

## Identification risk for microdata stemming from official statistics

Blien, Uwe; Wirth, Heike; Müller, Michael

Veröffentlichungsversion / Published Version

Forschungsbericht / research report

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Blien, U., Wirth, H., & Müller, M. (1991). *Identification risk for microdata stemming from official statistics*. (ZUMA-Arbeitsbericht, 1991/11). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-68945>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

**Identification Risk for Microdata Stemming  
from Official Statistics**

**Uwe Blien / Heike Wirth / Michael Müller**

ZUMA-Arbeitsbericht Nr. 91-11

Zentrum für Umfragen, Methoden und  
Analysen e. V. (ZUMA)

Postfach 12 21 55  
D-6800 Mannheim 1



Seit Juli 1983 sind die ZUMA-Arbeitsberichte in zwei Reihen aufgeteilt:

Die **ZUMA-Arbeitsberichte** (neue Folge) haben eine hausinterne Begutachtung durchlaufen und werden vom Geschäftsführenden Direktor zusammen mit den übrigen Wissenschaftlichen Leitern herausgegeben. Die Berichte dieser Reihe sind zur allgemeinen Weitergabe nach außen bestimmt.

Die **ZUMA-Technischen Berichte** dienen zur hausinternen Kommunikation bzw. zur Unterrichtung externer Kooperationspartner. Sie sind nicht zur allgemeinen Weitergabe bestimmt.



Uwe Blien  
Heike Wirth  
Michael Müller  
Mannheim/ Nürnberg  
3. 12. 1990

Identification Risk for Microdata Stemming from Official  
Statistics

1. Introduction
2. Using a social science scenario in testing factual anonymity
3. The data
4. The simple-matching method
5. Results (I)
6. Paass' method based on discriminant analysis
7. Results (II)
8. Conclusions

Addresses of the authors:

Uwe Blien

Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt  
für Arbeit (IAB - Institute for Employment Research of the  
Federal Office of Labour)  
Regensburger Str. 104  
D-8500 Nürnberg 30

Heike Wirth

Zentrum für Umfragen, Methoden und Analysen  
(ZUMA - Center for Survey, Methodology & Analysis)  
B2, 1  
D-6800 Mannheim

Michael Müller

University of Mannheim  
Fakultät für Soziologie  
A5  
D-6800 Mannheim



## 1. Introduction

If a number of legal restrictions are observed microdata generated by German official statistics can be released for purposes of social scientific research. The Federal Act on Statistics (§16 Abs. 6 BStatG 1987) stipulates factual anonymity rather than absolute anonymity of the respondents. In the case of factual anonymity the cost in terms of manpower, time, and money of identifying an individual in a released microdata file (MDF) exceeds the returns.

In order to clarify the conditions under which factual anonymity of released microdata can be guaranteed, the University of Mannheim (Prof. Walter Müller), the German Federal Office of Statistics (StBA, Wiesbaden) and ZUMA carried out a research project. The identification risk for individuals in a released microdata file was tested under realistic conditions by using empirical data.

Since the Federal Act on Statistics is based on the notion of identity disclosure (cf. Duncan, Lambert 1989, p. 208), this concept was used for the "Research Project on Data Anonymity". Identity disclosure or "Identification of an individual takes place when a one-to-one relationship between a record in released statistical information and a specific individual can be established" (Bethlehem, Keller, Pannekoek 1990, p. 38). For the identification of an individual whose data are in a file of microdata stemming from official statistics, a potential intruder needs so-called "prior knowledge" (or "identifying information") about persons which has some variables in common with the microdata file. The prior knowledge is organized in an identification file (IF). This has to include formal identifiers such as names and addresses of the individuals to which the data apply. The identification of an individual can only be achieved by creating a one-to-one association between unique records of the microdata file and the identification file.

## 2. Using a social science scenario in testing factual anonymity

Under the regulations of the German Federal Act on Statistics factual anonymous microdata may only be released for scientific purposes. Thus the aim of the Research Project on Data Anonymity was to determine the disclosure risk and the cost of identity disclosure in the framework of a scientific scenario. One main contribution of the project was the analysis of potential motives and benefits of identity disclosure and the information content of prior knowledge from sources the scientists have access to.

On the basis of these results several subscenarios were developed which were intended to cover a wide range of possible situations where there might be a danger for data anonymity in the scientific scenario. For two of these subscenarios identification experiments with empirical data were carried out. The research project has not been finished yet. So in this paper only a description of the results of one specific subscenario is given.

