung notwendig, um die nicht eins-zu-eins zuweisbaren Codes nachzuvercoden. Wenn die automatische Codierung nach der KldB-92 in einem deutlich besseren Maße als nach der ISCO-88 möglich sein sollte, so würde sich die Frage stellen, ob sich eine Berufsvercodung über den Umweg einer Vercodung nach KldB-92 anbietet. Das ist dann der Fall, wenn die Nachvercodung für die automatische KldB-Codierung zusammen mit der Nachbearbeitung der automatischen Umsetzung auf die ISCO-88-Codes deutlich weniger aufwendig ist als die bisherige ausschließliche ISCO-Codierung (automatische und manuelle Bearbeitung der umfangreichen Leftover-Fälle).

Diese Überlegungen werden Gegenstand des vierten und letzten Beitrags in dieser Reihe zur Berufsvercodung sein. Der dritte Beitrag widmet sich der Nutzung der Berufsvercodung für die Ermittlung von sozialem Prestige und/oder sozialem Status.

## Literatur

Geis, A.: Vercodung offener Berufsangaben. ZUMA-Technischer Bericht Nr. T 85/11. ZUMA, Mannheim 1985.

International Labour Office (ed.): International Standard Classification of Occupations. Revised Edition 1968. Genf 1968.

International Labour Office (ed.): International Standard Classification of Occupations: ISCO-88. Genf 1990.

Statistisches Bundesamt (Hg.): Internationale Standardklassifikation der Berufe. Übersetzung der „International Standard Classification of Occupations“ des Internationalen Arbeitsamtes, Genf 1968, deutsche Ausgabe 1968. Verlag W. Kohlhammer, Stuttgart und Mainz 1971.

Statistisches Bundesamt (Hg.): Gegenüberstellung der Kennziffern der Klassifikation der Berufe 1992 mit der Internationalen Standardklassifikation der Berufe, Fassung für EU-Zwecke (ISCO-88 COM). Wiesbaden 1994.

Statistisches Bundesamt (Hg.): Klassifizierung der Berufe, systematisches und alphabetisches Verzeichnis der Berufsbenennungen, Ausgabe 1992. Verlag Metzler Poeschel, Stuttgart 1992.

# **DIE ÜBERTRAGUNG DER MAGNITUDE- PRESTIGESKALA VON WEGENER AUF DIE KLASSIFIZIERUNG DER BERUFE**

*RAINER FRIETSCH & HEIKE WIRTH*

**E**in globaler Indikator für die Bestimmung der sozialen Lage von Personen ist das Berufsprestige. In Deutschland wird hierfür häufig die von Bernd Wegener entwickelte Magnitude-Prestigeskala (MPS) genutzt. Eine zentrale Voraussetzung für die Anwendung der MPS ist allerdings die Verschlüsselung von Berufen auf Basis der International Standard Classification of Occupations 1968 (ISCO-68). Während diese Bedingung in den großen, regelmäßig durchgeführten sozialwissenschaftlichen Erhebungen in der Regel gegeben ist, erfolgt die Berufsverkodung in der amtlichen Statistik nach der deutschen 'Klassifizierung der Berufe' (KldB). Ein 'Umsteigen' von der KldB auf ISCO und damit die Umsetzung der Magnitude-Prestigeskala für amtliche Mikrodaten, wie z.B. dem Mikrozensus, ist aufgrund der unterschiedlichen Ausrichtung dieser beiden Klassifikationssysteme (international versus national) nur auf einer sehr groben Basis möglich und mit entsprechenden Unschärfen verbunden. Eine Alternative hierzu, die Gegenstand dieser Arbeit ist, besteht darin, die MPS in Anlehnung an die Vorgehensweise von Wegener auf Basis der 'Klassifizierung der Berufe' zu rekonstruieren.

## **1. Einleitung**

In der Sozialstrukturanalyse haben Merkmale zur Erwerbstätigkeit eine zentrale Bedeutung, da die gesellschaftliche Verortung und Lebenschancen von Menschen in nicht unbeträchtlicher Weise durch ihre Position am Arbeitsmarkt beeinflusst werden (Mayer 1979; Müller/Noll 1997). Dementsprechend enthalten die meisten sozialwissenschaftlichen Erhebungen Informationen zur Erwerbstätigkeit der Befragten. In der Regel werden die Erwerbsangaben dabei anwendungsneutral erhoben, das heißt es werden sozio-ökonomische Grundinformationen (z. B. ausgeübte Tätigkeit, Stellung im Beruf) erfasst und in standardisierter Weise verschlüsselt, ohne

dass hiermit schon konkrete Hypothesen oder Fragestellungen verknüpft wären. Diese Vorgehensweise hat den Vorteil, dass die Datennutzer - insbesondere bei Sekundäranalysen - nicht auf die Verwendung eines spezifischen theoretischen Konstrukts festgelegt sind, sondern in Abhängigkeit des Forschungsinteresses unterschiedliche Konstrukte (Klassenlage, Schicht, sozio-ökonomischer Status, Berufsprestige) zur Verortung der erwerbstätigen Bevölkerung verwenden können. Allerdings ist die Konstruktion von soziologisch gehaltvollen Typologien oder Skalen für Berufs- bzw. Tätigkeitsmerkmale aufwendig und trägt häufig schon den Charakter eines eigenen Projektes. Eine Möglichkeit diesen Aufwand zu umgehen, die sich insbesondere dann anbietet, wenn Berufsattribute und Berufspositionen nicht der zentrale Untersuchungsgegenstand einer Studie sind und/oder Vergleichbarkeit von Forschungsergebnissen angestrebt wird, besteht in der Verwendung vorhandener, bereits validierter Instrumente. Standardmäßig werden daher in den großen, regelmäßig durchgeführten sozialwissenschaftlichen Erhebungen wie dem ALLBUS, dem Wohlfahrtssurvey oder dem SOEP neben den erwerbsstatistischen Basisinformationen zum Teil auch Umsetzungen weit verbreiteter Prestige- und Klassenskalen zur Verfügung gestellt, wie etwa die Internationale Berufsprestigeskala von Treiman, die Magnitude-Prestigeskala von Wegener (MPS) oder das Klassenschema von Erikson, Goldthorpe und Portocarero (EGP-Skala).

Ein anderes Bild zeigt sich für die Scientific Use Files<sup>1</sup> der amtlichen Statistik. So enthält insbesondere der Mikrozensus eine Vielzahl von erwerbsstatistischen Grundinformationen und ist insofern eine zentrale Datenquelle für die Sozialstrukturforschung (Lüttinger/Riede 1997, Lüttinger 1999). Die Übertragung gängiger sozialwissenschaftlicher Typologien und Skalen auf diese Datenbasis steht jedoch noch weitgehend aus.<sup>2</sup> Zum Teil resultiert diese Lücke vermutlich daraus, dass die Anwendung von Skalen auf amtliche Mikrodaten vor allem da auf Probleme stößt, wo der Indikator 'Beruf' von Relevanz ist. Denn während die Berufsverkodung und somit auch die Bildung von Skalen in sozialwissenschaftlichen Erhebungen im Allgemeinen auf der Internationalen Standardklassifikation der Berufe (ISCO) beruht, erfolgt die Verkodung in der amtlichen Statistik auf Basis der deutschen Klassifizierung der Berufe (KldB). Auch wenn beiden Klassifikationssystemen ähnliche Strukturierungsprinzipien

---

1 Bei Scientific Use Files handelt es sich um faktisch anonymisierte Mikrodaten der amtlichen Statistik (z.B. Mikrozensus, Zeitbudget, Einkommens- und Verbrauchsstichprobe, Europäisches Haushaltspanel), die für die Forschung nach §16 Abs.6, BStatG 1987 zugänglich sind.

2 Eine Ausnahme hiervon bildet das EGP-Klassenschema, dessen Operationalisierung für den Mikrozensus kürzlich von Brauns/Steinmann/Haun (2000) vorgestellt wurde.

zugrunde liegen, nämlich die Ordnung nach Tätigkeitsmerkmalen, ist eine Eins-zu-Eins-Übertragung der ISCO auf die KldB nicht möglich (Geis/Hoffmeyer-Zlotnik 2000).<sup>3</sup> Dementsprechend lassen sich anhand von ISCO operationalisierte Typologien oder Skalen nicht direkt auf die KldB übertragen, sondern müssen neu konstruiert oder rekonstruiert werden.

In der vorliegenden Arbeit wird eine solche KldB-Rekonstruktion für Wegeners Magnitude-Prestigeskala vorgestellt. Die Basiselemente dieser Skala sind die fast 300 Berufsgattungen der Internationalen Standardklassifikation der Berufe 1968 (ISCO-68), für welche Wegener (1985; 1988) über ein komplexes, mehrstufiges Schätzverfahren Prestigescores entwickelt hat. Da die Anwendung der Magnitude-Prestigeskala die Verschlüsselung der Berufsangaben nach ISCO-68 voraussetzt, konnte diese Skala bei Mikrozensusanalysen bislang nicht genutzt werden. Um diese Lücke zu schließen, wurde hier ausgehend von der Originalskala (im folgenden  $MPS_{ISCO}$ ) und soweit möglich in Anlehnung an die Vorgehensweise von Wegener, der Versuch unternommen, die Magnitude-Prestigeskala für die Klassifikation der Berufe (im Folgenden  $MPS_{KldB}$ ) zu rekonstruieren. Die Einzelschritte der Rekonstruktion sowie die hieraus resultierenden Skalen  $MPS_{KldB75}$  und  $MPS_{KldB92}$  werden im Folgenden dargestellt. Hierfür werden zunächst kurz die beiden Berufsklassifizierungssysteme ISCO und KldB vorgestellt. Anschließend werden die grundlegenden Schritte von Wegener bei der Konstruktion vom  $MPS_{ISCO}$  skizziert.<sup>4</sup> Schließlich wird die Umsetzung von MPS auf Basis der KldB vorgestellt und auf spezifische Probleme der Umsetzung hingewiesen. Die SPSS-Programme zum Umstieg von der KldB auf die  $MPS_{KldB}$  stehen im Internet<sup>5</sup> zur Verfügung.

---

**3** Ab dem Mikrozensus 1996 enthalten die MZ-Scientific-Use-Files sowohl KldB-Codes wie auch ISCO-Codes. Allerdings handelt es sich hierbei zum einen um ISCO-88-COM, das heißt um eine spezifisch auf die Bedürfnisse der amtlichen Statistik zugeschnittene Version. Zum anderen enthalten die Scientific-Use-Files nur die ISCO-Minor Gruppen mit maximal 116 Kategorien, während in den Sozialwissenschaften im Allgemeinen die Unit Group Ebene - mit maximal 390 Kategorien - genutzt wird.

**4** Sehr hilfreich für die Erschließung der Konstruktion der Magnitude-Prestigeskala ist ein Essay von Mueller (1990) über Wegeners 'Kritik des Prestiges' (1988).

**5** [http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten\\_tools](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools)

## 2. Berufsklassifikationen

### 2.1 International Standard Classification of Occupations (ISCO-68)

Die International Standard Classification of Occupations (ISCO) ist eine vom International Labour Office (ILO) in den 1950er-Jahren entwickelte systematische, international einheitliche Berufsklassifizierung für die zivile Erwerbsbevölkerung. In den 1960er-Jahren wurde die Erstversion überarbeitet und unter der Bezeichnung ISCO-68 bekannt. In den Folgejahren diente die ISCO-68 als Grundlage für die Konstruktion einer Reihe von Status- und Prestigeskalen, so z.B. für die Prestigeskala (SIOPS) von Treiman (1979), die Magnitude-Prestigeskala (MPS) von Wegener (1988) und die Internationale Skala des sozio-ökonomischen Status (ISEI) von Ganzeboom et al. (1992). Zugleich setzte sich die ISCO-68 als Standardverschlüsselung für Berufstätigkeiten bei sozialwissenschaftlichen Erhebungen durch. So entschied sich ZUMA bei Einführung der Standarddemographie für eine Verkodung der beruflichen Tätigkeiten nach ISCO-68 und gegen die deutsche Klassifizierung der Berufe, da für die KldB keine Prestigeskala vorlag (Pappi 1979; Schönbach 1979).<sup>6</sup>

Die Klassifizierung von Berufen nach ISCO-68 basiert auf Tätigkeitsmerkmalen, die nach dem Kriterium der Ähnlichkeit in Gruppen zusammengefasst werden. Die Kodierung ist in bis zu fünf Stellen aufgeteilt, wobei jede Ziffer eines Codes eine weitere Ausdifferenzierung der Gruppen beinhaltet. Die erste Ziffer, welche die größte Einteilung darstellt und an Beschäftigungssektoren orientiert ist, steht für die 'Berufshauptgruppen' bzw. 'major groups'. Insgesamt wird zwischen den folgenden acht Berufshauptgruppen unterschieden: 'Wissenschaftler, technische und verwandte Fachkräfte' (ISCO-Code 0/1), 'Leitende Tätigkeiten im Öffentlichen Dienst und in der Wirtschaft' (ISCO-Code 2), 'Bürokräfte und verwandte Berufe' (ISCO-Code 3), 'Handelsberufe' (ISCO-Code 4), 'Dienstleistungsberufe' (ISCO-Code 5), 'Berufe des Pflanzenbaues, der Tier-, Forst und Fischwirtschaft sowie der Jagd' (ISCO-Code 6), 'Gütererzeugende und verwandte Berufstätigkeiten, Bedienung von Transportmitteln und Handlangertätigkeiten' (ISCO-Code 7/8/9) und schließlich 'Beruflich nicht näher klassifizierbare Arbeitskräfte' (ISCO-Code X)<sup>7</sup> (Statistisches Bundesamt 1971). Die zweite Stelle des Codes bezieht sich auf die 'Berufsuntergruppen' bzw. 'minor groups': In Verbindung mit der ersten Stelle ent-

---

<sup>6</sup> Für ISCO-68 sprach auch das Kriterium der internationalen Vergleichbarkeit. Zudem erschien ISCO für berufssoziologische Analysen besser geeignet.

<sup>7</sup> Die Berufshauptgruppe X der 'Beruflich nicht näher klassifizierbaren Arbeitskräfte' wurde nicht in die Konstruktion der Magnitude-Prestigeskala einbezogen.

stehen so 83 Berufsuntergruppen, die jeweils ein sehr breites Spektrum an Tätigkeiten umfassen (Statistisches Bundesamt 1971). Die dritte Stelle des Codes bezieht sich auf die 'Berufsgattungen' bzw. 'unit groups'. Unter 'Berufsgattungen' werden vergleichbare oder auch sich ergänzende berufliche Tätigkeiten (z.B. Flugzeugführer, Navigator, Flugingenieur) zusammengefasst. Im Einzelnen wird zwischen 284 Berufsgattungen differenziert. Die Berufsgattungen sind nochmals in insgesamt 1.506 'Berufsfelder'<sup>8</sup> bzw. 'occupational categories' aufgeteilt. Eine solche feine Differenzierung von Berufstätigkeiten ergibt allerdings nur bei Erhebungen mit einer sehr großen Fallzahl einen Sinn, da sonst die Zellenbesetzungen schnell zu klein werden. In der empirischen Sozialforschung wird daher im Allgemeinen auf die dreistelligen Kennziffern, das heißt die Berufsgattungen der ISCO-68 zurückgegriffen.

Zwischenzeitlich wurde die ISCO-68 von der ILO einer grundlegenden Revision unterzogen und durch eine neue Klassifikation, die ISCO-88, abgelöst. Auch bei der ISCO-88 wird zwischen vier Gliederungsebenen unterschieden: zehn Berufshauptgruppen (Major Groups); 28 Berufsgruppen; 116 Berufsuntergruppen und 390 Berufsgattungen. Die Zusammenfassung von Einzelberufen innerhalb der einzelnen Hierarchieebenen erfolgt allerdings nicht mehr wie bei ISCO-68 vorwiegend anhand der Ähnlichkeit von Tätigkeitsmerkmalen, sondern über das 'Skill'-Konzept (Elias/Birch 1994). Unterschieden wird hierbei zwischen dem 'Skill-Level'<sup>9</sup> der Berufe und 'Skill-Specialisation'.<sup>10</sup> Ein zweiter wichtiger Unterschied zwischen ISCO-68 und ISCO-88 besteht darin, dass die berufliche bzw. betriebliche Stellung (selbständig versus abhängig beschäftigt, Sachbearbeiter, Führungskraft etc.) nicht mehr als Klassifikationskriterium herangezogen wird, wohingegen sie bei der ISCO-68 ein wichtiges Unterscheidungskriterium darstellt. Aufgrund dieser zum Teil erheblichen Abweichungen in der Klassifikationslogik, ist ein einfacher Umstieg zwischen diesen beiden Versionen nicht möglich (Ganzeboom/Treiman 1996; Geis/Hoffmeyer-Zlotnik 2000). Entsprechend lassen sich ISCO-68 basierte Typologien, Indizes und Skalen auch nicht einfach auf ISCO-88 übertragen, sondern müssen aufwendig rekonstruiert werden (Ganzeboom/Treiman 1996). Die meisten in der empirischen Sozialforschung verwendeten Status- und Prestigeskalen basieren daher nach wie

---

<sup>8</sup> Einen vierstelligen Code gibt es nicht.

<sup>9</sup> Bezieht sich auf den Grad der Komplexität einzelner Aufgaben.

<sup>10</sup> Bezieht sich auf das Gebiet, in welchem Kenntnisse erforderlich sind, um die Aufgaben kompetent zu erfüllen.

vor auf ISCO-68, so auch die Magnitude-Prestigeskala von Wegener,<sup>11</sup> weshalb ISCO-88 im Folgenden nicht mehr weiter betrachtet wird.

## **2.2 Die Klassifizierung der Berufe (KldB)**

Von einer auf internationale Vergleichbarkeit ausgerichteten Berufsklassifikation wie der ISCO kann nicht erwartet werden, dass sie die jeweiligen spezifischen Besonderheiten der Berufsgliederungen in den verschiedenen Ländern im Detail erfasst. Die Feindifferenzierung zwischen beruflichen Tätigkeiten, die für ein Land von besonderer Bedeutung sind, erfolgt daher im Allgemeinen auf Basis einer nationalen Berufsgliederung. In Deutschland ist dies die 'Klassifizierung der Berufe', die von der amtlichen Statistik für die Erhebung und Darstellung der nationalen beruflichen Gliederung herangezogen wird.

Ähnlich wie bei ISCO erfolgt auch bei der KldB die Einteilung der Berufe über eine hierarchische Gliederung. Hierbei wird unterschieden zwischen Berufsbereichen, Berufsabschnitten, Berufsgruppen, Berufsordnungen und Berufsklassen (Statistisches Bundesamt 1992). Die Berufsklassen stellen die unterste Einheit der Berufssystematik dar und sind durch eine vierstellige Kennziffer bezeichnet. Die nächsthöhere Ebene sind die Berufsordnungen, die durch dreistellige Kennziffern bezeichnet werden. Hierunter werden Berufe zusammengefasst, die in Hinsicht auf die Berufsaufgabe und die Tätigkeit als gleichartig einzustufen sind. Die Berufsgruppen weisen eine zweistellige Kennziffer auf und fassen bezüglich der Berufsaufgabe und Tätigkeit verwandte Berufsordnungen zusammen. Die Berufsabschnitte wiederum, stellen eine Zusammenfassung von Berufsgruppen dar, wobei sich die Zusammenfassung an der Ähnlichkeit des verarbeiteten Materials orientiert. Die Berufsbereiche schließlich stellen die oberste Einheit der KldB dar. Hier werden Berufe zusammengefasst, die sich in Bezug auf ihre Arbeitsaufgaben, der allgemeinen Art der Tätigkeit, berühren. Die Differenzierung zwischen den einzelnen Berufsbereichen orientiert sich an der traditionellen Unterscheidung zwischen Urproduktions-, Fertigungs- und Dienstleistungsberufen.

Wie die ISCO wurde auch die KldB mehreren Revisionen unterzogen. Die erste Nachkriegsfassung wurde 1961 veröffentlicht. Ende der 60er-Jahre erfolgte eine Überarbeitung, die jedoch bereits Mitte der 70er-Jahre korrigiert und ergänzt wurde und als KldB75-70 bezeichnet wird. Ende der 80er-Jahre erfolgte eine weitere Überarbeitung der Berufsklassifikation, die zu der derzeit gültigen KldB-92 führte. Um

---

<sup>11</sup> Für eine Aktualisierung von MPS<sub>ISCO</sub> auf Basis von ISCO-88 siehe Christoph (2000).



die zeitliche Vergleichbarkeit der Berufsstatistiken nicht unnötig zu beeinträchtigen, wurde das Gliederungs- und Kennziffersystem der KldB75-70 im Wesentlichen beibehalten (Statistisches Bundesamt 1992: 7f.). Zu Änderungen kam es auf der Ebene der Zweisteller (Berufsgruppen)<sup>12</sup> und der Dreisteller (Berufsordnungen), die teilweise zusammengefasst, zum Teil stärker aufgliedert wurden (vgl. Übersicht 1).

**Übersicht 1: Klassifikation der Berufe 1975/70 und 1992: Gliederungsebenen**

	KldB75-70	KldB-92
	Anzahl der Kategorien	
Berufsbereiche	6	6
Berufsabschnitte	33	33
Berufsgruppen	86	88
Berufsordnungen	*333	369
Berufsklassen	1672	2287

Quelle: Statistisches Bundesamt (1975:9; 1992:13).

\* Die Quelle nennt hier nur 328 Berufsordnungen. Offensichtlich wurde der Berufsbereich VI (Sonstige Arbeitskräfte) nicht berücksichtigt. In der KldB-92 ist der Berufsbereich VI hingegen bei den Berufsordnungen enthalten.

In Hinblick auf die Scientific Use Files<sup>13</sup> bedeutet dies, dass die Berufsverkodung bis zum Mikrozensus 1991 nach der KldB75-70, seit dem Mikrozensus 1992 nach der KldB-92 vorliegt. Hierbei erweist es sich als äußerst hilfreich, dass das Statistische Bundesamt (1992: 554ff.) einen Umsteigeschlüssel zwischen diesen beiden Versionen auf der Ebene des Dreistellers zur Verfügung gestellt hat.<sup>14</sup>

Eine nahe liegende Möglichkeit der Operationalisierung von MPS<sub>KldB</sub> wäre eine Verschlüsselung aller KldB-Berufsordnungen (Dreisteller) entsprechend der ISCO-68 Berufsgattungen (Dreisteller) mit anschließender Zuweisung der Magnitude-Prestigescores. In der Praxis erweist sich dieser Weg allerdings als wenig geeignet. So liegt zwar ein Umsteigeschlüssel zwischen der KldB75-70 und ISCO-68 auf Ebene

<sup>12</sup> Von den ursprünglich 86 Berufsgruppen blieben 55 inhaltlich unverändert. Die restlichen Berufsgruppen wurden in der einen oder anderen Form (Streichungen, Umstrukturierungen, Neubildungen) revidiert.

