

# “Interviewer” Effects in Face-to-Face Surveys: A Function of Sampling, Measurement Error, or Nonresponse?

*Brady T. West<sup>1</sup>, Frauke Kreuter<sup>2</sup>, and Ursula Jaenichen<sup>3</sup>*

Recent research has attempted to examine the proportion of interviewer variance that is due to interviewers systematically varying in their success in obtaining cooperation from respondents with varying characteristics (i.e., nonresponse error variance), rather than variance among interviewers in systematic measurement difficulties (i.e., measurement error variance) – that is, whether correlated responses within interviewers arise due to variance among interviewers in the pools of respondents recruited, or variance in interviewer-specific mean response biases. Unfortunately, work to date has only considered data from a CATI survey, and thus suffers from two limitations: Interviewer effects are commonly much smaller in CATI surveys, and, more importantly, sample units are often contacted by several CATI interviewers before a final outcome (response or final refusal) is achieved. The latter introduces difficulties in assigning nonrespondents to interviewers, and thus interviewer variance components are only estimable under strong assumptions. This study aims to replicate this initial work, analyzing data from a national CAPI survey in Germany where CAPI interviewers were responsible for working a fixed subset of cases.

*Key words:* Interviewer variance; nonresponse error variance; measurement error variance; face-to-face data collection; multilevel modeling; PASS study.

## 1. Introduction

Effects of interviewers on the measurement of survey variables have been clearly documented in surveys using face-to-face interviewing (Hansen et al. 1960; Kish 1962; Fellegi 1964; Freeman and Butler 1976; Collins and Butcher 1982; Mangione et al. 1992; O’Muircheartaigh and Campanelli 1998; Schnell and Kreuter 2005). In general, for certain survey items, responses collected by the same interviewer will tend to be more similar than responses collected by different interviewers. Assuming that random subsamples of the full sample have been assigned to the interviewers (interpenetrated assignment), one possible source of this between-interviewer variance is measurement error variance among the interviewers (e.g., Hansen et al. 1960; Biemer and Stokes 1991; Biemer and Trewin 1997; Groves 2004, ch. 8).

<sup>1</sup> Survey Methodology Program (SMP), Survey Research Center (SRC), Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48106, U.S.A. Email: bwest@umich.edu

<sup>2</sup> Joint Program in Survey Methodology (JPSM), Ludwig Maximilian University of Munich, Institute for Employment Research (IAB), Nuremberg, Germany. Email: fkreuter@survey.umd.edu

<sup>3</sup> Institute for Employment Research (IAB), Nuremberg, Germany. Email: Ursula.Jaenichen@iab.de

**Acknowledgments:** We are indebted to Gerrit Müller, Mark Trappmann and Daniela Hochfellner from the Institute for Employment Research (IAB) in Germany for their help in accessing and navigating the various data sets necessary to conduct this research. We are also grateful to Theresa Scholz from the RDC at IAB for providing the PSU-level data, and to the project “Nachbarschaftseffekte: Die Analyse individuell-rationalen Verhaltens im sozialen Kontext” conducted by the IAB, DIW Berlin and RWI Essen for granting us permission to use the data.

Various hypotheses have been proposed in the literature concerning the source of these intra-interviewer correlations (see Schaeffer et al. 2010 for a recent review), and many have been related to interactions between the interviewer and the respondents, such as differential interviewer probing (e.g., Mangione et al. 1992) or social desirability (Schnell 1997). Factual, non-sensitive survey items (e.g., age) would seemingly be immune to most of these issues, and often show lower intra-interviewer correlations than non-factual and sensitive items (Schnell and Kreuter 2005). However, the existing literature does show evidence of significant interviewer variance in responses to factual items despite interpenetrated designs (see West and Olson 2010 for a review). What has never been tested for face-to-face surveys is the alternative hypothesis that intra-interviewer correlations arise from variance among interviewers in the types of respondents successfully recruited (i.e., nonresponse error variance), which could provide an explanation for these unusual findings. We aim to test this hypothesis with this study.

Intra-interviewer correlations in survey responses have negative impacts on the quality of survey estimates, which makes the identification of mechanisms introducing the correlations important. Do the values on certain survey variables tend to be correlated within an interviewer due to sample assignment, the similarity of sample units recruited by an interviewer, or systematic measurement problems associated with an interviewer? Whatever its source, the intra-interviewer correlation has a multiplicative effect on the variance of an estimated mean in surveys that either achieve a 100% response rate, or have a plausible *missing completely at random* mechanism, where nonrespondents are a random subsample of the full sample. This effect is sometimes referred to as an *interviewer effect* (Davis and Scott 1995). This multiplicative interviewer effect on the variance is written as  $1 + (\bar{m} - 1)\rho_{int}$ , where  $\bar{m}$  represents the average sample workload for an interviewer and  $\rho_{int}$  represents an item-specific intra-interviewer correlation (Kish 1962). This effect, which is similar in definition to (and often larger than) the *design effect* due to cluster sampling (Davis and Scott 1995; Schnell and Kreuter 2005), is an undesirable product of the data collection process that can reduce the precision of survey estimates, limit the power of statistical analyses based on survey data, and increase costs of data collection, due to the need for a larger sample size to offset the losses in precision.

Unfortunately, face-to-face surveys rarely (if ever) achieve a 100% response rate, and interviewers play a key role in the recruitment of respondents. Repeatedly, interviewers have been shown to have differential recruitment success in face-to-face surveys; that is, response rates have consistently been shown to vary across interviewers (Wiggins et al. 1992; Morton-Williams 1993; Snijders et al. 1999; Campanelli and O'Muircheartaigh 1999; O'Muircheartaigh and Campanelli 1999; Pickery and Loosveldt 2002; Hox and de Leeuw 2002; Durrant et al. 2010). In light of these findings, West and Olson (2010) recently analyzed data from a telephone survey (the Wisconsin Divorce Study, or WDS) to test a hypothesis that variation across interviewers arose due to *nonresponse error variance*, or variance across interviewers of the nonresponse biases specific to each interviewer's subsample of cases when he or she achieves less than 100 percent response rates, rather than measurement error variance.

These authors found empirical support for the nonresponse error variance hypothesis for certain survey items. For example, in a CATI setting, with interpenetrated assignment of sample cases approximated (and demonstrated empirically) by performing separate

analyses for different calling shifts, the mean ages at times of divorce for WDS respondents (known from official records) were found to vary significantly across interviewers, and nearly 81% of the total interviewer variance in the reported age at divorce values arose from this nonresponse error variance. For some other items (e.g., months since divorce), measurement error variance among interviewers made up the vast majority of the total interviewer variance. The findings of West and Olson therefore suggested that interviewers may actually recruit respondents with different features according to certain variables (e.g., ages), and then measure them appropriately on those variables.

More generally, the ability to decompose interviewer variance into its various sources due to differential nonresponse error and measurement error across interviewers, given the right data set, provides survey researchers with a natural opportunity to simultaneously investigate two important error sources that define the larger *Total Survey Error* (TSE) framework, and recent publications in TSE have called for more investigations of this type (Biemer 2010; Groves and Lyberg 2010, p. 871). To date, the nonresponse error variance hypothesis tested by West and Olson (2010) has not been tested using data from a face-to-face survey. With the present study, we aim to test such a hypothesis, and contribute another empirical examination of simultaneous error sources to this growing area of research in Total Survey Error.