In this subscenario, it was assumed that the intention of the intruder was a mass identification. It was assumed that the intruder carried out a "fishing operation", i. e. tried to identify as much individuals as possible in a released microdata file irrespective of the persons identified. An additional assumption was that the intruder did not know whose data are in the MDF.

## 3. The data

As a microdata file (MDF) in the experiments the 1987 microcensus of the German state North Rhine-Westfalia could be used. The microcensus is a survey of 1 % of the total population (approximately 17 million persons). The file given to the project group contained 169 368 records with a selection of variables needed for the experiments. Only the formal identifiers (names and addresses) had been removed from the MDF. There had been no other measures to secure the anonymity of the respondents.

To find or to construct a corresponding IF a comprehensive list of sources of prior knowledge an intruder could use was compiled. It was ascertained that handbooks about professional groups like architects, jurists, scientists etc. might be a relatively

comfortable source of prior knowledge. These handbooks often comprise the whole subpopulation. They provide detailed information on attributes like region, profession, age, sex etc. As these variables are also included in the microcensus, they are common variables and can be regarded as key variables (cf. Dalenius 1986) in the MDF. For the experiments "Kürschners Deutscher Gelehrtenkalender", a handbook about German scientists and scholars, was selected. Nearly the total subpopulation of German university professors and parts of the subpopulation of other scientists were included.

The information included in this handbook comprised 10 attributes which were also in the microcensus. This was the highest number of common variables of all the relevant handbooks. The common variables were:

- region ("Bundesraumordnungsregion")
- size of place of residence
- sex
- year of birth
- half-year of birth
- industry
- occupation
- occupational status
- specification of the professional activity
- subject of the university degree

The information of the "Gelehrtenkalender" handbook was not formalized but in the form of uncoded text. The German Federal Office of Statistics did the coding for the identification experiments according to microcensus categories. Thus, a high comparability between the codes of the MDF and the IF was guaranteed.

A major problem arising during the coding work were the ambiguities of many statements. Not in all cases was it possible to define exactly the category of an individual. For example, there were heads of university hospitals who were university professors as well as physicians. It was unknown which these respondents would had regarded as their primary occupation if they had been interviewed in microcensus. In all these cases a second variable was coded. With four key variables ambiguities were important. Where there was no ambiguity the alternativ variables was set to

the "missing" value.

The IF contained 7 983 records from the German Land North Rhine-Westfalia.

Taking empirical data for the experiments did not involve making real identifications. To control the identification results a costly method ("double blind procedure") was developed. The project group in Mannheim (ZUMA/ University) received the data files without formal identifiers for the experiments. Each record had a definite number. After the construction of an association between single records in the IF and the MDF by means of an identification method these numbers were sent to a neutral person called "data trustee". He had a file with the numbers and the addresses of the respondents but no further information. This enabled him to control how many people had given data for both files and to control whether the "identifications" were correct.

#### 4. The simple-matching method

The simple-matching method is based on a comparison of the values of the key variables in the MDF and the IF. If an intruder succeeds in finding a unique value set in a MDF which is completely identical to a value set in his IF this record is regarded as identified.<sup>1</sup>

The simplicity of the method is an advantage from the standpoint of the intruder with regard to the necessary efforts, but it is also its weak point. From the analysis of Paass (Paass 1988, Paass, Wauschkuhn 1985) it is known, that there are two problems an hypothetical intruder has to deal with:

- There is a broad variety of incompatibilities between the data of the two used files. These deviations may be the result of "noise" in the data, of coding errors, reliability problems, different definitions of categories, and different points of time the data refer to.
- In most cases the intruder only has data based on samples. Thus he does not know whether there are cases in the unknown rest of

---

<sup>1</sup> A general description of the realization of the matching process is given by Dalenius (1986).

the population which have the same attributes as those he wants to identify.<sup>2</sup>

If the data files are not completely comparable and/or the files are only samples, a simple-matching method may fail. No statistical estimation is provided, which can deal with data errors or the possibility of "statistical twins" in the population when different people have the same values in the relevant variables. This method leaves an intruder uncertain about the reliability of his identification efforts.

### 5. Results (I)

For testing the simple-matching method the matching procedure of SPSS<sup>^</sup> was used. The IF contained 10 key variables and 4 alternative key variables. Since the SPSS<sup>^</sup> matching does not allow to take these alternative variables into account, 2<sup>4</sup> variations of the "Gelehrtenkalender" were created. Thus there were 16 different part-identification files. Each part-IF included only one of the alternatives in the relevant four variables respectively.