<sup>13</sup> Aktuell (Stand: März 2001) stehen die faktisch anonymisierten Mikrozensen 1989, 1991, 1993, 1995, 1996 und 1997 als Scientific Use Files zur Verfügung.

<sup>14</sup> Zusätzlich enthält diese Vergleichsübersicht auch einen Umsteigeschlüssel zu ISCO-88 (Viersteller).

der Dreisteller vor (Statistisches Bundesamt 1975: 316ff), häufig ist es aber nicht möglich, eine KldB-Kategorie in eindeutiger Weise einer ISCO-Kategorie zuzuordnen. Stattdessen treten viele Fälle auf, bei welchen auf einen KldB-Beruf zwei oder mehr ISCO-Berufe zutreffen. Bei Mehrdeutigkeit könnte ein KldB-Prestigescore nur über eine ungewichtete Mittelwertbildung erzeugt werden, da keine Informationen vorliegen mit welcher Gewichtung einzelne ISCO-Kategorien in eine KldB-Kategorie einfließen. Als zusätzliches Problem erweist sich, dass es unseres Wissens keinen offiziellen Umsteigeschlüssel zwischen der KldB-92 und ISCO-68 gibt. Für die Zuweisung von Prestigescores zu der KldB-92 ist also der Umweg über die KldB75-70 zu wählen. Die aufgrund von mehrdeutigen Zuordnungen schon bei der KldB75-70 auftretenden Unschärfen würden beim Umstieg auf die KldB92 demnach kumuliert.

Vor diesem Hintergrund haben wir im ersten Schritt für die Erzeugung der  $MPS_{KldB75}$  eine alternative Vorgehensweise gewählt. Die Prestigescores für die KldB75-70 werden nicht über ein Umschlüsselungsverfahren zugeordnet, sondern in Orientierung an Wegeners Vorgehensweise berechnet. Die Verknüpfung zwischen  $MPS_{KldB}$  und  $MPS_{ISCO}$  erfolgt über 140 Berufe, für welche eine eindeutige Zuordnung zwischen KldB75-70 und ISCO-68 und damit auch eine entsprechende Zuweisung der  $MPS_{ISCO}$ -Scores möglich war. Die Ermittlung der letztendlichen Scores der  $MPS_{KldB75}$  erfolgt dann ausschließlich auf Basis von Mikrozensusdaten, das heißt eine direkte Vergleichbarkeit zwischen  $MPS_{ISCO}$ - und  $MPS_{KldB}$ -Scores ist nicht gegeben. In einem zweiten Schritt wird die  $MPS_{KldB92}$  anhand des vom Statistischen Bundesamt vorgegebenen Umsteigeschlüssel und einer gewichteten Mittelbildung erzeugt.

Bevor unter Abschnitt 4 das Verfahren für die Konstruktion der  $MPS_{KldB}$  näher beschrieben wird, werden im Folgenden Abschnitt die wesentlichen Schritte Wegeners bei der Konstruktion von  $MPS_{ISCO}$  skizziert, da diese die Basis unserer Operationalisierung der  $MPS_{KldB}$  bilden. Die dem Prestigekonstrukt zugrunde liegenden theoretischen Überlegungen werden weitgehend ausgeblendet, der hieran interessierte Leser sei auf die Originalliteratur von Wegener (1985; 1988) verwiesen.

### **3. Die Konstruktion von $MPS_{ISCO}$**

Im Unterschied zu Statusskalen, bei welchen die soziale Einstufung im Allgemeinen durch den Forscher erfolgt, beruht die soziale Verortung von beruflichen Positionen bei Prestigeskalen nach Wegener typischerweise auf dem Reputationsansatz, das heißt auf der Wahrnehmung des gesellschaftlichen Ansehens von Berufspositionen

durch die Betroffenen. Die soziale Hierarchie wird hierbei durch eine direkte Befragung ermittelt (Wegener 1985:211f). Eine nicht unerhebliche Einschränkung erfährt der Reputationsansatz allerdings dadurch, dass bei sehr differenzierten Berufsklassifikationen, wie etwa ISCO, die Einordnung aller Berufe durch die Befragten mit einem immensen Zeit- und Kostenaufwand einhergehen würde. Vermutlich aus diesem Grund wurde bei der Konstruktion von  $MPS_{ISCO}$  nur die Reputation von 50 Berufen direkt erhoben. In einem zweiten Schritt wurde - ausgehend von dieser Direktmessung - mittels eines aufwendigen Verfahrens eine Übertragung der Magnitude-Prestigescores auf insgesamt 283 Berufsgattungen<sup>15</sup> des ISCO-Schemas vorgenommen. Im Folgenden wird die Reputationsmessung für die 50 Berufe, von Wegener auch als MAG-50 bezeichnet, in ihren Grundzügen dargestellt.

### 3.1 Die Magnitude-Messung (MAG-50)

Die MPS basiert - wie schon aus der Bezeichnung ersichtlich - auf einer Magnitude-Messung (Wegener 1978, 1980). Diese Technik wurde in der Psychophysik für die Messung subjektiv empfundener physikalischer Reizintensitäten (z.B. Temperatur oder Lautstärke) entwickelt, findet gelegentlich aber auch in der Einstellungsmessung Verwendung.<sup>16</sup> Die Magnitude-Messung beruht auf der Annahme, dass Menschen in der Lage sind, die wahrgenommene Intensität einer Empfindung anzugeben, bspw. durch eine numerische Quantifizierung. Den Befragten wird zunächst ein Standardstimulus vorgegeben, anschließend werden sie einer Reihe weiterer Stimuli ausgesetzt, die im Verhältnis zu dem Standardstimulus beurteilt werden sollen.<sup>17</sup> Bei der Magnitude-Einstellungsmessung finden in der Regel zwei Reaktionsmodalitäten Anwendung. Bei Wegener sind dies eine numerische Magnitude-schätzung sowie das Zeichnen von Linien proportional zur wahrgenommenen Sti-

---

<sup>15</sup> Wegener verwendet die ISCO-Kategorien 011 bis 999 einschließlich der ZUMA-Kategorien '001: Soldat (Wehrberuf)'; '002: Offizier (Wehrberuf)'. Keine Prestigezuweisung erfolgt für die Berufsgattungen der Berufshauptgruppe X, welche die beruflich nicht näher klassifizierbaren Arbeitskräfte umfasst.

<sup>16</sup> Wenngleich die Magnitude-Messung im Vergleich zur herkömmlichen Kategorialmessung gewisse Vorteile bezüglich der Messgenauigkeit aufweist, hat sie sich in den empirischen Sozialwissenschaften bislang nicht als Standardmethode in Bevölkerungsumfragen durchsetzen können. Dies hängt möglicherweise damit zusammen, dass sie im Vergleich zu konventionellen Meßverfahren sehr zeitaufwendig ist und nicht nur Interviewer, sondern auch Befragte zunächst eine Trainingsphase durchlaufen müssen.

<sup>17</sup> Ist als Standardstimulus etwa ein Würfel mit einem Inhalt von  $10 \text{ cm}^3$  vorgegeben, sollten die Befragten bei einem als viermal so groß wahrgenommen Vergleichswürfel den Wert 40 angeben.

mulusintensität. Bei der numerischen Schätzung werden die Versuchspersonen gebeten, eine Zahl zu nennen, welche die Höhe der Empfindungsstärke im Vergleich zu der Referenzzahl ausdrückt. Beim Linienzeichnen soll die Länge der gezeichneten Linie die Intensität im Vergleich zur Referenzlinie ausdrücken. Werden beide Reaktionsmodalitäten auf einen Stimulus angewandt, lässt sich hieraus unmittelbar ein Gütekriterium für die Messung berechnen (Wegener 1980: 7f).<sup>18</sup> Weitere Vorteile der Magnitude-Messung im Vergleich zu der klassischen - auf Kategorienskalen beruhenden - Reputationsmessung sieht Wegener (1978) u.a. darin, dass die Skalierung auf einer offenen Skala stattfindet. Damit ist gemeint, dass die Anfangs- bzw. Endpunkte nicht bereits festgelegt sind, wie dies beispielsweise bei einer Kategorienskala der Fall ist. Hierdurch lässt sich das Problem von 'Deckeneffekten' umgehen, das heißt die Kumulation von Antworten an den Skalenenden; zugleich ist eine weitaus feinere Differenzierung der Beurteilungen möglich.<sup>19</sup> Darüber hinaus können die bei einer Magnitudeschätzung ermittelten Werte im Unterschied zu einer Kategorialmessung metrisch interpretiert werden.

Ein Nachteil der Magnitude-Messung ist die Komplexität und der mit ihr verbundene Zeitaufwand, zumal der eigentlichen Skalierungsphase durch die Befragten eine Trainings- und Eichungsphase vorgeschaltet werden muss. Der Versuch von den Befragten eine numerische und/oder graphische Verhältnisschätzung für fast 300 Berufsangaben zu erhalten, dürfte vermutlich nicht nur an Zeitproblemen, sondern auch an der Motivation der Befragten selbst scheitern.

Entsprechend hat Wegener (1985: 222f.) für die Konstruktion seiner Magnitude-Prestigeskala aus den 284 Berufen des ISCO-Schemas insgesamt 50 Berufe<sup>20</sup> als

---

**18** Wenn den Verhältnisschätzungen sowohl numerische Angaben wie auch Linienzeichnungen zugrunde liegen, sollten die jeweiligen Schätzungen proportional erfolgen. Verdoppelt sich eine Zahl, sollte sich auch die Linie verdoppeln (Wegener 1980: 7f).

**19** Nach Wegener (1979) erlaubt eine Magnitude-Messung eine mehr als 100 mal feinere Differenzierung des Beurteilungsgegenstandes als eine Kategorienskala.

**20** Elektroinstallateur, Studienrat, Fabrikarbeiterin, Polizist, Bauingenieur, Volksschullehrerin, Taxiunternehmer, Maschinenschlosser, Maurer, Sekretärin, Arzt, Kraftfahrzeugmechaniker, Briefbote, Krankenschwester, Bauarbeiter, Industriekaufmann, Professor, Bankdirektor, Diplom-Psychologe, Apotheker, Finanzbuchhalter, Bäckermeister, Bankangestellter, Landwirt, Krankenwagenfahrer, Kindergärtnerin, Fotograf, Omnibusfahrer, Dreher, Textilweber, Schuhverkäuferin, Hilfsarbeiter, Autowäscher, Richter, Rechtsanwalt, Architekt, Journalist, Lebensmitteltechniker, Programmierer, Bibliothekarin, Feinmechaniker, Buchdrucker, Zahnärztin, Bundesbahnschaffner, Fernfahrer, Friseur, Eisengießer, Landarbeiter, Müllwerker, Lagerarbeiter.

Stimuli ausgewählt. Diese 50 Berufe wurden in drei Serien mit einmal 16 und zweimal 19 Berufen unterteilt. Zwei Berufsbezeichnungen (Elektroinstallateur, Kraftfahrzeugmechaniker) waren in allen drei Serien enthalten; sie dienten als Verbindungspunkte für die Konstruktion der gemeinsamen Skala. In den ZUMA-Bus-Erhebungen 1979 (16 Berufe) und 1980 (Split A und B zu je 19 Berufen) wurden die Befragten<sup>21</sup> gebeten, das gesellschaftliche Ansehen der ausgewählten Berufe relativ zu dem Beruf des Elektroinstallateurs (Standardstimulus, der später auf den Wert 50 standardisiert wurde) zum einen numerisch, zum anderen durch Linien darzustellen.<sup>22</sup> Zugleich sollten die Befragten das Ansehen der jeweiligen Berufe auf einer konventionellen Rating-Skala mit Werten zwischen 1 (sehr geringes Ansehen) und 9 (sehr hohes Ansehen) einstufen, um so eine weitere Validierungsmöglichkeit der Magnitude-Skala zu erhalten und einen direkten Vergleich der beiden Messmethoden zu ermöglichen. Die Berechnung der Prestige-Scores für die einzelnen Berufe erfolgte anschließend über eine arithmetische Mittelung der Einstufung aller Befragten. Wegener bezeichnet die Magnitude-Skala der 50 direkt bewerteten Berufe als MAG-50, die Kategorieskala als KAT-50. Im nächsten Schritt wurde die direkt gemessene MAG-50 auf das Gesamtspektrum des ISCO-68 Schemas ausgedehnt.

### 3.2 Die Übertragung von MAG-50 auf das Gesamtspektrum des ISCO-68 Schemas

Der Übergang von der 50 Berufe umfassenden Magnitude-Messung (MAG-50) zu der 283 Berufe umfassenden Magnitude-Prestigeskala (MPS) setzt sich aus einer Reihe von Einzelschritten zusammen, die hier nicht en detail dokumentiert werden können. Stattdessen betrachten wir jene Schritte, die für die spätere Rekonstruktion der MPS<sub>KldB</sub> von Interesse sind.

Zunächst konzentriert sich Wegener auf die Wahl eines 'Zielkontinuums', das heißt das Auffinden einer Ordnung, welche es ihm erlaubt, die Werte von MAG-50 auf das gesamte ISCO-Schema auszudehnen. Ausgehend von seinen theoretischen

---

<sup>21</sup> Insgesamt 4015 Personen (Wegener 1985: 222).

<sup>22</sup> Eine zentrale Voraussetzung für diese Vorgehensweise ist die Annahme, dass alle Befragten eine nach Transformationen entsprechende Metrik bei der Beurteilung von Prestige aufweisen. Beispielhaft kann man sich dies daran verdeutlichen, dass man das Gewicht einer Packung Kaffee mit 500 Gramm, einem Pfund oder 0,5 Kilo angeben kann. Zwei Packungen Kaffee entsprechen dennoch jeweils dem doppelten Gewicht – 1000 Gramm, zwei Pfund oder 1 Kilo. Das Prestige entspricht in diesem Beispiel dem Gewicht und die Relationen bleiben erhalten, egal in welcher Einheit das Prestige angegeben wird.

Überlegungen, dass Prestige eine gesellschaftliche Strukturkomponente abbildet, liegt es nahe, als Zielkontinuum eine andere Skala heranzuziehen, der gleichfalls gesellschaftliche Strukturkomponenten zugrunde liegen. Zu diesem Zweck korreliert Wegener (1985: 223) die MAG-50 mit vier alternativen Skalierungen für berufliche Tätigkeiten: (1) der Prestigeskala von Treiman, (2) einer neu konstruierten sozio-ökonomischen Statusskala (STAT), (3) einer Strukturskalierung (SAS), der eine Rangordnung der Tätigkeiten nach STAT zugrunde liegt, und (4) einer weiteren Strukturskala (SAS-Trei), welche auf der Rangordnung der Berufe nach der Prestigeskala von Treiman (SIOPS) basiert.<sup>23</sup> Die höchste Korrelation findet sich mit Punkt .901 zwischen MAG-50 und SAS. Nach Wegener (1984: 43f.) übernimmt daher SAS die Funktion des Zielkontinuums für die Konstruktion der Gesamtskala. Da STAT und SAS für die Konstruktion der MPS<sub>KIDB</sub> von Relevanz sind, werden sie im Folgenden kurz erläutert.

STAT ist eine Skala, die aggregierte Statusmaße für berufliche Tätigkeiten wiedergibt. Untersuchungseinheiten sind dabei nicht Individuen, sondern Berufe. Erzeugt wird STAT über eine Hauptkomponentenanalyse, in welche das Netto-Einkommen, das Ausbildungsniveau (in Jahren) und die subjektive Schichteinstufung der erwerbstätigen Befragten pro Berufskategorie einfließen. Der erste extrahierte Faktor erklärt eine Gesamtvarianz von 74 Prozent in den Variablen und hat einen Eigenwert von 1.703 (Wegener 1984: 45).

Mit der Staturerwerbsskala SAS sollen vertikale Schließungsprozesse in Bezug auf berufliche Tätigkeiten abgebildet werden. Als theoretisches Modell zieht Wegener das Vakanzkettenmodell von Sørensen (1979) heran. Nach diesem Modell hängt - stark vereinfacht dargestellt - die erreichbare Staturhöhe nicht primär von dem individuellen Humankapital ab, sondern vor allem von den auf dem Arbeitsmarkt zur Verfügung stehenden freien Positionen bzw. Vakanzen.<sup>24</sup> Neu geschaffene oder freigewordene Positionen werden von unten nach oben besetzt. Da durch jeden Aufstieg wiederum eine Vakanz entsteht, zieht sich eine Kette von Leerstellen nach unten hin durch, der eine Kette von Aufstiegen von unten nach oben entgegensteht.

---

**23** Datenbasis für die Korrelationen sowie die Konstruktion der endgültigen MPS war der kumulierte Standarddemographie-Datensatz (mit neun Einzelerhebungen), wobei nur männliche Befragte (N=5716) in die Analysen einbezogen wurden (Wegener 1984: 223).

**24** Nach diesem Modell weist der Status eines Berufes die Eigenschaft eines Positionsguts auf. Je weniger berufliche Positionen über einem bestimmten Beruf angeordnet sind, desto höher ist der Status des entsprechenden Berufes. Je mehr berufliche Positionen über einem bestimmten Beruf angeordnet sind, desto geringer ist der Status dieses Berufes.

Ausgangspunkt für die empirische Umsetzung dieses Modells ist die von STAT vorgegebene Rangordnung der Berufstätigkeiten, welche die unterschiedlichen Positionsniveaus im Vakanzkettenmodell abbildet. Die Transformation von STAT in SAS erfolgt nach der Gleichung

$$y = -\ln [1 - F(y)].$$

Der Term  $[1 - F(y)]$  steht für die von oben kumulierten Häufigkeiten der Berufe auf den verschiedenen Positionsniveaus, wobei die Häufigkeiten als Indikator für die Vakanzraten dienen. Mittels dieser Transformation wird die ordinale Statusskala STAT zu der intervallskalierten Strukturskala SAS. Wegener (1984: 46) weist hierbei auf eine spezifische Schwäche bei der Konstruktion von SAS hin. Diese besteht darin, dass die in STAT abgebildeten Rangordnungen der Berufspositionen bei einer schwachen Zellbesetzung ( $n < 10$ ) oder bei Nullbesetzungen, mit einer gewissen Unsicherheit behaftet sind. Aus diesem Grund bedient er sich für die Ausdehnung der MAG-50 auf das ISCO-Gesamtschema eines zweistufigen Verfahrens. In der ersten Stufe erfolgt zunächst eine Erweiterung von KAT-50 mittels der Treiman-Prestigeskala auf die 283 ISCO-Berufe. Die resultierende Skala wird als KTREI bezeichnet. Da stimulusgleiche Kategorienskalen eine additive Potenzfunktion zu Magnitudeskalen bilden, kann Wegener anschließend KAT-50 und MAG-50 in eine Relation bringen, die es ihm dann ermöglicht, aus KTREI eine vorläufige Magnitude-Prestigeskala (MPS1) für das ISCO-Schema zu bilden. In der zweiten Stufe erfolgt die Erweiterung von MAG-50 auf SAS. Hierbei werden alle SAS-Werte von Berufen mit einer Zellbesetzung größer zehn direkt in die Skala übernommen, SAS-Werte von Berufen mit einer Zellbesetzung kleiner oder gleich zehn werden durch linear transformierte MPS1-Werte ersetzt. Um sicherzustellen, dass die verschiedenen Übertragungsschritte nicht zu einer Verschiebung der 50 direkt gemessenen Magnitude-Werte führen, werden die ursprünglichen Werte von MAG-50 nach einer linearen Transformation wieder aufgenommen (Wegener 1984: 48). Die so erzeugte Skala wird abschließend über eine Potenztransformation so normiert, dass die niedrigste Berufsposition den Wert 20, die höchste Berufsposition den Wert 186,1 annimmt. Die resultierende Skala ist die Magnitude-Prestigeskala von Wegener, die hier verkürzt als  $MPS_{ISCO}$  bezeichnet wird.

#### 4. $MPS_{KldB}$ : Die Umsetzung von $MPS_{ISCO}$ auf die Klassifikation der Berufe

Wie dargestellt, erfolgt die Umsetzung der  $MPS_{KldB}$  in enger Anlehnung an die Vorgehensweise von Wegener bei der Konstruktion von  $MPS_{ISCO}$ . Dennoch handelt es

sich bei den resultierenden KldB-Prestigewerten nicht um eine Eins-zu-Eins Übertragung der ISCO-Prestigewerte. Denn, wie dargestellt, sind einerseits die KldB und die ISCO nicht deckungsgleich und auch die Datenquellen für die jeweilige Skalendesignkonstruktion sind unterschiedlich.<sup>25</sup> Andererseits gibt es bei der Rekonstruktion von  $MPS_{KldB}$  da Abweichungen, wo Wegener aufgrund schwacher Zellbesetzungen oder Nullbesetzungen Zwischenschritte einführen musste, die sich bei der hier zur Verfügung stehenden Datenbasis erübrigen.

#### 4.1 Datenbasis

Als Datenbasis für die Konstruktion der  $MPS_{KldB}$  dienen die kumulierten Mikrozensen der Jahre 1982 und 1985 (im folgenden MZ82/85), die in relativer zeitlicher Nähe zu Wegeners Daten liegen.<sup>26</sup> In Analogie zu Wegener fließen in die Berechnung von  $MPS_{KldB}$  nur die Angaben männlicher Erwerbstätiger ein. Hierbei werden nur Personen mit deutscher Staatsangehörigkeit und einer Wochenarbeitszeit zwischen 30 und 60 Stunden berücksichtigt.<sup>27</sup> Fälle, die nicht in allen benötigten Merkmalen eine gültige Angabe aufweisen, werden nicht in die Schätzungen einbezogen.<sup>28</sup> Nach Abschluss aller Selektionen verbleibt ein N von 215.191 Fällen. Die für die Berechnung von  $MPS_{KldB}$  herangezogenen Variablen werden im Folgenden kurz skizziert.

**Klassifikation der Berufe:** Als Ausgangspunkt für die Konstruktion der  $MPS_{KldB}$  dient der dreistellige Berufscode (Berufsordnungen) der KldB75-70, welche ein-

---

<sup>25</sup> Zu Verteilungsabweichungen zwischen amtlichen Mikrodaten und Daten aus der Umfrageforschung siehe Hartmann/Schimpl-Neimanns (1992).

