Effects of mixing modes on nonresponse and measurement error in an economic panel survey

Joseph W. Sakshaug[*], Jonas Beste and Mark Trappmann

Abstract

Numerous panel surveys around the world use multiple modes of data collection to recruit and interview respondents. Previous studies have shown that mixed-mode data collection can improve response rates, reduce nonresponse bias, and reduce survey costs. However, these advantages come at the expense of potential measurement differences between modes. A major challenge in survey research is disentangling measurement error biases from nonresponse biases in order to study how mixing modes affects the development of both error sources over time. In this article, we use linked administrative data to disentangle both nonresponse and measurement error biases in the long-running mixed-mode economic panel study "Labour Market and Social Security" (PASS). Through this study design we answer the question of whether mixing modes reduces nonresponse and measurement error biases compared to a single-mode design. In short, we find that mixing modes reduces nonresponse bias for most variables, particularly in later waves, with only small effects on measurement error bias. The total bias and mean-squared error are both reduced under the mixed-mode design compared to the counterfactual single-mode design, which is a reassuring finding for mixed-mode economic panel surveys.

Keywords: Longitudinal survey, Mixed-mode survey, PASS, Mode effects, Total survey error

JEL Classification: C83

1 Introduction

Panel surveys are indispensable tools for conducting labour market research and informing labour market policy. For example, the German Panel Study "Labour Market and Social Security" (PASS; Trappmann et al. 2019) has interviewed representative household samples of the general population and welfare recipients since 2006/07, generating insights into determinants of labour market participation (e.g. Abraham et al. 2019; Denzer et al. 2021; Lietzmann and Frodermann 2021) and the economic and social consequences of poverty and unemployment (e.g. Gundert and Hohendanner 2015; Krug et al. 2019; Pohlan 2019; Hetschko et al. 2020).

Like many panel surveys in Germany and elsewhere, the PASS survey data are collected using a mix of data collection modes. Since Wave 5, the initial default mode for new samples has been computer-assisted personal interviewing (CAPI) with nonresponse follow-ups carried out using computer-assisted telephone interviewing (CATI). In subsequent waves the successful mode of the previous wave then becomes the default mode and the other mode the follow-up mode. Numerous other panel studies implement sequential mixed-mode designs using a mix of CAPI and CATI, often alongside other modes (e.g. web), including the UK Next Steps cohort study (Calderwood et al. 2021), the US American Community Survey (US Census Bureau 2020), and The UK

*Correspondence: joe.sakshaug@iab.de

Institute for Employment Research, Regensburger Str. 104, 90478 Nuremberg, Germany

Household Longitudinal Study (University of Essex, Institute for Social and Economic Research 2021).

There are multiple reasons that drive panel surveys to mix modes (De Leeuw 2005). First, it can reduce systematic selection error, including noncoverage and non-response, as certain population units may be unable or unwilling to participate in a given mode, but may be able and/or more willing if an alternative mode is offered to them. For example, persons who are busy and often away from the household (e.g. due to employment obligations) may not be reachable in a traditional CAPI survey but may be more reachable if a CATI follow-up mode is offered. Conversely, persons who are often at home (e.g. due to unemployment) and weary of unsolicited telephone calls may be more likely to participate in a CAPI follow-up mode as opposed to an exclusively CATI survey. A second reason for mixing modes is that it may lower per-respondent costs if a cheaper mode is offered and a large fraction of sample units participate in that mode, as opposed to exclusively using a more expensive single mode (e.g. CAPI). This is the main motivation for offering less expensive survey modes (e.g. web, CATI) as alternatives to CAPI, as this approach has shown to yield significant cost savings when the modes are implemented sequentially with the less expensive mode offered first (Bianchi et al. 2017; Carpenter and Burton 2018).

However, mixing modes can also have undesirable effects on data quality. For example, it is well-established that modes have inherently different measurement error properties that, when mixed, can introduce measurement effects[1] (de Leeuw 1992; 2005). In other words, measurement effects can arise when the same respondent answers the same question differently depending on the mode of data collection. Measurement effects are undesirable because they can bias comparisons between respondents who answer in different modes and comparisons with single-mode surveys. Such effects are more likely to occur when mixing interviewer- and self-administered modes which differ strongly in their communication channel (aural vs. visual) and level of interviewer presence (Klausch et al. 2017). These mode differences have been linked to cognitive response processes that differentially affect multiple types of measurement error in surveys, including social desirability bias (the tendency for respondents to provide answers that conform to social or societal norms; Tourangeau and Yan 2007), primacy

effects (the tendency to select the first options presented in a visual list of response options), and recency effects (the tendency to select the last options presented in a spoken interview) (Schwarz et al. 1991; Krosnick and Alwin 1987). Although CAPI and CATI modes are both aural and interviewer-administered modes, CATI is thought to have greater "social distance" between respondent and interviewer than CAPI, which is generally viewed as one of the reasons for lower respondent engagement and larger social desirability effects in CATI surveys (Holbrook et al. 2003).[2]

Panel surveys that implement a sequential mixed-mode design at each wave of data collection, such as the PASS survey, are particularly susceptible to measurement effects, not only because different respondents may answer the same questions in different modes, but also because the same respondents may answer in different modes in different waves. While most respondents tend to answer in the same mode at each wave, there is a non-trivial group of respondents who switch between modes over multiple waves (Cernat and Sakshaug 2021a). The ways in which respondents can answer in different modes within- and between-waves has the potential to exacerbate measurement effects over time, causing respondents to answer the same questions differently at different time points depending on the mode they use. In panel surveys, where the primary goal is to collect repeated measures and observe important events over the life course (e.g. education, marriage, births, employment and unemployment, welfare benefit receipt), measurement effects have the potential to misrepresent the prevalence of these key events and distort time trends (Cernat and Sakshaug 2021b,c).