The data base revealed a high identification risk. Between 45 and 65 % of the records of the part-IFs were unique. The microdata file had 36 % unique records. If only the relevant subpopulation was regarded, the rate of unique cases was higher. To do this, only cases of the MDF were included which had values in the variables which were also in the IF (because in simple matching only these values are important). The reduced MDF had 3 099 records. The rate of uniqueness in this file was 79 % (see Table 1). Since the "Gelehrtenkalender" contained the total subpopulation of the German university professors, a high identification risk was to be expected provided that the information in the data files was completely comparable.

In matching all 16 part-IFs with the microdata file only 14 records corresponding without ambiguity on a one-to-one level were found (Table 1). The checking of the record numbers by the data trustee showed that only 4 of the 14 matches were correct. The simple method failed with a probability of 0.71.

---

<sup>2</sup> But there are statistical methods to estimate the rate of unique cases in the population (Bethlehem et al. 1990 and Mokken et al. 1990).

Table 1: Results of experiments realized with a simple-matching procedure using data of "Kürschners Deutschem Gelehrtenkalender" ("IF") and Microcensus of Nordrhein-Westfalen (MDF).

Rate of unique records in the part-IFs (a)  
and quantity of one-to-one associations (b)

	MDF		No. of the part-IFs															
	total	red.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N	169368	3099	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983	7983
a)	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	38.7	79.3	62.7	45.1	45.2	51.5	64.5	51.5	64.5	62.8	51.4	64.4	64.4	51.4	50.8	50.8	64.1	64.1
b)	14		1	-	3	-	-	-	-	7	-	3	-	-	-	-	-	-

4 of the total of 14 one-to-one associations were correct (28.6%)

The data trustee ascertained that the data of 53 individuals were present in both files. Only 4 of these could be related on a one-to-one level in the experiments.

In addition to these one-to-one associations there were ambiguous associations. 13 cases of the MDF were related to more than one case of the "Gelehrtenkalender". 2 were related to one case in the IF. Six of these additional associations were correct. But under the assumptions made there were no further identity disclosures. Because of the ambiguity of the associations the identity of the respondents was not uncovered. 81% of the persons whose data were in both files had incompatible values in the key variables. Thus these 43 records could not have been identified with a simple matching method. Even if there had been additional information (i. e. more variables) an identification with a simple-matching method would have been impossible.

The cost of the experiments carried out were high. The preparation of the identification file by the Federal Office of Statistics costed about 44 000 DM. The costs of developing the identification concept and of computing amounted to 17 000 DM. Hence it costed about 61 000 DM to determine 14 one-to-

one relations between records in the MDF and the IF.

#### 6. Paass' method based on discriminant analysis

To overcome the weaknesses of simple-matching methods and to carry out realistic tests of data anonymity a procedure based on discriminant analysis (PDA) was developed by the "German National Research Center for Computer Science" ("Gesellschaft für Mathematik und Datenverarbeitung" - GMD) in St. Augustin. The method was conceived to solve the mentioned problems of a hypothetical data aggressor.

To overcome the difficulties with "statistical twins" and data incompatibilities identification is regarded as a statistical decision problem. Thus, the method of Paass has the important advantage of explicit computing the probability of correctly relating two records, which makes it a very powerful tool. Only a short description of the method was published in English (Paass 1988). The report of the whole research project is only available in German language (Paass, Wauschkuhn 1985). The long version shows the enormous complexity of the method. Only a short overview is possible in this context.

- i Every case in the MDF is treated as a class of its own in discriminant analysis.
- ii To do this, probability distributions within classes are estimated. These distributions depend on assumptions about data incompatibilities. Information describing these deviations has to be used as input for the estimation.
- iii The method provides an estimation of unconditional probabilities to overcome the problem of "statistical twins".
- iv To compute the probability that one record from the MDF and one record from the IF stem from the same person, the results of steps ii and iii have to be related to each other. This is done in a Bayes approach typical for linear discriminant analysis.
- v The intruder has to compare this probability with an a priori chosen certainty level (e. g. 99 %). If the computed probability is higher he will accept the relation between the two cases as an identification.