<sup>26</sup> Den zeitlich näher liegenden Mikrozensen 1979 und 1981, die gleichfalls als ZUMA-Files zur Verfügung stehen, fehlen zum Teil die für die Konstruktion von MPS zentralen Informationen.

<sup>27</sup> Gleichfalls nicht in die Analysen einbezogen wurde der KldB-Berufsbereich VI 'Sonstige Arbeitskräfte'. In diesem Berufsbereich werden erfasst: 'Mithelfende Familienangehörige außerhalb der Landwirtschaft', soweit sie anderweitig nicht angeführt sind (Code 971); Lehrlinge mit noch nicht feststehendem Beruf (Code 981); Praktikanten, Volontäre mit noch nicht feststehendem Beruf (Code 982); Arbeitskräfte (arbeitsuchend) mit noch nicht bestimmtem Beruf (Code 983) sowie Arbeitskräfte ohne nähere Tätigkeitsangabe (991). Die Berechnung von  $MPS_{KldB75}$  erfolgt also nur für 328 der insgesamt 333 Berufsordnungen. Der KldB-Berufsbereich VI ist weitgehend deckungsgleich mit der ISCO-Berufshauptgruppe X (Statistisches Bundesamt 1975: 320), die bei der Berechnung von  $MPS_{ISCO}$  unberücksichtigt blieb.

<sup>28</sup> Beim Mikrozensus besteht Auskunftspflicht, weshalb sowohl der unit-nonresponse wie auch der item-nonresponse deutlich geringer als in sozialwissenschaftlichen Bevölkerungserhebungen ist.



schließlich bis zum Mikrozensus 1991 von der amtlichen Statistik für die Verschlüsselung der Berufstätigkeiten herangezogen wurde. Die Prestigewerte für die KldB-92 (ab Mikrozensus 1992) werden über ein einfacheres Umschlüsselungsverfahren gewonnen. Zur Unterscheidung zwischen diesen beiden Versionen sprechen wir im Folgenden von der  $MPS_{KldB75}$  und der  $MPS_{KldB92}$ .

**Bildung in Jahren:** Die Variable 'Bildung in Jahren' wurde aus den Mikrozensus-Variablen 'letzter allgemeinbildender Schulabschluss' und 'letzter beruflicher Ausbildungsabschluss' generiert (vgl. Anhang A1).

**Einkommensvariable:** Als Einkommensvariable diente das individuelle Nettoeinkommen. Da das Einkommen im Mikrozensus nur gruppiert weitergegeben wird, wurde eine Metrisierung über die Gruppenmittelwerte vorgenommen. Die Einkommensvariable floss in logarithmierter Form in die Analyse ein.

**Schichtzugehörigkeit:** Das Merkmal 'Schichtzugehörigkeit' wird im Mikrozensus nicht erhoben. Ersatzweise wurde daher aus den beiden Variablen 'Stellung im Beruf' und 'Stellung im Betrieb' eine Schichtvariable konstruiert (vgl. Anhang A2).

**MPS-140:** Wie dargestellt, ist der Kern von  $MPS_{ISCO}$  eine direkte Magnitude-Messung von 50 Berufen (MAG-50). Bei der Generierung der  $MPS_{KldB}$  ergibt sich das Problem, dass nicht alle MAG-50 Berufe in eindeutiger Weise einer KldB-Kategorie zugeordnet werden können. Wird jedoch  $MPS_{ISCO}$  als valide akzeptiert (Wegener 1985: 225ff.), können anstelle der Prestigescores dieser 50 Berufe auch die Scores jener KldB-Berufe für die Umsetzung der  $MPS_{KldB}$  herangezogen werden, die in eindeutiger Weise mit einer ISCO-Kategorie korrespondieren. Auf Basis eines Umsteigeschlüssels des Statistischen Bundesamts (1975: 316ff.) wurden 140 der 328 KldB-Berufsordnungen in eindeutiger Weise einem ISCO-Code zugeordnet,<sup>29</sup> diese werden im Folgenden als MPS-140 bezeichnet.<sup>30</sup> Die MPS-Scores dieser 140 Be-

---

<sup>29</sup> Aus Platzgründen konnten wir die Liste hier nicht veröffentlichen. Sie finden sie im Internet unter: [http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten\\_tools](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools).

<sup>30</sup> In die Kategorie der eindeutig zuordenbaren Berufe fallen auch Zahnärzte, Apotheker und Tierärzte. In  $MPS_{ISCO}$  weisen diese Berufe einen Score zwischen 110 und 139 auf und liegen damit weit hinter den Ärzten, die mit 186 den höchsten Score haben. Die ersten Berechnungen mit MZ-Daten ergaben ein völlig anderes Bild. Hier liegen die Zahnärzte gefolgt von den Apothekern deutlich vor den Ärzten am oberen Ende der Skala. Bei der Regressions-schätzung (vgl. Abschnitt 4.3) stellen die Zahnärzte, Apotheker und Tierärzte Ausreisser dar, die mehr als drei Standardabweichungen abweichen und mittels der errechneten Regressionsgeraden nicht entsprechend ihrer Originalwerte in der Prestigeverteilung geschätzt wer-

rufe dienen bei den folgenden Analysen als Verbindungsglied zwischen  $MPS_{ISCO}$  und  $MPS_{KldB}$ .

## 4.2 Die Konstruktion von $STAT_{KldB}$ und $SAS_{KldB}$

Für die Konstruktion der  $MPS_{KldB75}$  wird zunächst die sozio-ökonomische Statusskala  $STAT$  für die  $KldB75-70$  repliziert. Auf Basis von  $STAT$  kann dann die Strukturskala  $SAS$  generiert werden, die als Zielkontinuum für die Ausdehnung von  $MPS-140$  auf das gesamte Berufsspektrum der  $KldB$  dient.

### $STAT_{KldB}$

Die im Mikrozensus enthaltenen Informationen beziehen sich auf Individuen. Da bei  $MPS$  nicht das Prestige von Individuen, sondern von Berufen im Zentrum des Interesses steht, werden die Daten des  $MZ82/85$  zunächst nach Berufen der  $KldB$  aggregiert.<sup>31</sup> Die Merkmale 'Bildung in Jahren', 'Einkommen' und 'Schichtzugehörigkeit' wurden hierbei für jede einzelne Berufsordnung gemittelt. Anschließend gingen diese Merkmale in eine Faktorenanalyse ein, in welcher über ein Maximum-Likelihood-Verfahren ein Faktor extrahiert werden konnte, der einen Eigenwert von 2,61 und eine Varianzaufklärung von 87,3 Prozent aufweist. Die pro Berufskategorie gemittelten Factorscores bilden die Rangordnung der beruflichen Tätigkeiten ab, das heißt die Werte für die Statusskala  $STAT_{KldB}$ . Ausgehend von  $STAT_{KldB}$  wurde in einem nächsten Schritt  $SAS_{KldB}$  nach dem Modell von Sørensen berechnet, welche die 328 Berufsordnungen der  $KldB75-70$  in eine metrische Ordnung überführt.

### $SAS_{KldB}$

Wie erwähnt, bestand bei der ursprünglichen Generierung von  $SAS_{ISCO}$  das Problem, dass eine Reihe von Berufskategorien in  $STAT_{ISCO}$  entweder nur gering ( $n < 10$ ) oder nicht besetzt waren. Um diese Lücken zu füllen führte Wegener einen Zwischenschritt ein, bei welchem die fehlenden Werte mittels der Treiman-Prestigeskala geschätzt wurden. Die Fallzahl des hier zur Verfügung stehenden  $MZ82/85$  ist mit

---

den können. Der Einfluss dieser Berufe auf die Lage der Regressionsgeraden und damit auf die Schätzung aller Werte ist hierbei so groß, dass wir uns dafür entschieden haben, diese drei Berufe - obwohl eindeutig zuordenbar - nicht in  $MPS-140$  aufzunehmen, d.h. sie fließen nicht in die Schätzung der Regressionsgeraden ein.

<sup>31</sup> Auch für die Generierung von  $MPS_{ISCO}$  wurde  $STAT$  aggregiert. Allerdings ist aus Wegeners Ausführungen nicht ersichtlich, wann die Aggregation erfolgte - vor oder nach der Durchführung der Faktorenanalyse. Für die Generierung von  $MPS_{KldB}$  wurden die Daten vor der Faktorenanalyse aggregiert.

einem N über 200.000 allerdings um ein Vielfaches höher als bei Wegeners Daten. Dementsprechend sind alle 328 Klub-Kategorien besetzt und nur elf Berufsordnungen weisen eine Zellbesetzung kleiner 10 auf. Da einerseits die Treiman-Skala auf ISCO basiert und deshalb im Mikrozensus aus den oben angeführten Gründen nicht einfach reproduzierbar ist, andererseits die Zahl der schwach besetzten Kategorien sehr gering ist, wurde auf die Einführung eines Zwischenschrittes verzichtet. Im Unterschied zu der Vorgehensweise bei der ursprünglichen Generierung von MPS wurden also in die Berechnung von  $SAS_{KldB}$  einige gering besetzte Berufskategorien einbezogen. Die resultierenden Prestigescores der betreffenden Berufe<sup>32</sup> weisen daher weniger gesicherte Werte auf.

Die weitere Berechnung von  $SAS_{KldB}$  erfolgte in Analogie zu der Vorgehensweise von Wegener nach der Gleichung  $y = -\ln[1-F(y)]$ , wobei  $F(y)$  die Kumulationsfunktion der nach  $STAT_{KldB}$  geordneten Häufigkeiten darstellt. Diese Funktion nimmt für die letzte Ausprägung den Wert 1 an. Da der natürliche Logarithmus von Null nicht definiert ist, wurde der letzte Wert manuell auf 0.999 gesetzt, um eine Berechnung zu ermöglichen.<sup>33</sup> Der höchste Wert der resultierenden  $MPS_{KldB}$  stellt damit eine künstliche Obergrenze dar.

Der negative Logarithmus der von oben kumulierten Häufigkeiten entspricht dann  $SAS_{KldB}$ . Die Korrelation von  $SAS_{KldB}$  mit MPS-140 beträgt 0.977, zwischen  $STAT_{KldB}$  und MPS-140 findet sich eine Korrelation von 0.953 (vgl. Übersicht 2).  $SAS_{KldB}$  kann daher als Zielkontinuum für die Übertragung von MPS-140 auf die KldB75-70 herangezogen werden.<sup>34</sup>

---

**32** Es handelt sich um die Berufsordnungen 'Nieter' (Code 243); 'Metallkleber' (Code 244); Tuftingwarenmacher (Code 343); Filzmacher (Code 345); Wäscheschneider (Code 353); Sticker (Code 354); Hut-, Mützenmacher (Code 355); Handschuhmacher (Code 377); Fischverarbeiter (Code 403); Fertiggerichtekonservierer (Code 412); Angehörige geistlicher Orden (Code 892).

**33** Auf eine weitere Differenzierung der Nachkommastellen wurde verzichtet, da bei der Logarithmusfunktion geringfügige Veränderungen in der Nähe des Extremwertes Null mit starken Effekten einhergehen und sich daher der Abstand zwischen dem letzten und dem vorletzten Wert extrem vergrößert hätte.

**34** Eine ergänzende Überprüfung mit anderen von Wegener verwendeten Skalen ist nicht möglich, da diese zum Teil auf ISCO beruhen oder für die Konstruktion Variablen benötigt werden, die im MZ82/85 nicht enthalten sind. Es ist jedoch auch kaum zu erwarten, dass sich ein ähnlich hoher oder gar ein noch höherer Zusammenhang mit irgendeiner anderen Skala gezeigt hätte.

**Übersicht 2: Korrelationen zwischen MPS-140, STAT<sub>KIdB</sub> und SAS<sub>KIdB</sub>**

	MPS-140	STAT <sub>KIdB</sub>	SAS <sub>KIdB</sub>
MPS-140	1,000		
STAT <sub>KIdB</sub>	0,953	1,000	
SAS <sub>KIdB</sub>	0,977	0,957	1,000

Quelle: Eigene Berechnungen auf Basis des MZ82/85.

**4.3 Die Übertragung von MPS-140 auf die KIdB75-70**

Für die Berechnung von MPS<sub>KIdB</sub> wird zunächst eine lineare Regression auf Basis der 140 eindeutig zugeordneten Berufen mit SAS<sub>KIdB</sub> als unabhängiger und MPS-140 als abhängiger Variable durchgeführt. Die hieraus resultierende Regressionsgleichung lautet:

$$\text{MPS-140} = 30.036 + 26.921 \cdot \text{SAS}_{\text{KIdB}}$$

Der Standardfehler von SAS<sub>KIdB</sub> beträgt 0.0182, R<sup>2</sup> weist einen Wert von 0,954 auf. Die jeweiligen T-Werte sind hochsignifikant (p < 0.000).

Auf Basis der so bestimmten Regressionskoeffizienten können nun durch Einsetzen von SAS<sub>KIdB</sub> in die Regressionsgleichung die Werte aller 328 Berufsordnungen geschätzt werden. Die nach der Gleichung

$$30.036 + 26.921 \cdot \text{SAS}_{\text{KIdB}} = \text{MPS}_{\text{KIdB75}}$$

ermittelten Werte ergeben die Prestigescores der MPS<sub>KIdB75</sub>.

Wegener (1984: 48) hat im Anschluss an diese Berechnungsphase die geschätzten Prestigewerte der MAG-50-Berufe wieder durch die direkt erhobenen - linear angepassten - Magnitude-Werte ersetzt. Um eine numerische Vergleichbarkeit mit Treimans Prestigeskala zu erreichen, wurde weiterhin die resultierende Skala nochmals so normiert, dass der niedrigsten Berufsposition (ungelernte Arbeiter ohne nähere Bezeichnung: ISCO-Code 999) der Prestigescore 20 und der höchsten Berufsposition (Ärzte: ISCO-Code 061) der Prestigescore 186.8 zugewiesen wurde. Die endgültige Skala MPS<sub>ISCO</sub> hat einen Mittelwert von 63.8 und eine Standardabweichung von 30.8.

In Abweichung zu Wegeners Vorgehensweise wurden bei der MPS<sub>KIdB</sub> die für MPS-140 geschätzten Prestigescores nicht auf die Originalscores zurückgesetzt, obgleich die schätzungsbasierte Rangordnung dieser 140 Berufe nicht mehr vollständig deckungsgleich ist mit der durch MPS<sub>ISCO</sub> vorgegebenen Rangfolge. Dies verringert

zwar die direkte Vergleichbarkeit von  $MPS_{ISCO}$  und  $MPS_{KldB}$ , ist aber unabdingbar in Hinblick auf die interne Konsistenz der  $MPS_{KldB}$ . Denn die absoluten Prestigewerte erlauben für sich genommen keine sinnvollen Aussagen, sondern müssen in Relation zu den anderen Werten interpretiert werden. Ein Zurücksetzen der geschätzten MPS-140-Werte auf die Original-Werte hätte daher die empirisch ermittelte Ordnung von  $MPS_{KldB}$  nachhaltig gestört. Darüber hinaus wurden auch die Prestigescores der  $MPS_{KldB75}$  keiner abschließenden Normierung unterzogen. So wünschenswert eine numerische Vergleichbarkeit zwischen  $MPS_{ISCO}$  und  $MPS_{KldB75}$  wäre, wurde von einer Normierung abgesehen, da ja bereits über die Originalwerte von MPS-140 die von Wegener normierten Prestigescores in die Berechnung eingeflossen sind.

Die endgültige  $MPS_{KldB75}$  hat einen Range von 186.0, einen Mittelwert von 53.8 und eine Standardabweichung von 28.5.<sup>35</sup> Die höchste Skalenposition nehmen die Zahnärzte (Code 242) mit dem Wert 216.0 ein. Den niedrigsten Skalenwert mit 30.0 weisen mithelfende Landarbeitskräfte (Code 041), Korbmacher (Code 184) und Fischverarbeiter (Code 403) auf. Die Prestigewerte für alle 328 Berufsordnungen der KldB75-70 stehen im Internet ([www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten\\_tools](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools)).

## 5. Die Übertragung der $MPS_{KldB75}$ auf die KldB-92

Seit dem Mikrozensus 1992 erfolgt die Verschlüsselung der Berufsangaben auf Basis der Klassifikation der Berufe 1992. Im Hinblick auf die Ermittlung der MPS-Scores für die KldB-92 war hier abzuwägen zwischen einem einfachen Umschlüsselungsverfahren versus einer Berechnung in Analogie zur  $MPS_{KldB75}$ . Wir haben uns zugunsten des Umschlüsselungsverfahrens entschieden. Dies geschah erstens in Bezug auf die Vergleichbarkeit der beiden KldB-Versionen: Berufe, die in der 75-70er wie auch in der 92er-Version in identischer Weise erfasst werden, sollten auch die gleichen Prestigescores erhalten. Zweitens ist unseres Erachtens für eine Aktualisierung eine Neuberechnung auf Basis der Anfang der 1980er-Jahre ermittelten MPS-Scores nicht ausreichend. Eine wirkliche Aktualisierung würde auch eine neue Magnitude-Messung voraussetzen.<sup>36</sup> So wünschenswert eine aktuelle Magnitude-

---

<sup>35</sup> Vergleicht man diese Werte mit den von Wegener berichteten Werten, dann erkennt man einen leichten Unterschied dahingehend, dass es sich bei der  $MPS_{KldB}$  um eine linkssteilere Verteilung handelt.

<sup>36</sup> Zwar gibt es Untersuchungen, welche die Stabilität des Prestiges über die Zeit unterstreichen (Roeder 1985), die zugrundegelegten Variablen, so zeigen Sozialstrukturanalysen, haben

Messung auch prinzipiell wäre, war dies im Rahmen unseres Projektes nicht zu realisieren.

Als Basis für den Umstieg zwischen der KldB75-70 und der KldB-92 dient der Umsteigeschlüssel des Statistischen Bundesamts (1992: 560ff.) auf der Ebene des dreistelligen Berufscodes, wobei allerdings nicht immer eine umkehrbar eindeutige (1:1) Zuordnung möglich war (vgl. Übersicht 3).

### Übersicht 3: Verteilung der ein- und mehrdeutigen Berufszuordnungen beim Umstieg von der KldB75-70 auf die KldB-92

Art der Zuordnung	Anzahl der Kategorien in	
	KldB75-70	KldB-92
1:1	182	182
1:n	33	85
n:1	65	28
n:n	48	63
Kategorien Gesamt	328	358

Wie aus Übersicht 3 hervorgeht, können 182 der insgesamt 328 Berufsordnungen<sup>37</sup> der KldB75-70 in eindeutiger Weise einer entsprechenden Berufsordnung der KldB92 zugeordnet werden. Hier wird der Prestigescore der  $MPS_{KldB75}$  direkt für die  $MPS_{KldB92}$  übernommen.

Bei 33 Berufsordnungen der KldB75-70 erfolgt eine 1:n-Zuordnung. Das heißt, eine spezifische Berufsordnung der KldB75-70 (z.B. 071: Bergleute) ist in der 92er-Version in verschiedene Berufsordnungen aufgesplittet (070: Bergleute ohne nähere Tätigkeitsangabe; 071: Bergleute (Bergtechnik)). In diesem Fall wurde der Prestigescore der 75er-Version (071: 35.3) gleichermaßen den gesplitteten Berufsordnungen der 92er-Version zugewiesen (070: 35.3; 071: 35.3).

---

sich seit den 80er-Jahren jedoch deutlich verändert. Eine erneute Berechnung würde daher zu einer neuen Skala führen.

<sup>37</sup> Wie oben erläutert umfasst die KldB75-70 insgesamt 333 Berufsordnungen. Die MPS-Berechnung umfasst jedoch nur 328 Berufsordnungen, da der Berufsbereich VI 'Sonstige Arbeitskräfte' mit den dreistelligen Codes '971' bis einschließlich '991' nicht berücksichtigt wurde.

Für 65 Berufsordnungen liegt eine n:1-Zuordnung vor. Hierbei werden mehrere Berufsordnungen der KldB75-70 (z.B. 021: Tierzüchter; 022: Fischer; 042: Melker) zu einer Berufsordnung (023: Tier-, Pferde-, Fischwirte) in der KldB92 zusammengefasst. In diesem Fall wurde der  $MPS_{KldB92}$ -Score (023: 41.2) aus dem (mit Bezug auf die Häufigkeit der einfließenden Berufe) gewichteten arithmetischen Mittel der entsprechenden  $MPS_{KldB75}$ -Scores gebildet (z.B. 021:49.1; 022:38.5; 042:32.7).

Schließlich verbleiben noch 48 Berufsordnungen, bei welchen eine n:n-Zuordnung vorliegt, das heißt mehrere Kategorien der 75er-Version (z.B. 163: Buchbinderberufe; 177: Druckerhelfer) werden mehreren Kategorien der 92er-Version (178: Buchbinder; 179: Druckereiverarbeitungshelfer) zugewiesen. Hier wurde zunächst aus den Prestige-Scores der 75er-Version das gewichtete arithmetische Mittel gebildet und dieses dann den jeweiligen Kategorien der KldB92 zugeordnet (178, 179: 39.5).<sup>38</sup>

Für 91 der insgesamt 358 Berufsordnungen<sup>39</sup> der KldB-92 erfolgte die Prestigezuweisung demnach über die Zuweisung eines gewichteten Mittels auf Basis der KldB75-70 Prestigescores. Bei 44 Berufsordnungen geht dies mit lediglich geringfügigen Abweichungen (0 bis 5 Prestigepunkten) zu den in den Mittelwert eingeflossenen  $MPS_{KldB75}$ -Scores einher. Weitere 17 Berufsordnungen weisen eine Abweichung zwischen 6 und 10 Prestigepunkten zu den 'Originalscores' auf. Bei immerhin 30 Berufsordnungen beträgt die Abweichung zu den 'Originalscores' allerdings mehr als 10 Prestigepunkte, so dass der neu zugewiesene Wert eine 'künstliche' Verbindung der früheren Kategorien darstellt.

