

Reasons for refusals, their collection in surveys and interviewer impact

Natalja Menold and Cornelia Zuell

GESIS-Working Papers 2010|11

Reasons for refusals, their collection in surveys and interviewer impact

Natalja Menold and Cornelia Zuell

GESIS-Working Papers

GESIS – Leibniz-Institut für Sozialwissenschaften

Postfach 12 21 55

68072 Mannheim

Telefon: (0621) 1246 - 286

Telefax: (0621) 1246 - 100

E-Mail: natalja.menold@gesis.org

ISSN: 1869-0483 (Print)

ISSN: 1869-0491 (Online)

Herausgeber,

Druck und Vertrieb: GESIS - Leibniz-Institut für Sozialwissenschaften
Lennéstraße 30, 53113 Bonn

ABSTRACT

Refusals are a significant source of non-response in surveys. During field periods of some surveys reasons for refusals are collected in call record data (as part of para-data). This article presents a study employing a content analysis of open-ended comments on reasons for refusals collected by interviewers in a survey of the German population (ALLBUS). We analysed the reasons for refusals contained in these comments, as well as to what extent these comments include information about factors relevant to participation in surveys. Additionally, we analysed the impact of interviewer characteristics – gender, age, education and experience – on data collection using various multilevel multinomial models. The results show that interviewer comments provide typical reasons for refusals, as well as specific information about target persons, their environment and the survey process. Interviewers' age and education influenced the collection of reasons for refusals. At the same time interviewer variances (obtained through multinomial multilevel models) were very high, showing that interviewers prefer to report certain reasons for refusals. The highest interviewer level variances were obtained for providing no comments at all. To improve data quality and reduce high interviewer impact, we suggest using improved standardised instruments to collect reasons for refusals. Codings based on a categorisation scheme which we developed for our content analysis show high reliability ($\kappa = .81$). Thus, this scheme can be used as a basis for developing such standardised instruments.

1 Introduction

Over the last few decades, response rates in surveys have been declining (De Leeuw & de Heer 2002). At the same time, reluctance to take part in surveys has been increasing (Groves & Heeringa 2006). Refusals may constitute a large portion of non-response: in the European Social Survey 2008 (ESS, 3rd round)¹ for example, refusal rates were 40% or higher. In the German General Social Survey (ALLBUS), from which we obtained data for the current study, the refusal rate was close to 50% in 2008.

Research on refusers and efforts to reduce refusals are important, since refusers may differ from other groups, such as respondents, non-contacts and reluctant respondents (Neller 2005; Reuband & Blasius 2000; Stoop 2004), in that significant refusal rates can lead to survey bias (e.g. Esser 1973; Groves et al. 2004; Reuband 1975; Zeh 1976). In some surveys – in this article we will focus on face-to-face surveys – reasons for refusals are collected by interviewers as a part of para-data. Para-data are data which document the process of data collection (Blom 2009; Kreuter & Kohler 2009). As a rule they are available for each case (each sampled person). Reasons for refusals are actually relatively rare information which a survey may have about non-respondents. This information can be used in two ways:

- 1) To conduct studies on non-respondents: Some researchers already use reasons for refusals, for example to predict future participation in the case of follow up contacts in statistical models (Bates, Dahlhamer & Singer 2008; Kreuter & Kohler 2009).
- 2) To conduct field monitoring in order to prevent refusals or reduce refusal rates. Possibilities in this context are:
 - a) Usage for refusal conversion: In particular, reasons for refusals could indicate how easily refusal conversion can be achieved and – as a consequence – in which cases refusal conversion may be effective. The highest rates of refusal conversion were obtained in cases of “no time”, “no interest”, “general refusal” or “prompt ringing off” (Fuse & Xie 2007; Neller 2005; Reuband & Blasius 2000; Schnauber & Daschmann 2008). Considering data protection issues in Germany, target persons who refused any participation are not to be contacted a second time (ADM²). Here it is necessary to have information about which sampled persons refused and for what reason, in order to comply with data protection regulations.
 - b) Usage by interviewers to improve doorstep vs. contact behaviour (Durrant & Steele 2009; Neller 2005): for example, Neller (2005) used conversion guidelines in which interviewers received different persuasion instructions for different reasons for refusals. Such instructions would be very helpful for interviewers, since the first few minutes of contact are critical in determining the level of success (Groves & Couper 1998).

¹ For more details see <http://www.europeansocialsurvey.org/>. Accessed 30 November 2010

² Compare: http://www.adm-ev.de/fileadmin/user_upload/PDFS/R07_D.pdf. ADM is a business association which represents the interests of the private-sector market and social research agencies in Germany and defines standards of norms and ethics for these areas (<http://www.adm-ev.de/index.php?id=2&L=1>; Accessed 30 November 2010)

c) Information about reasons for refusals can be used in follow up letters to convince initial refusers to participate. Furthermore, it is reasonable to consider whether reasons for refusals could be used by surveyors to create flexible incentives or flexible applications of interview mode (Stoop 2004).

Thus, reasons for refusals could be used in different ways in studies of non-response, as well as for survey field monitoring. AAPOR (2008) suggests collecting reasons for refusals in para-data as standard procedure. In this regard, standardised and reliable instruments of data collection would be helpful for surveyors. But there is currently a lack of such instruments, as well as studies which evaluate their quality.

In some surveys, reasons for refusals are collected using a closed answer format with different categories, such as “no time”, “not interested”, “never do surveys”, and so on (e.g. ESS, compare table 1). Other surveys do not apply such categories at all. Reasons for refusals are collected here in an open answer format (e.g. ALLBUS). An appropriate analysis of interviewer comments – a content analysis – is time-consuming. This restricts the usage of collected data not only for survey monitoring but also for research purposes.

Additionally, interviewers involved in collecting reasons for refusals could be a significant source of systematic error variance, reducing the validity and objectivity of collected data (Groves & Couper 1998). This is a particular problem if such data are collected in an open-ended format (Groves et al. 2004). Since it is necessary to obtain interviewer impact on responses in questionnaires – as interviewer related variance (Groves et al. 2004) – obtaining this impact on the collection of para-data has not been established in survey research until now.

In our study we address the topic of para-data quality (in our case these data are reasons for refusals). We developed a scheme to categorise reasons for refusals in ALLBUS 2008 data. We then analysed interviewer impact on collected reasons for refusals using different multilevel models. On the basis of our results we provide suggestions regarding standardised collection of reasons for refusals in surveys.

2 Theoretical Background

2.1 What do Reasons for Refusals Tell Surveyors?

Frequently named reasons for refusals are “no time”, “no interest”, “invasion of privacy”, “bad experience with interviews” (DeMaio 1980; Erblöh & Koch 1988; Groves & Cooper 1998; Költringer 1992; Neller 2005).

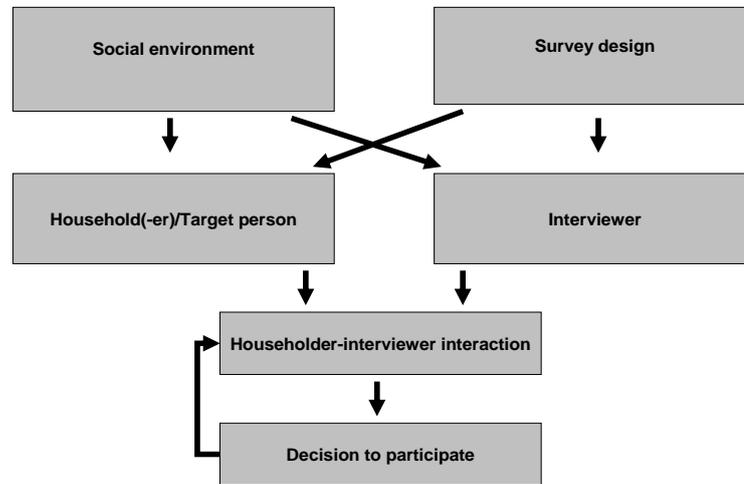
At first glance, one might think that such reasons say nothing to surveyors. This is particularly the case as reasons for refusals are often thought to be reactions which facilitate escape from participation by giving a general excuse which will clearly work. There is little research on reasons for refusals or on the trustfulness of information included in a particular comment. But the few studies that are available point out that reasons for refusals may actually be serious answers by respondents and not just general excuses (Schnauber & Daschman 2008).