The statistical idea of the procedure is shown by the chosen

Bayes approach. From the theory of discriminant analysis it is known that the probability of a case  $z_j$  belonging to a class  $y_a$  can be given by the following expression:

$$p(y_a | z_j) = \frac{p(z_j | y_a) p(y_a)}{\sum_i p(z_j | y_i) p(y_i)}$$

A  $z_j$  is a case from the IF and a  $y_a$  one from the MDF. The  $p(z_j | y_i)$  (the class probability) gives the prior probability of a case  $z_j$  stemming from the same person as  $y_i$ . This conditional probability is computed on the basis of the distributions<sup>3</sup> of data errors and data incompatibilities the intruder has to specify (step ii). Expression (1) is equivalent to (2), as can be shown using computing rules for conditional probabilities.

$$p(y_a | z_j) = \frac{p(z_j | y_a) p(y_a)}{p(z_j)} \quad (2)$$

The intruder has to compute  $p(z_j)$ , which is the unconditional probability to find a record  $z_j$ .

If the intruder has only samples serving as IF and as MDF he has to use an estimator for  $p(z_j)$ . It is shown at length in Paass, Wauschkuhn (1985, pp. 112ff.) that instead of (2) expression (3) can be used (if  $p(y_a)$  is equal for all cases a).

$$p(y_a | z_j) = \frac{p(z_j | y_a)}{p(z_j | y_a) + (N-1)p_w(z_j)} \quad (3)$$

To compute the conditional probabilities  $p(z_j | y_a)$  and especially the estimator  $p_w$  is a very tedious task. The GMD developed a computer program accomplishing it. The program combines many different statistical and numerical procedures to estimate the different probability densities.

The PDA was applied by Paass and Wauschkuhn in their disclosure-risk experiments. As an MDF they used the German Earnings and

---

<sup>3</sup> The word "distribution" is used to describe the case of discrete and continuous variables by only one term.

Consumption Sample (Einkommens- und Verbrauchsstichprobe - EVS). As an IF they used a sample synthetically generated out of these data by a modification with stochastic errors. For the distributions and the magnitudes of these errors they made assumptions based on estimations by experts from the Federal Office of Statistics and other institutions (Paaß, Wauschkuhn 1985, p. 177).

In the experiments done by the GMD the PDA was very effective. The simulations were based on different types of scenarios. In these, different assumptions concerning the motive of the intruder, the amount of common variables and the probability and the structure of data errors were tested. If there were about 40 common variables nearly every second of persons whose data were in both files could be identified. Even if there were only 15 variables the data were not safe.

In the international discussion on data confidentiality the results have been noticed. It has been taken for granted that an identification under a broad range of conditions is possible using methods like the PDA.

## 7. Results (II)

In the project on data anonymity the PDA was applied using two different files with empirical data as an MDF and as an IF. To keep the results comparable to those obtained by simple-matching procedures the same files were used: the Microcensus of Nordrhein-Westfalen of 1987 and "Kürschners deutscher Gelehrtenkalender".

To implement the PDA program on the available computer and to do adaptations for the structure of the used data we needed about a year because of the complexity of the method. The whole cost of computation during the time of implementation and application of the method, without the cost of working time and of the preparation of the data were about 140 000 DM. Thus the PDA includes no real danger for the factual anonymity of official data in the context of a social science scenario.

One problem arising in the application of the method was that double-valued variables could not be dealt with properly. It was impossible to check all possible combinations as we did in the case of the simple matching procedures. A run of the PDA program needs

too much computing time. So it was tried to find the most plausible combination of the alternative variables. Finally the combination No. 8 was chosen, which had had the most one-to-one relations in the experiments with the simple matching procedure. Thus the research design had two steps: it was assumed that a supposed intruder at first tries simple matching procedures and then the PDA.

To realize an experiment with the PDA we had to specify the properties of data errors. To find an empirical basis for this task, the data of a study on reliability (ALLBUS 1984) were analysed (cf. Porst, Zeifang 1987).<sup>4</sup> But besides reliability problems there are other sources of data incompatibilities. So assumptions had to be made. To control for the effects of these assumptions we worked with two levels of errors. Experiment No. 1 was done with error specifications related only to reliability problems. A higher level of errors based on plausible assumptions was the starting point of experiment No. 2.

Table 2 Results of experiments realized with Paass' method based on discriminant analysis (PDA) using data of "Kürschners Deutschem Gelehrtenkalender" ("IF") and of the Microcensus of North Rhine-Westfalia (MDF). Chosen certainty level of a correct "identification": 99%		
No. of Experiment	1	2
level of errors with respect to incompatibilities	reliability problems only	different sources of data
number of "identifications"	9	11
correct "identifications"	2	2
percentage of correct "ident."	22	18

<sup>4</sup> Special research on errors in the microcensus (cf. Deininger 1960, Herberger 1985, StBA 1978) was used additionally.

Table 2 shows the results obtained in the experiments. When it was worked with the simple matching method using file No. 8 as "IF" 7 one-to-one relationships between records of the "IF" and the MDF were found, three of these were correct. The "rate of success" was 42.9 %. Compared with this the PDA was less "successful".