Die resultierende  $MPS_{KldB92}$  hat einen Range von 186.0, einen Mittelwert von 57.7 und eine Standardabweichung von 29. Die höchste Skalenposition nehmen die Zahnärzte (Code 242) mit dem Wert 216.0 ein. Den niedrigsten Skalenwert mit 30.0 weisen mithelfende Landarbeitskräfte (Code 014) auf. Die Prestigewerte für alle 358 Berufs-

---

**38** Als Gewicht wurde wie schon oben die Häufigkeit des Auftretens der jeweiligen KldB75-70-Berufe herangezogen. Es konnte hierbei nicht berücksichtigt werden, zu welchen Anteilen die Einzelberufe in die verschiedenen KldB-Kategorien einfließen, da die entsprechenden Informationen nicht zur Verfügung stehen. Wenn also beispielsweise der Code 163 der Klassifikation von 1975/70 in die Codes 178 und 179 der Klassifikation aus dem Jahr 1992 eingeht, so kann nicht bestimmt werden, ob die Kategorie 163 zu gleichen Teilen in 178 und in 179 eingeht, oder ob beispielsweise lediglich 10 Prozent in 178 und 90 Prozent in 179 enthalten sind.

**39** Auch bei der KldB92 wird der Berufsbereich VI 'Sonstige Arbeitskräfte' mit den Berufsordnungen '971' bis einschließlich '997' ausgeklammert. Von den insgesamt 369 Berufsordnungen wird daher nur 358 Ausprägungen der KldB92 ein Prestigewert zugewiesen.

ordnungen der KldB92 stehen im Internet ([www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten\\_tools](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools)).<sup>40</sup>

## Fazit

In dieser Arbeit wurde ein Instrument zur Abbildung von Prestige für die Klassifizierung der Berufe vorgestellt. Die Vorgehensweise bei der Konstruktion wie auch der theoretische Hintergrund, der hier nicht ausführlich betrachtet wurde, orientiert sich an den Arbeiten von Wegener.

Es ist festzuhalten, dass eine umfassende Validierung der hier für die Klassifizierung der Berufe entwickelten Magnitude-Prestigeskalen noch aussteht. Erste Überprüfungen, bei welchen die MPS<sub>KldB</sub> am Kriterium der Originalskala von Wegener, wie sie im ALLBUS bereitgestellt wird, validiert wurde, finden sich bei Frietsch (2000). Dort finden sich auch Ausführungen zu Fragen der Übertragbarkeit des Prestiges auf Gesamtdeutschland sowie auf weibliche Erwerbstätige, die hier ebenso ausgeblendet wurden, wie die Frage nach der Anwendung auf Teilzeittätige.

Wenngleich die Validierung des vorgestellten Instruments noch nicht abgeschlossen ist, laden wir alle Mikrozensus-Nutzer ein, die MPS<sub>KldB</sub> schon jetzt bei ihren eigenen Analysen heranzuziehen. Letztlich muss sich jedes Instrument dieser Art immer wieder empirisch neu bewähren. Erst durch eine sinnvolle Verwendung und durch den erbrachten Erklärungsbeitrag in substantiellen Modellen, legitimiert sich eine solche Skala.

## Korrespondenzadresse

*Rainer Frietsch  
Fraunhofer Institut Systemtechnik und Innovationsforschung  
Abteilung Technikbewertung und Innovationsstrategien  
Breslauer Str. 48  
76 139 Karlsruhe*

---

<sup>40</sup> Die aus der Mittelwertsbildung hervorgegangenen KldB92 sind gekennzeichnet (\* für eine Abweichung zwischen 0 und 5 Prestigepunkte; \*\* für eine Abweichung zwischen 6 und 10 Prestigepunkten; \*\*\* für eine Abweichung von mehr als 10 Prestigepunkten).



## Literaturverzeichnis

Brauns, H./Steinmann, S./Haun, D., 2000: Die Konstruktion des Klassenschemas nach Erikson, Goldthorpe und Portocarero (EGP) am Beispiel nationaler Datenquellen aus Deutschland, Großbritannien und Frankreich. ZUMA-Nachrichten 46: 7-63.

Christoph, B., 2000: Aspekte der Berufsprestigemessung im vereinigten Deutschland. Theoretische Grundlagen, methodische Probleme und Aktualisierung der Magnitude-Prestigeskala. Diplomarbeit. Humboldt Universität Berlin. Philosophische Fakultät III.

Elias, P./Birch, M., 1994: Establishment of a Community-Wide Occupational Statistics. ISCO(88) COM. A Guide for User. Coventry:IER, University of Warwick.

Frietsch, R., 2000: Berufsprestige in Deutschland. Konstruktion und Validierung einer Prestigeskala. Diplomarbeit. Universität Mannheim. Lehrstuhl für Methoden der empirischen Sozialforschung und angewandte Soziologie.

Ganzeboom, H./De Graaf, P.; Treiman, D. J., 1992: A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21: 1-56.

Ganzeboom, H./Treiman, D. J., 1996: Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25: 201-239.

Geis, A. J./Hoffmeyer-Zlotnik, J., 2000: Stand der Berufsvercodung. ZUMA-Nachrichten 47: 89-114.

Hartmann, P./Schimpl-Neimanns, B., 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44: 315-340.

Lüttinger, P., (Hg.), 1999: Sozialstrukturanalysen mit dem Mikrozensus. ZUMA-Nachrichten SPEZIAL 6. Mannheim.

Lüttinger, P./Riede, T., 1997: Der Mikrozensus. Amtliche Daten für die Sozialforschung. ZUMA-Nachrichten 41: 19-43.

Mayer, K. U., 1979: Berufliche Tätigkeiten, berufliche Stellung und beruflicher Status. S. 79-123 in: Pappi, F. U. (Hg.): Sozialstrukturanalysen. Königstein/Ts: Athenäum.

Mueller, U., 1990: Kritik des Prestiges. *Soziologische Revue* 13, 133-140.

Müller, W./Noll, H.-H., 1997: Arbeit und Sozialstruktur. S. 553-571 in: Kahsnitz, D./Ropohl, G./Schmid, A., Handbuch zur Arbeitslehre. München: R.Oldenbourg Verlag.

Pappi, F. U., 1979: Der Beitrag der Umfrageforschung zur Sozialstrukturanalyse. S. 9-40 in: Pappi, F. U. (Hg.), Sozialstrukturanalysen. Königstein/Ts: Athenäum.

- Roeder, Burkhard, 1985: Über die Langzeitstabilität von Berufsprestigewerten. Zeitschrift für Experimentelle und Angewandte Psychologie, Band 32.
- Schönbach, K., 1979: Probleme der Verschlüsselung von Berufstätigkeiten. S. 71-78 in: Pappi, F. U. (Hg.), Sozialstrukturanalysen. Königstein/Ts: Athenäum.
- Sørensen, A. (1979): A model and a metric for the intragenerational status attainment process. American Journal of Sociology 85: 965-978.
- Statistisches Bundesamt, 1971: Internationale Standardklassifikation der Berufe. Stuttgart: Verlag W. Kohlhammer.
- Statistisches Bundesamt, 1975: Klassifizierung der Berufe 1975-70. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt, 1992: Klassifizierung der Berufe 1992. Stuttgart: Metzler-Poeschel.
- Treiman, D. J., 1979: Begriff und Messung des Berufsprestiges in der international vergleichenden Mobilitätsforschung. S. 124-167 in: Pappi, F. U. (Hg.), Sozialstrukturanalysen. Königstein/Ts: Athenäum.
- Wegener, B., 1978: Einstellungsmessung in Umfragen: Kategorische versus Magnitude-Skalen. ZUMA-Nachrichten 3: 3-27.
- Wegener, B., 1979: Magnitude-Messung beruflicher Einstellung. S. 125-184 in: Beck, U./Brater, M./Wegener, B., Berufswahl und Berufszuweisung. Zur sozialen Verwandtschaft von Ausbildungsberufen. Frankfurt/New York: Campus Verlag.
- Wegener, B., 1980: Magnitude-Messung in Umfragen: Kontexteffekte und Methode. ZUMA-Nachrichten 6: 4-40.
- Wegener, B., 1984: Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-Prestigeskala. ZUMA-Arbeitsbericht 84/02. Mannheim.
- Wegener, B., 1985: Gibt es Sozialprestige? Zeitschrift für Soziologie 14 (3): 209-235.
- Wegener, B., 1988: Kritik des Prestiges. Opladen: Westdeutscher Verlag.

## Anhang

### A1: Konstruktion der Variablen 'Bildung in Jahren'\* auf Basis der Mikrozensen 1982 und 1985

		Letzter allgemeinbildender Schulabschluss				
		Angabe fehlt	Hauptschule	Realschule	Fachhochschulreife	Abitur
Letzter beruflicher Ausbildungsabschluss	K.A.**	0	9	10	12	13
	Lehr-/Ablernausbildung	12	12	13	14	15
	Berufliches Praktikum	0	9	10	12	13
	Meister/Techniker	14	14	15	16	17
	Fachhochschule	16	16	16	16	17
	Hochschule	18	18	18	18	18

\* Die Zellbesetzungen stehen für die zugeordneten Bildungsjahre.

\*\* Kein Ausbildungsabschluß bzw. Angabe fehlt.

### A2: Konstruktion einer 'Schichtvariablen' auf Basis der Variablen 'Stellung im Betrieb' und Stellung im Beruf

Stellung im Betrieb	Stellung im Beruf				
	Selbst. ohne Beschäftigte	Selbst. mit Beschäftigte	Beamte, Richter, Soldat	Angestellte	Arbeiter
Selbst. bis 4 Besch.	-	2	2	2	1
Selbst. ab 5 Besch.	-	4	2	2	1
Bürokräfte, angel. Arb.	1	2	2	2	1
Verkäufer, Facharb.	1	2	2	2	1
Sach-, Vorarbeiter	1	2	2	2	1
Fachkraft, Meister	2	2	2	2	2
Referent, Hdlbevollm	3	3	3	3	2
Abtlgsleit., Prokurist	3	3	3	3	2
Direktor, Amtsleiter	4	4	4	4	2

Schichtestufung nach den von Wegener verwendeten Kategorien: Unterschicht (1); Mittelschicht (2); Obere Mittelschicht (3); Oberschicht (4).

## **ALLGEMEINE BEVÖLKERUNGSUMFRAGE DER SOZIALWISSENSCHAFTEN 2000: DATEN DES ALLBUS 2000 VERFÜGBAR**

Im Jahr 2000 wurde die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zum zwölften Mal durchgeführt. In persönlich-mündlichen Interviews wurden circa 2000 Personen in Westdeutschland (inklusive West-Berlin) und 1.100 Personen in Ostdeutschland (inklusive Ost-Berlin) befragt. Die Erhebung erfolgte erstmals computerunterstützt als CAPI-Interview. Wie bereits in den ALLBUS-Surveys 1994 und 1996 wurden für die Stichprobenbildung Adressen aus den Einwohnermelderegistern gezogen. Die Grundgesamtheit der Umfrage bildete die erwachsene Wohnbevölkerung in Privathaushalten. Ausländische Personen wurden allerdings nur dann befragt, wenn das Interview in deutscher Sprache durchgeführt werden konnte.

Das Fragenprogramm des ALLBUS 2000 ist thematisch sehr breit angelegt. Es bietet einen Querschnitt aus den Themen und Fragen des gesamten, nunmehr eine Laufzeit von 20 Jahren umfassenden ALLBUS-Programms. Neben den in jeder Erhebung standardmäßig erfassten demographischen Informationen wurden unter anderem Fragen repliziert zu den Bereichen:

- Regionale Mobilität: Immigration und Migration innerhalb Deutschlands; Bereitschaft zum Umzug in anderen Teil Deutschlands/in anderes Land der europäischen Union; Wohndauer am Wohnort; Entfernung zum vorherigen Wohnort.
- Ehe/Familie/Geschlechterrollen: Einstellungen zu Familie, Ehe, Erwerbstätigkeit von Frauen, Legalisierung des Schwangerschaftsabbruchs; Erziehungsziele; Alter beim Auszug aus Elternhaus; Zusammenleben vor Heirat; Ehebiographie; Arbeitsteilung im Haushalt.
- Arbeit/Beruf: Berufswerte; Furcht vor Arbeitslosigkeit beziehungsweise dem Verlust der beruflichen Existenz; Wochenarbeitsstunden; Dauer der Beschäftigung beim derzeitigen Arbeitgeber.
- Soziale Ungleichheit: Einstellungen zu sozialer Ungleichheit und zum Wohlfahrtsstaat; Aufstiegskriterien; Bewertung von Bildungsmöglichkeiten; Erfüllung persönlicher Lebensvorstellungen; subjektive Schichtestufung.

- Soziale Sicherung: Persönliche Alterssicherung; soziales Sicherungssystem einschränken beziehungsweise ausbauen.
- Subjektive Befindlichkeiten und Lageeinschätzungen: Wirtschaftliche Lage; gerechter Anteil; persönliche Lebensvorstellungen erfüllt; Anomie; Vertrauen in Mitmenschen.
- Gruppen/soziale Netzwerke Freundschaftsbeziehungen; Einstellungen und Kontakte zu Ausländern; Einstellungen zu den Zuzugsmöglichkeiten für verschiedene Immigrantengruppen.
- Politik: Politisches Interesse; Links-Rechts-Einstufung; Inglehart-Index; Zufriedenheit mit Demokratie; Stolz auf nationale Errungenschaften; Stolz, Deutscher zu sein; Parteiloyalität; politische Partizipation.
- Institutionen/Organisationen: Vertrauen in Institutionen; Vertrauen in europäische Institutionen; affektive Bindung an politische Gemeinwesen; Einstellungen zum Prozess der deutschen Vereinigung.
- Religion: Religiositätsskala; Taufe der Kinder; kirchliche Trauung; frühere Kirchenmitgliedschaft.
- Sanktion und abweichendes Verhalten: Moralische Bewertung und Wunsch nach gesetzlichem Verbot für verschiedene Verhaltensweisen; eigene Delinquenz und Entdeckungswahrscheinlichkeit für ausgewählte Delikte; Wunsch nach Strafe zusätzlich zu Wiedergutmachung; Opfer eines Diebstahls; Einstellung zur Befolgung von Gesetzen; abschreckende Wirkung von harten Strafen; Kriminalitätsfurcht.

Neben diesen ALLBUS-internen Replikationen wurden in der Erhebung 2000 neue Themenvorschläge aufgegriffen, die im bisherigen Fragenprogramm nicht oder in nicht so ausführlicher Form vertreten waren. Dabei ging es zum einen um das Thema „Kinderwunsch“ (Geschlecht und Geburtsjahr von Kindern; Wunsch nach weiteren Kindern; Geschlechtspräferenz und gewünschter Zeitpunkt für weitere Kinder), zum anderen um „Wirtschaftliche Befindlichkeiten und Perzeptionen“ (wirtschaftliche Lage im Vergleich zu vor ein paar Jahren; geschätzte Zahl von Arbeitslosen; erwartete Entwicklung in Zukunft). Um das umfangreiche Fragenprogramm im Rahmen der geplanten Befragungszeit von durchschnittlich 50 Minuten realisieren zu können, wurde ein Teil der Fragen lediglich bei der Hälfte der Befragten im Split erhoben.

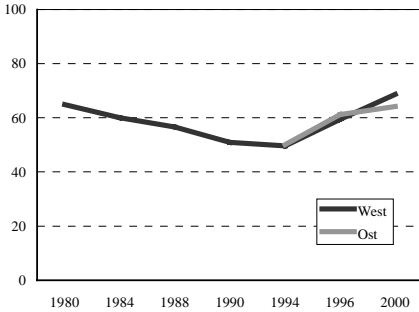
Die thematisch breite Anlage des ALLBUS2000 erlaubt es, einer Vielzahl von Zwischenbezüge zwischen einzelnen Fragen nachzugehen. So kann die Beziehung zwischen Institutionenvertrauen und Einstellungen zum Wohlfahrtsstaat untersucht werden oder auch die zwischen Religiosität, Postmaterialismus, Erziehungszielen und Be-

rufswerten. Obwohl die Fragen überwiegend dem bestehenden ALLBUS-Fragenprogramm entnommen sind, wurden viele erstmals gemeinsam in einer Studie gestellt, so dass sich teilweise gänzlich neue Analysemöglichkeiten eröffnen. Dies gilt beispielsweise für den Zusammenhang zwischen Einstellungen zur Rolle der Frau, Kinderwunsch und praktizierter Arbeitsteilung im Haushalt oder die Beziehung zwischen der affektiven Bindung an Deutschland, dem Stolz auf bestimmte nationale Errungenschaften und dem generellen Nationalstolz. Darüber hinaus bieten insbesondere die theoretisch aufeinander bezogenen Fragen des Themenbereichs „Sanktion und abweichendes Verhalten“ – eine gekürzte Replikation des 1990er ALLBUS-Schwerpunkts – die Möglichkeit zu Kausalanalysen.

Aber nicht nur im Querschnitt eröffnet der ALLBUS2000 vielfältige Analysemöglichkeiten. Da für fast alle Fragen mindestens zwei Messzeitpunkte im ALLBUS zur Verfügung stehen, ergeben sich auch für viele Themenbereiche Möglichkeiten, zeitvergleichende Analysen mit ALLBUS-Daten durchzuführen. So lässt sich beispielsweise die Entwicklung der Einstellungen gegenüber Ausländern mit Hilfe einer vier Items umfassenden Frage verfolgen, die seit dem ersten ALLBUS 1980 insgesamt sieben Mal Bestandteil des ALLBUS-Frageprogramms war. Sah es 1996 noch so aus, als gäbe es im Osten gegenüber 1994 einen teilweise besorgniserregend deutlichen Anstieg diskriminierender Einstellungen in der Bevölkerung und als wäre der im Westen seit 1980 zu beobachtende Trend hin zu ausländerfreundlicheren Einstellungen gestoppt, gilt es vor dem Hintergrund der 2000er Ergebnisse diese Interpretationen teilweise zu revidieren. In der ostdeutschen Bevölkerung sind die fremdenfeindlichen Einstellungen im Jahr 2000 fast wieder auf dem Niveau der ersten Messung von 1994, mit Ausnahme des relativ „weichen“ Items zur Lebensstilanpassung. Und auch im Westen ist die Trendwende jetzt nur noch bei zwei der vier Items zu beobachten, bei dem Wunsch nach Remigration der Ausländer bei Arbeitsplatzknappheit und – wie im Osten – bei der Forderung nach etwas stärkerer Anpassung der Ausländer an den deutschen Lebensstil. Überspitzt ausgedrückt: die Tendenz geht zwar im Großen und Ganzen in die Richtung, dass die Diskriminierung von Ausländern zwar zunehmend abgelehnt wird (wenn gleich die Zeitreihe für das Arbeitsmarktitem diesen Trend doch – teilweise – als eine Art „Schönwetterliberalismus“ zu entlarven scheint), aber die Forderung nach Assimilation an die „deutsche Leitkultur“ scheint eine gewisse Renaissance zu erleben (vergleiche Abbildung 1).

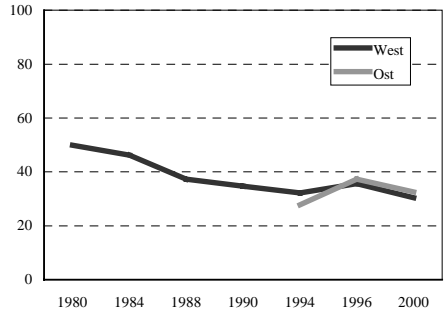
Abbildung 1: Meinungsklima zur Diskriminierung von Ausländern 1980 bis 2000<sup>1</sup>

Prozent Zustimmung



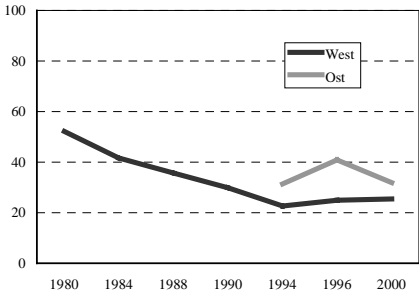
... sollten ihren Lebensstil ein bisschen besser an den der Deutschen anpassen

Prozent Zustimmung



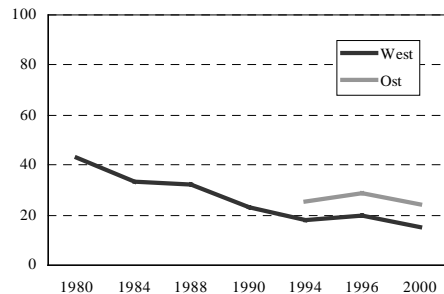
... sollte jede politische Betätigung in Deutschland untersagt werden

Prozent Zustimmung



... sollte man wieder in ihre Heimatländer zurückschicken, wenn Arbeitsplätze knapp werden

Prozent Zustimmung



... sollten sich ihre Ehepartner unter ihren eigenen Landsleuten auswählen

<sup>1</sup> Bis 1994 wurde statt „Ausländer“ der Begriff „Gastarbeiter“ in der Frage verwendet; 1994 wurden - jeweils im Split - beide Formulierungen verwendet und erwiesen sich als äquivalent; aufgeführt sind die Anteilswerte für die Skalenpunkte 5, 6 und 7 einer siebenstufigen Skala von 1 'stimme überhaupt nicht zu' bis 7 'stimme voll und ganz zu'.

Auch in vielen anderen Themenbereichen dürfte die zeitvergleichende Perspektive besonders interessant sein. So kann zum Beispiel das Zusammenwachsen des vereinten Deutschlands nun über einen beinahe 10-jährigen Zeitraum verfolgt werden, wobei solchen Analysen der disproportionale Stichprobenansatz des ALLBUS, bei dem die Ostdeutschen überrepräsentiert sind, zugute kommt. Und auch verschiedene Analysen sozialstruktureller Veränderungen sind dank der zahlreichen, zum Teil auch retrospektiven Faktenfragen zu Beruf, Wohnen, Familie, etc. möglich.