Additionally, reasons for refusals are concomitants of refusal as behaviour. Refusal is not always a trait (stable personal behaviour), but rather involves spontaneous reactions which are highly dependent on situational factors. This is apparent since a portion of target persons who refused at one point in time often accept participation when re-contacted (e.g. Esser 1973; Reuband & Blasius 2000; Schnauber & Daschmann 2008; Stoop 2004; Zeh 1976). In a follow up study by Stoop (2004) a participation rate of 70% by former refusers was obtained. Reasons for refusals, such as “no time” and “no interest”, as well as very spontaneous reactions (prompt ring off), are associated with higher success rates in the case of re-contact than other reactions on the part of target persons (Fuse & Xie 2007; Neller 2005; Reuband & Blasius 2000; Schnauber & Daschmann 2008).

A decision to cooperate depends on a number of factors (Groves & Couper 1998), including the social environment of selected persons, structure of the household, characteristics of householders (socio-demographic characteristics, psychological predispositions), survey factors (e.g. topic, mode of administration), interviewer characteristics and householder-interviewer interaction (figure 1). Generally speaking, reasons for refusals may be connected with any of these factors. Regarding survey factors, for example, reasons for refusals may communicate the saliency of the survey topic, incentives and length of the interview (related to “leverage-saliency theory”, Groves, Singer & Corning 2000). Regarding characteristics of householders or their personal situation, reasons for refusals such as “waste of time” or “invasion of privacy” express a burden for these respondents. In this way, knowledge about these special factors (saliency of survey characteristics and burden on the part of householders) would be helpful in developing adaptive methods in order to prevent or reduce refusals.

In our study we will especially look at factors involved in decisions about survey participation, as defined by Groves and Cooper (1998, figure 1), in order to find out whether reasons for refusals in ALLBUS data provide information which could be useful for surveyors and researchers.

Figure 1: Factors involved in a decision to participate (adapted from Groves & Cooper 1998, p. 30)



2.2 What Is the Effect of Interviewers on the Collection of Reasons for Refusals?

Interviewers involved in collecting reasons for refusals may have an impact on the data, thus reducing data quality. In such a way interviewers could be a source of systematic measurement error in surveys, for example regarding substantial data (opinions and behaviours of target persons) and item non-response. This is particularly the case if standardised instruments of data collection are not used (see Groves et al. 2004).

Secondly, interviewers may have an impact on unit non-response. For example, it has been shown that several interviewer characteristics are correlated with non-response, particularly interviewer experience (e.g. Couper & Groves 1992; de Leeuw & Hox 1996; Durbin & Stuart 1951; Durrant et al. 2010) or attitudes, for example toward persuasion strategies (de Leeuw et al. 1997; Durrant et al. 2010; Lehtonen 1996). Regarding gender and age of interviewers, ambiguous results can be found in the literature: while some studies show that females and older persons provide the best response rates (Költringer 1992; Neller 2005; O’Muicheartaigh & Campanelli 1999), other researchers were not able to find any associations (Pickery & Loosveldt 2002; Schnauber & Daschmann 2008).

In general, Hox and colleagues (Hox 1994; Hox, de Leeuw & Kreft 1991; but also Durrant & Steele 2009) suggest that, in cases in which an interviewer was involved in data collection, interviewer impact should be taken into account using hierarchical multilevel regressions in order to accurately deal with the hierarchical data structure. In our analysis interviewer impact on collecting reasons for refusals was obtained by applying hierarchical models.

Prior to our analyses we considered the level of interviewer impact on the collection of para-data (reasons for refusals). In order to do this, we developed the following theoretical considerations regarding interviewer motivation to collect these data.

According to Crespi (1945) two groups of factors influence the motivation of interviewers to produce accurate survey data. Crespi (1945) assumed that interviewer motivation is influenced by ballot and administrative demoralisers. Ballot demoralisers are factors related purely to questionnaire design (e.g. length of interview, difficult and long questions). Administrative demoralisers are related to work

conditions (e.g. part-time work, payment) and factors associated with the surroundings. Applying this theory to collecting reasons for refusals, it can be concluded that poorly designed instruments of data collection (ballot factor) are associated with high interviewer error variance.

If we look at a more generalised process model of work motivation (Vroom 1964), effort toward or performance of a work task depends on 1) the expectation that an action (performed task) will bring a result, and 2) the valence (importance and usefulness) assumed by the person working toward this result. Applying this model to the task of collecting reasons for refusals in contact protocols, it is of central importance for the interviewer to consider the valence of this result – for themselves or for their organisation.

To summarise, Crespi's model (1945) suggests in particular that the quality of data collection instruments impacts interviewer performance. Vroom's model (1964) suggests that interviewers should know the importance of this task (collecting reasons for refusals) for their own benefit (e.g. payment, status) or for the benefit of surveyors (e.g. data quality). If the quality of the data collection instrument (here contact forms) is poor, and additionally if interviewers do not have any idea about the usability of these data, low task motivation would result and consequently high interviewer impact on the collected data would be expected.

3 Research Questions and Hypotheses

The first research question for our study is to determine which factors (social environment, personal situation of target persons, survey decisions, householder-interviewer interaction) were related to reasons for refusals. This helps to indicate how informative collected reasons are (for surveyors or for researchers interested in using these data). Regarding this question our hypothesis is:

H1: Interviewer comments regarding reasons for refusals contain information related to factors which influence decisions to participate.

The second question regards interviewer impact on collected data. We expected that collected reasons for refusals are correlated with interviewer gender, age, experience and education, and we expected to obtain high interviewer related variances. These are influenced by a non-standardised situation for data collection (open answer format) and – possibly – little knowledge about the usability of the collected data on the part of interviewers. Our hypotheses in this regard are:

H2: Interviewer characteristics (gender, age, experience and education) correlate with the collection of reasons for refusals.

H3: Interviewer impact on the collection of reasons for refusals is high.

4 Data

Our study is based on the ALLBUS 2008 survey. ALLBUS has been conducted since 1980 and is a biennial survey in which data is collected on attitudes, behaviour and the social structure of persons residing in Germany. A representative cross-section of the population is questioned using face-to-face interviews³. The sample in 2008 was a representative cross-section of the population based on a population register. A total of 3469 respondents participated in the survey, which was conducted between March and August of 2008. The refusal rate was high (48.7%). Thus, we used interviewer call records from this survey for our analysis of reasons for refusals.

Interviewers had to report all contact attempts in weekly contact protocols using CAPI. In cases of the outcome was no interview the interviewer had to provide one of ten possible contact attempt outcomes in the contact protocol: wrong address, respondent died, respondent moved, address is not a private household, no one at home, respondent not at home (but another person), refusal because of time concerns, refusal, don't speak German, not able to follow the interview. In cases of the outcome "refusal", interviewers were asked to write down the reason given for refusal as an open-ended comment in the CAPI protocol. In the protocols no information about the usability of open-ended comments was provided. Additionally, the provided list of outcomes was poorly structured: the various kinds of outcomes were not logically ordered/grouped here (e.g. a grouping ineligible address, non-contact, refusal would be possible). To collect reasons for refusals two possibilities were provided: a) a category for refusal due to time concerns, and b) a field to provide written comments in cases in which persons refused for another reason.

210 interviewers wrote call records including comments about the reasons given for refusals. The database contained 6061 CAPI protocols at the end of data collection. We used these open-ended comments for the analysis. All sampled persons in this data set were refusers who refused after one or three contacts.

³ For more details about the survey see <http://www.gesis.org/en/services/data/survey-data/allbus/>. Accessed 30 November 2010

5 Methods

5.1 Development of a Categorisation Scheme to Analyse Reasons for Refusals

As a first step to develop our categorization scheme we looked at the categories of reasons for refusals used in other surveys. Only ESS provides such data (survey para-data) for secondary research. Here data related to reasons for refusals are freely available. In contrast to ALLBUS, ESS uses categories, such as “bad timing”, “not interested”, “waste of money”, “waste of time” etc. (see table 1) when collecting reasons for refusals. Our analysis of data from the first three ESS rounds showed that, overall, a high percentage of the residual category “other” (ESS 1: 12.3%; ESS 2: 10%; ESS 3: 11.7%) is apparent. In the third ESS round in Germany, the category “other” amounts to 19% (430 cases). At the same time in Germany there are very low frequencies for other categories (e.g. 2.1% for the category “cooperated too often”, 1.4% for “do not trust surveys” and 0.1% for “previous bad experiences”). Comparable results can be seen for Finland and Norway (ESS 3) in table 1. High percentages for the category “other” and only marginal usage of other categories indicate low quality for these categories as a data collection instrument (compare Dillman 2007).