Another undesirable effect of mixing modes is differential selection, or composition effects of mode. Differential selection error may increase in sequential mixed-mode surveys if certain types of respondents are strongly over-represented in the follow-up mode. For example, CATI surveys have been shown to overrepresent households with higher socio-economic status compared to CAPI surveys (Lipps 2016; Holbrook et al. 2003). Although the intention of mixed-mode surveys is to bring in different types of respondents to achieve a more balanced respondent pool, a strong overrepresentation in the follow-up mode (e.g. CATI) may actually produce a greater imbalance relative to the starting mode (CAPI). The problem can be further worsened by selective attrition where respondents with higher socioeconomic status

---

[1] In much of the literature, these effects of survey mode on measurement are denoted as "mode effects". As mode can affect sample composition as well as measurement (as we argued above), we prefer a terminology that allows to clearly distinguish between the two and therefore refer to mode effects on measurement as measurement effects of mode and mode effects on sample composition as composition effects of mode.

[2] Another important difference between CATI and CAPI with respect to measurement is that CAPI interviews can make use of visual stimuli by presenting showcards to respondents or showing them the interviewer's screen.

who are overrepresented in one mode (CATI) are more likely to stay in the panel, while respondents with lower socioeconomic status and who are better represented in the alternative mode (CAPI) are more likely to drop-out of the panel. The implication here is that persons with higher socioeconomic status (SES) will be increasingly overrepresented, resulting in an overestimation of SES time trends. Coupled with measurement effects in the CATI mode, the overestimation may be amplified by the remaining respondents who are reluctant to report episodes of unemployment, benefit receipt, or other socially undesirable events.

Given that mixing modes in a panel study can have both positive and negative effects on data quality, a key question is whether the net effect of offering a follow-up mode is beneficial from a total error (or mean-squared error) perspective, relative to a single-mode design, and whether the net effect changes over multiple waves of the panel. A reduction in total error would indicate that a sequential mixed-mode design is advantageous for data quality, whereas an increase in MSE would indicate that the mixed-mode design is not advantageous over a single-mode design.

In this article, we address this question by utilizing gold standard measures from administrative data that are available for both respondents and nonrespondents of the PASS survey to assess the effect of implementing the secondary CATI mode on nonresponse bias, measurement error bias, and total bias, relative to the counterfactual single-mode CAPI design. The remainder of the article is as follows. In "Background" section we review the relevant literature, identify the specific research gaps that are addressed in the study, and outline our expectations of the results. "Data Sources" section describes the data sources used and "Methods" section presents the methodological framework for the study. The study results are presented in "Results" section and further discussed in "Discussion" section. The main study conclusions and their practical implications are summarized in "Conclusion" section.

2 Background

Previous research has shown that CATI and CAPI modes can differentially affect both selection error and measurement error, though most studies analyze only one source of error rather than both. For instance, some studies find that CAPI households tend to have a lower income, are less likely to own their home, are more socially disadvantaged and suffer from more deprivation, and better reflect population sociodemographic characteristics than CATI households (Lipps 2016; Holbrook et al. 2003; Fessler et al. 2018; Klausch et al. 2015). Following up CATI nonrespondents with CAPI tends to increase the

representativeness of sociodemographic characteristics (Klausch et al. 2015), though differences can be minimal (Lynn 2013). Klausch et al. (2015) also report a backfiring effect where following up CATI nonrespondents with CAPI yielded a larger selection error on non-sociodemographic variables concerning crime victimization. These results challenge the current practice of using sequential mixed-mode designs for the purpose of reducing selection error.

As additional contact attempts are much cheaper and thus more frequent, CATI follow-ups to CAPI are likely to bring in larger proportions of respondents who are often away from home (e.g. employed or younger persons) and thus underrepresented in CAPI. In the context of a panel survey, residentially mobile households should be easier to follow by CATI than by CAPI as mobile phone numbers are likely to remain stable after moving. Finally, interviews conducted in other languages than the majority language can in many studies only be offered in CATI mode because interviewers capable of conducting interviews in foreign languages will rarely be available in all geographic areas covered by the survey.

On the measurement error side, several studies have reported greater measurement error (including social desirability bias) in CATI surveys compared to CAPI surveys (St-Pierre and Beland 2004; Revilla 2010; De Leeuw and van der Zouwen 1988; Aquilino 1994; Holbrook et al. 2003; Hope et al. 2014), though some studies report no or few measurement differences between these modes (Scherpenzeel 2001; Klausch et al. 2017; Schouten et al. 2013). Studies tend to find only few measurement differences between single-mode CAPI and sequential mixed-mode designs involving CATI and CAPI, when either is used as the starting mode or follow-up mode (Revilla 2010; Cernat 2015; Klausch et al. 2017). Only Cernat (2015) has investigated measurement effects in a longitudinal setting, finding only few measurement differences between a sequential mixed-mode design (CATI-CAPI) and a single-mode CAPI design, either in the wave that the mixed-mode design was implemented or in subsequent waves. These results suggest that measurement effects may not be a concern for longitudinal mixed-mode studies.