## ISSP-Erhebung für Deutschland

Im Rahmen der ALLBUS-Befragung 2000 wurde auch wieder die International Social Survey (ISSP)-Erhebung durchgeführt. Inzwischen hat das ISSP 37 Mitgliedsländer weltweit ([www.issp.org](http://www.issp.org)). Neu bei der gemeinsamen Durchführung von ALLBUS und ISSP beim ALLBUS 2000 war, dass zum ersten Mal *zwei* ISSP-Module - nämlich zum Thema „Soziale Ungleichheit“ aus dem Jahr 1999 und zum Thema „Umwelt“ aus dem Jahr 2000 – im Split bei jeweils der Hälfte der Befragten erhoben wurden. Beide Module waren als schriftlicher Selbstausfüllbogen konzipiert, den der Befragte im Anschluss an das mündliche ALLBUS-Interview bearbeiten sollte. Das Modul „Umwelt“ stellt eine Replikation aus dem Jahr 1993 dar; das Modul „Soziale Ungleichheit“ wurde zum größten Teil bereits 1987 und 1992 im Rahmen des ISSP-Programms erhoben. Das Fragenprogramm „Umwelt“ umfasst unter anderem Fragen zu folgenden Themen:

- Einstellungen zur Verbindung zwischen Natur und Religion
- Wissensfragen zu Umwelt und Wissenschaft
- Perzipierte Umweltgefährdungen
- Eigenes umweltrelevantes Verhalten
- Unterstützung von Aktionen/Organisationen

Das Modul „Soziale Ungleichheit“ enthält unter anderem Fragen zu folgenden Themen:

- Aufstiegskriterien
- Verdienst und Gerechtigkeit (Perzeptionen des Ist-Zustands und Vorstellungen des Soll-Zustands)
- Wahrgenommene Konflikte in der Gesellschaft
- Akzeptanz von Ungleichheit
- Gesellschaftsstruktur (Ist- und Soll-Vorstellungen)

Die Daten und das Codebuch des ALLBUS 2000 (ZA-Nr. 3451) sind auf CD-Rom beim Zentralarchiv für Empirische Sozialforschung (DR. MICHAEL TERWEY, Adresse



siehe unten) erhältlich und können dort gegen eine geringe Gebühr bezogen werden. Informationen zum ALLBUS 2000 erteilen MARTINA WASMER und ACHIM KOCH bei ZUMA; Ansprechpartnerin für das ISSP ist JANET HARKNESS (ZUMA).

Besuchen Sie auch die Web-Sites zum ALLBUS und ISSP. Sie finden dort Informationen zu aktuellen Entwicklungen bei beiden Umfrageprogrammen.

<http://www.gesis.org/Dauerbeobachtung/Allbus/index.htm>

<http://www.gesis.org/Datenservice/Allbus/index.htm>

<http://www.issp.org/>

## **Ansprechpartner**

### **Zentralarchiv für Empirische Sozialforschung in Köln**

Postfach 41 09 60

50 869 Köln

Fax: 0221/47694-44

Dr. Michael Terwey

Tel. 0221/47694-22

E-Mail: [terwey@za.uni-koeln.de](mailto:terwey@za.uni-koeln.de)

### **Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim**

Postfach 12 21 55

68 072 Mannheim

Fax: 0621/1246-100

Dr. Janet Harkness

Tel. 0621/1246-284

E-Mail: [harkness@zuma-mannheim.de](mailto:harkness@zuma-mannheim.de)

Achim Koch

Tel. 0621/1246-280

E-Mail: [koch@zuma-mannheim.de](mailto:koch@zuma-mannheim.de)

Martina Wasmer

Tel. 0621/1246-273

E-Mail: [wasmer@zuma-mannheim.de](mailto:wasmer@zuma-mannheim.de)

## **FORSCHUNGSPREIS DER UNIVERSITÄT FRANKFURT FÜR ZUMA-DISSERTATION**

**S**tefan Weick, Dipl. Soziologe am Zentrum für Umfragen, Methoden und Analysen wurde für seine Doktorarbeit zum Thema: „Relative Einkommensarmut bei Kindern. Untersuchungen zu Lebensbedingungen und Lebensqualität in Deutschland von 1984 bis 1996“ mit dem Forschungspreis 2001 der Josef-Popper-Nährpflicht-Stiftung an der Johann Wolfgang Goethe-Universität in Frankfurt am Main ausgezeichnet. Die Preisverleihung erfolgte im Rahmen einer Festveranstaltung am Freitag, dem 26.1.2001.

Der Josef-Popper-Forschungspreis wird alle zwei Jahre für herausragende Examensarbeiten, Dissertationen oder Habilitationsschriften, die sich mit der Weiterentwicklung der Popperschen Idee einer allgemeinen Grundsicherung wissenschaftlich fundiert auseinandersetzen, vergeben. Die Auswahlkommission, bestehend aus den Professoren Dr. Malcolm Dunn, Dr. Roland Eisen, Dr. Wolfgang Glatzer und Dr. Richard Hauser, hat in diesem Jahr – wie zuletzt 1998 - zwei Examensarbeiten und eine Dissertation für preiswürdig befunden.

### **Relative Einkommensarmut bei Kindern. Untersuchungen zu Lebensbedingungen und Lebensqualität in Deutschland von 1984 bis 1996.**

Die Diskussion um Kinderarmut in Deutschland basiert in erster Linie auf einer Zunahme des Anteils von Sozialhilfebeziehern in der Altersgruppe unter achtzehn Jahren. Diese Betrachtung gründet allerdings auf Daten der Sozialhilfestatistik, die staatliche Maßnahmen zur Armutsbekämpfung widerspiegeln. Die weitere empirische Grundlage zu diesem Themenfeld ist für Deutschland schmal. Einzelne Analysen mit Umfragedaten zur Verteilung von Haushaltseinkommen führen zu unterschiedlichen Ergebnissen hinsichtlich der Verbreitung von Kinderarmut. Die Dissertation erörtert vor diesem Hintergrund nationale und internationale Studien zur Einkommensarmut und liefert mit eigenen empirischen Analysen eine breitere empirische Basis für Aussagen zur Verbreitung, Entwicklung und Dauer von Einkommensarmut bei Minderjährigen sowie zu Lebensbedingungen von Kindern oberhalb und unterhalb der Armutsschwelle.

Der erste Teil der Arbeit stellt bisherige Untersuchungen zur Kinderarmut in den Zusammenhang mit Studien zur Sozialstruktur für diese Altersgruppe und der wissenschaftlichen Diskussion zur Armutsforschung. Ausgehend von der jüngeren Entwicklung der internationalen Forschungstätigkeit zur Lebensqualität von Kindern, gibt die Dissertation einen Überblick über neuere Aktivitäten zur Etablierung einer Sozialberichterstattung über Kinder in Deutschland. Anschließend wird die kontrovers geführte, öffentliche und wissenschaftliche Diskussion zur Armutsforschung dargestellt und erörtert.

Im zweiten Teil der Arbeit werden eigene Analysen zur Berechnungsweise, Verteilung, Entwicklung und dem zeitlichem Verlauf von relativer Einkommensarmut bei Kindern in den alten und neuen Bundesländern präsentiert und durch Auswertungen zu Lebensbedingungen und Lebensqualität im Haushaltskontext ergänzt. Im Bereich der Verbreitung und Entwicklung von Einkommensarmut von Kindern in Deutschland wurden bisher in der Literatur eher heterogene Einzelergebnisse präsentiert, die sich hinsichtlich der Armutskonzepte, Datenbasen und Operationalisierungen erheblich unterscheiden. In der Dissertation wird angesichts dieser Ausgangslage demonstriert, wie sich unterschiedliche Äquivalenzskalen und Abgrenzungen von Armutsschwellen auf Einkommenspositionen und Quoten relativer Einkommensarmut auswirken. Dabei wird die besondere Sensitivität der Teilpopulation von Kindern für unterschiedliche Berechnungsverfahren empirisch untermauert. Auf der Grundlage von drei ausgewählten Äquivalenzskalen, die sich bezüglich der Personengewichtung deutlich unterscheiden, wird die Entwicklung der Einkommensarmut von 1984 bis 1996 in differenzierter Form im Kontext weiterer Maße zur Einkommensungleichheit herausgearbeitet: Bei getrennter Berechnung für Ost- und Westdeutschland, kann in den alten Bundesländern kein anhaltender Trend zur weiteren Verbreitung relativer Einkommensarmut bei Kindern identifiziert werden, wie er auf Basis der Sozialhilfestatistik zu vermuten wäre. Im Vergleich zu anderen westlichen Industrieländern liegt die Kinderarmutsquote im unteren Mittelfeld. In den neuen Bundesländern hat sich mit der Zunahme der Einkommensungleichheit und der Erhöhung des Einkommensniveaus im Transformationsprozess auch der Bevölkerungsanteil unter der Armutsschwelle deutlich erhöht, insbesondere bei Kindern. Seit 1994 hat sich diese Tendenz allerdings abgeschwächt und 1996 ist erstmals ein Rückgang zu erkennen. Die Analysen werden durch die Untersuchung individueller Verläufe von Einkommensarmut und der relativen Einkommenspositionen über fünf Panelwellen ergänzt.

In der Dissertation werden nun in einem weiteren Kapitel die objektiven Lebensbedingungen von Kindern in Einkommensarmut mit Minderjährigen als Untersu-

chungseinheit vergleichend analysiert. Hierzu werden Strukturdaten der Bereiche Haushalt und Familie, Wohnen, Transferleistungen, Erwerbstätigkeit und Bildung im Trendverlauf für mehrere Zeitpunkte präsentiert. Das letzte Kapitel der Dissertation behandelt subjektives Wohlbefinden im Haushaltskontext von einkommensarmen Kindern. Einen Schwerpunkt des hier besprochenen Kapitels der Dissertation stellt in diesem Zusammenhang die Frage dar, ob Kinder in Abhängigkeit von Einkommensarmut mit erhöhten psychischen Belastungen in der Familie konfrontiert sind, was anhand der Angaben der Mütter untersucht wird und zu folgenden Ergebnissen führt: Kinder in einkommensschwachen Haushalten sind bei ihren Müttern häufig mit einem geringen subjektiven Wohlbefinden konfrontiert. Das äußert sich in spezifischen Sorgen und vermehrten Problemen der Mütter, sich mit den Verhältnissen zurechtzufinden. Mütter in einkommensschwachen Haushalten weisen unterdurchschnittliche Zufriedenheiten mit verschiedenen Lebensbereichen auf - bei erhöhter Heterogenität der Zufriedenheitsbewertungen. Eine zunehmende Adaption an die objektive Mangelsituation bei zunehmender Armutsdauer kann bei dieser spezifischen Zufriedenheitsbewertung somit nicht identifiziert werden.

# AKTIVITÄTEN IM FORSCHUNGSBEREICH ERNÄHRUNGSVERHALTEN

*ERIKA CLAUPEIN, KRISTINA NEBEL UND ULRICH OLTERS DORF<sup>1</sup>*

Das Institut für Ernährungsökonomie und -soziologie hat 1999 in Zusammenarbeit mit der Arbeitsgemeinschaft Ernährungsverhalten e.V. (AGEV) und mit Unterstützung von ZUMA eine Untersuchung über die Forschungsaktivitäten im Bereich des Ernährungsverhaltens durchgeführt. Über die Ergebnisse dieser Erhebung, die als Replikation einer Umfrage von 1977 angelegt war, gibt es einen Bericht, der hier zusammengefasst wird. Außerdem wurde eine Datenbank erstellt, die die erfassten Forschungsprojekte und Forscher des Ernährungsverhaltens enthält und online recherchierbar ist. Diese Ressourcen sind allgemein zugänglich unter der website [www.ernaehrungsverhaltensforschung.de](http://www.ernaehrungsverhaltensforschung.de).

## 1. Ziel und Begründung der Untersuchung

Gesundheit und Wohlbefinden der Bevölkerung hängen eng mit der Ernährung zusammen. Wie in allen westlichen Industrieländern werden auch in Deutschland ernährungsbedingte Problemlagen hauptsächlich durch problematisches Ernährungsverhalten verursacht. Insofern nimmt die Verhaltensforschung eine wichtige Schlüsselposition im Rahmen der Ernährungsforschung ein. Es ist daher für Forscher und Politiker wichtig und für Ausbildungs- und Beratungskräfte sicher interessant zu erfahren, welche Institutionen in welchem Bereich der Ernährungsverhaltensforschung tätig sind. So wurde bereits 1977 von der Bundesforschungsanstalt für Ernährung eine „Untersuchung über Aktivitäten in der Bundesrepublik Deutschland im Bereich der Erforschung des Ernährungsverhaltens der Bevölkerung“ vorgenommen. Im Sommer 1999 wurde diese Untersuchung mit genau demselben Forschungsdesign wiederholt, um nicht nur Aussagen zur Situation sondern auch zur Entwicklung der Ernährungsverhaltensforschung machen zu können, und dies so-

---

<sup>1</sup> Institut für Ernährungsökonomie und -soziologie der Bundesforschungsanstalt für Ernährung, Haid- und Neustr. 9, 76 131 Karlsruhe ([www.bfa-ernaehrung.de](http://www.bfa-ernaehrung.de)).

wohl im Hinblick auf die Präsenz bestimmter Themen und Gebiete als auch auf feststellbare Forschungslücken.

## 2. Methodik und Durchführung

Bereits 1977 basierte der Aufbau des Fragebogens auf der Erkenntnis, dass Ernährungsverhalten zwar stark von biologischen und physiologischen Bedürfnissen des Menschen beeinflusst wird, dass aber soziologische, psychologische, sozioökonomische sowie kulturelle Determinanten ebenso eine wichtige Rolle spielen. In unserer heutigen Wohlstands- und Erlebnisgesellschaft, in der die individuelle und vielgestaltige Lebensführung so wichtig geworden ist, dürften diese Faktoren an Bedeutung gewonnen haben.

Was, wann, wo, wieviel und mit wem gegessen wird ist Bestandteil des Ernährungs- und Lebensstils und bestimmt somit das Sein und Selbstbewußtsein der Menschen ganz entscheidend mit. Ernährungsverhalten als Teil des Konsumverhaltens differenziert sich jenseits der herkömmlichen sozioökonomischen Merkmale immer weiter aus. Ernährung birgt wie Mode und Freizeitgestaltung ein großes und bedeutsames Inszenierungspotential.

Um einen Überblick über Stand und Entwicklung der Ernährungsverhaltensforschung zu bekommen, haben wir uns jedoch entschlossen, auf den relativ einfachen Fragebogen von 1977 in seiner Grobstruktur zurückzugreifen und auch in der Darstellung die Gliederung in die vier Hauptbereiche Ökonomie, sozialkulturelle und psychologische sowie medizinische und physiologische Determinanten und Verbraucherarbeit zu übernehmen. Die Untergruppen wurden, dem Trend zur Spezialisierung folgend, entsprechend ausdifferenziert und ergänzt.

1. Ökonomie
  - 1.1 Ernährungs- und Kostformen
  - 1.2 Verzehrsmengen
  - 1.3 Absatz- und Verbraucherprobleme des Lebensmittelmarktes
  - 1.4 Globale Ernährungssituation und Entwicklungsländerproblematik
2. Sozialkulturelle und psychologische Determinanten
  - 2.1. Soziale und kulturelle Determinanten (allgemein)
    - 2.1.1 Kinder und Jugendliche
    - 2.1.2 Ältere Menschen
    - 2.1.3 Spezielle Gruppen (Berufsgruppen, Häftlinge)
    - 2.1.4 Obdachlose, Arme
    - 2.1.5 Ethnien

- 2.2 Psychologische Determinanten
- 2.3 Methodik
3. Medizinische und physiologische Determinanten
- 3.1. Gesunderhaltung, Leistungserhaltung, Leistungssteigerung
- 3.2 Ernährungsepidemiologie, Public Health
- 3.3 Ernährungskrankheiten: Ursachen, Therapien
4. Verbraucherarbeit (Aufklärung, Bildung, Information, Beratung, Erziehung)
5. Sonstiges

Ebenso wie 1977 wurde nach Projekten im Bereich der Ernährungsverhaltensforschung gefragt, deren Ausarbeitung und Planung nicht länger als fünf Jahre zurückbeziehungsweise vorauslag.

Die Adressen der potentiell in Frage kommenden Institute wurden weitgehend in unserem Hause ermittelt. Ziel war es, ein möglichst breites Spektrum des Forschungsbereichs „Ernährungsverhalten“ zu erfassen, um auch Randbereiche einbeziehen zu können. Daher hat sich die Zahl der angeschriebenen Einrichtungen deutlich von 142 auf 624 erhöht. Für den Bereich der Marktforschung haben wir auf den Adressenkatalog von ESOMAR (European Society for Opinion and Marketing Research, Amsterdam) zurückgegriffen.

Insgesamt wurden knapp 1300 Institutionen angeschrieben, 959 haben geantwortet. Hiervon haben 273 Institutionen ein oder mehrere laufende und/oder abgeschlossene Projekte benannt und beschrieben, 624 Aktivitäten wurden insgesamt gemeldet. Die Rücklaufquote lag insgesamt bei 75 %.

Erwartungsgemäß konzentrieren sich die Forschungsaktivitäten zum Ernährungsverhalten auf den Bereich der Hochschulen: 182 Institutionen von Hochschulen haben mindestens ein Projekt gemeldet. Diese Institute sind in einem breiten Spektrum unterschiedlicher Provenienzen beheimatet, die Mehrheitlich mit ihrer Fachrichtung multi- bzw. interdisziplinär orientiert sind, beispielsweise sind dies haus- und ernährungswissenschaftliche und geographische Institute (76 gemeldete Aktivitäten). Es folgen eher sozialwissenschaftlich ausgerichtete Institute mit 42 gemeldeten Aktivitäten und medizinisch ausgerichtete Institute mit 33 Nennungen.

Mit weitem Abstand in der Zahl der benannten Forschungsaktivitäten folgen Marktforschungsinstitute und andere Forschungseinrichtungen. Allerdings ist ein Rückschluss auf die tatsächliche Zahl ihrer Aktivitäten kaum möglich, denn gerade die Marktforschungsinstitute waren in ihrer Auskunftsfähigkeit und –willigkeit hinsichtlich Zahl und Art ihrer einschlägigen Forschungsaktivitäten sehr zurückhaltend, da

sie weitgehend private Auftragsforschung durchführen, deren Inhalt und Ergebnisse nicht für die breite Öffentlichkeit bestimmt ist. Insofern war ihre Rückmeldung sehr unterschiedlich, einige gaben lediglich an, dass sie in diesem Bereich tätig sind, andere führten eine große Zahl von Projekten an, ohne diese jedoch näher zu charakterisieren. Unsere Zahlen sind daher nur als grober Annäherungswert zu verstehen.

### **3. Auswertung der Umfrage**

Allen Schwierigkeiten bei der Zählung und Zuordnung zum Trotz hat die Zuordnung der gemeldeten Aktivitäten zu den verschiedenen Forschungsrichtungen ergeben, dass die Mehrzahl der Aktivitäten (39%) soziokulturelle und psychische Fragestellungen aufweist. Mit jeweils 26% folgen ökonomische und medizinische Themen. Der Bereich der Verbraucherarbeit ist mit lediglich 8% und sonstige Aktivitäten mit 1% vertreten. Im Folgenden werden die Ergebnisse für die einzelnen Forschungsbereiche kurz zusammengefasst.

#### **3.1 Bereich: Ökonomie**

Dem Forschungsbereich Ökonomie sind 26% der gemeldeten Aktivitäten zugeordnet worden. Der Schwerpunkt liegt mit 15% bei Forschungsaktivitäten zu „Absatz- und Verbraucherproblemen des Lebensmittelhandels“. Hier sind insbesondere die Mehrzahl der Marktforschungsinstitute, aber auch einige Hochschulinstitute tätig. Der Anteil der Forschungsprojekte zu „Ernährungs- und Kostformen“ beträgt 8%. Insgesamt 11 (2%) Studien zu „Verzehrmengen“ und vier Studien zur „globalen Ernährungssituation und Entwicklungsländer“ sind genannt worden.

#### **3.2 Bereich: Soziokulturelle und psychische Determinanten des Ernährungsverhaltens**

Die Erforschung des Ernährungsverhaltens in seinen soziokulturellen und psychischen Dimensionen bildete in der Befragung von 1999 mit knapp 40% den Schwerpunkt aller gemeldeten Forschungsaktivitäten, 1977 waren noch die ökonomischen Dimensionen vorherrschend. Neben der eher allgemein orientierten Studien zur Erforschung sozialer und kultureller Determinanten (16%), waren 7% auf Kinder und Jugendliche, 5% auf Ethnien und 3% auf ältere Menschen bezogen.

#### **3.3 Bereich: Medizinische und physiologische Determinanten**

Die 164 Studien mit medizinisch-physiologischem Fokus haben einen Anteil von 26%, vor 22 Jahren lag der Anteil noch bei 12%. Dieser Anstieg um mehr als das Doppelte mag an der Zunahme ernährungsbedingter Krankheiten liegen. Die Zah-



len scheinen diese These zu betätigen, denn immerhin 14% der gemeldeten Aktivitäten beziehen sich auf „Ernährungskrankheiten: Ursachen und Therapien“. Mit 7% folgen Studien zum neu eingeführten Bereich „Ernährungsepidemiologie/Public Health“ und mit 5% Studien zur „Gesunderhaltung; Leistungserhaltung und –steigerung“.

### **3.4 Bereich: Verbraucherarbeit**

Der Bereich der Verbraucherarbeit umfasst die klassischen Instrumente Beratung, Information, Erziehung und Aufklärung. Der Anteil der gemeldeten Aktivitäten liegt heute – wie vor 22 Jahren – bei 8% und ist damit nach wie vor der Bereich, in dem die wenigsten Aktivitäten gemeldet wurden. Hier wirken sich sicherlich u.a. die Mittelkürzungen im Bereich der Verbraucherpolitik und im Bereich der Gesundheitsprävention negativ aus. Gemeldet wurden sowohl methodisch als auch inhaltlich orientierte Projekte. Methodische Fragestellungen waren beispielsweise die Qualitätssicherung der Ernährungsberatung, die Entwicklung von Modellen zur Informationsverarbeitung oder von Schulungskonzepten. Inhaltlich ausgerichtete Arbeiten des Bereichs Verbraucheraufklärung konzentrieren sich hauptsächlich auf aktuelle Themen wie Lebensmittelskandale und Gentechnik, im Bereich der Erziehung die Ernährungserziehung von Kindern, im Bereich Weiterbildung Schulungsprogramme bei bestimmten ernährungsbedingten Krankheiten.

## **4. Fazit und Ausblick**

Obwohl versucht wurde an die erste Erhebung dieser Art aus dem Jahr 1977 anzuknüpfen, kann keine eindeutige Antwort darauf gegeben werden, wie sich die Aktivitäten im Forschungsbereich Ernährungsverhalten in bezug auf heute (1999) entwickelt haben. Das „Aktivitätsgebiet“ ist 1999 größer, nicht nur dadurch dass Deutschland wieder vereinigt ist, sondern dass Ernährungsverhalten in mehr Fachdisziplinen wahrgenommen wird. Vordergründig betrachtet gibt es heute deutlich mehr Aktivitäten. Bezogen auf die Ernährungsproblematik in unserer Gesellschaft fällt jedoch auf, dass wohl die Beachtung der Verhaltensaspekte von ernährungsabhängigen Erkrankungen zugenommen hat; auch die Untersuchung von psychosozialen, aber auch medizinisch-physiologischer Determinanten, dass aber die Aspekte der Prävention bzw. die Determinanten des „normalen“ Ernährungsverhaltens und der Verbraucherarbeit bestenfalls stagnieren.