The literature on interviewer effects has shown that interviewer effects are much smaller in telephone surveys than in face-to-face surveys (Groves and Magilavy 1986), motivating additional research of this phenomenon in a personal interview setting. In face-to-face surveys, the social distance between the interviewer and the respondent is much smaller than for telephone surveys, introducing more opportunities for complex respondent-interviewer interactions than would be possible over the telephone. Existing studies have shown that respondents to telephone surveys are more likely to satisfice, less engaged and cooperative, more suspicious about the interview process and confidentiality, and more likely to present socially desirable responses (Beland and St. Pierre 2008; Holbrook et al. 2003). Interviewers in face-to-face surveys can address these concerns in a more personal manner, and differential ability to address these concerns and/or issues could impact both decisions to participate and measurement errors in a differential manner across interviewers, resulting in greater interviewer variance (Brunton-Smith et al. 2012). Differences in nonresponse error variance across interviewers may be one of the primary contributors to the unexpected interviewer variance in more objective survey items (such as age) that has been reported previously for face-to-face surveys, given that the interviewer plays a larger and more personal role in securing cooperation and establishing rapport with the respondent in these surveys. However, no existing studies have demonstrated this empirically, and we aim to contribute to this gap in the literature with the present study.

Testing the nonresponse error variance hypothesis using data from a telephone survey is also difficult due to cases being worked by multiple interviewers. Face-to-face surveys offer an advantage in this respect, in that cases are typically only worked by one interviewer, and there are typically few or no refusal conversion activities. Unfortunately, the interviewer-related nonresponse error variance hypothesis cannot be tested in the absence of interpenetrated assignment of subsamples to interviewers. Interpenetrated

assignment is much more difficult in personal interview surveys, primarily for cost reasons: Interviewers are often assigned to work in a single primary sampling unit, or PSU. The availability of records containing “true” values for key survey items for both the entire sample and the respondents, however, would enable estimation of interviewer variance in the:

- (1) means of true values for assigned sample units;
- (2) nonresponse errors, or differences between the means of true values for respondents and the means of assigned sample values for each interviewer; and
- (3) mean response deviations, or mean differences between reported values and true values for each interviewer.

The additional contributions of the last two variance components to total interviewer variance, above and beyond any variance that may be attributed to sampling or PSU-level features, can still be examined even in the absence of truly interpenetrated assignment of subsamples. The objective of this study is to estimate these components of interviewer variance in a face-to-face survey conducted in Germany, where a sample was drawn from administrative records containing true values for selected survey variables.

## 2. Data and Methods

### 2.1. Overview of the Labour Market and Social Security (PASS) Study

The Labour Market and Social Security (PASS, or “Panel Arbeitsmarkt und Soziale Sicherung” in German) study is a panel survey conducted by the Institute for Employment Research (IAB) in Nuremberg, Germany, and uses both CAPI and CATI interviewing techniques to collect labor market, household income, and unemployment benefit receipt data from a nationally representative sample of the German population, covering more than 12,000 households annually. This study has a stated purpose of providing “a new database which will allow social processes and the nonintended side-effects of labour market reforms to be assessed empirically” (<http://www.iab.de/en/befragungen.aspx>). Two samples of roughly equal size (initially 6,000 households each) have data collected from them annually: a sample of households receiving unemployment benefits (as recorded in registers of the German Federal Employment Agency), hereafter referred to as the UB sample, and a sample of households from the general German population with low-income households oversampled, hereafter referred to as the GP sample. The UB sample is refreshed each year by a sample of new entries to the population, and the GP sample was refreshed in the fifth wave (2011) with a cluster (municipality) sample from the German population register. Due to the availability of administrative information for the UB samples, we focus exclusively on selected UB samples from PASS for this study.

To date, five waves of PASS data collection have been completed (2006–2011). Using the AAPOR RR1 calculation, response rates at the household level in the initial wave of each UB sample have varied between 26.3% for the refreshment UB sample in Wave 2 (Gebhardt et al. 2009) and 31.3% for the refreshment UB sample in Wave 3 (Büngeler et al. 2010). While these response rates could be seen as relatively low due to the difficulty of collecting data in a face-to-face setting from this particular population (recipients of

unemployment benefits), we note that response rates for face-to-face surveys in Germany are generally lower than in the U.S. and most northern European countries. For example, response rates to the European Social Survey (ESS) have varied between 30% and 56% across waves in Germany (see <http://ess.nsd.uib.no> and Schnell 1997). In addition, strong refusal conversions are uncommon in Germany (Schnell 2012, p. 223).

The relatively low response rates in PASS and possible interviewer variance in the response rates have the potential to introduce large amounts of nonresponse error variance across interviewers, but recent studies have shown that nonresponse rates and nonresponse errors tend to have a very weak association (Groves and Peytcheva 2008), and the relative contributions of nonresponse error variance and measurement error variance to total interviewer variance have never been demonstrated empirically in a face-to-face survey, let alone in face-to-face surveys with different response rates. Replications of this study using face-to-face surveys with higher response rates will certainly be needed in the future to see whether the same patterns of results emerge. For additional details on the general design of the PASS study, readers can refer to Christoph et al. (2008) or Trappmann et al. (2010). Additional sampling details can be found in the German publication by Rudolph and Trappmann (2007).

In the PASS study, as in many large area probability sample surveys, CAPI interviewers are assigned to work in a single sampling point (or primary sampling unit). This introduces concerns regarding lack of interpenetration that we will address in our data analysis. To avoid interviewer effects that may arise due to repeated interviews conducted with the same household in the panel sample, we consider all sample cases attempted using CAPI in the original Wave 1 sample, and only refreshment sample households attempted using CAPI in Wave 2. At the time of this writing, the Wave 3 data lacked sufficient contact protocol data to be included in the analyses presented here, Wave 4 administrative data were not yet available, and Wave 5 data collection was still ongoing.

## 2.2. Data Set Construction

Before we describe the specific PASS data sets that we constructed for this investigation, it is worth describing the “ideal” data set that would be needed for this type of analysis. This ideal data set, collected in a face-to-face survey, would include:

- A variable containing interviewer ID codes, with a “large” random subsample of all sampled units assigned to each of a “large” number of interviewers (i.e., interpenetrated sampling).
- A response indicator, equal to 1 for sampled units cooperating with the survey request and 0 otherwise, with non-contacts preferably separated from refusals.
- A “large” number of respondents (say, at least 20; see Hox 1998), nested within each of a “large” number of interviewers (say, at least 50) and providing responses on a variety of objective and subjective variables.
- True values for selected objective (and possibly subjective) variables for the *full sample*, which are required for estimation of nonresponse error variance across interviewers, linked from administrative records or some other external source of record information, and

- Covariates describing characteristics of the geographic areas from which the respondents were sampled and where the interviewers are assigned to work.

We note that the availability of “true” values for selected survey variables for the full sample, from administrative records or some other external source, enables computation of nonresponse errors for each interviewer, in addition to the measurement errors associated with reported values on the survey variables. Given these data, one could then analyze the interviewer variance in the nonresponse errors and the measurement errors simultaneously.