Table 1: Frequencies (N) and Percentages for Reasons for Refusal Collected in the ESS 3 for Finland, Norway and Germany

Reason	Finland		Norway		Germany	
	N	%	N	%	N	%
Bad timing, otherwise engaged	164	16.6%	71	6.4%	435	19.1%
Not interested	349	35.4%	581	52.1%	657	28.9%
Do not know subject. too difficult	30	3.0%	33	3.0%	-	-
Waste of time	113	11.5%	59	5.3%	-	-
Waste of money	12	1.2%	3	0.3%	-	-
Interferes with my privacy	42	4.3%	31	2.8%	94	4.1%
Never do surveys	69	7.0%	93	8.3%	405	17.8%
Cooperated too often	8	0.8%	17	1.5%	48	2.1%
Do not trust surveys	21	2.1%	6	0.5%	31	1.4%
Previous bad experience	5	0.5%	9	0.8%	2	0.1%
Do not like subject	39	4.0%	11	1.0%	79	3.5%
No approval to cooperate	11	1.1%	6	0.5%	91	4.0%
Other	122	12.4%	195	17.5%	430	18.9%

Note: Data source: <http://ess.nsd.uib.no/> Accessed 30 November 2010; Table presents results for first refusal

We coded ALLBUS 2008 data using ESS categories. We found that the ESS categories were not sufficient to code interviewer comments in ALLBUS (we found the same problems with ESS categories in Germany: high percentages for the category “other” and marginal percentages in some categories, e.g. “waste of time”, “cooperated too often”). But in ALLBUS 2008 data we found new categories and improved ESS categories by merging some categories together.

All open-ended call records (n = 6061 CAPI protocols) from ALLBUS 2008 were coded by a coder based on these improved categories. Up to three reasons for refusals could be coded per call record. Next, a 10% sample was coded by a second coder in order to calculate coding reliability (using the

formula proposed by Früh 2007). The intercoder reliability was $r = 0.85$, which is high considering the large number of categories and the quality of interviewer comments. Cohens' kappa was $k = 0.81$ (a high intercoder reliability as well).

Table 2: Frequencies (N) and Percentages of Reasons for Refusals in the ALLBUS 2008, Multiple Sets for Three Reasons

Main Categories	Subcategories	N	%
General (1)	a. General denial	2150	31.2%
	b. Not interested	1734	25.2%
	c. No time	1285	18.7%
	d. Participation interdicted/proxy refusal	312	4.5%
Vitality (2)	a. Age of target person (too old)	200	2.9%
	b. Health	181	2.6%
Political situation (3)	a. Dissatisfied with the political situation	81	1.2%
	b. Refusal due to being a foreigner	24	0.3%
Negative feeling about interviews (4)	a. Surveys are useless	179	2.6%
	b. Bad previous experience with interviews	61	0.9%
	c. Too many surveys nowadays	60	0.9%
Survey process (5)	a. Data protection and invasion of privacy	132	1.9%
	b. Methodology of the survey	85	1.2%
	c. Do not participate because participation is voluntary	88	1.3%
No comment (6)		280	4.1%
Other (7)	Other reasons	14	0.2%
	Not codable	23	0.3%
Total		6889	100%

Note: "No time" includes 444 openly given statements. Table presents results for all – first to third refusals

The final categorization scheme is comprised of 5 main categories, each composed of 14 subcategories (table 2). We added the following categories to those of ESS:

- General denial: all statements that refer to a general denial of interviews, for example the refusal of any interview. This includes statements like "is generally not willing to do the interview" and "generally no surveys". Comments by interviewers regarding interviewer-respondent interaction, such as "did not open the door" or "hung up on the intercom", were coded here as well.

- Dissatisfaction with the political situation: coded in this way if the respondent explicitly points out his or her dissatisfaction with politics, politicians or the state, for example, “not interested in this state anymore”, “government is incompetent”, or “they just raised the expense allowance”⁴.
- Data protection: statements expressing doubts about the confidentiality of survey data, such as “don’t want to give any data”, “distrust data protection”, or “this is too personal”.
- Voluntary participation: coded in this way if the respondent emphasises the volunteer aspect of participation, for example, “if I don’t have to”, or “it’s no obligation”.
- Age of target person/health condition: the respondent is too old (interviewer’s estimation) or feels too old (own declaration). Additionally, statements based on state of health are coded here, for example “is ill”; “terminal cancer”.
- Methodology of surveys: coded whenever survey methodology is the reason for refusal, for example “a respondent would participate in a written census but denies access”, “incentives”, or “the length of the interview”.

Figure 2: Examples of Category Definitions in the Categorisation Scheme for Description of Reasons for Refusals

General denial

Coded are all statements that refer to a **general denial** of interviews, i.e. the refusal of any interview. This includes statements like “is generally not willing to do the interview”, and „generally no surveys“. Statements such as „I don’t want to do this“ or “I don’t feel like doing this” or denial of access are coded here as well.

Specifications such as e.g. „refused, not interested“ are always coded with 101 „no interest“.

Examples:

- Denies access;
- Hung up the phone;
- Burst the appointment;
- General denial;
- Doesn’t want to be bothered;
- Even not for money;
- Is generally not about to do so;

Methodology of Surveys

To code is, if whenever the methodology of the survey is the reason of refusal. This includes the kind of questioning: for example a respondent would participate in a written census but denies access. Incentives can be coded here as well: a respondent would participate if a (higher) incentive/ gratification were offered. A third methodological effect is the length of the interview. The respondent does not like long interviews.

Examples:

- Interview lasts too long;
- 1 hour is too much time;
- Without presents or gifts, no participation;
- If at all then certainly only with payment;
- Participation in online-surveys only;
- Only written;
- Not personally.
- „We would do this if you pay back our debt (€ 3000)“

⁴ The main topics of ALLBUS 2008 were political issues: political attitudes (such as political interest), political participation (such as participation in elections), party work, citizens’ initiatives, the frequency of political discussions with family, friends, acquaintances or others, attitudes towards the political system or social inequality, and national pride.

Categories such as “waste of time” or “waste of money” (used in ESS) were merged into the category “surveys are useless”. Two examples of our category definitions can be found in figure 2. The full categorization scheme, definitions of categories and basic examples have been provided by Menold and Zuell (2010).

Additionally, we found comments (9.5% of all categorised statements) which did not refer to a refusal (table 3): target persons did not refuse, but rather signalled a willingness to participate, clearly indicating that there was no relation between the ticked refusal and the comment. Other comments unrelated to reasons for refusals indicated problems with survey organization, for example, double addresses; persons who had already participated in the survey or an appointment that had already been made by another interviewer. We added the categories, “problems of survey organisation/performance” and “no refusal, but willing or accessible later” in order to code such comments (see table 3). The high rate of misclassification (9.5% of all categories) by interviewers shows that contact forms used in ALLBUS are not optimal.

Table 3: Frequencies (*N*) in Categories in the ALLBUS 2008 with no Relation to a Reason for Refusal. Multiple Sets for Three Subcategories

Main category	Subcategory	N
Non contact/ Accessibility later	There is no refusal, interviewer could not contact the sampled person or sampled person signalled accessibility later	282
Problems of survey organisation/ performance	Address error	63
	New contact, but person generally refused before (e.g. called to survey field institute)	194
	Performance problems (e.g. no cover letter; interviewed anyway)	97
Total		636

Note: Table presents results for all – first to third refusals

5.2 Specification of Multinomial Multilevel Models to Obtain Interviewer Effect

To obtain interviewer effect, multinomial multilevel regressions were conducted with MLwiN 2.14. “The advantage of using multinomial models, rather than fitting separate binary logistic models for each type of non-response, is that the ...effects ...could be evaluated simultaneously and tested for equivalence” (Durrant & Steele 2009, p. 368). The procedure applied in the current analysis was described by Rasbash et al. (2009), and adapted by O’Muircheartaigh and Campanelli (1999) as well as by Durrant and Steele (2009).