Das weite Feld der Ernährungsverhaltensforschung wird an erfreulich vielen Stellen bearbeitet. Die Schwachpunkte oder „Forschungslücken“ sind aber ebenso sichtbar.

Wir wissen, dass viel methodische Grundlagenforschung notwendig wäre, diese ist nur marginal vertreten. Ebenso sind die entsprechend ihrer Bedeutung für die Erreichung von Ernährungszielen einer Gesellschaft notwendigen Bereiche, wie „Public Health Nutrition“ und Evaluierung unterrepräsentiert. Die Hinweise könnten in viele Richtungen erweitert werden, wie mangelnde Beachtung der Ernährungsverhaltensprobleme bei solchen Gruppen wie ältere Menschen, Behinderte und sozial Benachteiligte. Auffallend ist auch die „Vereinzelung“ der vielen Perlen der Ernährungsverhaltensforschung.

# VERBRAUCHERPANELDATEN FÜR DIE EMPIRISCHE SOZIAL- UND WIRTSCHAFTSFORSCHUNG

*Neuer ZUMA-Service:*

*Einzeldaten von Verbraucherpanels für wissenschaftliche Zwecke*

Erstmals stehen Einzeldaten eines kommerziellen Verbraucherpanels, (ConsumerScan Haushaltspanel der GfK, Nürnberg) für die sozial- und wirtschaftswissenschaftliche Forschung standardmäßig zur Verfügung. Die Daten können nach Einsendung einer Einverständniserklärung zur wissenschaftlichen Nutzung und gegen eine geringe Gebühr bei ZUMA direkt angefordert werden.

Bislang gab es für die empirische Sozial- und Wirtschaftsforschung keinen einfachen Weg zur wissenschaftlichen Analyse des immensen Forschungspotentials von Verbraucherpaneldaten, die von Marktforschungsinstituten mit einem hohen Aufwand an Pflege und Investitionen zur Produktion von Marktinformationen erhoben werden. Der überaus große Aufbereitungsaufwand machte nur punktuelle Weitergaben an die Forschung möglich. Insbesondere für eine breite, empirische Analyse des Konsumverhaltens in einem longitudinalen und verhaltensorientierten Ansatz schien es in Deutschland keine Perspektive zu geben.

ZUMA hat nun diese gravierende Versorgungslücke der empirischen Sozial- und Wirtschaftsforschung mit der Entwicklung eines neuen GESIS-Services geschlossen. Nach intensiven Vorbereitungen und Verhandlungen mit der Gesellschaft für Konsumforschung (GfK, Nürnberg), hat ZUMA die Rahmenbedingungen für einen neuen wissenschaftlichen Service mit Verbraucherpaneldaten aufgebaut. Ab sofort kann ein aufbereiteter Datensatz von Verbraucherpaneldaten der GfK aus dem Jahr 1995 direkt von ZUMA bezogen werden.

Das besondere Analysepotential dieser ZUMA-Verbraucherpaneldaten besteht darin, dass detaillierte Einzeldaten über die Produktkäufe privater Haushalte (80 Warengruppen mit Marktuntergruppen) als in-time erfasste event-history-Daten vorliegen.

Außerdem sind aus mehreren Panelwellen sowohl Daten der sozio-ökonomischen und demographischen Haushaltsstruktur, als auch die Einstellungen einer Referenzperson des Haushaltes verfügbar. Und schließlich sind diese beiden Datentypen über eine Haushaltsidentifikationsnummer miteinander verknüpfbar, so dass nun ein einzigartiges, sozial- und wirtschaftswissenschaftliches Forschungspotential zur Analyse des Konsums privater Haushalte bereitsteht.

Die ZUMA-Verbraucherpaneldaten basieren auf dem ConsumerScan Haushaltspanel der GfK, das 12000 Haushalte umfasst und als Quotenstichprobe angelegt ist. Der ZUMA-Datensatz ist eine Unterstichprobe der GfK-Daten und besteht aus jenen 9064 Haushalten, die im Jahr 1995 durchgängig ein Haushaltsbuch geführt haben.

Die ZUMA-Verbraucherpaneldaten, die als scientific use file für die Forschung zur Verfügung gestellt werden, umfassen zwei Datenarten: a) Ereignis-Daten über die Produktkäufe der Haushalte im Jahr 1995, b) Querschnitts-Daten über den sozio-ökonomischen und demographischen Status der Haushalte jeweils zu Anfang der Jahre 1994, 1995 und 1996, sowie über die Einstellungen der haushaltsführenden Person zu Ernährung, Umwelt und Konsum im Oktober 1994 und Oktober 1995. Die beiden Datenarten können mittels einer Haushaltsidentifikationsnummer auf Haushaltsebene verknüpft werden. Im Folgenden werden die je Datenart verfügbaren Informationen näher beschrieben.

#### **a) Einzeldaten über Produktkäufe des Haushaltes 1995**

Kern des Erhebungsprogramms des GfK ConsumerScan Haushaltspanels ist die möglichst fortlaufende Erfassung der Käufe eines Haushaltes bzw. von Personen mit genauen Informationen über Zeitpunkt des Kaufs und Art der gekauften Produkte. Die am Panel teilnehmenden Haushalte tragen ihre Käufe in ein Haushaltsbuch ein, das sie in regelmäßigen Abständen (wöchentlich) zum Marktforschungsinstitut schicken. Der Schwerpunkt des GfK ConsumerScan Haushaltspanels liegt auf sog. fast moving consumer goods, also insbesondere verpackten Lebensmitteln und Produkten der Haushaltsführung.

Aus methodischer Sicht handelt es sich hierbei um Ereignisdaten, mit denen in kontinuierlicher Zeit diskrete und kontinuierliche Informationen über den Wareneingangsprozess gemessen werden. Bei diesen Daten stellt das gekaufte Produkt die Beobachtungseinheit dar. Es handelt sich somit um eine asymmetrische Datensatzstruktur, da je Haushalt i.d.R. unterschiedlich viele Produktkäufe vorliegen, deren Zuordnung zu einem spezifischen Haushalt über eine Haushaltsidentifikationsnummer erfolgt.

Je Produktkauf liegen folgende Merkmale vor:

- Datum des Einkaufs,
- Wochentag des Einkaufs,
- Warengruppe des Produktes (insgesamt werden im Verbraucherpanel 81 Warengruppen erfasst),
- Teilmarkt des Produktes (dies entspricht i.d.R. der Geschmacksrichtung),
- Einkaufsstätte,
- Anzahl der gekauften Produkte,
- Normal-/Sonderpreis,
- Gesamtmenge,
- Gesamtausgaben für den Kauf,
- Zeitdauer seit letztem Kauf,
- Sortiernummer des Einkaufs (diese gibt an, um den wievielten Kauf des jeweiligen Produktes seit Beginn des Jahres es sich bei dem jeweiligen Kauf handelt),
- sowie warengruppenspezifische Attribute, wie z.B. Verpackungsart, Sorte, Konzentration usw.

Neben diesen kontinuierlich erfassten Informationen werden pro Haushalt zu bestimmten Zeitpunkten sozio-ökonomische und einstellungsbezogene Daten erfasst.

#### **b) Einzeldaten über die sozio-ökonomische und demographische Lage des Haushaltes und über Einstellungen der haushaltsführenden Person**

Die sozio-ökonomische und demographische Lage des Haushaltes wird mittels einer schriftlichen Befragung jeweils zum Anfang eines Jahres erfasst. Im ZUMA-Datensatz sind aus drei Wellen (1994, 1995 und 1996) folgende Haushaltsmerkmale verfügbar:

- Bundesland,
- Ortsgröße,
- Alter der haushaltsführenden Person,
- Anzahl der Kinder (bis 6, unter 14, unter 18 Jahre),
- Berufstätigkeit der HH-führenden Person,
- z. Zt. ausgeübter Beruf des Hauptverdieners,
- früherer Beruf des Hauptverdieners,
- Schulbildung des Hauptverdieners,
- Hausgröße,

- Ausstattung des Haushaltes (Waschmaschine, Geschirrspülmaschine, Mikrowelle, Videorekorder etc.), Gartenbesitz/-nutzung, Gartengröße, Haustiere (Katze, Hund, Wellensittich etc.).

Die Daten der drei Panelwellen sind mittels der Haushaltsidentifikationsnummer miteinander verknüpfbar und erlauben es, auch haushaltsspezifische Veränderungen der sozialen Lage des Haushaltes zu erfassen.

Einstellungen zu Ernährungs- und Umweltfragen werden im Verbraucherpanel der GfK durch Einschaltung von postalischen Meinungsumfragen erfasst. Da i.d.R. die haushaltsführende Person für die Einkäufe eine Art gate-keeper darstellt, ist sie es, deren konsumrelevante Einstellungen erhoben werden. Im ZUMA-Datensatz befinden sich die Einstellungsdaten von zwei eingeschalteten Umfragen, und zwar vom Oktober 1994 und vom Oktober 1995. Diese Datensätze befinden sich im gleichen Datensatz wie die sozio-ökonomischen Daten und sind ebenso über die Haushaltsidentifikationsnummer mit den Kaufdaten verknüpfbar.

Die Einstellungsdaten umfassen folgende Items:

- *Ansichten zur Ernährung:* Hierbei handelt es sich um 41 Items, in denen die Präferenzen beim Nahrungsmittelkauf erfasst werden (z.B. ob medizinisch gesunde oder naturbelassene Produkte bevorzugt werden, in welcher Weise Kriterien wie Entdeckerfreude, Herkunft der Produkte, Vollwertkost Frische, Hausmannskost, Markenartikel, Vitamine/Mineralstoffe oder Zubereitungskomfort etc eine Rolle spielen).
- *Ansichten zu Dingen des täglichen Lebens:* Hierzu gehören 20 Items, mit denen das Ausmaß des Interesses an neuen Produkten, der traditionellen Lebensführung, der Erlebnisorientierung, der Qualität etc. erhoben wird.
- *Ansichten zum Thema Umwelt:* Mit 13 Items werden Umweltbewusstsein, Umweltsverhalten und die Einstellung zum Verhältnis von Umwelt, Staat und Industrie erfasst.
- *Preisbewusstsein der haushaltsführenden Person,* summativer Index aus den Werten von drei Itempaaren wie z.B. „Ich suche beim Einkauf gezielt Sonderangebote, weil ich dadurch doch ganz erheblich sparen kann – Es ist mir zu umständlich, beim Einkauf nach Sonderangeboten zu suchen, das lohnt letztlich den Aufwand nicht“.

Eine interdisziplinäre Arbeitsgruppe hat sich mit dem Forschungspotential dieser Daten auseinandergesetzt und die Resultate 1999 in einem Symposium diskutiert. Die daraus entstandenen Analysen sind im Sonderheft 7 der ZUMA-Nachrichten veröffentlicht<sup>1</sup>. Neben einem detaillierten Einblick in die Praxis und das Datenerhebungsprogramm von Verbraucherpanels, wie sie z.B. bei der Marktforschung der GfK unterhalten werden, enthält der Band Untersuchungen zu Fragen der Flexibilität von Preisbildungsvorgängen, des Lebensstils im alltäglichen Konsum, der Gesundheitsorientierung im Konsumverhalten, der Umweltorientierung und ihrer Umsetzung im Kauf alltäglicher Haushaltungsprodukte, der Gewohnheitsbildung und Risikorientierung bei Kaufentscheidungen sowie der sozio-ökonomischen Einbettung des Kaufs spezieller Produkte, wie z.B. alkoholischer Getränke. Diese Untersuchungen bieten zahlreiche Anknüpfungspunkte für grundlagen- und anwendungsorientierte Fragen der Konsumforschung mit Verbraucherpaneldaten.

Die ZUMA-Verbraucherpaneldaten 1995 können mit untenstehendem Bestellformular direkt bei ZUMA angefordert werden, und zwar mit Einsendung der untenstehenden Einverständniserklärung zur wissenschaftlichen Nutzung der Daten. Die Gebühr für die Nutzungsrechte an den Daten beträgt 50.- Euro. Im Rahmen einer bis 31.12.2001 befristeten Subskription gibt es die Möglichkeit, zum Preis von Euro 50.- Daten und Sonderheft 7 der ZUMA-Nachrichten gemeinsam zu erwerben. Regulär kostet das Sonderheft 7 der ZUMA-Nachrichten 15.- Euro und enthält eine CD mit Erhebungsdokumentationen und Code-Büchern, aber ohne die ZUMA-Verbraucherpaneldaten 1995.

---

<sup>1</sup> Georgios Papastefanou, Peter Schmidt, Axel Börsch-Supan, Hartmut Lütke, Ulrich Oltersdorf (eds.): *Social and Economic Analyses of Consumer Panel Data*, ZUMA, Mannheim 2001, ISBN 3-924220-21-2, Preis: 25 DM (15 Euro). Bestellung direkt bei ZUMA.

## EINVERSTÄNDNISERKLÄRUNG

Der Datenempfänger verpflichtet sich durch Unterschrift, nach Erhalt der ZUMA-Verbraucherpaneldaten 1995 zur Einhaltung folgender Bedingungen:

- 1 Der Datenempfänger und alle Personen, denen die Daten zugänglich gemacht werden, verpflichten sich, keine De-Anonymisierungsversuche zu unternehmen und die Daten nicht an andere Personen - außer den Mitarbeitern des Datenempfängers - weiterzugeben oder sie ihnen zugänglich zu machen. Dies gilt auch für modifizierte Daten.
- 2 Die Daten dürfen nur in der eigenen wissenschaftlichen Forschung und der universitären Lehre des Datenempfängers eingesetzt werden. Eine Nutzung für gewerbliche Zwecke ist nicht gestattet.
- 3 Für die Überlassung des Nutzungsrechts an den Daten wird eine Gebühr bezahlt. Sie beträgt 50.- Euro (incl. MWST). Im Rahmen einer bis zum 31.12.2001 befristeten Subskription kann mit den Daten auch das Sonderheft 7 der ZUMA-Nachrichten erworben werden. Der Betrag wird mit meinem Einverständnis von meinem Konto abgebucht.
- 4 In die Publikationen, in die Ergebnisse der ZUMA-Verbraucherpaneldaten 1995 eingehen, wird folgender Quellenverweis aufgenommen:

*Die in diesem Beitrag benutzten Daten entstammen dem Datensatz, den die GfK, Nürnberg als Unterstichprobe des ConsumerScan Haushaltspanels von 1995 ZUMA zur Verfügung gestellt hat. Dieser ZUMA-Datensatz enthält alle Haushalte, für die im Jahr 1995 durchgehend Kaufdaten gesammelt worden sind. Für eine nähere Beschreibung der Verbraucherpaneldaten siehe: Papastefanou, Georgios: The ZUMA scientific use file of the GfK ConsumerScan Household Panel 1995, in: Papastefanou, Georgios/Schmidt, Peter/Börsch-Supan, Axel/Lüdtke, Hartmut/Oltersdorf, Ulrich (eds.), 2001, Social and Economic Analyses of Consumer Panel Data, Zentrum für Umfragen, Meinungen und Analysen (ZUMA), Mannheim.*

Mit meiner Unterschrift verpflichte ich mich die oben aufgeführten Bedingungen einzuhalten.

Adresse des Datenempfängers

Ort, Datum

Unterschrift



# ZUMA-VERBRAUCHERPANELDATEN 1995

## Bestellformular

Hiermit bestelle ich

- ..... Stück der Daten-CD „ZUMA-Verbraucherpaneldaten 1995“ inklusive ZUMA-Spezial Band 7 „Social and Economic Research with Consumer Panel Data“, zum Subskriptionspreis von € 50 (einschl. 7% Mehrwertsteuer) (gültig bis 31.12.2001)

..... Stück des ZUMA-Spezial Band 7 „Social and Economic Research with Consumer Panel Data“, zum Preis von € 15 (einschl. 7% Mehrwertsteuer)

<b>Gewünschte Zahlungsart:</b>	<input type="checkbox"/> Bankeinzug	<input type="checkbox"/> Kreditkarte
<b>Lieferanschrift:</b>		
Titel		
Vorname		
Nachname		
Institution		
Abteilung		
Straße und Hausnummer bzw. Postfach		
Postleitzahl		
Stadt		
Land		

Wie können wir Sie bei Rückfragen erreichen?	
E-Mail	
Phone	
Fax	

Bitte geben Sie die gewünschte Stückzahl des entsprechenden Angebots an.

Bei Bestellungen aus dem Ausland erhöht sich der Preis um € 5 für Porto und Versandkosten. Die Zahlung erfolgt per Bankeinzug (nur innerhalb Deutschlands) oder Kreditkarte (VISA oder MasterCard). Die Rechnung erhalten Sie zusammen mit der CD-ROM, die Ihnen per Post zugeschickt wird.

Bitte kreuzen Sie die von Ihnen gewünschte Zahlungsart an und füllen Sie die entsprechenden Felder am Ende des Formulars aus:

### Bezahlung per Kreditkarte

Bitte füllen Sie den folgenden Abschnitt nur aus, wenn Sie per Kreditkarte bezahlen wollen. Beachten Sie bitte, dass wir bei Bestellungen aus dem Ausland keinen anderen Zahlungsmodus akzeptieren.

Hiermit ermächtige ich Sie zur Abbuchung des Betrags von	<input type="checkbox"/> € 50	<input type="checkbox"/> € 15
Von der folgenden Kreditkarte	<input type="checkbox"/> VISA	<input type="checkbox"/> MasterCard
Kreditkartennummer		
Karteninhaber		
Gültig bis Ende		

### Bezahlung per Bankeinzug

Bitte füllen Sie den folgenden Abschnitt nur aus, wenn Sie per Bankeinzug bezahlen wollen. Bitte beachten Sie, dass es bei Bestellungen aus dem Ausland nicht möglich ist, per Bankeinzug zu bezahlen.

Hiermit ermächtige ich Sie, folgenden Betrag zu Lasten meines Kontos durch Lastschrift einzuziehen	<input type="checkbox"/> € 50
	<input type="checkbox"/> € 15
Kontonummer	
Bank	
Bankleitzahl	

Datum, Unterschrift

## ZUR ERINNERUNG AN AAGE B. SØRENSEN

Am 18. April verbreitete sich in dem großen Netz der Freunde von Aage B. Sørensen die erschütternde Nachricht, dass sich Aage nie mehr von den Folgen seines Sturzes auf glattem Eis erholen würde. Größte Behinderungen hatten ihn die letzten 14 Monate seines Lebens von Rollstuhl und Pflege abhängig gemacht. Gewiss hat er in dieser schweren Zeit durch seine Familie, die um seine Genesung rang, und in den spontanen Freundschaftsbeweisen, Hilfsangeboten und Ehrungen durch seine Kollegen und viele, die ihm nahe standen, aber auch die außergewöhnliche menschliche und berufliche Wertschätzung gespürt, die er genoss.

Aage Sørensen wurde am 13. Mai 1941 in Silkeborg, Dänemark, geboren und studierte in Kopenhagen unter anderem bei Kaare Svalastoga eine Soziologie, in der Fragen der sozialen Ungleichheit und Mobilität und ihre mathematische Formalisierung im Vordergrund standen. Nach dem Studium blieb er zunächst für zwei Jahre als Postgraduate Fellow in Kopenhagen und wurde dann (1970) Graduate Student an der Johns Hopkins University, Baltimore, und Forschungsassistent in der dort von James Coleman und Peter M. Rossi geleiteten Lebensverlaufsstudie. Nach seiner Promotion zum Ph.D. (1971) begann er seine glanzvolle akademische Laufbahn am Department of Sociology an der University of Wisconsin, Madison. Sie führte ihn in nur acht Jahren vom Assistant Professor zum Chairman eines der bedeutendsten und größten Sociology Departments der Welt. 1984 wechselte er nach Harvard, wo er – wiederum als Chairman (bis 1992) – das dortige Department in einem schwierigen Reorganisationsprozess zu neuem Ansehen brachte. Bis zu seinem tragischen Unfall wirkte er in Harvard als engagierter Lehrer von Graduate Students, innovativer theoretischer Denker, produktiver Forscher, erfolgreicher Manager und in der Disziplin einflussreicher Ratgeber, Reviewer und Mentor.

Bereits mit den ersten, auf seiner Kopenhagener Magisterarbeit basierenden Zeitschriftenveröffentlichungen kündigt sich für die damalige Zeit programmatisch Neues mit weitreichenden Implikationen an: In einem Beitrag zu *Sociology of Education* (1970) entwickelt Sørensen eine allgemeine, theoretisch und empirisch sehr fruchtbare Konzeptualisierung unterschiedlicher Varianten der organisatorischen Differenzierung von Bildungssystemen und formuliert Hypothesen zu den sozialen Selektionswirkungen, die sich als Folge der mit der Form der Differenzierung einhergehenden, unterschiedlichen Bildungsgelegenheiten ergeben. In „Equality of

Educational Opportunity in an Expanding Educational System“ (Acta Sociologica 1971) folgt dann für die empirische Analyse sozial ungleicher Bildungsbeteiligung, lange vor ähnlichen Arbeiten anderer, ein erster Versuch, durch eine methodisch innovative Modellierung die Auswirkungen der mit der Expansion verbesserten Bildungsgelegenheiten von den mit der sozialen Gruppenzugehörigkeit verbundenen Ressourcen, Restriktionen und Motivationen zu trennen. An der theoretisch und methodisch angemessenen Behandlung dieses Fundamentalproblems von ressourcengebundenem individuellem Handeln in Gelegenheitsstrukturen, die durch strukturelle oder institutionelle Bedingungen geprägt sind, richtet Sørensen einen zentralen Teil seines Lebenswerkes aus. Im Grenzgebiet zwischen Soziologie und Ökonomie überrascht er in seinen zahlreichen bahnbrechenden Aufsätzen Mal für Mal die Forschergemeinde mit Weiterentwicklungen der Theorie und methodisch ingeniosen Lösungen für ihren Test. Sein wichtigstes Studienobjekt sind die Strukturen und das Funktionieren von Arbeitsmärkten, die Mobilität von Individuen in ihnen und die aus dem Zusammenwirken von Strukturen und Ressourcen im Handeln der Individuen resultierenden Ungleichheiten. Gegen die Analysen von Mobilitätstabellen und gegen die Statuserwerbsforschung treibt er die Modellierung von Ereignisdaten und ihre Nutzung bei der Untersuchung individueller Erwerbsverläufe voran und ist einer der zentralen Akteure in der Durchsetzung dieses Paradigmas in einer neuen prozessorientierten Sozialstrukturforschung. In der Verbindung von Theorieformulierung, ihrer mathematischen Formalisierung und der mikroanalytischen empirischen Forschung erreicht Aage Sørensen früh eine kaum übertroffene Meisterschaft. Er bleibt darin für viele im Fach, auch für mich persönlich, ein bewundertes, aber nicht erreichtes Vorbild.