One source of measurement error that is much more prevalent in CAPI than in CATI interviews is underreporting on filter questions (Kreuter et al. 2011; Eckman et al. 2014). While motivated underreporting as a means for respondents to shorten the survey is likely to occur in both modes, interviewer bias in answers to filter question in the direction of the shorter path through the questionnaire has repeatedly been shown to occur in CAPI rather than CATI modes (Matschinger et al. 2005; Kosyakova et al. 2015; Josten and Trappmann 2016).

2.1 Research gaps

What is missing from this literature are assessments of mixing CAPI and CATI modes on both selection and measurement error jointly in a panel setting. One of the major challenges of studying both error sources in mixed-mode settings is that they are completely confounded. Typically, mode-specific selection and measurement error are confounded because the same mode is used for both recruitment and data collection. Usually it is not possible to separate selection and measurement from each mode without making some fairly strong assumptions about the selection or measurement mechanisms. Multiple approaches have been put forth to deal with this problem, including "back door" methods that rely on mode-insensitive covariates to control for selection with any remaining differences attributed to measurement error (Vannieuwenhuyze 2014), "front door" methods that control for measurement through statistical modeling and ascribe residual differences to selection (Vannieuwenhuyze 2014), and "benchmark mode" methods that designate one mode as the gold standard that is used to assess selection or measurement error bias for the other modes (or mode designs) (Klausch et al. 2017). Each of these methods comes with assumptions regarding the extent to which the selection and measurement error mechanisms are explained by the covariate information or statistical modeling, as well as the validity of the "gold standard" mode which is also subject to selection and measurement error.

In the present study, we apply a different approach for separating selection and measurement effects that avoids many of the above assumptions. Specifically, we utilize linked administrative "gold standard" data available for both respondents and nonrespondents in Waves 5–10 of the PASS study's general population refreshment sample who were initially recruited in the 5th wave of the study (Sakshaug et al. 2017). While administrative data are rarely available for nonrespondents in practice, we capitalize on the fact that these data contain several key demographic and economic variables that overlap with those collected in the PASS survey, allowing us to estimate nonresponse bias, measurement error bias, and total bias simultaneously at each wave and for each phase of the sequential mixed-mode design.

The present study addresses the following research question: What are the effects of implementing a CATI follow-up mode in an otherwise single-mode CAPI panel survey on nonresponse bias, measurement error bias, and total bias?

2.2 Expectations

From the literature review we derive several expectations concerning nonresponse and measurement error bias by mode.

ME1. For questions that serve as filter questions the answer category that triggers (more) additional questions will be downward biased in both modes, but more so in CAPI than in CATI. Thus, a negative measurement bias is expected for CAPI that is attenuated when CATI is added. Filter questions in PASS that can be investigated relate to current and past welfare benefit receipt, employment subject to social insurance contributions, and foreign nationality. These traits should all show a downward bias in CAPI as they each trigger several follow-up questions.

ME2. Socially undesirable traits are likely to be downward biased in both modes, but more so in CATI than in CAPI. Thus, a negative bias is expected for CAPI that increases when the CATI mode is added. Socially undesirable items in PASS include past and present welfare benefit receipt as well as not being employed.

Combining the two arguments, we get a clear expectation for past and present welfare benefit receipt. There should be a downward measurement bias in CAPI. However, by adding CATI, this downward bias may either increase due to increased social desirability effects in CATI or decrease due to less underreporting in filter questions in CATI. For employment, the expectation if we combine both arguments, is that there might either be a downward measurement bias in CAPI due to underreporting in filter questions or an upward measurement bias due to social desirability. In both events, we expect the addition of CATI to shift the estimate upwards because social desirability should be more pronounced in CATI while underreporting of filter questions should be less pronounced. For foreign nationality we expect a downward measurement bias in CAPI due to underreporting in filter questions that decreases when CATI is added.

NR1. The literature review suggests that we should expect a downward nonresponse bias for socially disadvantaged groups in both modes and that this downward bias should be more pronounced in CAPI than in CATI. Thus, CATI follow-up should reduce nonresponse bias with respect to socially disadvantaged groups. Socially disadvantaged groups in PASS include welfare benefit recipients, those not employed, those with lower income and those with foreign nationality.

NR2. Residentially mobile groups and those who are often away from home are difficult to contact and track in CAPI surveys where contact attempts are costly and

Table 1 Summary of expectations for each variable

Variable	Measurement error bias		Nonresponse bias	
	CAPI	Mixed-mode	CAPI	Mixed-mode
Unemployment benefit (current)	–	unclear	–	↑
Unemployment benefit (past)	–	unclear	–	↑
Employed (current)	unclear	↑	+	↓
Employed (past)	unclear	↑	+	↓
Income (in Euro)	unclear	unclear	+	↓
Non-German citizenship	–	↑	–	↑
Age (in years)	unclear	unclear	+	↓
Sex (male)	unclear	unclear	unclear	unclear

addresses can become outdated. Although CAPI mode offers additional opportunities for interviewer tracking (Couper and Ofstedal 2009), we expect that following mobile groups is easier in CATI mode, where a high quality of telephone numbers, including mobile phone numbers is maintained. Thus, we expect a downward nonresponse bias for such groups in CAPI that is likely to get attenuated when CATI is added. People with foreign nationality and young people have been shown to be residentially more mobile (Clark et al. 2000).