As mentioned above we coded up to three reasons for refusals per case. For the third reason only 43 cases were obtained. There were about 400 cases for the second reason (see table 4). As a result we used only data related to the first reason for refusal to analyse interviewer effects, since there were not enough data for the second and the third reasons for refusal to fit multinomial models. Additionally, information about second refusals (in cases of follow up contacts) was available in the data set⁵. To

⁵ In addition there were $n = 30$ cases which were contacted a third time. We could not use these in our analysis due to low numbers in cases of third contact attempts (second follow-up contact).

consider reasons for refusals during first contact (1st refusal) and follow up contact (2nd refusal) we conducted separate analyses.

Table 4: Numbers of Comments for 1st and 2nd Reasons for Refusals during First (1st refusal) and Second (2nd refusal, follow up contact) Contact Attempts in the ALLBUS 2008

Main Category	1st reason		2nd reason	
	1 st refusal	2 nd refusal	1 st refusal	2 nd refusal
General denial	1140	1029	-	-
Not interested	1086	686	-	-
No time (open)	197	133	74	32
No time (closed)	507	285	-	-
Vitality	156	92	96	50
Negative feeling about interviews	157	79	34	17
Survey process	153	84	40	20
Participation interdicted	109	89	58	34
Political situation	65	22	-	-
No comment	119	136	-	-
Total	3689	2635	302	153

Note: At the second contact attempt 24% of cases was assigned to the same interviewer

In our analysis we used the categories “general denial”, “no time” and “no interest” as non-ordered categories of the multinomial dependent variable “reason for refusals”. The response “no time” was split in accordance with data collection in open (as comment) and closed answer formats (using the category “refused because of time concerns” in the list of outcomes). In doing so, we aimed to take into account different interviewer impacts, if data regarding the reason “no time” were collected in both open and closed answer formats. The categories “participation interdicted”, “vitality”, “political situation”, “negative feeling about interviews” and “survey process” were combined into one category SUM_SPEC in order to avoid numerical problems (which were possible if numbers in the dependent variable categories were too low). Additionally, with the combined category SUM_SPEC we obtained a category with more specific information about reasons for refusals, as compared to the more typical “general denial”, “no interest” and “no time”. Finally, we also included not providing comments in the case of one refusal (“no comments”) in our analysis.

In the multinomial model applied y_{ij} is the reason for refusal categorized as i by interviewer j . It is coded as:

$$y_{ij} = \begin{cases} 0 & \text{SUM_SPEC} \\ 1 & \text{general denial} \\ 2 & \text{not interested} \\ 3 & \text{no time (open)} \quad \dots \\ 4 & \text{no time (closed)} \\ 5 & \text{no comment} \end{cases}$$

The probabilities of reasons for refusal are denoted by $\pi_{ij}^{(s)} = Pr (y_{ij} = s)$, $s = 0, 1, 2, 3, 4, 5$. For “SUM_SPEC” as the reference category, the applied multinomial model is:

$$\log \left(\frac{\pi_{ij}^{(s)}}{\pi_{ij}^{(0)}} \right) = \beta^{(s)T} x_{ij}^{(s)} + u_{ij}^{(s)}, \quad s = 1, 2, 3, 4, 5.$$

whereby $x_{ij}^{(s)}$ is a vector of individual level covariates, $\beta^{(s)}$ is a vector of coefficients and $u_{ij}^{(s)}$ presents a random effect of unobserved interviewer characteristics. Thereby, effects of $x_{ij}^{(s)}$ do not vary for different interviewers (random intercept model).

Independent variables in the models were:

- 1) interviewer gender (1 = man; 2 = woman);
- 2) interviewer age (numerical, starting with 18 years);
- 3) interviewer experience, based on the number of years at the survey research institute (1 = less than 4 years; 2 = 5 to 9 years; 3 = 10 years or longer);
- 4) interviewer education (1 = basic; 2 = secondary/vocational school; 3 = university entrance diploma/college of higher education/university without diploma)

Multilevel models were estimated by Markov Chain Monte Carlo Methods (MCMC, compare Browne 2009). The procedure, as suggested by Rasbasch et al. (2009), was that the starting values for MCMC models were estimated with the aid of the MQL-procedure of MLwiN. To run MCMC the MLwiN method default was used (burn in length of 500, chain length of 5000).

A problem of interviewer variance is that it can be confounded with spatial homogeneity (cluster-related design effect), meaning the homogeneity of persons living within a geographical area. An interviewer normally works within a geographical unit (sampling point) and these units, as well as secondary sample units, are not randomly assigned to interviewers (for an overview see Groves et al. 2004). A method of separating interviewer variance from geographical homogeneity is through interpenetrated design. This means that two or more interviewers work within a sampling point, and ad-

addresses of target persons are randomly assigned to each interviewer. But interpenetrated studies did not reveal a substantial amount of cluster related variance (spatial homogeneity) on interviewer related variance (O'Muircheartaigh 1998; Schnell & Kreuter 2003). Additionally, for the sample design applied in ALLBUS, only a small amount of variance ($\rho = 0.05$) is expected to be related to spatial homogeneity⁶.

⁶ This ρ value was reported to us by the GESIS team responsible for sampling questions. This value applies to the German sample of ESS, which is identical in procedure to the sample used in ALLBUS.

6 Results

6.1 Which Reasons for Refusals were Reported in the ALLBUS?

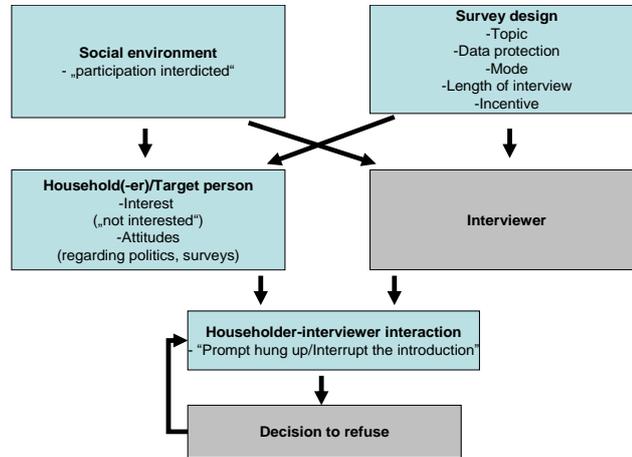
Table 2 (chapter 5.1) shows frequencies obtained for each category as defined in the categorisation scheme of the ALLBUS 2008 dataset. Here, up to three reasons for refusals were coded (multiple sets) for each contact attempt. The new category “general denial” is the largest (30% of all categorized statements). This category contains statements regarding general denial of surveys or denial of access. General denial comprises 470 cases (22%) of comments which represent refusal of any interview (in the sense of German law regarding data protection, ADM). In such cases it is not allowed to contact target persons a second time, at least in Germany. Similar to ESS and in accordance with the results of previous research, the next largest categories are “not interested” (25%) and “no time” (19%). Other meaningful categories were “vitality” (5.4%) – refusal due to inability based on advanced age or restricted health conditions – and “survey process” (4.4%). The latter provides feedback about appropriate survey design from the point of view of target persons (e.g. appropriate mode, length of interviews, privacy). Interestingly, in 4% of the comments interviewers did not provide information about reasons for refusals. Using the categorisation scheme described above, only 0.2% of “other” was identified.

Our first hypothesis (H1) expected to find information related to factors which predict survey participation (as defined by Groves & Couper 1998, see figure 1 in chapter 2.1). We found that interviewers’ comments were informative firstly regarding householders’ situations. Apart from “not interested” and “no time” interviewers’ comments included information related to political attitudes (“dissatisfied with the political situation”), as well as attitudes about surveys (“negative feeling about surveys”). Another type of statement concerns the health and ability of householders to take part in an interview (category “vitality”). In addition, “general denial” is linked to norms regarding survey participation, for example “I do not do such things”. Other comments were related to survey related decisions (category “survey process” with subcategories “data protection”, “methodology of surveys”, “voluntary participation”). The next group of statements is related to householders’ social environment, in particular that relatives refuse or interdict participation (category “participation interdicted/refusal by third person”). Regarding householder-interviewer interaction, “hung up on the intercom”, and sometimes comments such as “interrupted my introduction”, “did not let me say any introductory words”, were recorded (15 % of “general denial” in our data).

Our results show that, apart from the reasons “not interested” and “no time”, a large proportion of comments regarding reasons for refusals (20% in our data) is informative and provides surveyors with quite specific information, for example about personal situations (abilities, opinions) of householders, as well as about how householders perceive survey design features. However, comments such as “no time” and “not interested” are useful as well. Particularly in cases of “no time” householders can be contacted once again, and in cases of “not interested” special instructions or additional contacts (e.g. one call per phone) can be provided.