At the time of this writing, we are not aware of the existence of any such “ideal” data sets for potential secondary analyses. The design of a study producing such a data set would generally be cost prohibitive, but this description represents a target for future studies. We therefore attempted to construct an approximation of this ideal data set using available information from the first two waves of the PASS data. For purposes of the analyses presented in this article, we constructed a pooled data set of households ever attempted using CAPI in the first two PASS waves. We focus specifically on the heads of households in the UB samples in these two waves, because the “true” values on several PASS study variables are only available for households that were sampled from the administrative records of unemployment benefit recipients. Additional persons responding to the survey in these households may have been recruited by the head of the household rather than the interviewer, so we do not include these individuals, who are also less likely to have administrative records, in the data sets.

The resulting data set contains 4,829 heads of households nested within 211 professional CAPI interviewers, amounting to nearly 23 households per interviewer, along with the following variables:

- The ID of the CAPI interviewer working the sampled household.
- A response indicator (1 = household responded; 0 = nonrespondent).
- “True” values representing official administrative data for the *heads of households* on eleven variables; these variables are extracted from the Integrated Employment Biographies (IEB) data, which are provided by the Research Data Center (FDZ) of the German Federal Employment Agency, and include:
  - Age (based on the date of birth in the IEB data and the date of the interview; for nonrespondents, the median interview date of the respondents in the same sample release is used).
  - Gender.
  - Foreign nationality status.
  - An indicator of whether the household was receiving unemployment benefits at the time of the interview.
  - Employment status at the time of the interview (used to create indicators of employment and unemployment), and
  - Five indicators for educational background and vocational status, including 1) no educational degree up to intermediate secondary school degree, without vocational training; 2) up to intermediate secondary school degree, together with completed vocational training; 3) completed secondary school (entry level for university); 4) technical college or other college for applied sciences; and 5) university degree.



- Aggregate information available from the administrative employment and unemployment records for the postal code areas in Germany, which are the PSUs for the PASS sample; following Bauer et al. (2011a, 2011b), these PSU-level covariates included:
  - Local Unemployment Rate (%).
  - Local Long-Term Unemployment Rate (%).
  - Size of workforce (employees subject to social security and the registered unemployed).
  - Percent of workers with a college or university degree.
  - Percent of workforce that is foreign workers.
  - Percent of workforce that is untrained workers.
  - Percent of workers age 20–30.
  - Percent of workers age 50–65.

Given the area probability sample selected for the PASS study, interpenetrated assignment of sampled households to interviewers was not cost efficient, and a single interviewer was typically assigned to work in a given primary sampling area. We therefore include fixed effects of the PSU-level covariates described above in the models used for the eleven PASS variables in our interviewer variance decomposition analyses, to account for any interviewer variance that may arise from variance among PSUs in these characteristics. While we feel that these PSU characteristics are relevant for the eleven specific survey variables analyzed in this study, there could certainly be additional PSU characteristics not available to us that introduced variance in the features of sample assignments across PSUs, and we acknowledge this as a limitation of our approach. Future research needs to consider more elegant techniques for estimating interviewer variance in non-interpenetrated designs (see Biemer 2010), and we expand on this suggestion in our Discussion.

We also constructed a second data set containing survey reports on the eleven variables mentioned above for the 1,472 heads of households cooperating with the PASS study request, and including interviewer ID codes and the PSU covariates. As noted earlier, response rates to face-to-face surveys in Germany tend to be lower than those in the United States and other northern European countries, and were likely lower in this case because we are working with the UB samples. This limited the power of our multilevel analyses of the respondents, as there were only about seven respondents per interviewer. However, we still find evidence of significant total interviewer variance components based on analyses of the respondents, which will be described in the upcoming Results section. This second data set did not include true values of the survey variables from the administrative data; this kind of linking required respondent consent, and only about 80 percent of the PASS respondents consented. As a consequence, we could not compute response deviations for the responding households without reducing the size of the respondent subset.

Item-missing data among the respondents were considered on a variable-by-variable basis, and heads of households were treated as nonrespondents if failing to provide data on a particular variable. Based on the item-missing data patterns, only interviewers with at least two survey respondents for a particular variable were considered in the eventual interviewer variance decomposition analyses, given that interviewer effects cannot be estimated based on a sample of size 1. As a result, the number of interviewers and the

number of full sample and responding cases vary slightly depending on the variable being analyzed. We acknowledge that future research in this area should aim to use larger samples in face-to-face surveys with higher response rates to minimize concerns about statistical power when analyzing interviewer variance among the respondents.

All analyses of these data were conducted in the IAB Research Data Center (RDC) at the University of Michigan-Ann Arbor (Bender and Heining 2011).

### 2.3. Statistical Analysis

Initial descriptive analyses focused on computation of response rates (using the AAPOR RR1 calculation, and including noncontact as a form of nonresponse) for each of the 211 CAPI interviewers in the pooled PASS data set, and plotting the distribution of the response rates. The intra-interviewer correlation in the response indicators was estimated by using the `xtmelogit` command in Stata (Version 12.1) to fit a simple multilevel logistic regression model to the PASS response indicator, including a fixed intercept and random interviewer effects.

For each of the eleven analysis variables described in Section 2.2, the three-step estimation methodology described by West and Olson (2010) was then applied to decompose total interviewer variance into variance in means of true values for full subsamples, nonresponse error variance, and measurement error variance. Let  $y_{ij}$  be the true value of survey variable  $Y$  for sample unit  $j$  assigned to interviewer  $i$ , and let  $x_{ij}$  be the reported value for  $Y$  for that sample unit if the unit responds to the survey. Further, let  $R$  denote respondents, and  $NR$  denote nonrespondents. Assuming interpenetrated assignment of subsamples to interviewers, the expectation of the mean of respondent reports  $\bar{x}_i$  for interviewer  $i$  can be written as

$$E(\bar{x}_i|i) = \bar{Y} + Bias_{NR,i} + Bias_{ME,i} = \bar{Y} + (\bar{y}_{R,i} - \bar{y}_i) + (\bar{x}_i - \bar{y}_{R,i}) \quad (1)$$

That is, the expected value of the respondent mean for interviewer  $i$  is the sum of 1) the population mean of the true values,  $\bar{Y}$ ; 2) the difference between the mean of the true values for all *respondents* interviewed by interviewer  $i$ ,  $\bar{y}_{R,i}$ , and the mean of the true values for all *units* assigned to interviewer  $i$ ,  $\bar{y}_i$  (nonresponse error); and 3) the difference between the mean of the *reported* values for all respondents interviewed by interviewer  $i$ ,  $\bar{x}_i$ , and  $\bar{y}_{R,i}$  (measurement error).

Assuming interpenetration and negligible covariance between the two bias terms, the variance of the expectation in (1) is defined by the sum of two variance components:  $Var(Bias_{NR,i})$  and  $Var(Bias_{ME,i})$ . We were unable to compute the empirical correlation of these two error sources in the PASS data set, given that not all respondents consented to having their administrative data *linked* to their survey reports. West and Olson (2010) suggest that these correlations are small in a CATI survey, but the correlations could be different in a face-to-face survey; more work is certainly needed to examine this assumption. We sought unbiased estimates of these two variance components along with the relative proportions of the total interviewer variance contributed by these two sources of variance among interviewers. Estimation of these variance components using closed-form estimators is possible for very simple design and response scenarios (e.g., equal assignment sizes and response rates across interviewers) typically not experienced in



practice. Given the unequal assignment sizes and respondent counts for each CAPI interviewer in the PASS survey, we estimated these components for each variable using three distinct steps. The three steps below apply to the continuous variable *age*, and outline the general analytic approach; subsequent remarks address the estimation steps for the ten binary indicators. Item nonresponse on each variable introduced slight differences in terms of the respondent sample sizes for the variables.