In the experiments of the GMD the method was very reliable. There were only a few cases which were wrongly associated. Tests realized in Mannheim with the same data constellation, i. e. using a synthetic IF, yielded similar results. But the outcome of the experiments with two different files were completely different. What are the causes of these differences? A definite answer cannot be given, because there are four possibilities.

- The PDA was not designed to deal with ambiguous information resulting in the construction of double-valued variables. Taking a specific combination of alternatives for granted leads to an overestimation of the probability of a correct relation.
- The method was originally tested with variables most of which were quantitative. Most of the microcensus variables were, however, quantitative. Under the used common variables there was only one with a metric level of scale.<sup>5</sup>
- The level of errors is higher as was expected in the experiments of the GMD. This can be shown from studies on reliability.
- The structure of the incompatibilities is completely different from those assumed. The mentioned studies on reliability indicate that the distributions of data errors are often not of a simple type (for example normal distribution). But there is a lack of information about the parameters of complex multivariate distributions.

---

<sup>5</sup> In the report of Paass and Wauschkuhn it was argued, referring to experiments of Titterington et al. (1981), that the linear discriminant analysis shows good results in the case of nominal variables too. This was confirmed by the tests with a synthetic IF in Mannheim.

## 8. Conclusions

What kind of conclusions can be drawn from the results obtained in the experiments in Mannheim?

The experiments show the unexpected finding that identity disclosure is not as easy to do as it is often assumed (cf. for instance Steinmüller 1984). Most of the one-to-one associations of records established in the simulations of "fishing operations" were simply wrong.

Thus it can be concluded that an identification is possible under special circumstances. But following the definition of factual anonymity there was no single case of identity disclosure. The one-to-one relationships of records gotten from the simulations cannot be regarded as identifications by a rational intruder in a social science context. He cannot distinguish these from the overwhelming majority of wrong associations.

The costs of the used methods and experiments were considerable. In the assumed scenario it would be cheaper to try an alternative way to get the wanted information. Thus a rational person would pay a detective for the job or try another method (for instance organize a survey if he really wants the data of a lot of persons).

The calculation of an intruder was analysed in the project on data anonymity. It cannot be outlined in this paper. But the mentioned costs make it plausible that the factual anonymity of microdata is secured under a range of conditions.

The results of the experiments show that assessments of the risk of data confidentiality based on the counting of unique cases in the sample or on an estimation of the rate of unique cases in the population should be supplemented by an estimation of the frequency of errors in the data and of other incompatibilities. The empirical experiments indicate that these deviations are a "natural barrier" against intruders. They have the same effects as deliberately chosen measures to secure the anonymity of the data.

## References:

- Bethlehem, Jelke G.; Keller, Wouter J.; Pannekoek, Jeroen (1990): "Disclosure Control of Microdata", in: Journal of the American Statistical Association 85/409, pp. 38 - 45
- Dalenius, T. (1986): "Finding a Needle in a Haystack or Identifying Anonymous Census Records", in: Journal of Official Statistics 2, pp. 329 - 336
- Deininger, Rolf (1960): "Repräsentativstatistik der Bevölkerung und des Erwerbslebens", in: Statistisches Bundesamt (1960) (ed.): "Stichproben in der amtlichen Statistik", Wiesbaden
- Duncan, George; Lambert, Diane (1989): "The Risk of Disclosure for Microdata", in: Journal of Business and Economic Statistics 7/2, pp. 207 - 217
- Herberger, Lothar (1985): "Aktualität und Genauigkeit der repräsentativen Statistik der Bevölkerung und des Erwerbslebens", in: Allgemeines Statistisches Archiv 69/1, pp. 16 - 55
- Mokken, Robert J.; Pannekoek, Jeroen; Willenborg, Leon (1990): "Microdata and disclosure risks", unpublished paper presented on the annual research conference of the Bureau of the Census, Washington, D. C.
- Paaß, Gerhard (1988): "Disclosure Risk and Disclosure Avoidance for Microdata", in: Journal of Business and Economic Statistics 6/4, pp. 487 - 500
- Paaß, Gerhard; Wauschkuhn, Udo (1985): "Datenzugang, Datenschutz und Anonymisierung. Analysepotential und Identifizierbarkeit von anonymisierten Individualdaten", München, Wien: R. Oldenbourg
- Porst, Rolf; Zeifang, Klaus (1987): "A Description of the German General Social Survey Test-Retest Study and a Report on the Stabilities of the Sociodemographic Variables", in: Bohrnstedt, George W.; Mohler, Peter P.; Müller, Walter (1987) (ed.): "An Empirical Study of the Reliability and Stability of Survey Research Items", Sociological Methods & Research 15/3 (Special Issue)
- Statistisches Bundesamt (1978): "Volkszählung vom 27. Mai 1970; Untersuchungen zur Methode und Genauigkeit der Volkszählung 1970; Fachserie 1, Heft 26": Wiesbaden
- Steinmüller, Wilhelm (1984): "Amtliche Statistik und Technisierung der Verwaltung. Nachtrag zur Volkszählung 1983", in: Leviathan, pp. 176 - 194
- Titterington, D. M. et al. (1981): "Comparison of Discrimination Techniques Applied to a Complex Data Set of Head Injured Patients", in: Journal of the Royal Statistical Society A 144, Part 2, pp. 145 - 175

## ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler  
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-  
Arbeitstagung vom 25.09. - 05.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter  
Schmidt  
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman  
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler  
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll  
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler  
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler  
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst  
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu  
einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler  
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulationspro-  
gramm
- 82/14 Cornelia Krauth  
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer  
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener  
Two approaches to the analysis of judgments of prestige: Interindivi-  
dual differences and the general scale
- 83/11 Rolf Porst  
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragenpro-  
gramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und ALLBUS  
1982
- 84/01 Manfred Küchler, Peter Ph. Mohler  
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen  
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,  
Teil I + II
- 84/02 Bernd Wegener  
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-  
Prestige-Skala

- 84/03 Peter Prüfer, Margrit Rexroth  
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum  
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik  
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler  
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique
- 85/01 Hartmut Esser  
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang  
Methodenbericht zum "ALLBUS 1984"
- 86/01 Dagmar Krebs  
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich
- 86/02 Hartmut Esser  
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung
- 86/03 Bernd Wegener  
Prestige and Status as Function of Unit Size
- 86/04 Frank Faulbaum  
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels
- 86/05 Peter Prüfer, Margrit Rexroth (Übersetzung: Dorothy Duncan)  
On the Use of the Interaction Coding Technique
- 86/06 Hans-Peter Kirschner  
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung
- 86/07 Georg Erdmann  
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle
- 86/09 Heiner Ritter  
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS
- 86/11 Günter Rothe  
Bootstrap in generalisierten linearen Modellen
- 87/01 Klaus Zeifang  
Die Test-Retest-Studie zum ALLBUS 1984 - Tabellenband

- 87/02 Klaus Zeifang  
Die Test-Retest-Studie zum ALLBUS 1984 - Abschlußbericht
- 87/04 Barbara Erbslöh, Michael Wiedenbeck  
Methodenbericht zum "ALLBUS 1986"
- 87/05 Norbert Schwarz, Julia Bienias  
What Mediates the Impact of Response Alternatives on Behavioral Reports?
- 87/06 Norbert Schwarz, Fritz Strack, Gesine Müller, Brigitte Chassein  
The Range of Response Alternatives May Determine the Meaning of the Question: Further Evidence on Informative Functions of Response Alternatives
- 87/07 Fritz Strack, Leonard L. Martin, Norbert Schwarz  
The Context Paradox in Attitude Surveys: Assimilation or Contrast?
- 87/08 Gudmund R. Iversen  
Introduction to Contextual Analysis
- 87/09 Seymour Sudman, Norbert Schwarz  
Contributions of Cognitive Psychology to Data Collection in Marketing Research
- 87/10 Norbert Schwarz, Fritz Strack, Denis Hilton, Gabi Naderer  
Base-Rates, Representativeness, and the Logic of Conversation
- 87/11 George F. Bishop, Hans-Jürgen Hippler, Norbert Schwarz, Fritz Strack  
A Comparison of Response Effects in Self-Administered and Telephone Surveys
- 87/12 Norbert Schwarz  
Stimmung als Information. Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile
- 88/01 Antje Nebel, Fritz Strack, Norbert Schwarz  
Tests als Treatment: Wie die psychologische Messung ihren Gegenstand verändert
- 88/02 Gerd Bohner, Herbert Bless, Norbert Schwarz, Fritz Strack  
What Triggers Causal Attributions? The Impact of Valence and Subjective Probability
- 88/03 Norbert Schwarz, Fritz Strack  
The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction
- 88/04 Hans-Jürgen Hippler, Norbert Schwarz  
"No Opinion"-Filters: A Cognitive Perspective
- 88/05 Norbert Schwarz, Fritz Strack  
Evaluating One's Life: A Judgment of Subjective Well-Being
- 88/06 Norbert Schwarz, Herbert Bless, Gerd Bohner, Uwe Harlacher, Margit Kellenbenz  
Response Scales as Frames of Reference:  
The Impact of Frequency Range on Diagnostic Judgments