We assigned the information we found in interviewers’ comments to the factors defined by Groves and Couper (1998). Figure 3 illustrates this assignment.

Figure 3: Reasons for Refusals in the ALLBUS 2008 and Their Relationship to Factors defined by Groves and Cooper (1998).



If we apply the leverage-salience theory to the given reasons for refusals, time restrictions prove to be the most pronounced burden, but the comments also tell us that contacted persons are unsure of the benefits of participation (“surveys are useless”, “too many surveys nowadays”, “incentives”) or they are concerned about a possible negative impact of participation (“bad previous experiences”, “invasion of privacy”).

In summary, the results support H1 and show that apart from typical information, such as “not interested” and “no time”, interviewers’ comments contained a relevant amount of specific information (20% of comments) regarding factors which can predict a decision to participate (compare Groves & Cooper 1998): there is information about the survey process (incentive, mode, length of interview), about sampled persons (abilities to participate and opinions regarding surveys, vitality), and about the social environment of sampled persons who refused.

6.2 How High is Interviewer Effect?

In the following section, results related to interviewer effect are reported in order to address our second research question and evaluate H2 and H3 (compare chapter 3). H2 assumes correlations between the gender, age, experience and education of interviewers and collected reasons of refusals. H3 assumes obtaining high interviewer effect (in the form of unexplained interviewer variances in multi-level models) when collecting reasons for refusals due to poor quality of data collection instruments.

We analysed interviewer effect upon the collection of data using the main categories “general denial”, “not interested”, “no time” (separately analysed for open and closed answer formats), as well as on providing “no comments”. The categories “participation interdicted”, “vitality”, “political situation”, “negative feelings about surveys” and “survey process” were merged into the category SUM_SPEC, which contains special reasons for refusals. We conducted separate analyses for first and second (follow up contact) refusals.

6.2.1 Descriptive Results for the Relationship between Interviewer Characteristics and Reasons for Refusals

For the analysis first refusal data collected by 154 interviewers and second refusal data collected by 164 interviewers were available. The majority of interviewers were men (see for example 1st refusal in table 5, here 60% of interviewers were men; 40% women). Additionally, 52% of interviewers were 60 years or older (see 1st refusal in table 5). Most of the interviewers had been working at the field institute for 5 years or more (63% in the case of 1st refusal in table 5). Furthermore, a large percentage of interviewers had secondary/vocational or higher education (81%, see 1st refusal in table 5).

Table 5 shows descriptive statistics (differences in percentages) for reasons for refusals that were collected by interviewers with different socio-demographic characteristics. It can be seen that all interviewer characteristics are weakly but significantly correlated with the collection of reasons for refusals (compare *phi*-coefficients in table 5, which range from *phi* = .06 ($p < .05$) to *phi* = .14 ($p < .001$)).

Table 5: Percentages of Reasons for Refusals Dependent on Gender, Age, Experience (Years Working at the Institute) and Education of Interviewers in the ALLBUS 2008.

	gender		age		years working at the institute			education			n (comments)
	male	female	18-59	>=60	0 - 4	5 - 9	>=10	basic	sec.	higher.	
1st refusal											
General denial	30.7	31.2	32.1	29.8	33.5	29.9	29.8	29.8	30.4	32.1	1,140
not interested	29.8	29.0	28.9	29.9	24.3	29.2	33.2	35.0	27.1	29.2	1,086
no time (open)	4.4	6.8	4.7	5.9	6.3	5.2	4.8	4.9	5.3	5.6	197
no time (closed)	14.3	12.9	15.8	11.9	14.6	15.9	12.4	12.4	13.9	14.3	507
SUM_SPEC	17.2	17.6	15.8	18.8	17.9	16.8	17.2	14.6	19.3	19.4	640
no comment	3.6	2.6	2.7	3.7	3.3	3.0	3.3	3.2	3.9	2.4	119
<i>n interviewers</i>	92	62	76	78	57	30	67	30	67	57	
<i>phi</i>		.06*		.08**			.09**			.08**	
2nd refusal											
General denial	38.2	40.4	44.4	34.9	41.8	40.4	36.5	31.4	38.2	44.2	1,029
not interested	26.4	25.4	23.4	28.1	26.9	23.8	26.4	34.6	25.7	22.1	686
no time (open)	4.3	6.2	4.7	5.3	5.6	4.8	4.7	6.0	5.0	4.5	133
no time (closed)	11.6	9.6	10.7	10.9	7.9	12.2	12.4	11.0	9.9	12.0	285
SUM_SPEC	13.5	14.5	5.9	8.0	13.4	14.1	14.2	9.5	16.0	13.2	344
no comment	6.0	3.9	3.1	6.7	4.5	4.6	5.9	7.5	5.1	4.0	136
<i>n interviewers</i>	97	66	78	85	58	31	74	25	77	61	
<i>phi</i>		.07*		.12***			.09*			.14***	

6.2.2 Results of Multilevel Analyses regarding Interviewer Effect

In the next step we tested which of the differences in socio-demographic variables of interviewers regarding the collection of a particular reason (shown in table 5) were significant. We calculated different multilevel multinomial regressions (for specifications see chapter 5.2).

Firstly, we estimated a model without interviewer level and without dependent variables (0 Model). We obtained the following DIC⁷ statistics: DIC = 11,564,704 for first refusal and DIC = 8,099,617 for second refusal. Secondly, we added interviewer level to the model (Random-Intercept-Only-Model; R-I-O-M). DIC statistics decreased remarkably with these models: to DIC = 10,021,567 for first refusal and DIC = 6,822,169 for second refusal (see table 6).

Thirdly, we included independent variables describing gender, age, experience and education of interviewers in the models (Random-Intercept-Model; R-I-M). This did not decrease DIC statistics remarkably: DIC = 10,020,345 (first refusal) and DIC = 6,828,528 (second refusal) (see table 6). This indicates that socio-demographic variables do not significantly improve model fit. In contrast, taking into account variation between interviewers contributes remarkably to a better model fit.

7 DIC (Deviance Information Criterion) is the model fit characteristic - an informative basis for appropriateness of a model. It is comparable with the $-2\ln L$ coefficient for model fit in logistic and ordinal regressions. In models with MCMC samplings, DIC is a generalization of Akaike's Information Criterion (AIC) (Spiegelhalter et al. 2002). The DIC diagnostic involves calculating deviance from the expected values of unknown parameters at each iteration of a specified model (see Browne 2009).