NR3. In PASS, foreign language interviews are unavailable in CAPI, while two foreign languages are offered in CATI. For respondents with foreign nationality we expect a downward nonresponse bias due to the unavailability of foreign language interviews in both modes but more so in CAPI than in CATI. Thus, a negative nonresponse bias is expected for CAPI that is attenuated when CATI is added.

Combining the three arguments, we get a clear picture for welfare benefit recipients and those not employed. Based on the literature review, we expect negative nonresponse bias for the proportion of these socially disadvantaged groups in CAPI. The addition of CATI should reduce this bias. Likewise, for income, we expect negative nonresponse bias in CAPI that is reduced by CATI follow-up due to the same reasons. For foreign nationality we get a consistent picture from all three arguments: The proportion of foreign nationals is likely to be severely downward biased in CAPI due to foreigners being a socially disadvantaged group in Germany (Kogan 2011), due to foreigners being more residentially mobile (Clark et al. 2000), and due to difficulties in conducting CAPI interviews in German, which is the only available language in that mode. This downward bias should be reduced by the CATI follow-up that facilitates contacting mobile groups and offers interviews in multiple foreign languages.

Table 1 summarizes our expectations. A minus sign ("-") in the CAPI columns denotes a negative expected bias, and a plus sign ("+") a positive expected bias. In the mixed columns, an upward arrow ("↑") corresponds to an expectation that the initial CAPI bias (whatever its sign is) will be shifted in a positive direction, while a downward arrow ("↓") corresponds to an expectation that the initial CAPI bias (whatever its sign is) will be shifted in a negative direction.

3 Data sources

3.1 Survey data

This study uses survey data from the German Panel study "Labour Market and Social Security" (PASS; Trappmann et al. 2019). The PASS is an annual, longitudinal household survey of the German residential population oversampling households receiving welfare benefits (called unemployment benefit II, abbreviated UB II). It was initiated in 2006 by the Institute for Employment Research (IAB) of the Federal Employment Agency (BA) in response to the reorganization of the welfare and unemployment benefit system in Germany. It is a central dataset for research on the labour market, poverty, and means-tested income support in Germany (German Social Code Book II). Information on labour market outcomes, household income, and unemployment benefit receipt are collected from about 10,000 households annually. In addition to household interviews with the heads of the households, person interviews are conducted with all household members starting from the age of 15 years.

The first PASS wave was composed of two samples, a sample of the German residential general population and a sample of the welfare-benefit recipient population. The benefit recipient sample was drawn from national recipient registers at the Federal Employment Agency. The general population sample was drawn from address

lists held by a commercial provider. The welfare-benefit recipient sample is refreshed annually by a sample of new entries to welfare benefits. The general population sample was refreshed in Waves 5 and 11 to compensate for loss of statistical power due to panel attrition. These refreshments were drawn from official population registers. While addresses are available from all sampling frames for all target households, they may prove to be outdated or invalid in the course of the fieldwork. PASS uses a multitude of proactive and reactive techniques from official sources and commercial registers or interviewers asking neighbors to locate households. In contrast, telephone numbers are only available in the sampling frame for welfare benefit recipient samples and even there only in about 80 percent of the cases. For all other cases, they have to be searched in commercial and official registers with a success rate of about 50 percent. Once respondents have been recruited, all kinds of up-to-date contact information (address, landline number, mobile phone numbers) is collected at the end of the interview and respondents are reminded between waves to send updates in case of changes. A detailed documentation of the fieldwork can be found in the wave-specific field reports[3] (Jesske and Schulz 2012) for Wave 5.

Data collection is conducted using a sequential-mixed mode design of computer assisted-personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI). In Waves 1 to 4 CATI was the initial mode for all new samples, but since Wave 5 for new samples this has been changed to CAPI. In the course of the panel, the previous wave interview mode becomes the default mode for the subsequent wave. Mode switches are initiated if no contact information is available for the default mode, if the household cannot be contacted after a certain amount of contact attempts, or if respondents express their wish to switch modes. In the latter case, mode switches between different persons in the same household in the same wave are allowed.

Refusal conversion is generally done in CATI where it can be more efficiently organized that only interviewers who are very successful at recruiting make the calls. In addition, if an interview in German is not possible a switch to CATI is standard. The interviews in PASS are conducted in different languages: German, Russian, Arabic (since Wave 10), and Turkish (until Wave 9). Almost all foreign-language interviews are conducted in CATI mode by a native speaker interviewer. Further details about the PASS study design can be found in Trappmann et al. 2013; 2019).

For the present study, we only use the PASS Wave 5 refreshment sample for the general population (which

is denoted as sample 6 in the PASS dataset). The gross refreshment sample consists of 6,237 persons aged 18 or older who were drawn from population registers and issued for fieldwork. Population registers in Germany contain no household information. Therefore, initially a person sample had to be drawn, but the goal is to interview all household members aged 15 or older. All households were first assigned to CAPI mode.