- 88/07 Michael Braun  
Allbus-Bibliographie (7. Fassung, Stand: 30.6.88)
- 88/08 Günter Rothe  
Ein Ansatz zur Konstruktion inferenzstatistisch verwertbarer Indices
- 88/09 Ute Hauck, Reiner Trometer  
Methodenbericht  
International Social Survey Program - ISSP 1987
- 88/10 Norbert Schwarz  
Assessing frequency reports of mundane behaviors:  
Contributions of cognitive psychology to questionnaire  
construction
- 88/11 Norbert Schwarz, B. Scheuring (sub.)  
Judgments of relationship satisfaction: Inter- and intraindividual  
comparison strategies as a function of questionnaire structure
- 88/12 Rolf Porst, Michael Schneid  
Ausfälle und Verweigerungen bei Panelbefragungen  
- Ein Beispiel -
- 88/13 Cornelia Züll  
SPSS-X. Anmerkungen zur Siemens BS2000 Version
- 88/14 Michael Schneid  
Datenerhebung am PC - Vergleich der Interviewprogramme "interv<sup>+</sup>"  
und "THIS"
- 88/15 Norbert Schwarz, Bettina Scheuring  
Die Vergleichsrichtung bestimmt das Ergebnis  
von Vergleichsprozessen:  
Ist - Idealdiskrepanzen in der Partnerwahrnehmung
- 89/01 Norbert Schwarz, George F. Bishop, Hans-J. Hippler, Fritz Strack  
Psychological Sources Of Response Effects in Self-Administered  
And Telephone Surveys
- 89/02 Michael Braun, Reiner Trometer, Michael Wiedenbeck,  
Methodenbericht. Allgemeine Bevölkerungsumfrage der  
Sozialwissenschaften - ALLBUS 1988 -
- 89/03 Norbert Schwarz  
Feelings as Information:  
Informational and Motivational Functions of Affective States
- 89/04 Günter Rothe  
Jackknife and Bootstrap:  
Resampling-Verfahren zur Genauigkeitsschätzung  
von Parameterschätzungen
- 89/05 Herbert Bless, Gerd Bohner, Norbert Schwarz und Fritz Strack  
Happy and Mindless?  
Moods and the Processing of Persuasive Communications
- 89/06 Gerd Bohner, Norbert Schwarz und Stefan E. Hormuth  
Die Stimmungs-Skala: Eine deutsche Version des "Mood Survey"  
von Underwood und Froming

- 89/07 Ulrich Mueller  
Evolutionary Fundamentals of Social Inequality, Dominance  
and Cooperation
- 89/08 Robert Huckfeldt  
Noncompliance and the Limits of Coercion:  
The Problematic Enforcement of Unpopular Laws
- 89/09 Peter Ph. Mohler, Katja Frehsen und Ute Hauck  
CUI - Computerunterstützte Inhaltsanalyse  
Grundzüge und Auswahlbibliographie zu neueren Anwendungen
- 89/10 Cornelia Züll, Peter Ph. Mohler  
Der General Inquirer III -  
Ein Dinosaurier für die historische Forschung
- 89/11 Fritz Strack, Norbert Schwarz, Brigitte Chassein, Dieter Kern,  
Dirk Wagner  
The Saliency of Comparison Standards and the Activation of  
Social Norms: Consequences for Judgments of Happiness and their  
Communication
- 89/12 Jutta Kreiselmaier, Rolf Forst  
Methodische Probleme bei der Durchführung telefonischer  
Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen,  
Ausschöpfung und Nonresponse, Qualität der Daten.
- 89/13 Rainer Mathes  
Modulsystem und Netzwerktechnik.  
Neuere inhaltsanalytische Verfahren zur Analyse von  
Kommunikationsinhalten.
- 89/14 Jutta Kreiselmaier, Peter Prüfer, Margrit Rexroth  
Der Interviewer im Pretest.  
Evaluation der Interviewerleistung und Entwurf eines  
neuen Pretestkonzepts. April 1989.
- 89/15 Henrik Tham  
Crime as a Social Indicator.
- 89/16 Ulrich Mueller  
Expanding the Theoretical and Methodological Framework of  
Social Dilemma Research
- 89/17 Hans-J. Hippler, Norbert Schwarz, Elisabeth Noelle-Neumann  
Response Order Effects in Dichotomous Questions:  
The Impact of Administration Mode
- 89/18 Norbert Schwarz, Hans-J. Hippler, Elisabeth Noelle-Neumann,  
Thomas Münkel  
Response Order Effects in Long Lists:  
Primacy, Recency, and Asymmetric Contrast Effects
- 89/19 Wolfgang Meyer  
Umweltberichterstattung in der Bundesrepublik Deutschland
- 89/20 Michael Braun, Reiner Trometer  
ALLBUS Bibliographie (8. Fassung, Stand: 30.6. 1989)