Table 6: Results of Multi-level Analyses Regarding the Interviewer Impact

	R-I-O-M		1 st refusal R-I-M		R-I-O-M		2 nd refusal R-I-M	
	par.	s.e.	par.	s.e.	par.	s.e.	Par.	s.e.
Fixed Part								
constant								
general	0.72***	0.11	3.22***	0.50	1.02***	0.11	2.61***	0.49
not interested	0.32**	0.15	3.95***	0.88	0.26	0.16	1.38	1.49
no time (open)	-	0.12	-0.20	0.82	-	0.16	-1.69	1.05
no time (closed)	-0.25	0.16	2.54***	0.90	-0.44**	0.16	0.63	0.94
no comment	-	0.31	0.19	1.30	-	0.48	-5.37**	2.29
gender (male)								
general denial (female)			-0.08	0.19			0.02	0.18
not interested (female)			-0.15	0.26			0.14	0.30
no time (open) (female)			0.36*	0.21			0.50*	0.27
no time (closed) (female)			-0.34	0.31			-0.15	0.30
no comment (female)			-0.93	0.57			-0.31	0.74
age (continuous)								
general denial			-0.04***	0.01			-0.02**	0.01
not interested			-0.06***	0.01			0.00	0.02
no time (open)			-0.02	0.01			0.02	0.01
no time (closed)			-0.05***	0.01			-0.02	0.01
no comment			-0.05***	0.02			0.04	0.03
experience (0-4 Years)								
2 general denial (5-9 Years)			-0.08	0.29			0.08	0.26
3 general denial (>= 10 Years)			0.07	0.24			-0.05	0.19
2 not interested (5-9 Years)			0.32	0.36			-0.35	0.37
3 not interested (>= 10 Years)			0.92***	0.28			-0.08	0.30
2 no time (open) (5-9 Years)			-0.05	0.32			-0.28	0.35
3 no time (open) (>= 10 Years)			-0.02	0.25			-0.12	0.28
2 no time (closed) (5-9 Years)			0.23	0.32			0.31	0.38
3 no time (closed) (>= 10 Years)			0.25	0.25			0.39	0.30
2 no comment (5-9 Years)			0.44	0.97			-0.11	0.97
3 no comment (>= 10 Years)			0.64	0.65			0.08	1.03
education (basic)								
2 general denial (secondary)			-0.33	0.27			-0.61**	0.29
3 general denial (higher)			0.02	0.30			-0.17	0.30
2 not interested (secondary)			-0.76**	0.34			-1.02*	0.55
3 not interested (higher)			-0.41	0.35			-1.16**	0.55
2 no time (open) (secondary)			-0.34	0.25			-0.77*	0.40
3 no time (open) (higher)			0.02	0.28			-0.70*	0.38
2 no time (closed) (secondary)			-0.12	0.38			-0.64	0.45
3 no time (closed) (higher)			0.18	0.39			-0.24	0.42
2 no comment (secondary)			-0.13	0.78			-1.38	1.04
3 no comment (higher)			-0.57	0.79			-0.49	1.12
Random Part								
<i>Level 2 - Interviewer</i>								
σ^2 general denial	0.88***	0.18	0.91***	0.19	0.80***	0.20	0.77***	0.20
σ general denial/not interested	0.23	0.18	0.17	0.19	-0.09	0.24	-0.14	0.22
σ^2 not interested	1.92***	0.32	1.90***	0.33	2.14***	0.46	2.13***	0.47
σ no time (open)/ general denial	0.19	0.12	0.18	0.13	-0.04	0.15	-0.09	0.15
σ no time (open)/ not interested	0.29	0.18	0.29*	0.16	0.17	0.29	0.10	0.27
σ^2 no time (open)	0.41***	0.14	0.30***	0.11	0.58***	0.21	0.56**	0.21
σ no time (closed)/ general denial	0.45**	0.18	0.46**	0.19	0.47**	0.21	0.41*	0.20
σ no time (closed)/ not interested	0.76**	0.24	0.78***	0.25	0.47	0.30	0.43	0.30
σ no time (closed)/ no time (open)	0.10	0.17	0.16	0.16	-0.11	0.22	-0.15	0.20
σ^2 no time (closed)	1.46***	0.29	1.52***	0.33	1.39***	0.33	1.33***	0.36
σ no comment/ general denial	0.45	0.38	0.63	0.41	-0.68	0.48	-0.98*	0.57
σ no comment/ not interested	0.15	0.48	0.23	0.56	-0.18	0.75	-0.36	0.88
σ no comment/ no time (open)	-0.40	0.34	-0.09	0.36	0.91	0.67	1.09*	0.63
σ no comment/ no time (closed)	0.64	0.48	0.74	0.53	-0.04	0.67	-0.28	0.80
σ^2 no comment	5.55***	1.23	6.37***	1.68	9.77***	2.83	10.98**	3.09
Units interviewers			154				163	
Units target persons	3,689		3,689				2,635	
DIC		10,021,567		10,020,345		6,822,169		6,828,521

Note: * $p < 0.05$, ** $p < 0.01$; *** $p < .001$ σ^2 : Interviewer variances; Reference category "SUM_SPEC"; R-I-O-M: interviewer level only; R-I-M: interviewer level and independent variables.

6.2.2.1 Effect of Interviewers' Demographic Characteristics

Even if the addition of interviewers' demographic characteristics cannot remarkably alter the predictive overall model fit (see DIC statistics), some effects of interviewers' characteristics are significantly different from zero (table 6).

Gender

There is only one significant effect related to gender. Female interviewers collected fewer "no time" reasons in open answer format in the case of both first (*parameter* = 0.36, $p < .05$) and second refusals (*parameter* = 0.50, $p < .05$) than male interviewers, compared to the collection of specific reasons for refusals in the summarised category SUM_SPEC.

Age

For age a numerical variable was used in the models, in contrast to the categories used in table 5. Results show that there are very small but significant effects for "general denial" for both first and second refusals (1st refusal: *parameter* = -.04, $p < .01$; 2nd refusal: *parameter* = -.02, $p < .01$). Additionally, for first refusals significant effects of age are seen for "not interested" (*parameter* = -.06, $p < .01$), "no time" (*parameter* = -.05, $p < .01$) and "no comment" (*parameter* = -.05, $p < .01$). Negative significant parameters show that – compared with younger interviewers - older interviewers provided significantly fewer typical comments in nearly all categories and more specific reasons for refusals in the category "SUM_SPEC". This effect is particularly apparent and significant for first refusals.

Experience

Here only one significant effect for first refusal can be seen: more experienced interviewers (10 years or longer with the field research institute) provided the reason "not interested" more often than specific comments, when compared to less experienced interviewers (less than 5 years with the field research institute; *parameter* = 0.92, $p < .001$).

Education

Regarding education, there are numerous significant results in the case of second refusals. Effects related to collecting typical reasons for refusals, compared to more specific information, show the same trend for all categories (nearly all parameters are negative in table 6). This indicates that increasing interviewer education is associated with the provision of more specific comments and – in contrast to this – fewer comments of other types. These differences become significant in cases of "general denial" (*parameter* = - 0.61, $p < .001$), "not interested" and "no time" (collected in open answer format). In cases of "general denial" interviewers with secondary level education differ from those with only a basic education. In cases of "not interested" and "no time" interviewers with secondary and higher level education differ from those with basic education as well (see table 6 for coefficients).

The results pertaining to interviewers' demographic characteristics support H1. In particular, age and education significantly correlate with the provision of particular reasons for refusals. Older interviewers provided more specific comments and fewer other typical comments (such as "general denial" and "not interested" and "no time"). In cases of second contacts, higher educated interviewers provided specific comments more often than typical comments ones, as compared to less educated interviewers.

6.2.2.2 Interviewer Related Variances

In the R-I-O-M and R-I-M models the random interviewer effect (interviewer variances σ^2) is remarkable high and very significant (table 6, random part). The highest effect can be seen for “no comments” for both first and second refusals (first refusal, R-I-M: $\sigma^2 = 6.37$ ($p < .001$); second refusal, R-I-M: $\sigma^2 = 10.98$ ($p < .001$)). The lowest effect can be seen for “no time” in open answer format (first refusal, R-I-M: $\sigma^2 = 0.30$ ($p < .001$); second refusal, R-I-M: $\sigma^2 = 0.56$ ($p < .01$)).

The similarity between sampled persons in the same group (defined by one interviewer) is measured by intra-class correlation (ρ). “It may also be interpreted as the proportion of the total residual variation that is due to differences between groups, and is referred to as the variance partition coefficient (VPC) (see Goldstein (2003), pp 16-17)” (Rasbash et al. 2009, p. 28).

For first refusals intra-class correlations (ρ)⁸ for interviewer effects range from 0.8 to 0.66 (no comments), implicating that 8% (no time, open) to 66% (no comments) of the entire variation is accounted for by interviewer impact (table 7). For second refusals interviewer effect becomes somewhat higher, ranging from 15% (no time, open) to 77% (no comment). Compared to the spatial design effect (deffc) for Germany ($\rho = 0.05$) interviewer effect is very high. Here spatial design effects may present only a negligible part (5%) of obtained interviewer level variances.

These results imply that variations in one interviewer’s comments are significantly lower than those between interviewers. In other words, one interviewer produces more similar reasons for refusals than different interviewers do. At the same time, providing the reason “no time” (both in closed and open answer formats) for first refusals is positively related to providing “general denial” and “not interested” (covariances between categories shown by σ in random part in table 6). Positive and significant covariances can be interpreted as follows: although an interviewer named several specific reasons more often than others (showing high interviewer related variances), at the same time he or she also recorded additional reasons for refusals. Interestingly, “no comments” is not related (for first refusal) or is only weakly related (for second refusal) to providing other reasons for refusals (see accordant coefficients (σ) in random part in table 6). So, providing “no comments” is not associated with most of the other categories. This means that interviewers who do not provide any comments generally do so as their typical behavior.

⁸ Is calculated as $\rho = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \left(\frac{\sigma^2}{3}\right)}$ (see e.g. Pickery 2002). $\sigma^2_{\text{between}}$ is σ^2 related to interviewer variances in table 6.