3.2 Administrative data

Household- and person-level interviews were conducted with 1,510 persons from the refreshment sample in Wave 5. For all persons in the gross sample – respondents and nonrespondents – a probabilistic record linkage to administrative employment data of the IAB based on name, sex, address, and date of birth was attempted. A total of 3,668 persons (58.8 percent) could be linked with sufficient quality. Further details about the linkage can be found in Sakshaug et al. (2017).

The administrative data are referred to as the "Integrated Employment Biographies (IEB)" and consists of administrative employment data obtained from different administrative proceedings of the German Federal Employment Agency. The main sources are mandatory social security notifications of the employers regarding their employees and longitudinal information on registered unemployment, job search, participation in active labor market programs, or welfare-benefit receipt. Further details about the IEB can be found in Jacobebbinghaus and Seth (2007).

In addition to these administrative data, rich paradata are available for the PASS gross sample in each wave. These paradata include detailed information on timing and outcome of each contact attempt (including refusal and mode switches). While a linkage between survey data and administrative data is contingent on informed consent, we may link these paradata to survey data as well as to administrative data without consent. This linkage can be exploited to estimate total bias and separate it into nonresponse bias and measurement error bias as has first been shown by Kreuter et al. (2010).

4 Methods
4.1 Case selection

In Table 2 the case selection for the forthcoming analyses is shown. The PASS Wave 5 refreshment sample for the general population consists of 6237 persons in separate households. After excluding the "non-eligible" cases, 6120 cases remain. These cases serve as the base for the calculation of the response, non-contact and refusal rates. A total of 4799 cases could be linked to the IEB administrative data using an inclusive matching criterion. A more restrictive matching criterion yielded 3602

Table 2 Case numbers in Wave 5 refreshment sample

Sample	6237
Excl. "not Eligible"	6120
Linked cases (inclusive)	4799
Linked cases (restrictive)	3602
Household interviews	857
Person interviews	842

Data: Panel study "Labour Market and Social Security" (PASS), PASS_0618_v1, Sample 6, Waves 5–10 & PASS Paradata, Waves 5–10

matched cases. We use these 3602 cases as the baseline sample to estimate nonresponse and measurement biases from Wave 5 onward. For 857 cases a household interview was completed and for 842 cases an additional person interview was conducted.

4.2 Variables of interest
To estimate the biases (nonresponse bias, measurement error bias, and total bias) we treat the administrative data as the "gold standard". Therefore, we only use variables for which the quality of administrative data is high enough to be considered as a true score for the same variables collected in the PASS survey. This reduces the number of suitable variables in the administrative data. Eight variables on basic demographics and employment history were selected for the final analyses: sex, age, foreign nationality, monthly earned income, unemployment benefit receipt (at the time of the interview and within the 2 years preceding it) and employment subject to social security (at the time of the interview and within the 2 years preceding it).

While some of these variables (sex, age, foreign nationality) come from multiple administrative sources, others (income, unemployment benefit receipt, and employment subject to social security) are central to the calculation of monetary payouts. For these reasons, the selected variables can be presumed to have a high level of quality. Age is a metric variable that we derive in both sources from date of birth. Also, earned income from employment subject to social insurance contribution is a metric variable.[4] All other information is generated as binary variables.

In both, survey and administrative data, the information on duration of unemployment benefit receipt and employment is stored in spell format. However, the generation of the status at the time of the interview and during the previous 2 years is straightforward.

In PASS, the information on sex is collected in the household matrix (in which all persons in a household are listed by the respondent to the household questionnaire). The information from the initial survey is transferred and checked for accuracy in each of the following interviews. The age of the respondent is collected in each wave. In the case of re-interviews, a comparison with the pre-wave information takes place. If there are noticeable differences, a plausibility check is carried out during the interview. The questions on earned income and German citizenship are asked anew for each wave without using previous-wave information. Information on a person's employment is collected as part of the employment module which is collected in spell format. In the first interview, respondents are asked to report all employment that is subject to social security contributions in the previous 2 years. In addition to employment, other activities relevant to the labour market such as unemployment are collected in the same way. In the follow-up interviews, reference is made to the activities reported in the last interview (dependent interviewing) and the question is asked whether these activities are still ongoing or whether there have been any changes or additional activities. In a similar way, during the household interview, questions are asked about the durations during which unemployment benefit II was received. Here, too, the current information from the last interview is used in the follow-up interview. From this detailed employment, unemployment, and benefit receipt information originally collected in spell format, indicators for current and past receipt can easily be derived.

4.3 Counterfactual single-mode (CAPI) design
The analysis is designed to simulate the counterfactual design of PASS as a CAPI single-mode study and compare this to the actual mixed-mode CAPI/CATI design. For the single-mode scenario, we must make an assumption regarding what would have happened to cases that switched modes in the actual mixed-mode design in the counterfactual single-mode design. In our analysis, we assume that they would have dropped out of the study at the time of the first mode switch from CAPI to CATI. This assumption makes sense because at the time of a mode switch considerable energy has already been invested in the initial mode (at least six contact attempts in CAPI or twelve contact attempts in CATI, address and phone number searches in different registers including municipal resident registers, and Federal Employment

---

[4] This variable is truncated in both data sources at the so called "social security contribution assessment ceiling". In Germany, social security contributions only have to be paid up to a certain amount that differs between regions and years (roughly between 55,000 and 80,000 euros annual income in the years and regions in our dataset). As any income above this ceiling is not relevant for social security payments, employers only have to report incomes up to this ceiling. By truncating at the region- and year-specific ceiling in both datasets, we make income data comparable in both sources.