Von der intellektuelle Größe von Aage Sørensen zeugen gerade auch seine letzten Arbeiten. Sie lösten bereits mit ihrem Erscheinen lebhaft Diskussionen in der Disziplin aus. In „Toward a Sounder Basis for Class Analysis“ (American Journal of Sociology, 2000) begegnet er mit einer allgemeinen Theorie sozialer Ungleichheit dem Theoriedefizit in diesem Bereich und bringt damit die Ungleichheitssoziologie einen wichtigen Schritt voran. Seine Theorie erklärt die vielfältigen Erscheinungsformen und Ausprägungen von Ungleichheit mit einem generellen Mechanismus: der Monopolisierung von Renten. Die brillante historische Studie zur Entwicklung des skandinavischen Wohlfahrtsstaates „On Kings, Pietism and Rent-Seeking in Scandinavian Welfare States“ (Acta Sociologica 1998) belegt nicht nur die Tragweite dieser Theorie, sondern bleibt auch ein lebendiges Zeugnis für die umfassende Bildung und – besonders in der Antwort auf seine Kritiker – die selten souveräne professionelle Statur ihres Autors.

Neben dem wegweisenden und bleibenden wissenschaftlichen Werk kommt Aage Sørensen eminenten Verdienst als Vermittler zwischen der amerikanischen und der europäischen Soziologie zu, in Europa vor allem in Skandinavien und Deutschland. Nur für kurze Zeit lehrte er als Professor für Soziologie in Oslo. Über fast dreißig Jahre war er aber in zahllosen Fachkonferenzen in Europa präsent und waren seine Beiträge fast immer ein Höhepunkt, wenn sie auch nicht immer im ersten Anhieb verstanden wurden. Er hinterlässt maßgebliche Spuren in der Entwicklung der Soziologie als internationaler Gutachter in Berufungsverfahren und mehreren hochrangigen Evaluations- und Steuerungsgremien (u.a. für viele Jahre im Direktorium der Danish National Science Foundation), als Mitglied von Beiräten, Gastdozent in der Nachwuchsförderung und durch wiederholte Forschungsaufenthalte in namhaften Einrichtungen in Wien, Mannheim, Berlin, Stockholm, Oslo, Bergen und Oxford. Vielen, gerade jungen Forschern aus Europa hat er den Weg für kürzere oder längere Aufenthalte in seinen amerikanischen Departments geebnet und war auch darin ein Brückenbauer über den Atlantik. U.a. hat seine frühere Hilfskraft bei ZUMA und gegenwärtige Präsidentin der Deutschen Gesellschaft für Soziologie, Jutta Allmendinger, unter seiner Betreuung in Harvard promoviert.

Ein wichtiger Teil der innovativsten Entwicklungen in der empirischen Sozialforschung und Soziologie Deutschlands ist eng mit seinem Namen verbunden. Die vor allem durch die Arbeiten von Sørensen bekannt gewordene Johns Hopkins Life History Study war das prägende Vorbild für die entsprechende deutsche Studie. Karl Ulrich Mayer ist es gelungen, Aage Sørensen als Inspirator und Diskussionspartner in den 80er Jahren Jahr für Jahr und dann nochmals 1994 für längere Forschungsaufenthalte in Deutschland zu gewinnen. Dies führte ihn in den Jahren 1981-1983 und 1987 für insgesamt fast ein Jahr zu ZUMA, und dann zu noch regelmäßigeren und längeren Aufenthalten an das MPI für Bildungsforschung nach Berlin. Während einer der Aufenthalte bei ZUMA ist einer seiner berühmten Aufsätze – „Processes of Allocation to Open and Closed Positions in Social Structure“, erschienen 1983 in der Zeitschrift für Soziologie – entstanden. In Anknüpfung an die Weberschen Grundkategorien von offenen und geschlossenen sozialen Beziehungen und Simmels Idee der Objektivierung des Handelns und der sozialen Beziehungen in „Stellungen“ systematisiert er dort zentrale theoretische Grundlagen für die neuere soziologische Analyse von Bildungssystemen, Arbeitsmärkten und beruflichen Karrieren. Integriert ist auch die von Sørensen schon in früheren Arbeiten modellierte Idee der Positionsgüter von Fred Hirsch. Diese Modellierung wird zu einer der tragenden Säulen der Magnitude-Prestige-Skala für Deutschland, die in dieser Zeit von Bernd Wegener bei ZUMA entwickelt wurde und zu einem

der am häufigsten genutzten Instrumente in der deutschen Sozialstrukturforschung avancierte.

Dem MPI für Bildungsforschung und ZUMA diene Sørensen für jeweils fast ein Jahrzehnt als Mitglied ihrer wissenschaftlichen Beiräte. Er war ein großer Freund beider Institute. Mit seiner scharfen Beobachtung in der Diagnose von Schwächen, ihrer unkompromittierten und schonungslosen Aufdeckung auch gegenüber Freunden und mit der Eröffnung tragfähiger Perspektiven und Anregungen war er einer der hilfreichsten kritischen Begleiter und Förderer, den eine Institution haben konnte. Ich kann hier nur im Namen von ZUMA sagen, dass ihm das Institut viele wichtige Impulse verdankt. Wenn nicht nach seinen hohen Maßstäben gehandelt wurde oder gehandelt werden konnte, ist dies natürlich nicht ihm zuzuschreiben.

Viele in aller Welt haben mit Aage Sørensen einen unvergesslichen treuen Freund verloren. Welch ein glückliches Privileg im eigenen Leben war es, mit ihm und seiner Frau Mette durch den Odenwald zu gehen, ein Weinfest in der Pfalz oder die Liedertafel in Haardt zu besuchen, ihre Gastfreundschaft in Cambridge zu erfahren und dabei zwischen irren Geschichten und Gossip voller Witz und Lebensfreude sich unverhofft in tiefsinnige Argumente verwickelt zu sehen, und dann viel später – und doch als wäre es gestern gewesen – Aage irgendwo auf der Welt wieder zu begegnen, seinem zwinkernden Blick, der alles sagte, und seinem ureigenen aus dem versunkenen Innenleben plötzlich aufblitzenden freudigen und einnehmenden Lachen im Gesicht. Wir haben mit ihm allzu früh viel verloren: einen herausragenden Wissenschaftler, einen unschätzbaren Kollegen und einen zugewandten Menschen voll reichster persönlicher Qualitäten.

Walter Müller, Universität Mannheim

P.S. Aage Sørensens Familie und das Sociology Department haben an der Harvard University zu seinem Gedenken und in Anerkennung seiner besonderen Verdienste bei der Ausbildung junger Wissenschaftler einen Fond zur Förderung von Graduate Students eingerichtet. Spenden können am besten per Scheck gesandt werden an den Aage B. Sørensen Memorial Fund, c/o Ms. Jackie Piracini, Department of Sociology, Harvard University, Cambridge MA 02138 (making checks payable to Harvard University and noting that they are for the Aage B. Sørensen Memorial Fund).

---

## BUCHBESPRECHUNGEN

---

### INTERNATIONALE FORSCHUNG ZUR ZEITVERWENDUNG

„Zeit ist Geld“ – dies ist ein Zitat, das von Benjamin Franklin stammen soll, und das in den vergangenen Jahren häufig erweitert wird um „Zeit ist Macht“. Zeit als eine der kostbarsten Ressourcen, die weder anhäufbar noch abbaubar ist – wie wir u.a. bei Michael Endes Momo und den Zeitdieben gelernt haben –, hat in den Sozialwissenschaften als Forschungsobjekt zunehmend Bedeutung gewonnen. Neben den theoretischen Arbeiten zum Umgang mit Zeit in verschiedenen differenzierten Gesellschaftssystemen, zur Bedeutung von Zeit im Rahmen der Strukturierungstheorie oder zu den Funktionen von Zeit innerhalb einer Gesellschaft sind in den vergangenen Jahren – nicht zuletzt im Zusammenhang mit der Lebensstilforschung – zahlreiche empirische Untersuchungen zur Zeitverwendung der Bevölkerung entstanden. Dazu bietet das hier vorgestellte Buch „Time Use – Research, Data and Policy“, herausgegeben von JOACHIM MERZ (Universität Lüneburg, Forschungsinstitut Freie Berufe) und MANFRED EHLING (Statistisches Bundesamt, Wiesbaden) einen umfassenden Überblick.

---

JOACHIM MERZ/MANFRED EHLING (EDS.)

**Time Use – Research, Data and Policy**

Schriften des Forschungsinstituts Freie Berufe – FFB, Vol./Bd. 10

Nomos Verlagsgesellschaft, Baden-Baden, 1998

572 S., 84 Mark, ISBN 3-7890-6244-8

---

In dem 572 Seiten mächtigen Werk sind die Ergebnisse einer internationalen Konferenz zur Zeitverwendung (International Conference on Time Use (ICTU)) zusammen gestellt, die 1998 in Lüneburg gemeinsam von MERZ (Universität Lüneburg) und EHLING (Statistisches Bundesamt) organisiert wurde. Ziel dieser Konferenz war, eine umfassende ganzheitliche Betrachtung der individuellen Zeitverwendung anzubieten, die über die bisherigen, an einzelnen Bereichen, wie Arbeitszeit oder Hausarbeitszeit orientierten, Ansätze hinaus geht. Für die an Zeitverwendung interessierte Leserschaft erweist sich als besonders nützlich, dass über die inhaltliche

Beschäftigung mit dem Thema hinaus auch zentrale Datensätze vorgestellt werden sowie methodische Aspekte Berücksichtigung finden.

Nach dem ersten Kapitel der Eröffnung des Bandes, in dem GERSHUNY (Essex, UK) die Bedeutung der Studien zur Zeitverwendung für die Erfassung des informellen Arbeitsmarktes hervorhebt, werden im zweiten Kapitel den Leserinnen und Lesern verschiedene internationale und nationale Zeitverwendungsstudien vorgestellt. Diese basieren z.T. auf den frühen Arbeiten der russischen Zeitforscher, zum großen Teil sind sie jedoch in der Tradition der - mittlerweile als Klassiker zu bezeichnenden - europäischen Studie „The Use of Time“ (1972) von Szalai zu verstehen. Es wird über Studien berichtet, die von der OECD zur Haushaltsproduktion und unbezahlter Hausarbeit (CHADEAU) durchgeführt werden, über ein Projekt von Eurostat zur Harmonisierung von Statistiken zur Zeitverwendung (RYDENSTAM), über vergleichende Zeitbudgetstudien von 1920-1990 in Russland (ARTEMOV/ARTEMOVA), über einen Vergleich von europäischen Ländern hinsichtlich ihrer sozialen Zeitstrukturen (GARHAMMER), und in zwei weiteren Beiträgen dieses Kapitels werden Ergebnisse der bundesdeutschen Zeitbudgeterhebung zu deren Erhebungsdesign und Ergebnissen zum Zeitstress (EHLING) sowie zu zeitlichen Verpflichtungen in Privathaushalten (KÜSTER/MEYER) präsentiert.

Das dritte Kapitel, das sich mit der Forschung zur Zeitverwendung und Erhebungsmethoden beschäftigt, wird durch einen Beitrag von HARVEY (Halifax, CAN), eröffnet. Dieser Aufsatz bietet einen ausgezeichneten Überblick über die Entstehung dieser Forschungsrichtung seit Beginn des 20. Jahrhunderts sowie über ihre aktuelle Anwendungen. In den weiteren Beiträgen werden methodische Probleme, wie z.B. die Klassifizierung der Aktivitäten (BEDIAKO/VANEK sowie HOFFMANN/Mata), Möglichkeiten der Kombination von Zeitverwendungsdaten mit amtlicher Statistik (WENKE) oder Probleme des Datenschutzes vs. Datenverfügbarkeit (HOLZ) erörtert.

Die nachfolgenden Kapitel widmen sich nun thematischen Schwerpunkten, wie Arbeit (Kapitel 4), Haushaltsproduktion (Kapitel 5), Aktivitäten im Markt und außerhalb des Marktes (Kapitel 6) sowie der Soziologie der Zeitverwendung und sozialpolitischen Aspekten (Kapitel 7). Innerhalb dieser thematischen Kapitel, die knapp zwei Drittel des Buchs umfassen, werden nicht nur verschiedene Ansätze und Modelle verwendet, sondern es werden zudem weitere Datengrundlagen, wie das IAB Beschäftigten Panel oder das Sozio-oekonomische Panel vorgestellt, mit Hilfe derer auch Analysen zu Zeitverwendung durchgeführt werden können.



Die Themenbereiche innerhalb des Kapitels 4 „Arbeit“ reichen von dem Verhältnis zwischen gewünschter und tatsächlicher Arbeitszeit über Scheinselbständigkeit bis hin zu Modellen, in denen steuerliche Veränderungen in die Gestaltung der Arbeitszeiten einbezogen werden. Im Kapitel 5 „Haushaltsproduktion“ steht die innerhalb der Haushalte geleistete unbezahlte Arbeit im Vordergrund. Dieser Bereich zählt zu den Themen, die erst durch umfangreiche Zeitbudgetstudien überhaupt erfassbar und bewertbar geworden sind. Dabei werden ebenfalls unterschiedliche Modelle, wie Modelle zur Haushaltsproduktion oder spieltheoretische Modelle im Rahmen von Analysen zur Verteilung von Zeit und Geld innerhalb der Haushalte getestet. Darüber hinaus werden Ansätze zur Bewertung der unbezahlten Arbeit entwickelt, die im nachfolgenden Kapitel 6 der kombinierten Betrachtung von Arbeitsmarkt (Markt) und Haushaltsproduktion (Nicht-Markt) weiter vertieft werden. Dabei liefert KLEVMARKEN in seinem Beitrag zur Mikroökonomie und ihre Beschäftigung mit Zeitverwendung einen guten Überblick über die Arbeiten auf diesem Gebiet und ermuntert zur intensiveren Nutzung der Zeitbudgetdaten. Ein Beitrag innerhalb dieses umfangreichen Kapitels verdient m.E. besondere Erwähnung, nämlich der Artikel von OCCELLI, die ein Konzept der Erreichbarkeit (accessibility) und Zeitverwendung in einer Stadtregion vorstellt. Gestützt durch eine Analyse der Wegezeiten und des Angebots an „Gelegenheiten“ (Schulen, Geschäfte usw.) entwickelt sie ein Modell, in dem sowohl zeitliche als auch räumliche Aspekte der Erreichbarkeit ebenso wie subjektives „Wissen“ über Erreichbarkeit einbezogen werden. Da in der Forschung zur Zeitverwendung bisher die Dimension Raum nur selten Berücksichtigung findet, stellt dieser Beitrag einen gelungenen Ansatz zu diesem Thema dar.

In Kapitel 7 zur Soziologie der Zeitverwendung und sozialpolitischen Aspekten wird das Thema Ungleichheit in der Verfügbarkeit von Zeit zwischen den Geschlechtern, das bereits in zahlreichen o.g. Beiträgen angeklungen war, am Beispiel von Verkehr und Wegemustern, wie z.B. den typisch weiblichen „Wegeketten“, vertieft (GRIECO/TURNER). Im Ausblick auf zukünftige Forschung zur Zeitverwendung wird deutlich, wie wichtig kontinuierliche systematische und harmonisierte Programme zur Erfassung der individuellen Zeitverwendung sind, um den immer schneller werdenden Wandel der Flexibilisierung im Arbeitsleben, aber auch in anderen Bereichen, erfassen zu können. Mit dieser Forderung nach sowohl international als auch über die Zeitintervalle vergleichbaren Erhebungen - die man in Eurostat-Projekten bereits umzusetzen versucht - endet der umfangreiche Konferenzband.

Die an „Time Use“ interessierten Leserinnen und Leser erwartet ein spannendes und facettenreiches Buch, in dem deutlich wird, welche Vielfalt von gesellschaftlich und politisch relevanten Themen mit der Analyse von individueller Zeitverwen-

dung erschlossen werden kann. Es ist m.E. gelungen, den „bunten Strauß“ der Studien übersichtlich zu „binden“ und sowohl allgemein Interessierten einen Einstieg zu bieten (ihnen seien besonderes die Beiträge von HARVEY und GERSHUNY zu empfehlen) als auch dem engeren Forscher/-innen-Kreis neue Aspekte, Methoden und Daten aus diesem Forschungsgebiet vorzustellen.

Neben den zahlreichen in diesem Band vorgestellten Bereichen sind auf dem Gebiet der Forschung zur Zeitverwendung m.E. noch zwei weitere große Themen zu nennen, die zwar anklingen, aber derzeit nur selten bearbeitet werden. Dies ist zum einen die subjektive Wahrnehmung und Bewertung der – wie auch immer – verwendeten Zeit; dies ist ein Aspekt, der bei Erhebungen zur Zeitverwendung mehr Beachtung finden sollte. Und zum anderen verdient die räumliche Komponente, die Lokalisierung der Akteure in ihrem – nicht zuletzt räumlichen – Kontext m.E. mehr Aufmerksamkeit. Hier müsste bereits im Erhebungsdesign die Dimension Raum neben der zentralen Dimension Zeit Berücksichtigung finden. Auch interkulturelle Vergleiche bieten sicher aufschlussreiche Ergebnisse. Zum Abschluss dieser Besprechung dazu eine Kostprobe: „Die Europäer haben die Uhr, wir haben die Zeit.“ (Orientalisches Sprichwort)

Weitere Informationen zum Forschungsnetzwerk „Research Network on Time Use (RNTU) sind zu finden unter [www.uni-lueneburg.de/timeuse](http://www.uni-lueneburg.de/timeuse).

*Dr. Caroline Kramer (Geographisches Institut der Universität Heidelberg)*

Das Paket Statistik; Interaktive Lernsoftware zur beschreibenden und schließenden Statistik von A. GODENZI und R. STADLER enthält eine CD-ROM mit der von den Autoren entwickelten Lernsoftware „STAT inside“, sowie verschiedene Datensätze aus Deutschland, Österreich und der Schweiz. Außerdem befindet sich auf der CD-ROM die Hypertextversion des beiliegenden Lehrbuchs.

---

ALBERTO GODENZI, ROGER STADLER

**Statistik. Interaktive Lernsoftware zur beschreibenden und schließenden Statistik**

[mit einer bearbeiteten Übersetzung des „Statistics for Social Workers“, von Weinbach/Grinnell, 4. Auflage, 1998]

Luchterhand Verlag GmbH, Neuwied, 2000, 89,- DM

---

Das Lehrbuch ist eine überarbeitete und auf deutschsprachige Verhältnisse angepasste Übersetzung des Statistikbuchs von WEINBACH/GRINNELL „Statistics for Social Workers“, 4. Auflage (Mittlerweile liegt das amerikanische Buch in fünfter Auflage vor). Das Statistikpaket ist sowohl für Studenten sozialwissenschaftlicher Disziplinen als auch für sozialwissenschaftlich Tätige konzipiert. Es bietet eine Einführung in statistische Grundkenntnisse.

„STAT inside v1.0“ läuft auf den Betriebssystemen Windows 95+, Windows NT+ oder Linux 2.2.14+ mit mindestens 32 MB RAM, oder auf Macintosh 8.1+ mit mindestens 48 MB RAM. Die Installation verläuft unkompliziert. Homepage und Hotline sind notfalls angegeben.

Die Programmoberfläche ist übersichtlich und einfach zu bedienen. Es ist in zwei Frames unterteilt: Inhalts-, bzw. Navigationsseite, Textseite. Bei Bedarf öffnen sich weitere Fenster.

„STAT inside“ ist ein computergestütztes Lernsystem, das in Java geschrieben wurde. Das didaktische Konzept von „STAT inside“ beruht auf Verbindung von Lerntext, Simulationen (Animationen), statistischen Daten und Analyseprogrammen.

Der Lerntext beinhaltet den vollständigen Text des beiliegenden Lehrbuchs mit Hyperlink-Verweisen. Darüber hinaus enthält es weiterführende Texte, sowie einen Multiple-Choice-Test.

„STAT inside“ hat eine eigene Klassenbibliothek, welche statistische Analysen ermöglicht. Dazu gehören beispielsweise Klassen für das Einlesen von Rohdaten, die Berechnung statistischer Kennwerte, die grafische Darstellung der Daten, sowie die Berechnungen und grafische Darstellung verschiedener Verteilungsfunktionen (Normal-

verteilung, t-Verteilung, F-Verteilung, Binomialverteilung, chi-quadrat-Verteilung). Es enthält insgesamt 22 verschiedene uni- oder bivariate Tests und bietet Analysemöglichkeiten anhand eigener Datenbeispiele oder unter Verwendung der beiliegenden Datensätze. Diese sind u.a. ISSP-Datenbanken entnommen.

„STAT inside“ ist ein didaktisches Lehrmittel und erhebt nicht den Anspruch, eine Alternative zu kommerziellen Statistikprogrammen zu sein.

Das Lehrbuch aus dem Amerikanischen ist unter dem Aspekt einfacher Verständlichkeit ausgesucht worden. Es beschränkt mathematische Grundlagen auf das Notwendige zugunsten im Sozialbereich häufig eingesetzter statistischer Verfahren. Das Buch ist im Taschenbuchformat, circa 270 Seiten stark und besteht aus zwölf Kapiteln.

### **Das Lehrbuch**

Der Inhalt ist klar gegliedert und leicht verständlich geschrieben. Das Buch verzichtet weitgehend sowohl auf theoretische Konzepte als auch auf mathematische Formeln, was der erklärten Absicht der Autoren entspricht, einen praktisch orientierten Grundlagentext heranzuziehen. Konzepte und Definitionen werden an vielen Beispielen erläutert; Formeln sind in Worte gefasst oder an Beispielen durchgerechnet. Dadurch verlieren sie allerdings an Eleganz, wirken umständlich und unhandlich.

Der erste Teil befasst sich mit deskriptiver Statistik. Er zeigt den Gebrauch und Einsatzbereich grundlegender tabellarischer, grafischer Datendarstellungen, sowie verbreteter Kennzahlen.