*Estimation Step 1: Estimate the interviewer variance in the means of the true values for the assigned sample cases.* First, we estimated the variance among interviewers in the means of the true values for all CAPI sample cases assigned to each interviewer. Assuming interpenetrated assignment of cases to interviewers, this variance component should be negligible, as all interviewers should have a full sample mean of true values equal, on average, to the population mean. We estimated this component using a one-way random effects model, assuming that the interviewers were a random subsample from a larger hypothetical population of interviewers:

$$y_{ij} = \bar{Y} + b_i + e_{ij} \quad (2)$$

In this notation,  $y_{ij}$  is the true value of variable  $Y$  for sample unit  $j$  assigned to interviewer  $i$ ,  $b_i$  is the random deviation of interviewer  $i$ 's assignment mean from the population mean of the true values,  $\bar{Y}$ , and  $e_{ij}$  is a normally distributed random error with mean 0 and constant variance (the element variance within each assignment). We estimated the variance of the  $b_i$ , or  $\text{Var}(b_i) = \sigma_{\text{int},\text{full}}^2$ , using restricted maximum likelihood (REML) estimation to obtain an unbiased estimate of this variance component given unequal interviewer workloads (Patterson and Thompson 1971).

We tested a null hypothesis that the interviewer variance component is equal to zero,  $H_0 : \sigma_{\text{int},\text{full}}^2 = 0$ , versus the alternative that assignment means vary across interviewers,  $H_A : \sigma_{\text{int},\text{full}}^2 > 0$ ; see West and Olson (2010) for a discussion of appropriate methods for testing this null hypothesis. We performed asymptotic likelihood ratio tests of this null hypothesis, referring the standard likelihood ratio chi-square statistic to a mixture of chi-square distributions, with degrees of freedom 0 and 1 and equal weight 0.5 (Zhang and Lin 2008). We report  $p$ -values for the test statistics under this null distribution. The `xtmixed` and `xtmelogit` commands in Stata facilitate these tests.

*Estimation Step 2: Estimate the interviewer variance in the means of the true values for the responding sample cases.* Second, when assignments are interpenetrated, each interviewer has the same  $\bar{y}_i$  in expectation, and the variance of the nonresponse biases simplifies to  $\text{Var}(\text{Bias}_{NR,i}) = \text{Var}(\bar{y}_{R,i} - \bar{y}_i) = \text{Var}(\bar{y}_{R,i})$ . We thus estimated the non-response error variance component by estimating the variance across interviewers in the means of the true values for *respondents*. We again used a one-way random effects model for the true values of *respondents* to the survey request,  $y_{R,ij}$ :

$$y_{R,ij} = \bar{Y}_R + b'_i + e'_{ij} \quad (3)$$

Here,  $b'_i$  captures the random deviation of each interviewer's mean for their recruited respondents' true values from the expected value of the mean of the true values for respondents over all possible sample assignments to interviewers ( $\bar{Y}_R$ ). We note that in the absence of overall nonresponse bias,  $\bar{Y}_R$  will be equal to the population mean of the true

values,  $\bar{Y}$ . We estimated the variance of these random effects,  $\text{Var}(b'_i) = \sigma^2_{\text{int}, \text{resp}}$ , using REML to obtain an unbiased estimate of  $\text{Var}(\text{Bias}_{NR,i})$ . We tested this component of variance against zero using the likelihood ratio test described above.

*Estimation Step 3: Estimate the interviewer variance in the means of the reported values for the responding sample cases (the total interviewer variance).* Third, under an assumption of interpenetrated assignment of subsamples to interviewers, Equation 1 can be rewritten as:

$$E(\bar{x}_i|i) = \bar{Y} + \text{Bias}_{NR,i} + \text{Bias}_{ME,i} = \bar{Y} + (\bar{y}_{R,i} - \bar{Y}) + (\bar{x}_i - \bar{y}_{R,i}) = \bar{x}_i \quad (4)$$

Using the sample mean of the respondent reports for interviewer  $i$  as an estimate of this expectation, we then computed an unbiased estimate of the variance in the means of the reported values across interviewers (Equation 4) using a one-way random effects model and REML estimation:

$$x_{ij} = \bar{X}_R + b''_i + e''_{ij} \quad (5)$$

This is the interviewer variance model that is often estimated in practice using respondent data only. The variance of the random interviewer deviations ( $b''_i$ ) around the expected value of the mean of the respondent reports over all possible sample assignments to interviewers ( $\bar{X}_R$ ), or  $\text{Var}(b''_i) = \sigma^2_{\text{int}, \text{resp}, \text{obs}}$ , captures measurement error variance and nonresponse error variance introduced by the interviewers, along with sampling variance in the absence of interpenetration, and is thus the “total variance” due to interviewers. We tested this variance component for significance using appropriate likelihood ratio tests. We then subtracted the estimate of the nonresponse error variance from Estimation Step 2 to get an estimate of the measurement error variance across interviewers. The estimated proportion of variance introduced by interviewers due to nonresponse error variance was then computed as:

$$\frac{\hat{\sigma}^2_{\text{int}, \text{resp}} - \hat{\sigma}^2_{\text{int}, \text{full}}}{\hat{\sigma}^2_{\text{int}, \text{resp}, \text{obs}} - \hat{\sigma}^2_{\text{int}, \text{full}}} \quad (6)$$

Evidence of successful interpenetration from Estimation Step 1 implies that  $\sigma^2_{\text{int}, \text{full}} = 0$  (i.e., interviewer-level means of true values for their full assignments do not vary). We subtracted estimates of  $\sigma^2_{\text{int}, \text{full}}$  from the numerator and denominator to remove components of variance that were not due to the interviewer from this calculation.

Given that ten of the eleven PASS variables were binary in nature (1 = yes, 0 = no), multilevel logistic regression models with random interviewer effects were fitted in each estimation step to estimate the components of variance due to interviewers in the *log-odds* of each binary variable being equal to 1, rather than variance in the means of outcomes in the standard one-way random effects models for the age variable, which followed a roughly normal distribution. These models were fitted using approximate maximum likelihood methods, based on the adaptive Gauss-Hermite approximation to the log-likelihood of the model, as implemented in the `xtmelogit` command in the Stata software.

Finally, to address the confounding of interviewer ID and PSU (or sampling area) in the PASS data set, we repeated the analyses described above after including fixed effects of the PSU-level covariates in the models used for each analysis step. Estimated components

of variance at each analysis step may arise due to variance in PSU-level features rather than interviewers; for example, nonresponse error variance among interviewers for a specific variable may be due to between-PSU variance in features that are correlated with both the variable of interest and response propensity. These analyses were performed to estimate components of variance due to the interviewers after removing any sources of interviewer variance (at any step) due to between-PSU variance in the values of the available PSU-level covariates.

3. Results

Figure 1 below presents a histogram showing the distribution of unweighted response rates across the 211 CAPI interviewers.