Table 7: Intraclass Correlations (ρ) for Interviewer Related Variances for the Categories in the ALLBUS Presented for First (First Contact) and Second (Recontact) Refusal

Category	1 st refusal	2 nd refusal
General denial	0.22	0.19
Not interested	0.37	0.39
No time (open)	0.08	0.15
No time (closed)	0.32	0.29
No comment	0.66	0.77

The results of multilevel analyses support H3: very high random interviewer effects were obtained for collecting reasons for refusals. The highest interviewer impact was apparent for not providing any comments at all.

7 Discussion

The first aim of our study was to analyse which reasons for refusals were reported by interviewers in ALLBUS 2008 para-data. For their reports interviewers used an open answer format to comment on the outcome “refusal”. We analysed information related to factors influencing survey participation (compare Groves & Cooper 1998). With the help of content analysis we developed a reliable categorization scheme to characterise the reasons for refusals contained in interviewers’ comments.

As with previous research (DeMaio 1980; Erblöh & Koch 1988; Groves & Cooper 1998; Költringer 1992; Neller 2005) our data showed high frequencies of the reasons “no time” and “not interested”. But the largest category was “general denial” (30%). This category was not used in ESS, which may explain high frequencies of the category “other” in ESS data. The category “general denial” is important particularly for Germany, since those refusing participation in an interview cannot be contacted again (ADM). In our data 15% of “general denial” was due to rejecting participation in any interview. In addition – compared to other studies available in the literature – the category “survey process” was new in our data (4% of all reasons for refusals). As expected, interviewer comments about reasons for refusals in ALLBUS were especially informative regarding the factors which influence survey participation (Groves & Couper 1998), in particular the factors “householder/target person” and “survey process”.

The second aim of our study was to analyze interviewer impact on collected reasons for refusals. As expected, we found a high interviewer impact in providing the comments “no time”, “not interested”, and “general denial”, in comparison to more specific comments (e.g. “vitality”, “survey process”, “political situation”, “negative feeling about interviews”). But the highest interviewer impact was found in cases of not providing any comments at all, which seems to be consistent behaviour on the part of some interviewers (regarding the absence of covariance or small covariances with providing other comments). Interviewer variances were very high, accounting for 8 to 77% of the entire variance. These high interviewer level variances could not be fully explained by spatial homogeneity (which can be expected to be less than 5% in the ALLBUS sample). The obtained results regarding interviewer impact are comparable with results from studies which have found high interviewer related variances in explaining response and non-response via multinomial multilevel regressions (Durrant & Steele 2009; Pickery & Loosveldt 2002).

Additionally, we found significant relationships between providing comments on reasons for refusals and the age and education of interviewers. Providing more specific comments (in comparison to providing more typical ones such as “no time” and so on) is positively correlated with higher age and higher education of interviewers. These results are comparable to the findings of previous research in which older and more highly educated interviewers were found to have success regarding response rates (Költringer 1992; Neller 2005; O’Muicheartaigh & Campanelli 1999). In contrast to previous research, we did not find significant systematic associations with interviewer experience (Couper & Groves 1992; de Leeuw & Hox 1996; Durbin & Stuart 1951; Durrant et al. 2010).

The high effect of interviewers on collected reasons for refusals reduces the objectivity of collected data. This problem can be explained by the fact that reasons for refusals were collected in ALLBUS 2008 using a non-standardized form for data collection. Prior to our analysis, we assessed contact forms used in ALLBUS as being poorly designed: categories regarding contact attempt outcomes were not ordered (structured) and reason for refusals (no time) could be collected in two different ways. According to Dillman and colleges (Dillman 2007; Dillman, Smyth & Christian 2009) such problems are typical for poorly designed questionnaires. Poorly designed instruments of data collection – a ballot demoraliser – decrease interviewer motivation (Crespi 1945). The next explanation of

high interviewer impact is that interviewers did not know how their comments would be used by surveyors, an additional source of demoralisation for interviewers (Vroom 1964).

We have found some evidence in our data that contact forms used in ALLBUS should be improved. Our results show that some interviewers constantly used open answer format to collect “no time” and ignored the provided category for this outcome (covariances between these two possibilities do not significantly differ from zero). Additionally, we found a large number of incorrect comments (9.5% of all comments): interviewers recorded “refusal” as the category but wrote an incompatible comment, for example that the target person could be contacted later or that the address was illegible (despite the fact that an appropriate category was provided in the list).

Our results show that it is necessary to improve contact forms used for ALLBUS: both as an instrument of data collection, and to foster interviewer motivation to collect reasons for refusals. This would be a way to reduce high interviewer impact on data collection. In this regard we suggest the following:

- 1) Usage of a better structured list to document contact attempt outcomes.
- 2) Providing categories for interviewers to record reasons for refusals. To do this definitions of categories and representative examples should be provided and these should be flexibly used (for example with computerisation) to foster an improved understanding of these categories on the part of interviewers, as well as to provide help with categorization difficulties.
- 3) Providing a short explanation to interviewers (in contact forms) regarding the usability of collected information about reasons for refusals.

The results of the study support our theoretical assumptions regarding interviewer motivation. But the limitation of our study is that we did not directly test the effects of a better design of contact forms and little information about the usability of collecting reasons for refusals as demoralizers (Crespi 1945; Vroom 1964). An explicit test, for example with the help of an experimental design, is needed in this regard, as well as further research and evaluative studies on different methods suggested in our article in order to reduce both refusal rates and interviewer impact on data collection. Nevertheless, our study shows that there is a large potential to use information related to reasons for refusals.

8 Considerations Regarding the Standardised Collection of Reasons for Refusals

Our categorisation scheme – which allows a high intercoder-reliability – can currently be used for the analysis of reasons of refusals collected in open format, for example in ALLBUS (ESS may use them as well, e.g. to analyse comments in the category “other”). Furthermore, our categorization scheme can be used as a basis to develop standardized instruments for interviewers to collect reasons for refusals. AAPOR (2008) suggests collecting reasons for refusals as standard procedure. But there are no suggestions/deliberations about how reasons for refusals should be collected in order to obtain reliable and valid data. In this section we will provide some considerations regarding standardised collection of reasons for refusals through applying the categories we have developed for our analysis. Since we used ALLBUS data, we will discuss the possible consequences of standardised data collection for ALLBUS as a particular example.

When using these categories during a field period by interviewers, we expect an increase in the collection of more specific reasons for refusals (e.g. regarding survey process) and a reduction of typical reasons for refusals (including general denial). If categories for special reasons for refusals are provided, it should become obvious to interviewers that surveyors are interested in this information. As a result, interviewers will pay more attention to such information and document it in contact protocols. To improve the informative content of collected data, categories describing interviewer-householder interaction should be included (compare Groves and Cooper 1998).

Providing categories for interviewers and – in general – providing an improved instrument of data collection would reduce mistakes made by interviewers (false declaration of an outcome as refusal). Such mistakes have consequences when calculating response and refusal rates in surveys. Particularly for ALLBUS, in which refusal rates are very high, it is important to carry out such improvements.

Collecting reasons for refusals with the help of a standardised instrument would make their usage for the purpose of reducing refusal rates more practicable. The possibilities to use reasons for refusals during the data collection period (e.g. in ALLBUS) would be:

- 1) Usage for follow-up contacts: especially sampled persons who provide typical reasons for refusals, for example by saying “no time” vs. “not interested”, seem to be easy to convert for participation by means of new contact attempts (Fuse & Xie 2007; Schnauber & Daschmann 2008; Neller 2005; Reuband & Blasius 2000). The corresponding groups were large in our data set (“no time” and “not interested” amount to nearly 33%).
- 2) Adapting methods of follow-up contacts depending on the given reason for refusal: according to the “tailored design method” of Dillman and collaborators (Dillman 2007; Dillman, Smyth & Christian 2009) follow up contacts will be successful if the contact strategy is changed. Reasons for refusals provide information about which survey properties were salient for sampled persons, as well as information about the salient burden of participation perceived by sampled persons. In the case of follow-up contacts, providing specific information during the first few minutes, depending on the previously mentioned reason for refusal, would decrease concerns on the part of the respondent regarding the perceived burden of participation. For example, in the case of “no interest”, a more meaningful explanation of the survey topic should be provided during the first few minutes of contact. In the case of “too old” as reason for refusal information about the importance of participation by older target persons should be provided. In the case of “political situation” as reason for refusal the relevance of survey results for society (not only for politics) should be provided. Regarding concerns about survey methods and data protection related information should be given as well. Conducting short interviews in the case of concerns about the “length of interview” (with basic questions)

or a special incentive in the case of such reasons for refusals should be the next strategy to reduce refusal rates in ALLBUS.