Der zweite Teil beschreibt statistische Testverfahren, u.a. Korrelations-, Regressionsanalyse,  $\chi^2$ -Test, t-Test, diverse nichtparametrische Tests. Ein gesondertes Kapitel diskutiert Auswahlkriterien für statistische Tests. Die Abläufe der wichtigsten Tests sind rezeptartig zusammengefasst und mit Rechenbeispielen unterlegt. Da angenommen wird, dass Berechnungen mit Statistikprogrammen durchgeführt werden, liegt der Schwerpunkt auf Anwendungsbeispielen und auf Interpretation der Ergebnisse. Zusätzlich werden die wichtigsten multivariaten Testmethoden kurz vorgestellt.

Im Anhang sind vier statistische Tabellen abgedruckt, sowie eine Literaturempfehlung; ein Stichwortverzeichnis fehlt (die CD-Version jedoch hat eine Suchfunktion).

### **Ergänzungen zum Lehrbuch**

Auf der CD haben die Autoren einzelne Themen des Lehrbuchs eingehender dargestellt und einen Abschnitt über Kombinatorik hinzugefügt. Simulationsprogramme il-

lustrieren Probleme aus Bereichen wie z.B. Datendarstellung, Zufallsexperimente, Häufigkeitsverteilungen. Außerdem sind nützliche Adressen oder Links zu Statistikprogrammen, ZIS, Datenbanken, etc. angegeben, sowie zu einer Seite mit Statistiker-Witzen!

### **Der Multiple-Choice Test**

Im Anschluss an das Lehrbuch befindet sich ein Multiple-Choice-Test zur Selbstkontrolle. Der MC-Test enthält über 100 Fragen zu allen Kapiteln des Lehrbuchs. Jede eingegebene Antwort wird prompt auf Richtigkeit überprüft.

### **Die Statistikprogramme**

„STAT inside“ enthält eine Sammlung mit 22 statistischen Tests. Ein Assistent erfragt die Informationen, die für die Auswahl eines Tests notwendig sind. Die Daten können entweder als Rohdaten in eine Tabelle eingegeben werden oder als aggregierte Kennzahlen. Alternativ besteht die Möglichkeit, Beispieldaten zu generieren. Fehlermeldungen sind selten (z.B. einfache Varianzanalyse mit Beispieldaten). Daten aus den mitgelieferten Datensätzen können eingelesen werden.

Für den Import externer Daten ist ein Assistent eingerichtet, der das Üben mit eigenen Daten ermöglichen soll. Dieses Zubehör erweist sich indessen als unausgereift und fehleranfällig.

Insgesamt ist die Testsammlung ein ausgezeichneter Sandkasten, um statistische Tests durchzuspielen und mit Analyseergebnissen vertraut zu werden.

### **Resümee**

Das Paket ist für Anfänger ohne Vorkenntnisse – aber nicht nur für Soziologen geeignet. Es vermittelt auf unterhaltsame Art Basiskenntnisse in Statistik. Der autodidaktische Nutzer wird befähigt, mittels eines Statistikprogramms einfache Analysen durchzuführen. Das Paket ist, wie es die Autoren auch betonen, für den Praxisgebrauch zusammengestellt und soll kein akademisches Lehrwerk ersetzen. Alles in allem hat das Statistikpaket Buch+Lernsoftware ein faires Preis/Leistungsverhältnis.

*Dr. Ferda Koymen*

Bei der Sammlung von Umfragedaten über Personen („Befragte“) ist es aus inhaltlichen Gründen häufig erforderlich, Informationen über ihren sozialen Kontext zu berücksichtigen. Nicht immer können die benötigten Informationen direkt erfragt werden. Selbst wo dies möglich ist, sollte zunächst nach vorhandenen Daten aus anderen Quellen gesucht werden, die zumindest „im Prinzip“ den Befragungsdaten hinzugefügt werden können. Soweit es sich bei den benötigten Informationen über den sozialen Kontext der Befragten um „Regionaldaten“ handelt, gibt Jürgen H.P. Hoffmeyer-Zlotnik mit dem vorliegenden Buch einen umfassenden Überblick über die Möglichkeiten einer solchen Ergänzung von Umfragedaten.

---

JUERGEN H.P. HOFFMEYER-ZLOTNIK:  
**Regionalisierung sozialwissenschaftlicher Umfragedaten.  
Siedlungsstruktur und Wohnquartier.**  
Wiesbaden 2000, Westdeutscher Verlag  
249 Seiten, 49.80 DM, ISBN 3-531-13538-4

---

Im ersten Kapitel werden einführend einige Grundlagen der Regionalisierung von Umfragedaten behandelt: Welche Bedeutung haben Regionaldaten für die Entwicklung und das Verhalten von Individuen? Inwieweit kennzeichnen sie - über „soziale Milieus“ - Eigenschaften der jeweils ablaufenden sozialen Prozesse und tragen somit in Verbindung mit den je individuellen Eigenschaften der dort lebenden Personen zum Verständnis ihrer Verhaltensweisen bei? In knapper Form fasst Hoffmeyer-Zlotnik in diesem Kapitel theoretische Ansätze der Regionalforschung zusammen, skizziert mögliche Verwendungen von Regionaldaten bei der Durchführung und Interpretation von Umfragedaten und beschreibt die wichtigsten Quellen für Regionaldaten.

Kapitel 2 beschreibt verschiedene Abgrenzungen von Regionen und klärt ihre ursprünglichen Entstehungsgrundlagen ebenso wie ihre nachträglich (zur Ergänzung von Umfragedaten) sinnvoll erscheinende Verwendungen. Hier wie in den folgenden Kapiteln zieht Hoffmeyer-Zlotnik einen weiten Bogen und stellt vielfältige Formen möglicher wie auch realisierter Abgrenzungen dar:

(1) Administrative Grundeinheiten wie (in Deutschland) Gemeinde, Kreis, Regierungsbezirk und Bundesland, Untergliederungen zu bestimmten Zwecken (wie u.a. Ortsteile, Statistische Bezirke) und Zusammenfassungen kleinerer Einheiten zu Planungs-, Raumordnungs- und Vergleichszwecken.

(2) Aus wirtschaftlichen Gründen festgelegte Grundeinheiten (wie etwa Nielsen-Gebiete) zur Belegung mit Werbung über Massenmedien und eine entsprechende Analyse in der Marktforschung.

(3) Verflechtungsgebiete aufgrund vor allem des wirtschaftlichen Austausches von Gütern und des Pendelns von Arbeitskräften über die Grenzen der vorher genannten regionalen Grundeinheiten hinweg. Daneben werden viele andere Abgrenzungsformen erwähnt. Auch in den folgenden Kapiteln wird offenkundig Wert auf eine möglichst vollständige Darstellung gelegt wird. Besonders hilfreich ist dies vor allem für Leser, die das Buch als Übersichts- und Nachschlagewerk nutzen wollen.

Die Kapitel 3 bis 5 sind der ausführlichen Darstellung von Einheiten unterschiedlicher regionaler Ebenen gewidmet: Verschiedenen Formen von Stadtregionen und den mit ihnen durch wirtschaftlichen Austausch oder durch Pendlerströme verflochtenen Gebieten (Kapitel 3), der Stadt in unterschiedlichen Entwicklungsformen (Kapitel 4) sowie Untergliederungen der Städte in (intern möglichst homogene) Teilgebiete oder Wohnquartiere (Kapitel 5).

Bei der Darstellung besticht, dass eine weitgehend einheitliche Gliederung in allen drei Kapiteln (und damit über Regionaleinheiten aller drei Ebenen) aufrecht erhalten wird: Nach einführenden Hinweisen auf die Bedeutung der jeweiligen Regionaldaten im Verwendungszusammenhang werden Regionaleinheiten und ihre Eigenschaften ausführlich dargestellt, Datenquellen zu ihrer Beschreibung sowohl in ihrer aktuell verfügbaren wie voraussichtlich künftig zu erwartenden Form beschrieben und in einem abschließenden Beispiel (ebenfalls einheitlich über alle drei Ebenen) der Nutzen der jeweiligen Regionaldaten zur ergänzenden Analyse von Umfragedaten illustriert. Die drei Kapitel (3-5) geben einen zusammenfassenden Überblick über die derzeit allgemein verfügbaren Regionaldaten und einige ihrer Verwendungsmöglichkeiten zur Beschreibung individueller Kontexte.

Kapitel 6 schließlich zielt auf die Primärerhebung von Regionaldaten bei der Umfrage und beschreibt das vom Autor selbst entwickelte und mehrfach geprüfte Instrument zur standardisierten Beschreibung von Wohnquartieren.

Wie schon erwähnt, trägt das Buch trotz seines relativ knappen Umfangs die Züge eines Handbuchs. Es liefert eine umfassende und aktuelle Übersicht über die Standardinstrumente zur regionalen Beschreibung in Deutschland.

Wer die Primärerhebung von Umfragedaten und die Analyse des Verhaltens von Personen „im sozialen Kontext“ plant, sollte es möglichst bereits bei der Vorbereitung konsultieren, um Hinweise auf eine effiziente Realisierung der bestehenden

Pläne zu erhalten. Noch eindringlicher ist das Buch denjenigen zu empfehlen, die sich auf die Analyse der Einstellungen und Verhaltensweisen einzelner (isolierter) Individuen konzentrieren: Ihnen gibt es eine Fülle von Anregungen, wie Umfragedaten über einzelne Personen auf sinnvolle Weise mit verfügbaren Daten über regionale Kontexte verbunden und damit der Kreis in Frage kommender Deutungsmöglichkeiten erweitert werden kann.

*Prof. Dr. Wolfgang Sodeur (Wolfgang.Sodeur@uni-koeln.de)*



---

## PUBLIKATIONEN

---

### ZUMA-ARBEITSBERICHTE

Die ZUMA-Arbeitsberichte informieren über das gesamte Spektrum der Arbeit bei ZUMA und enthalten überwiegend Beiträge, die zur Veröffentlichung in Fachzeitschriften vorgesehen sind. Nachfolgend werden die ZUMA-Arbeitsberichte, die seit der letzten Ausgabe der ZUMA-Nachrichten publiziert worden sind, in Form von Abstracts kurz vorgestellt. ZUMA-Arbeitsberichte sind auch im Internet als PDF-Datei abrufbar ([www.gesis.org/publikationen/berichte/zuma\\_arbeitsberichte/](http://www.gesis.org/publikationen/berichte/zuma_arbeitsberichte/)). Einzelhefte können unter folgender Adresse bestellt werden:

Zentrum für Umfragen, Methoden und Analysen  
ZUMA-Publikationen  
Postfach 12 21 55  
68072 Mannheim

\* \* \* \* \*

*Rüdiger Schmitt-Beck: Mass Communication, Personal Communication and Vote Choice. ZUMA-Arbeitsbericht 2001/01.*

In jüngerer Zeit hat die Frage, ob die Massenkommunikation und die persönliche Kommunikation persuasive Einflüsse auf Wahlentscheidungen ausüben können, verstärktes Interesse gefunden. Jedoch hat es bisher kaum Versuche gegeben zu untersuchen, wie diese beiden Formen politischer Kommunikation im Rahmen von Einflussprozessen interagieren. Die auf Katz/Lazarsfeld (1955) zurückgehende „Filterhypothese“ postuliert, dass die persönliche Kommunikation der Wähler Einflüsse der Massenkommunikation moderiert. Sie verstärkt oder blockiert solche Einflüsse, in Abhängigkeit von der Richtung der persuasiven Medienbotschaften und der politischen Zusammensetzung der Diskutantennetzwerke der Wähler, die von diesen Botschaften erreicht werden. Der Arbeitsbericht untersucht diese These anhand repräsentativer Wählerbefragungen aus Großbritannien, Spanien, den USA und Westdeutschland. Auf der Grundlage detaillierter Informationen über Struktu-

ren und Inhalte der politischen Gespräche von Wählern, über ihre Mediennutzung und über die Inhalte der genutzten Medien kann die Filterthese empirisch bestätigt werden. Das Papier schließt mit Überlegungen zu den demokratietheoretischen Implikationen der Befunde im Hinblick auf die Strukturen von Parteiensystemen.

\*\*\*

*Frank Kalter/Nadia Granato: Recent Trends of Assimilation in Germany. ZUMA-Arbeitsbericht 2001/02.*

This study investigates whether the life circumstances of immigrants and natives in Germany have converged within the last decades. Theoretically, the idea of social production functions is suggested as a general framework to integrate different assimilation approaches. Empirically, a method is used that combines the multivariate regression approach with common measures of segregation. This technique allows an easy assessment of assimilation trends taking into account relevant structural changes, in our case altering demographic compositions and the educational expansion. Immigrants and Germans are compared regarding central typologies of education, work, family, and place of residence through analyzing population censuses of 1970, 1989, and 1996. In spite of the fact that the educational gap has clearly widened over the years under observation, it will turn out that in respect of the other aspects of life the general trend appears to be towards assimilation, especially for the second generation of the 'classical' labor migrants.

## ZUMA-METHODENBERICHTE

In der ZUMA-Reihe „Methodenberichte“ werden Beiträge publiziert, die wichtige Aspekte der Arbeit von ZUMA dokumentieren, aber in der Regel nicht für eine Veröffentlichung in einer Fachzeitschrift geeignet sind, wie z.B. Dokumentationen, Bibliographien, Literaturberichte und Datenbeschreibungen. Die Berichte sind im Internet als PDF-Datei abrufbar ([www.gesis.org/publikationen/berichte/zuma\\_methodenberichte/](http://www.gesis.org/publikationen/berichte/zuma_methodenberichte/)). Einzelhefte können auch bei folgender Adresse bestellt werden:

Zentrum für Umfragen, Methoden und Analysen  
ZUMA-Publikationen  
Postfach 12 21 55  
68072 Mannheim

\* \* \* \*

ZUMA-Methodenbericht 2001/01

*Janet Harkness*: ISSP 1998: Religion II. ZUMA-Report on the German Study.

ZUMA-Methodenbericht 2001/02

*Janet Harkness*: ISSP 1997: Work Orientations. ZUMA-Report on the German Study.

ZUMA-Methodenbericht 2001/03

*Janet Harkness & Martina Wasmer*: German General Social Survey 1998. English Translation of the German “ALLBUS”-Questionnaire.

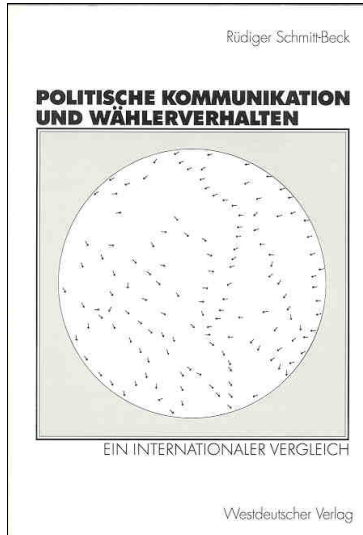
## ZUMA-HOW-TO-REIHE

In der ZUMA-Reihe „How-to“ werden ausgewählte Themen aus dem Bereich der sozialwissenschaftlichen Methodik anwendungsorientiert dargestellt. Der Schwerpunkt liegt auf der Vermittlung von Know-how, das für den Anwender relevant, wichtig und nützlich ist. Die Berichte sind im Internet als PDF-Datei abrufbar ([www.gesis.org/publikationen/berichte/zuma\\_how\\_to/](http://www.gesis.org/publikationen/berichte/zuma_how_to/)). Einzelhefte können auch bei den jeweiligen Autoren bestellt werden.

\* \* \* \* \*

- How-to Nr. 1: Ingwer Borg: Explorative Multidimensionale Skalierung.
- How-to Nr. 2: Rolf Porst: Question Wording - Zur Formulierung von Fragebogen-Fragen.
- How-to Nr. 3: Michael Braun: Evaluation der Äquivalenz eines gemeinsamen Satzes an Indikatoren in der interkulturell vergleichenden Sozialforschung.
- How-to Nr. 4: Jürgen H.P. Hoffmeyer-Zlotnik: Regionalisierung von Umfragedaten. Eine kleine Handlungsanleitung.
- How-to Nr. 5: Michael Häder: Die Expertenauswahl bei Delphi-Befragungen.
- How-to Nr. 6: Sabine Häder: Telefonstichproben.
- How-to Nr. 7: Jürgen H.P. Hoffmeyer-Zlotnik: Wohnquartiersbeschreibung: Ein Instrument zur Regionalisierung von Nachbarschaften.

# POLITISCHE KOMMUNIKATION UND WÄHLERVERHALTEN



**G**espräche über politische Themen (interpersonale Kommunikation) und die politische Berichterstattung von Medien wie Presse und Fernsehen (Massenkommunikation) stellen die wichtigsten Quellen dar, aus denen die Wähler politische Informationen erhalten. Die Studie untersucht, auf welche Weise diese beiden Formen politischer Kommunikation individuelle Wahlentscheidungen beeinflussen und welche Konsequenzen sich hieraus für die Ergebnisse von Wahlen ergeben. Analysiert werden die Entscheidungen von Wählern aus fünf Gesellschaften in vier Ländern: West- und Ostdeutschland, Großbritannien, Spanien und USA.

*Politische Kommunikation und Wählerverhalten. Ein internationaler Vergleich.*

Rüdiger Schmitt-Beck

Westdeutscher Verlag 2000

448 Seiten, 68 DM, ISBN 3-531-13526-0

## SOCIAL AND ECONOMIC ANALYSIS OF CONSUMER PANEL DATA



Eine von der Abteilung Einkommen und Verbrauch von ZUMA organisierte Arbeitsgruppe hat sich mit datentechnischem Handling und Analysepotential von komplexen Verbraucherpaneldaten, am Beispiel des ConsumerScan Haushaltspanels der Gesellschaft für Marktforschung (GfK, Nürnberg) beschäftigt und die Ergebnisse in einem Symposium im Oktober 1999 vorgestellt. Die überwiegende Zahl der vorgetragenen Arbeiten, die man als Werkstattberichte ansehen kann, sind in diesem Band 7 der ZUMA-Nachrichten Spezial abgedruckt. Der Band enthält zudem eine CD-ROM mit Dokumenten und Codebüchern der aufbereiteten ZUMA-Verbraucherpaneldaten 1995. Ein Bestellformular findet sich im Internet unter [http://www.gesis.org/Publikationen/Zeitschriften/ZUMA\\_Nachrichten\\_spezial/index.htm](http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/index.htm).

SOCIAL AND ECONOMIC ANALYSIS OF CONSUMER PANEL DATA

Georgios Papastefanou, Peter Schmidt, Axel Börsch-Supan,

Hartmut Lüdtke, Ulrich Oltersdorf (Eds.)

Mannheim: ZUMA, 2001, 216 Seiten, 15 €, ISBN 3-924220-21-2

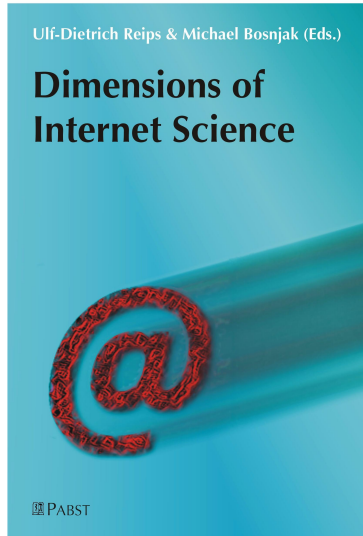
## BLICKPUNKT GESELLSCHAFT BAND 6: POLITISCHE PARTIZIPATION IN DER BUNDESREPUBLIK DEUTSCHLAND



Der sechste Band der Buchreihe „Blickpunkt Gesellschaft“ befasst sich mit dem Thema „Politische Partizipation“. Auf der Grundlage der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1998 wird ein breites Spektrum unterschiedlicher Formen politischer Beteiligung analysiert: von der Teilnahme an Wahlen über parteibezogene Partizipation und politisches Protestverhalten bis zu gewalttätigen politischen Handlungen. Verschiedene theoretische Modelle zur Erklärung politischer Partizipation werden dabei herangezogen. Daneben wird gezielt die Rolle spezifischer Determinanten politischer Beteiligung - wie z.B. der Mediennutzung - untersucht.

BLICKPUNKT GESELLSCHAFT 6  
POLITISCHE PARTIZIPATION IN DER BUNDESREPUBLIK DEUTSCHLAND  
Achim Koch, Martina Wasmer, Peter Schmidt (Hrsg.)  
Opladen: Leske + Budrich 2001  
324 Seiten, 48 DM, ISBN 3-8100-3062-7

## DIMENSIONS OF INTERNET SCIENCE



Internet Science is a new and exciting interdisciplinary field. Its purpose is the conduct of empirical studies which examine the Internet as both an instrument for, and an object of, scientific investigation. This book is the first comprehensive collection of contributions to Internet Science appearing in English, written by experts from seven different countries. *Dimensions of Internet Science* consists of twenty contributions, covering the following topics: Issues in Net-based Survey Research, Psychological Web Experiments and Web Questionnaire Studies, Online Communication Research and E-Commerce, Knowledge Acquisition and Learning with the Net, Studying Perception on the Net.

DIMENSIONS OF INTERNET SCIENCE  
Ulf-Dietrich Reips & Michael Bosnjak (Eds.)  
Lengerich: Pabst Science Publishers, 2001  
347 pages, 48 DM (24.- €), ISBN 3-935357-52-4



## STATISTIK FÜR DIE SOZIALWISSENSCHAFTEN



Dieses Lehrbuch bietet eine anwendungsorientierte Einführung in die sozialwissenschaftliche Statistik und stellt die Grundzüge der statistischen Datenanalyse dar. Um die Voraussetzungen und Interpretationsmöglichkeiten von Analyseverfahren zu schaffen, wird stets die Beziehung zu der zugrunde liegenden Logik statistischer Argumentation hergestellt. Statt mit komplexen mathematischen Ableitungen wird mit Analogien und Beispielen gearbeitet. Da Datenanalysen letztlich generelle Aussagen anstreben, steht die Vermittlung umfassender Informationen über die Möglichkeiten und Grenzen der Generalisierung von Stichprobendaten auf Eigenschaften einer Population im Mittelpunkt. Die Methoden zur Datenanalyse werden anhand von Anwendungsbeispielen aus der sozialwissenschaftlichen Umfrageforschung erläutert, wobei die gleichen Daten unter unterschiedlichen Fragestellungen und mit unterschiedlichen Auswertungsverfahren wiederholt analysiert werden. Abschließend wird ein Überblick über multivariate Analysemodelle gegeben.

*STATISTIK FÜR DIE SOZIALWISSENSCHAFTEN.*

Steffen Kühnel & Dagmar Krebs

Reinbeck bei Hamburg: Rowohlt Verlag, 2001

654 Seiten, 34,90 DM (17,90 €), ISBN 3-499-55639-1