Figure 1 illustrates the substantial variance among these interviewers in terms of their unweighted response rates, which is consistent with the existing literature (Wiggins et al. 1992; Morton-Williams 1993; Snijkers et al. 1999; Campanelli and O’Muircheartaigh 1999; O’Muircheartaigh and Campanelli 1999; Pickery and Loosveldt 2002; Hox and de Leeuw 2002; Durrant et al. 2010) and provides motivation for the hypothesis being examined in this study. The estimated intra-interviewer correlation in the binary response indicators for the sampled households assigned to the 211 interviewers was computed as follows:

$$\hat{\rho}_{\text{int},R} = \frac{\hat{\sigma}_{\text{int},R}^2}{\hat{\sigma}_{\text{int},R}^2 + \pi^2/3} = \frac{0.385}{0.385 + \pi^2/3} = 0.105$$

The estimated between-interviewer variance in the log-odds of responding (0.385) was found to be significantly greater than zero based on an asymptotic likelihood ratio test

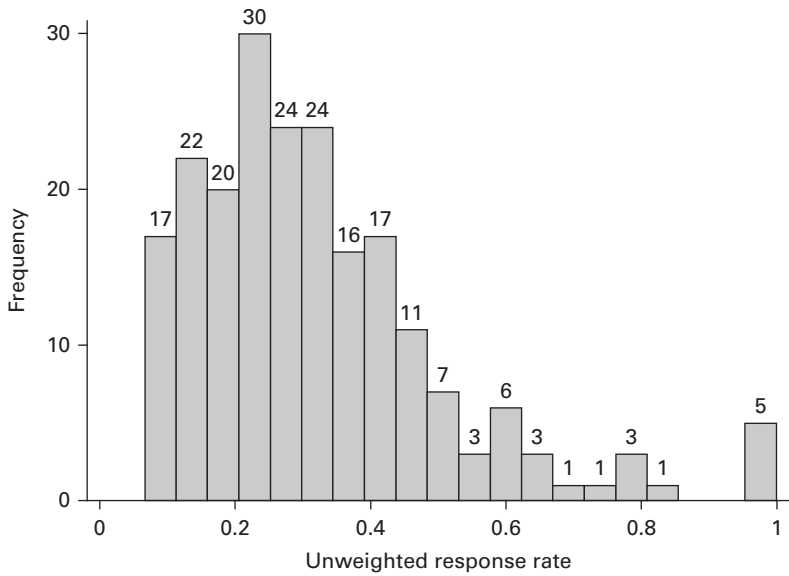


Fig. 1. Distribution of unweighted response rates across the 211 CAPI interviewers in PASS Waves 1 and 2

using a mixture of chi-square distributions ( $p < .001$ ), providing strong support for the conclusion that the within-interviewer correlation in response indicators was significant. The existence of significant between-interviewer variance in response rates introduces the possibility of varying nonresponse error between interviewers, which we examine for the eleven PASS variables below. The estimated interviewer variance component for the log-odds of responding dropped to 0.343 after including the fixed effects of all of the PSU-level covariates in the logistic regression model, meaning that the PSU-level features accounted for about 10.9% of the between-interviewer variance in the response rates, but this component of variance remained significantly greater than zero. Small amounts of item-missing data on each of these eleven survey variables did not substantially alter the interviewer variance components for the response indicators when repeating the same analysis on a variable-by-variable basis.

Previous studies of interviewer effects on survey variables (Schnell and Kreuter 2005) have noted that interviewer variance components tend to be larger than variance components due to sampling areas. The findings above, which show significant interviewer variance in the response rates, suggest that while the same may be true in terms of response indicators, it would still be important to adjust for the relationships of area-level features with response before inferring how much of the variability in cooperation rates is truly due to interviewers. We find that there is still variance among interviewers in response probabilities, even after accounting for the relationships of the PSU-level covariates with the probability of response.

We now turn to our interviewer variance decomposition analyses. Table 1 below provides estimates of the variance components of interest at each step of the estimation process. Table 1 also provides estimates of intra-interviewer correlations ( $\hat{\rho}_{\text{int}}$ ) based on the estimated total interviewer variance components, computed using the respondent reports.

Examining the estimates of the total interviewer variance components in row 8 of Table 1, we find evidence that five of the total interviewer variance components are greater than zero (using  $p < .10$  for marginal significance, and  $p < .05$  for significance) after adjusting for fixed effects of the PSU covariates: age of household head, household receipt of unemployment benefits, employment of household head, having up to an intermediate secondary degree together with vocational training, and having a degree from a technical college or from another college for applied sciences. Notably, given the fairly low response rates for the households in the UB samples in Waves 1 and 2 of the PASS survey, we have limited power to detect significant components of variance due to the interviewers based on the *respondent* data. We have close to 23 sampled households per CAPI interviewer in the full sample (row 3 of Table 1), and only about seven responding households per interviewer in the respondent subset (row 6 of Table 1). We therefore have adequate power for detecting moderate interviewer variance components based on the full sample, having more than 50 interviewers with more than 20 households per interviewer (see Hox 1998), but limited power based on the respondent subset. Despite this limitation, we still find five significant total interviewer variance components, and our research interest lies in the sources of these significant variance components. We focus on the estimates and tests of significance in Table 1 *after* adjusting for the effects of the PSU-level covariates, to account for any variance among interviewers arising from the areas of Germany in which they are working.

Table 1. Decompositions of interviewer variance components for eleven PASS survey variables with available validation data, with estimated interviewer variance components presented before and after adjustment for selected PSU-level covariates

	Age	Benefit Receipt	Gender (Female)	Foreign Status	EMP	UNEMP	Education and Vocational Training				
							Up to Inter-mediate Sec Degree, no VOC	Up to Inter-mediate Sec Degree, VOC	SEC School, Complete	TECH College	UNIV
1 Interviews	210	209	211	210	211	211	211	211	211	211	211
Step 1: Full Sample True Values											
2 Full sample $n$	4790	4787	4803	4801	4803	4803	4803	4803	4803	4803	4803
3 Full $n_{\text{int}}$	22.8	22.9	22.8	22.9	22.8	22.8	22.8	22.8	22.8	22.8	22.8
4 $\sigma^2_{\text{int,full}}$	3.03***/ 2.12***	0.07***/ 0.04*	0.04**/ 0.02	0.49***/ 0.02	0.06***/ 0.04**	0.19***/ 0.07***	0.26***/ 0.11***	0.20***/ 0.12***	0.18***/ 0.10**	<0.01/ <0.01	0.61***/ 0.22***
Step 2: Respondent True Values											
5 Respondent $n$	1377	1351	1383	1378	1383	1383	1368	1368	1368	1368	1368
6 Resp. $n_{\text{int}}$	6.6	6.5	6.6	6.6	6.6	6.6	6.5	6.5	6.5	6.5	6.5
7 $\sigma^2_{\text{int,resp}}$	4.62***/ 3.54*	0.31***/ 0.15	0.06/ 0.01	0.48***/ 0.05	<0.01/ <0.01	0.79***/ 0.44***	0.42***/ 0.17**	0.33***/ 0.19***	0.08/ <0.01	<0.01/ <0.01	1.53***/ 1.00***
Step 3: Respondent Reported Values											
8 $\sigma^2_{\text{int,resp,obs}}$	5.37***/ 3.58*	0.44***/ 0.33***	0.08*/ 0.03	0.64***/ 0.09	0.51***/ 0.32**	0.12**/ <0.01	0.31***/ 0.08	0.30***/ 0.13**	0.35***/ 0.13	1.56***/ 1.59**	0.300.09
9 $\hat{\rho}_{\text{int}}$	0.03	0.09	0.01	0.03	0.09	<0.01	0.02	0.04	0.04	0.33	0.03

NOTES: Cells presenting estimated variance components (rows 4, 7, and 8) include the estimates and tests of significance before/after adjustment for fixed effects of PSU-level covariates. Likelihood Ratio Test results: \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$ . Variable Abbreviations: EMP = Employed; UNEMP = Unemployed; VOC = Vocational Training; SEC = Secondary; TECH = Technical; UNIV = University.