- 89/21      Günter Rothe  
Gewichtungen zur Anpassung an Statusvariablen.  
Eine Untersuchung am ALLBUS 1986
- 89/22      Norbert Schwarz, Thomas Münkel, Hans-J. Hippler  
What determines a "Perspective"?  
Contrast Effects as a Function of the Dimension  
Tapped by Preceding Questions
- 89/23      Norbert Schwarz, Andreas Bayer  
Variationen der Fragenreihenfolge als Instrument  
der Kausalitätsprüfung: Eine Untersuchung zur Neu-  
tralisationstheorie devianten Verhaltens
- 90/01      Norbert Schwarz, Fritz Strack, Hans-Peter Mai  
Assimilation and Contrast Effects in Part-Whole  
Question Sequences:  
A Conversational Logic Analysis
- 90/02      Norbert Schwarz, Fritz Strack, Hans-J. Hippler, George Bishop  
The Impact of Administration Mode on Response Effects in  
Survey Measurement
- 90/03      Norbert Schwarz, Herbert Bless, Gerd Bohner  
Mood and Persuasion: Affective States Influence the  
Processing of Persuasive Communications
- 90/04      Michael Braun, Reiner Trometer  
ALLBUS-Bibliographie 90
- 90/05      Norbert Schwarz, Fritz Strack  
Context Effects in Attitude Surveys:  
Applying Cognitive Theory to Social Research
- 90/06      Norbert Schwarz, Herbert Bless, Fritz Strack,  
Gisela Klumpp, Annette Simons  
Ease of Retrieval as Information:  
Another Look at the Availability Heuristic
- 90/07      Norbert Schwarz, Fritz Strack, Hans-J. Hippler  
Kognitionspsychologie und Umfrageforschung:  
Themen und Befunde eines interdisziplinären Forschungsgebietes
- 90/08      Norbert Schwarz, Hans-J. Hippler  
Response Alternatives:  
The Impact of their Choice and Presentation Order
- 90/09      Achim Koch  
Externe Vergleichsdaten zum ALLBUS 1984, 1986, 1988.
- 90/10      Norbert Schwarz, Bärbel Knäuper, Hans-J. Hippler,  
Elisabeth Noelle-Neumann, Leslie Clark  
Rating Scales:  
Numeric Values May Change the Meaning of Scale Labels
- 91/01      Denis J. Hilton  
Conversational Inference and Rational Judgment

- 91/02 Denis J. Hilton  
A Conversational Model of Causal Explanation
- 91/03 Joseph P. Forgas  
Mood Effects on Interpersonal Preferences:  
Evidence for Motivated Processing Strategies
- 91/04 Joseph P. Forgas  
Affective Influences on Interpersonal Perception
- 91/05 Norbert Schwarz, Herbert Bless  
Constructing Reality and Its Alternatives:  
An Inclusion / Exclusion Model of  
Assimilation and Contrast Effects in Social Judgment
- 91/06 Herbert Bless, Roland F. Fellhauer, Gerd Bohner, Norbert Schwarz  
Need for Cognition: Eine Skala zur Erfassung von Engagement und  
Freude bei Denkaufgaben
- 91/07 Norbert Schwarz, Bärbel Knäuper, E. Tory Higgins  
Der Einfluß von Rangordnungsaufgaben auf nachfolgende Denkprozesse:  
Zur Aktivierung prozeduraler Sets
- 91/08 Bettina Scheuring, Norbert Schwarz  
Selbstberichtete Verhaltens- und Symptommhäufigkeiten:  
Was Befragte aus Antwortvorgaben des Fragebogens lernen
- 91/09 Norbert Schwarz, Herbert Bless  
Scandals and the Public's Trust in Politicians:  
Assimilation and Contrast Effects
- 91/10 Rolf Porst  
Ausfälle und Verweigerungen bei einer telefonischen Befragung