Agency registers) and it seems unlikely that more of the same would have been successful in many of these cases. Thus, in the single-mode scenario, we treat everyone who switched modes at least once from CAPI to CATI as a nonrespondent in all subsequent waves. No assumptions are made about the mixed-mode scenario since this was the actual mode design of the PASS.

4.4 Bias estimation

While theoretically total survey error can be split up into several error sources (e.g. coverage error, sampling error, nonresponse error, adjustment error, specification error, measurement error, editing error; Biemer 2010; Groves et al. 2011), we can neglect most of these error sources since we look at unadjusted data and take the sample as given and identical in both modes. Therefore, we can focus on measurement error bias and nonresponse bias as well as sampling variance. We then compute the total bias as the sum of the nonresponse bias and the measurement error bias. We also consider different types of nonresponse biases, including non-contact and refusal,[5] as well as item nonresponse bias.

To compute the different biases, we use the "true value" from the register data, the self-reported value from survey data, and the information on the response status from paradata. We introduce the following notation and formulas to compute the biases.

- $\bar{y}_{s,k,w}$ denotes the mean reported value from the survey (s) for sample subgroup k in Wave w
- $\bar{y}_{t,k,w}$ denotes the mean true value from the register data for sample subgroup k in Wave w
- $\bar{y}_{t,w}$ denotes the mean true value for the complete (or gross) sample from the register data in Wave w

In the mixed-mode design we denote the subgroup of respondents in the default mode CAPI as *r1* and the respondents in mixed-mode design as *r2*. Based on this notation, estimates of the biases for each variable are computed as follows:

- Nonresponse bias, CAPI: $\bar{y}_{t,r1,w} - \bar{y}_{t,w}$
- Nonresponse bias, Mixed Mode: $\bar{y}_{t,r2,w} - \bar{y}_{t,w}$
- Measurement error bias, CAPI: $\bar{y}_{s,r1,w} - \bar{y}_{t,r1,w}$
- Measurement error bias, Mixed Mode: $\bar{y}_{s,r2,w} - \bar{y}_{t,r2,w}$
- Total bias, CAPI: $(\bar{y}_{t,r1,w} - \bar{y}_{t,w}) + (\bar{y}_{s,r1,w} - \bar{y}_{t,r1,w}) = \bar{y}_{s,r1,w} - \bar{y}_{t,w}$
- Total bias, Mixed Mode: $(\bar{y}_{t,r2,w} - \bar{y}_{t,w}) + (\bar{y}_{s,r2,w} - \bar{y}_{t,r2,w}) = \bar{y}_{s,r2,w} - \bar{y}_{t,w}$

---

[5] Nonrespondents that were not contacted again (i.e. drop outs), keep their most recent exit status, either non-contact or refusal, in all subsequent waves. Respondents who dropped out of the study prior to the subsequent wave are set to refusals.

To compare the biases for the different binary and metric variables we compute the relative biases by dividing the different biases by the mean true value of the base sample. Additionally, we compute the Mean Squared Error (MSE) as the sum of the sampling variance and squared bias for each variable of the two mode designs. We adjusted for the stratified two-stage sampling design to estimate the variance properly. 95% confidence intervals are provided for all bias estimates using bootstrap variance estimation with 500 bootstrap samples.

5 Results

The wave-specific and overall response rates (AAPOR RR1) as well as the number of cases with realized household and person interviews for the CAPI and mixed-mode scenarios are shown in Table 3. In the initial Wave 5 refreshment sample the response rate in the CAPI mode was 18.7 percent compared to 24.7 percent in the mixed-mode scenario. The higher rate also results from the fact that the CAPI cases are transferred to the CATI field if no face-to-face contact is possible and a valid phone number is available. The higher response rate in the mixed-mode scenario (for both the wave-specific and cumulative rates) can be seen in all of the following waves.

In Table 4, the rate of non-eligible cases in the sample is reported and the non-contact and refusal rates are shown. In the CAPI mode the non-contact rates are clearly higher. This again is caused by the process that cases which could not be contacted in CAPI are transferred to the CATI mode. The differences between the different mode-design scenarios slightly increases over time. Looking at the refusal rates, they are higher in the mixed-mode scenario and tend to diverge slightly further over the waves.

In Table 5, the distributions of the variables of interest are shown for the initial Wave 5 gross sample. A slight majority of the sample is male (53 percent) with an average age of 43 years. About 10 percent of the sample consists of non-German citizens. The average income (including the zero values) is around 1,171 euros per month. Around 11 percent are receiving unemployment benefit II (UB II) at the time of the interview, and about 18 percent have received UB II at least once in the last two full calendar years. About 50 percent are currently in dependent employment subject to social security contributions and 61 percent have been in dependent employment within the last two calendar years.