First considering the age variable, we see evidence of between-interviewer variance in the average ages of the heads of households for the full sample assignments, both before and after controlling for fixed effects of the available PSU-level covariates in the multilevel model. This apparent lack of interpenetrated assignment of households to interviewers based on the ages of the heads of households precludes use of the aforementioned three-step estimation methodology to decompose the interviewer variance. Nevertheless, we note a 67% increase in the between-interviewer variance of the true age values when considering respondents only, from an estimate of 2.12 for the full sample to an estimate of 3.54 for the respondent subset (controlling for the PSU-level covariates), followed by a very slight 1% increase in the between-interviewer variance when considering the *reported* age values, rather than the true age values, from an estimate of 3.54 to an estimate of 3.58. While we cannot definitively decompose the total interviewer variance component for age based on these estimates, due to the lack of interpenetrated assignments, these findings suggest that the marginally significant ( $p < .10$ ) total interviewer variance for the ages of the heads of households arises from a combination of variance in the full sample assignments *and* variance in the ages of the household heads for the recruited respondents, rather than variance in the measurement errors. This is to be expected for a very objective survey item such as age, where extensive cognitive processing or interviewer clarification is generally not needed.

We see different results for the survey variable measuring current household receipt of unemployment benefits (yes/no). Analyzing the true values of this indicator variable for the sampled households assigned to each interviewer, we find marginal evidence of variance among interviewers ( $p < .10$ ) after controlling for the fixed effects of the PSU-level covariates, suggesting that interpenetrated assignment of households is at least a plausible assumption after accounting for the PSU-level covariates. For this variable, the estimate of the total interviewer variance component is 0.33 ( $p < .001$ ). Further assuming negligible correlations of the interviewer-specific nonresponse errors and measurement errors, which we cannot estimate using this data set, we can subtract the small portion of the variance due to the assigned samples (0.04) from the estimated interviewer variance in true values among respondents (0.15, resulting in 0.11) and from the estimated total variance (0.33, resulting in 0.29). We can then estimate that  $0.11/0.29 = 0.38 = 38\%$  of the additional variance in unemployment benefit receipt indicator values introduced by the interviewers is due to nonresponse error variance, with the remaining portion (62%) due to measurement error variance. These results suggest that interviewers are tending to recruit households with different current benefit receipt status, and then (potentially) having differential difficulty accurately measuring this variable in the interview (to a greater extent than their respondent pools are differing). After adjusting for the PSU-level covariates, the total interviewer variance for this variable corresponds to an estimated intra-interviewer correlation of  $\hat{\rho}_{\text{int}} = 0.33/(0.33 + \pi^2/3) = 0.09$ .

We find yet another pattern of results when considering the employment indicator for the head of the household. We only find weak evidence of interpenetrated assignment of sampled households to interviewers, with the variance in the proportions of households with the heads employed among the full interviewer samples significant at the 0.05 level after accounting for the fixed effects of the PSU-level covariates. While this once again precludes the use of our decomposition approach in theory, we note that the estimate of



between-interviewer variance when considering true values of the employment indicator for respondents only is extremely small. This finding suggests that little of the original variance in the full sample assignments remains, and that the sets of respondents recruited by the interviewers now vary in a way that eliminated the between-interviewer variance in the original assignments. In other words, the proportions of household heads that are employed are similar across interviewers in their recruited respondents. We then find evidence of significant between-interviewer variance in terms of the *reported* employment values for the respondents; the adjusted estimate of the total interviewer variance component (0.32) is actually 700% larger than the original variance component arising from the full sample assignments (0.04). Although the interviewers did appear to be working different samples to begin with, these differences became much more pronounced when considering the reported values among respondents, and these differences did not appear to arise due to differential recruitment. After adjusting for the PSU-level covariates, the total interviewer variance for this variable corresponds to an estimated intra-interviewer correlation of  $\hat{\rho}_{\text{int}} = 0.32 / (0.32 + \pi^2/3) = 0.09$ .

Finally, we find evidence of significant total interviewer variance for two of the indicators of educational level and vocational training of the head of the household. For the variable indicating up to an intermediate secondary degree together with completed vocational training, which is the modal category in this population, we once again find evidence of a lack of interpenetrated assignment of households. After adjusting for the fixed effects of the PSU-level covariates, the estimated interviewer variance in the indicator values for this category increased somewhat when considering true values for respondents only, and then fell back to the level of the original full sample when considering *reported* values for respondents only. Unusual patterns of results like this, where the total interviewer variance is smaller than the variance in the true values for respondents among interviewers, could arise from a *negative covariance* between nonresponse errors and measurement errors among interviewers. For example, interviewers may recruit sets of respondents with different educational profiles (relative to their assigned samples), with certain interviewers recruiting higher than expected proportions of individuals with up to an intermediate secondary degree and completed vocational training (i.e., a positive nonresponse error). However, these interviewers may then collect reported values of education that are lower than the truth (i.e., a negative measurement error), resulting in proportions that are lower than those based on the true values for their subset of recruited respondents. Unfortunately, we are unable to estimate the covariance of these two error sources for all interviewers using the PASS data sets, as the measurement errors require linking of the survey data with the administrative data, and this is only possible if the respondents consent to this process (and not all PASS respondents consented).

When considering the education/vocation indicator for technical college or other colleges for applied sciences, there was evidence of interpenetrated assignment, no evidence of a change in the interviewer variance given the features of the recruited samples, and then a substantial increase in the interviewer variance when considering the reported values on this indicator. These results point to possible differential misunderstanding among the interviewers, the respondents, or both when considering what exactly this category corresponds to.

In general, we note that in nearly all cases in Table 1, the estimated interviewer variance components in the various multilevel models are substantially decreased when controlling for the fixed effects of the PSU-level covariates. This was also noted earlier when examining interviewer variance in the response rates for each individual variable. Schnell and Kreuter (2005) showed that interviewer effects generally tend to be larger than area effects, which would suggest that it is not always necessary to disentangle area effects and interviewer effects when studying interviewer variance. The results in this study indicate a general need to adjust for the effects of area-level covariates when using multilevel models to quantify interviewer variance in a face-to-face survey, where interviewers are not assigned random subsets of the full sample; this is often the case in practice when interviewers only work in a single sampled area. While area effects may be smaller than interviewer effects, they can cause interviewer variance components to seem larger than they really are, and that was noted in this study. In this case, variance in the features of sampled areas could be introducing what seems like variance among interviewers in both the features of the recruited respondents and the collected survey responses.