- 3) Providing manuals and supervision for interviewers' doorstep behaviour: Knowing the reason for refusal, interviewers can adapt their doorstep behaviour or apply different conversion guidelines (Neller, 2005). Here, information as in the case of follow-up contacts could be used.

Our study about reasons for refusals should be seen as a first step toward more effective collection of these data and more effective usage of the information they provide for field monitoring and to reduce refusal rates. Additionally, our reliable categorization scheme (Menold & Zuell 2010) can be used as a basis for the development of improved standardised instruments to collect reasons for refusals. This is critical in reducing high interviewer impact on data collection, and increasing data quality.

References

AAPOR: Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (2008):

http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814. Accessed 30.10.2010

ADM: Richtlinie zum Umgang mit Adressen in der Markt- und Sozialforschung: http://www.adm-ev.de/fileadmin/user_upload/PDFS/R07_D.pdf. Accessed 30.10.2010

Bates, N., Dahlhamer, J., Singer, E.: Privacy Concerns, Too busy, or just not interested: using door-step concerns to predict survey nonresponse. *Journal of Official Statistics* 24, 591-612 (2008).

Blom, A.: Measuring, Explaining and Adjusting for Cross-Country Differences in Unit Nonresponse: What Can Process Data Contribute?, Essex University (2009).

Browne, W. J.: MCMC Estimation in MLwiN - Version 2.13. Centre for Multilevel Modelling, University of Bristol (2009).

Couper, M. P., R. M. Groves: The role of the interviewer in survey participation. *Survey Methodology* 18, 263-77 (1992).

Crespi, L. P. The cheater problem in polling. *Public Opinion Quarterly* 9, 431-445 (1945).

De Leeuw, E. D., Hox, J. J.: The effect of the interviewer on the decision to cooperate in a survey of the elderly. In: Laaksonen, S. (eds.) *International Perspectives on Nonresponse*, pp. 46-52. Statistics Finland, Helsinki (1996).

De Leeuw, E. D., Hox, J. J., Snijders, G., De Heer, W.: Interviewer opinions, attitudes and strategies regarding survey participation and their effect on response. *Nonresponse in survey research, ZUMA Special 4*, Mannheim (1997).

De Leeuw, E.D., De Heer, W.: Trends in household survey nonresponse: A longitudinal and international comparison. In: Groves, R. M., Dillman, D., Eltinge, J., Little, L. (eds.) *Survey Nonresponse*, pp. 41-54. Wiley, New York (2002).

DeMaio, T. J.: Refusals: Who, where and why. *Public Opinion Quarterly* 44, 223-233 (1980).

Dillman, D. A.: *Mail and Internet Surveys. The Tailored Design Method*. Wiley, New Jersey (2007, 2nd edition).

Dillman, D. A., Smyth, J. D., Christian, L. M.: *Internet, Mail, and Mixed-Mode Surveys. The Tailored Design Method*. Wiley, New Jersey (2009).

Durbin, J., Stuart, A.: Differences in response rates of experienced and inexperienced interviewers. *Journal of the Royal Statistical Society Series A* 114, 163-206 (1951).

Durrant, G. B., Steele, F.: Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK Government surveys. *Journal of the Royal Statistical Society Series A* 172, 361-81 (2009).

Durrant, G. B., Groves, R. M., Staetsky, L., Steele, F.: Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly* 74(1), 1-36 (2010).

Esser, H.: Kooperation und Verweigerung beim Interview. In: Erbsloeh, E., Esser, H., Reschka, W., Schoene, D. (eds.) *Studien zum Interview*, pp. 69-142. Hain, Meisenheim am Glan (1973).

- Erbslöh, B., Koch, A.: Die Non-Response-Studie zum ALLBUS 1986: Problemstellung, Design, erste Ergebnisse. ZUMA Nachrichten 22, 29-44 (1988).
- Früh, W.: Inhaltsanalyse: Theorie und Praxis. UKV, Konstanz (2007).
- Fuse, K., Xie, D.: A successful conversion or double refusal: A study of the process of refusal conversion in telephone survey research. *Social Science Journal* 44, 434-46 (2007).
- Goldstein, H.: *Multilevel Statistical Models*. London: Arnold, 3rd Edition (2003).
- Groves, R. M., M. P. Couper: *Nonresponse in Household Interview Surveys*. Wiley, New York (1998).
- Groves, R. M., Singer, E., Corning, A.: Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly* 64, 299-308 (2000).
- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R.: *Survey Methodology*, Wiley, New Jersey (2004).
- Groves, R. M., Heeringa, S. G.: Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A* 169, 439-457 (2006).
- Hox, J. J., de Leeuw, E. D., Kreft, G. G.: The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In: Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., Sudman, S. (eds.) *Measurement Errors in Surveys*. Wiley, New York (1991).
- Hox, J. J.: Hierarchical regression models for interviewer and respondent effects. *Sociological Methods Research* 22, 300-318 (1994).
- Költringer, R.: *Die Interviewer in der Markt- und Meinungsforschung*. Service Fachverlag, Wien (1992).
- Kreuter, F., Kohler, U.: Analyzing contact sequences in call record data: potential and limitation of sequence indicators for nonresponse adjustment in the European Social Survey. *Journal of Official Statistics* 25, 203-226 (2009).
- Lehtonen, R.: Interviewer attitudes and unit nonresponse in two different interviewing schemes. In: Laaksonen, S. (eds.) *International Perspectives on Nonresponse; Proceedings of the Sixth International Workshop on Household Survey Nonresponse*, Statistics Finland, Helsinki (1996).
- Neller, K.: Kooperation und Verweigerung: Eine Non-Response-Studie. ZUMA Nachrichten 57, 9-36 (2005).
- O'Muircheartaigh, C.: The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A* 161, 63-77 (1998).
- O'Muircheartaigh, C., Campanelli, P.: A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, Series A* 162, 437-46 (1999).
- Pickery, J.: *Contextual Effects on the Vote in Germany A Multilevel Analysis*. Wissenschaftszentrum Berlin für Sozialforschung gGmbH (WZB) (2002), <http://bibliothek.wz-berlin.de/pdf/2002/iii02-202.pdf>. Accessed 30 November 2010.
- Pickery, J., Loosveldt, G.: A multilevel multinomial analysis of interviewer effects on various components of unit nonresponse. *Quality & Quantity* 36, 427-37 (2002).
- Rasbash, J., Steele, F., Browne, W. J., Goldstein, H.: *A User's Guide to MLwiN - Version 2.10*. Centre for Multilevel Modelling. University of Bristol (2009).

Reuband, K.-H.: Ausfälle in einer mündlichen Befragung. Unveröffentlichtes Manuskript. Köln (1975).

Reuband, K.-H., Blasius, J.: Situative Bedingungen des Interviews, Kooperationsverhalten und Sozialprofil konvertierter Verweigerer. Ein Vergleich von telefonischen und face-to-face Befragungen. In: Hüfken, V. (eds.) *Methoden in Telefonumfragen*, pp. 139-167. Westdeutscher Verlag, Opladen (2000).

Schnauber, A., Daschmann, G.: States oder Traits? Was beeinflusst die Teilnahmebereitschaft an telefonischen Interviews? *Zeitschrift für empirische Sozialforschung* 2, 97-123 (2008).

Schnell, R., Kreuter, F.: Separating interviewer and sampling-point effects. Department of Statistics, UC Los Angeles (2003). <http://www.escholarship.org/uc/item/7d48q754>. Accessed 30 November 2010

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64:191-232 (2002).

Stoop, I. A. L.: Surveying nonrespondents. *Field Methods* 16, 23-54 (2004).

Vroom, V. H.: *Work and Motivation*, Wiley, New York (1964).

Zuell, C., Menold, Natalja. Codierung von Gründen der Verweigerung der Teilnahme an Interviews: ein Kategorienschema. *GESIS-Technical Reports 2010|11*. GESIS, Mannheim (2010).

Zeh, J.: *Der Verzerrungsfehler durch Ausfälle bei Meinungsbefragungen*. Dissertation, Bonn (1976).