5.1 Nonresponse bias: CAPI vs. Mixed mode

Figure 1 shows the main results for nonresponse bias including 95% confidence intervals (a tabular version is available in Additional file 1: Table S1 of the online

Table 3 Response Rate (RR) and Number of Completed Personal Interviews, Waves 5–10

	# of cases in the PASS gross sample	CAPI			Mixed mode		
		RR wave specific (in %)	RR overall (in %)	# of Interviews	RR wave specific (in %)	RR overall (in %)	# of Interviews
Wave 5	6120	18.7	18.7	1146	24.7	24.7	1510
Wave 6	1293	66.2	14.0	856	82.0	17.3	1060
Wave 7	1204	60.3	11.9	726	76.8	15.1	925
Wave 8	1007	61.4	10.1	618	79.3	13.0	798
Wave 9	858	62.7	8.8	538	83.2	11.7	714
Wave 10	663	63.2	6.9	419	86.3	9.3	572

Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10

Table 4 Outcome rates and numbers of non-contacts and refusals, Wave 5–10

	CAPI				Mixed			
	Non-contact		Refusal		Non-contact		Refusal	
	Rate, %	# of cases	Rate, %	# of cases	Rate, %	# of cases	Rate, %	# of cases
Wave 5	27.2	1665	54.1	3309	16.6	1013	58.8	3597
Wave 6	28.2	1724	57.8	3536	17.1	1048	65.5	4008
Wave 7	29.0	1777	58.9	3604	17.2	1050	67.5	4133
Wave 8	29.4	1799	60.2	3683	17.1	1047	69.5	4256
Wave 9	29.9	1831	60.8	3722	17.2	1053	70.7	4325
Wave 10	30.4	1858	62.3	3815	17.3	1060	72.9	4461

Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10

Table 5 Distribution of variables in gross sample—Wave 5

Variable	Level	Frequency	Scale level	Mean/%
Male (in %)	Person	First interview with household matrix, checked every reinterview	Binary	53.4
Age (in years)	Person	Every interview	Metric	43.1
Non-German citizen (in %)	Person	Every interview	Binary	9.6
Income (in Euro)	Person	Every interview	Metric	1171
Unemployment benefit receipt (actual in %)	Household	Every interview (dependent interviewing)	Binary	11.2
Unemployment benefit receipt (last 2 years in %)	Household	Every interview (dependent interviewing)	Binary	18.1
Employment subject to social security benefit receipt (actual in %)	Person	Every interview (dependent interviewing)	Binary	49.8
Employment subject to social security benefit receipt (last 2 years in %)	Person	Every interview (dependent interviewing)	Binary	60.7

Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10 & IAB Integrierte Erwerbsbiografien (IEB) V13.00.00

supplemental materials; absolute and absolute relative nonresponse biases are shown in Additional file 1: Figures S2 and S5, respectively). A separation into its components: non-contact bias, refusal bias, and item nonresponse bias can be found in Additional file 1: Figure S1 (or Table S2) of the online supplemental materials. Substantial, statistically significant and increasing nonresponse bias can be found with respect to three of the variables that were investigated. As we expected (Table 1), there is a substantial downward bias for non-German citizenship. Relative bias starts at about −30 percent for CAPI in Wave 5 and then grows to about −60 to −70 percent in Waves 7 to 10. The mixed-mode design exhibits an even larger initial negative bias in Wave 5 of
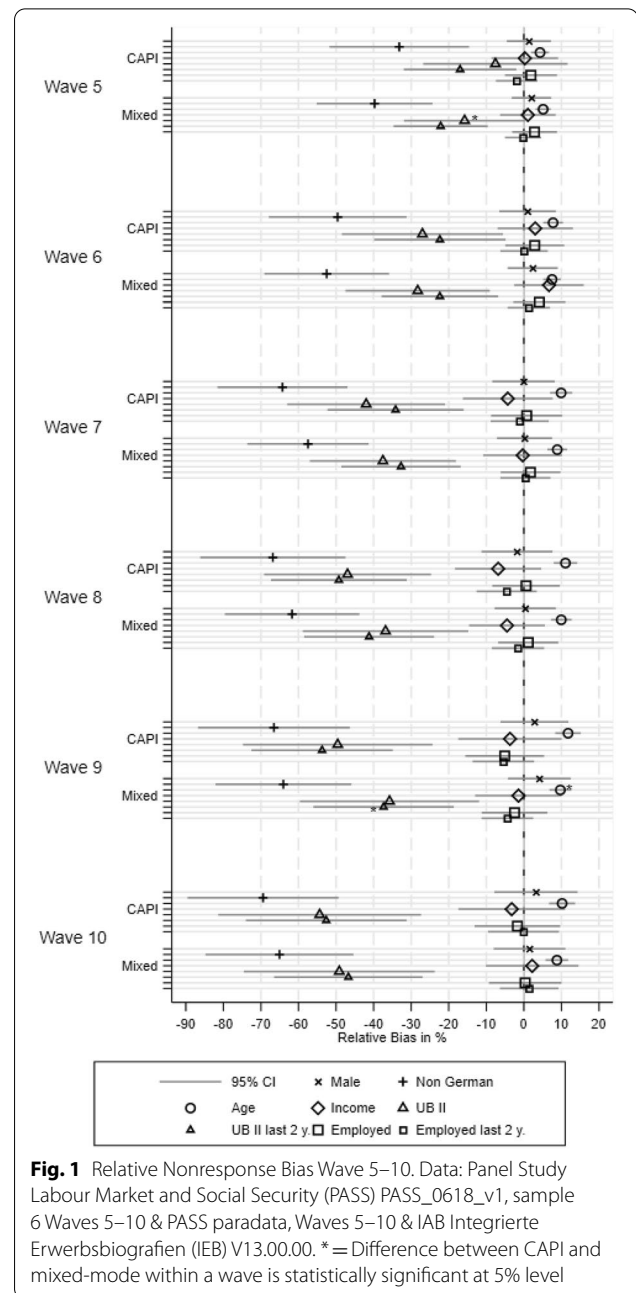
about −40 percent, but from Wave 7 onward, relative bias in the mixed-mode design is slightly smaller than for CAPI at about −60 percent. Differences between mode designs are however not significant in any of the waves.