#### **4. Discussion**

This is the first study to examine the contributions of nonresponse error variance and measurement error variance among interviewers to total interviewer variance in a face-to-face survey setting. The presence of administrative records in the PASS study, which collects data annually from a national sample of households receiving unemployment benefits in Germany, enabled the decomposition of total interviewer variance estimates for selected variables into nonresponse error variance and measurement error variance components. We first found significant variance among CAPI interviewers in terms of response rates for newly recruited cases, even after adjusting for the relationships of several features of the sampling areas with the response indicators. This finding is consistent with the prior literature in this area, and provides further evidence that interviewers do tend to vary in terms of their response rates. This widely reported result introduces the possibility of variance among the interviewers in nonresponse errors.

Second, we found significant (or marginally significant) total interviewer variance components for five of eleven PASS variables with available administrative data. For two of these variables, there was strong evidence ( $p < .01$ ) of a lack of interpenetrated assignment of sampled households to the interviewers, even after adjusting for the relationships of PSU-level covariates with these variables, which prevented the application of our decomposition methodology. Nevertheless, we did find evidence that significant interviewer variance in the ages of the heads of the assigned households was substantially increased when considering the responding heads of households, suggesting that the demographic features of recruited respondents may vary among interviewers to an even greater extent than their assigned samples vary. This finding was consistent with the findings of West and Olson (2010), who analyzed interpenetrated assignments in a telephone survey.

For the remaining three variables exhibiting evidence of significant ( $p < .05$ ) interviewer variance in the respondent reports in combination with at least marginal evidence of originally interpenetrated assignments, in terms of the true values on these

three variables, we found that the bulk of the total interviewer variance appeared to be driven by measurement error variance. Notably, for the indicator of whether a sampled household is currently receiving unemployment benefits, there appeared to be a mixture of nonresponse error variance and measurement error variance driving the total interviewer variance, with the majority of the variance stemming from measurement error variance. For the remaining two variables, an employment indicator and an indicator of completing technical college, measurement error variance appeared to be the primary source of the interviewer variance. These results therefore suggest that variance among interviewers in measurement difficulties tends to be the primary driver of total interviewer variance, at least for these PASS variables.

This study was certainly not without limitations. First, as indicated above, we were limited in this investigation by a lack of interpenetrated assignment of sampled households to the CAPI interviewers. Even after controlling for the relationships of PSU-level covariates with the variables in question, we still found significant variability among the interviewers in the true values of the assigned sample households for seven of the eleven PASS variables. While this could have been due to measurement error in the administrative records, especially for the education indicators (given the population of interest), this is a general problem with studying interviewer variance in face-to-face surveys. We can monitor changes in interviewer variance components from the full assigned samples to the respondents in a descriptive sense, but the estimation methodology outlined in this article relies on interpenetrated assignments for decomposing the interviewer variance. While the design of a future face-to-face survey of a large population with random subsamples of the full population sample assigned to interviewers would be ideal for future studies of this problem, such a design is generally not an economic reality for most survey organizations. Future research needs to focus on the development of methods for both estimating and decomposing interviewer variance in face-to-face surveys where interviewers were not originally assigned random subsamples of the full sample. For example, subsets of interviewers might be matched in terms of similarities of the sampled areas in which they are working, and then interviewer variance could be estimated based on the random samples assigned to the interviewers in these matched areas. However, interviewer sample sizes could continue to be an issue here.

Second, in the estimation steps presented in this article, we examined components of variance assuming that sampling errors, nonresponse errors, and measurement errors associated with the interviewers were independent. Non-zero covariances between these error sources may be important components of total interviewer variance, but we did not have the type of data needed to estimate these covariances; the survey data could only be linked to the administrative data with consent from the PASS respondents. More empirical work is certainly needed to examine assumptions of negligible covariance between these three error sources, and this work will require data sets where the true values for selected survey variables are available for the entire sample and can be linked to the respondent reports on those survey variables. West and Olson (2010) did examine such correlations among these interviewer-specific error sources and found that they were generally small, but this work needs to be repeated using data from a face-to-face survey.

Third, we had limited statistical power to detect components of variance due to interviewers among respondents, given the relatively low response rates found in the PASS samples of German households receiving unemployment benefits. This tends to be a difficult population to survey, and response rates for face-to-face surveys in Germany also tend to be lower than those found in the U.S. and other northern European countries. The ideal study of this problem would require CAPI interviewing, interpenetrated assignment of large subsamples to a large number of interviewers, a relatively high response rate, and available record values of good quality on respondents and nonrespondents for variables measured in the survey. Although potentially expensive, studies with these features will greatly enhance our understanding of these issues.

There are promising extensions of this work that we leave to future research. At present, estimates of intra-interviewer correlations are based on respondent reports only, and do not recognize possible contributions of nonresponse error variance to interviewer variance. The same is true for estimates of intra-cluster correlations, which are often used to estimate design effects in complex area probability samples. Collectively, the empirical work in this study and the study by West and Olson (2010) provide motivation for the analytical development of estimators of intra-interviewer (and intra-cluster) correlations that recognize contributions of sampling variance, refusal error variance, noncontact error variance, and measurement error variance to the total variance among interviewers (or clusters). Existing work by Biemer (1980) and Groves and Magilavy (1984) provides some possible avenues for this analytical development.

Alternative applications of multilevel modeling may also prove useful for studying this problem in the future. For example, given a data set with true values available for a full sample, a multilevel modeling framework incorporating multiple imputations of measurement errors for survey nonrespondents would enable estimation of interviewer variance components due to nonresponse error variance, via random interviewer coefficients for a response indicator with possible values  $1 = \text{respondent}$ ,  $-1 = \text{nonrespondent}$  (predicting actual and imputed respondent reports), and measurement error variance, via random interviewer intercepts, in addition to the *covariance* of these two errors sources at the interviewer level. Future applications of these types would also enhance our understanding of these issues.

Finally, if administrative records are available for selected items, field supervisors could monitor empirical best linear unbiased predictors (EBLUPs) of the random interviewer effects in the models that we propose, and identify interviewers with extremely unusual random effects in the models for either nonresponse error variance or total variance. For example, the EBLUPs computed from the nonresponse error variance estimation step for the active interviewers in a given survey could be ordered from highest to lowest, where the highest values for a particular survey variable would indicate interviewers with mean values for respondents that are substantially higher than the mean values for their assigned sample; this would mean that they only tend to recruit respondents with higher values on the particular variable. Interviewers having one of the five highest or one of the five lowest EBLUPs could have their recruited cases examined in more detail to see whether the supervisors would need to intervene and alter the recruiting strategies being used by these interviewers. Future studies could then evaluate the ability of this type of intervention to reduce total interviewer variance.

## 5. References

- Bauer, T.K., Fertig, M., and Vorrell, M. (2011a). Neighborhood Effects and Individual Unemployment. SOEP paper 409, DIW Berlin.
- Bauer, T.K., Flake, R., and Sinning, M.G. (2011b). Labor Market Effects of Immigration: Evidence from Neighborhood Data. Ruhr Economic Papers #257, Bochum, Dortmund, Duisburg, Essen.
- Beland, Y. and St-Pierre, M. (2008). Mode Effects in the Canadian Community Health Survey: A Comparison of CATI and CAPI. *Advances in Telephone Survey Methodology*, Chapter 14, J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japac, P.J. Lavrakas, M.W. Link, R.L. Sangsler (eds). New York: Wiley.
- Bender, S. and Heining, J. (2011). The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing. FDZ Methodenreport, 07/2011, Nürnberg.
- Biemer, P.P. (1980). A Survey Error Model which Includes Edit and Imputation Error. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 616–621.
- Biemer, P.P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74, 817–848.
- Biemer, P.P. and Stokes, S.L. (1991). Approaches to the Modeling of Measurement Error. *Measurement Errors in Surveys*, P.P. Biemer, R.M. Groves, L. Lyberg, N.A. Mathiowetz, and S. Sudman (eds). New York: Wiley.
- Biemer, P.P. and Trewin, D. (1997). A Review of Measurement Error Effects on the Analysis of Survey Data. In *Survey Measurement and Process Quality*, L. Lyberg, P.P. Biemer, M. Collins, E.D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds). New York: Wiley-Interscience.
- Brunton-Smith, I., Sturgis, P., and Williams, J. (2012). Is Success in Obtaining Contact and Cooperation Correlated with the Magnitude of Interviewer Variance? *Public Opinion Quarterly*, 76, 265–286.
- Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R., and Tschersich, N. (2010). IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/2009). Methoden- und Feldbericht. FDZ Methodenreport, 10/2010, Nürnberg.
- Campanelli, P. and O’Muircheartaigh, C. (1999). Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Journal of Official Statistics*, 2, 303–314.
- Christoph, B., Müller, G., Gebhardt, D., Wenzig, C., Trappmann, M., Achatz, J., Tisch, A., and Gayer, C. (2008). Codebook and Documentation of the Panel Study “Labour Market and Social Security” (PASS). Volume 1: Introduction and Overview, Wave 1 (2006/2007). FDZ Datenreport, 05/2008, Nürnberg.
- Collins, M. and Butcher, B. (1982). Interviewer and Clustering Effects in an Attitude Survey. *Journal of the Market Research Society*, 25, 39–58.
- Davis, P. and Scott, A. (1995). The Effect of Interviewer Variance on Domain Comparisons. *Survey Methodology*, 21, 99–106.
- Durrant, G.B., Groves, R.M., Staetsky, L., and Steele, F. (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly*, 74, 1–36.

- Fellegi, I.P. (1964). Response Variance and its Estimation. *Journal of the American Statistical Association*, 59, 1016–1041.
- Freeman, J. and Butler, E.W. (1976). Some Sources of Interviewer Variance in Surveys. *Public Opinion Quarterly*, 40, 79–91.
- Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H., Graf, T., and Biedermann, A. (2009). Codebook and Documentation of the Panel Study “Labour Market and Social Security” (PASS). Volume 1: Introduction and Overview, Wave 2 (2007/2008). FDZ Datenreport, 06/2009, Nürnberg.
- Groves, R.M. (2004). The Interviewer as a Source of Survey Measurement Error. *Survey Errors and Survey Costs*, (Second Edition). New York: Wiley-Interscience.
- Groves, R.M. and Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, 74, 849–879.
- Groves, R.M. and Magilavy, L.J. (1984). An Experimental Measurement of Total Survey Error. *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Section on Survey Research Methods*, 698–703.
- Groves, R.M. and Magilavy, L.J. (1986). Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys. *Public Opinion Quarterly*, 50, 251–266.
- Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72, 167–189.
- Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1960). Measurement Errors in Censuses and Surveys. *Bulletin of the International Statistical Institute*, 32nd Session, Vol. 38, Part 2, 359–374.
- Holbrook, A.L., Green, M.C., and Krosnick, J.A. (2003). Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Bias. *Public Opinion Quarterly*, 67, 79–125.
- Hox, J.J. (1998). Multilevel Modeling: When and Why. In *Classification, Data Analysis, and Data Highways*, I. Balderjahn, R. Mathar, and M. Schader (eds). New York: Springer.
- Hox, J.J. and de Leeuw, E.D. (2002). The Influence of Interviewers’ Attitude and Behavior on Household Survey Nonresponse: An International Comparison. *Survey Non-response*, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: Wiley.
- Kish, L. (1962). Studies of Interviewer Variance for Attitudinal Variables. *Journal of the American Statistical Association*, 57, 92–115.
- Mangione, T.W., Fowler, F.J., and Louis, T.A. (1992). Question Characteristics and Interviewer Effects. *Journal of Official Statistics*, 8, 293–307.
- Morton-Williams, J. (1993). *Interviewer Approaches*. Aldershot: Dartmouth Publishing Company Limited.
- O’Muircheartaigh, C. and Campanelli, P. (1998). The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision. *Journal of the Royal Statistical Society, Series A*, 161, 63–77.



- O'Muircheartaigh, C. and Campanelli, P. (1999). A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *Journal of the Royal Statistical Society, Series A*, 162(Part 3), 437–446.
- Patterson, H.D. and Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58, 545–554.
- Pickery, J. and Loosveldt, G. (2002). A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. *Quality and Quantity*, 36, 427–437.
- Rudolph, H. and Trappmann, M. (2007). Design und Stichprobe des Panels “Arbeitsmarkt und Soziale Sicherung” (PASS). In *Neue Daten für die Sozialstaatsforschung. Zur Konzeption der IAB-Panelerhebung “Arbeitsmarkt und Soziale Sicherung”*, (ed.) M. Promberger, IAB-Forschungsbericht, 12/2007, Nürnberg.
- Schaeffer, N.C., Dykema, J., and Maynard, D.W. (2010). Interviewers and Interviewing. In *Handbook of Survey Research*, J.D. Wright and P.V. Marsden (eds). (Second Edition). Bingley, U.K. Emerald Group Publishing Limited.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.
- Schnell, R. (2012). *Survey-Interviews. Standardisierte Befragungen in den Sozialwissenschaften*. Wiesbaden: VS-Verlag.
- Schnell, R. and Kreuter, F. (2005). Separating Interviewer and Sampling-Point Effects. *Journal of Official Statistics*, 21, 389–410.
- Snijders, G., Hox, J.J., and de Leeuw, E.D. (1999). Interviewers' Tactics for Fighting Survey Nonresponse. *Journal of Official Statistics*, 15, 185–198.
- Trappmann, M., Gundert, S., Wenzig, C., and Gebhardt, D. (2010). PASS: a Household Panel Survey for Research on Unemployment and Poverty. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 130, 609–622.
- West, B.T. and Olson, K. (2010). How Much of Interviewer Variance is Really Nonresponse Error Variance? *Public Opinion Quarterly*, 74, 1004–1026.
- Wiggins, R.D., Longford, N.T., and O'Muircheartaigh, C.A. (1992). A Variance Components Approach to Interviewer Effects. In *Survey and Statistical Computing*, A. Westlake, R. Banks, C. Payne, and T. Orchard (eds). Amsterdam: North-Holland.
- Zhang, D. and Lin, X. (2008). Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and Other Related Topics. *Random Effect and Latent Variable Model Selection*, D.B. Dunson (ed.), Springer Lecture Notes in Statistics, 192.

Received November 2012

Revised February 2013

Accepted March 2013