Current and past welfare benefit receipt show a similar pattern (again as expected; Table 1). Current welfare benefit recipients are underrepresented by about 10 percent (not yet statistically significant) in Wave 5 when only CAPI is used and grows to about -50 percent in Wave 10. Again, the mixed-mode design shows larger initial bias of about -15 percent, but again, from Wave 7 onward nonresponse bias is larger for CAPI than for mixed-mode. Like for foreign nationality, differences between mode designs are not statistically significant in any wave.

Welfare benefit receipt in the past 2 years is downward biased by about 20 percent in Waves 5 and 6 in CAPI as well as in mixed-mode and increases to about -50 percent by Wave 10. Like for the two previous indicators, nonresponse bias is larger for mixed-mode than for CAPI in Waves 5 and 6, while mixed-mode performs slightly better from Wave 7 onward. The only significant difference between mode designs is in Wave 5.

Apart from these three variables, relative nonresponse bias is strongest and consistently statistically significant for age (also as expected; Table 1). While it starts out lower, it reaches about 10 percent relative bias at the time of Wave 7 and then remains stable. Again, we find the pattern that while bias is initially somewhat larger for mixed-mode, it is larger for CAPI in later waves (starting from Wave 6). In Wave 9, the difference between mode designs is statistically significant. For the remaining variables, past and present employment, sex, and earned income, relative nonresponse bias remains statistically insignificant and on a level well below 10 percent relative bias across all observed waves.

In summary, these findings suggest that socially disadvantaged groups, namely UB II recipients and foreign nationals, are more difficult to recruit than other groups in each single wave. Consequently, a substantial initial nonresponse bias increases wave by wave and reaches severe levels of relative bias of up to −50 to −70 percent by the time of Wave 10. Contrary to the intention and to the expectation that the mixed-mode design would attenuate any biases from CAPI only (Table 1), the mixed-mode design increases nonresponse bias in the early Waves 5 and 6. In contrast, the CATI follow-up mode does reduce the initial nonresponse bias from the CAPI starting mode in later waves. This pattern proves to be stable across different variables, although differences between mode designs are only in two cases statistically significant (larger bias for age in CAPI compared

Fig. 1 Relative Nonresponse Bias Wave 5–10. Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10 & IAB Integrierte Erwerbsbiografien (IEB) V13.00.00. * = Difference between CAPI and mixed-mode within a wave is statistically significant at 5% level

to mixed-mode and larger bias for past UB II receipt in mixed-mode compared to CAPI in Wave 5).

As mentioned, Additional file 1: Figure S1 in the online supplemental materials breaks down nonresponse bias into its three components: non-contact bias, refusal bias, and item-nonresponse bias. The most striking finding here is that nonresponse bias is clearly dominated by refusal bias. An interesting combination can be found for foreign nationals. Here, positive non-contact bias

(foreign nationals are more likely to be contacted) offsets an even larger negative refusal bias.

5.2 Measurement error bias: CAPI vs. Mixed mode

In Fig. 2, measurement error biases and corresponding confidence intervals for each variable, mode design scenario, and wave are shown (a tabular version is available in Additional file 1: Table S1 of the online supplemental materials; absolute and absolute relative measurement error biases are shown in Additional file 1: Figures S3 and S6, respectively). Here we do not see big differences between both mode scenarios. The amount of measurement error bias is generally modest (in particular when compared to nonresponse bias), and only in rare occasions becoming statistically significant or exceeding 10 percent relative bias.

In Wave 5 and all following waves, contrary to our expectations (Table 1), bias for non-German citizenship has a positive sign. However, it only becomes significant in Waves 6 and 7 and there are no differences between mode designs. Age is also unaffected by mode design. Measurement error bias for income tends to have a negative sign, but is not statistically different from zero in most waves and there is no effect of mode design. While bias for current employment is statistically insignificant in all waves and both mode designs, for employment in the last 2 years biases are significantly negative in most of the waves, but neither tend to be affected by the mode design. For males, only slight negative biases show up. For these four variables (current and past employment, income, age, sex), we had no clear expectations about the sign of measurement error bias (Table 1).

Apart from foreign nationality, we find the largest measurement error bias for current and past welfare benefit receipt in Wave 8, which is statistically significant only for the latter. Contrary to our expectations (Table 1), respondents seem to overreport benefit receipt which is usually viewed as a socially undesirable trait. However, there is no clear trend and in other waves measurement error bias for benefit receipt is close to zero or even negative. Note that confidence intervals for this estimate are extremely large due to the low proportion of benefit recipients in the samples and only the Wave 8 estimate for past welfare receipt is statistically significant.[6] In nearly all waves, introducing the CATI follow-up mode attenuates this upward bias or even reverses it, though again, differences between mode designs are not statistically significant. Only in Wave 10 do we observe an increase in the positive bias of current receipt after the CATI mode is introduced.

In summary, levels of measurement error bias are rather small in comparison to nonresponse bias and not much affected by CATI follow-up, which is good news.
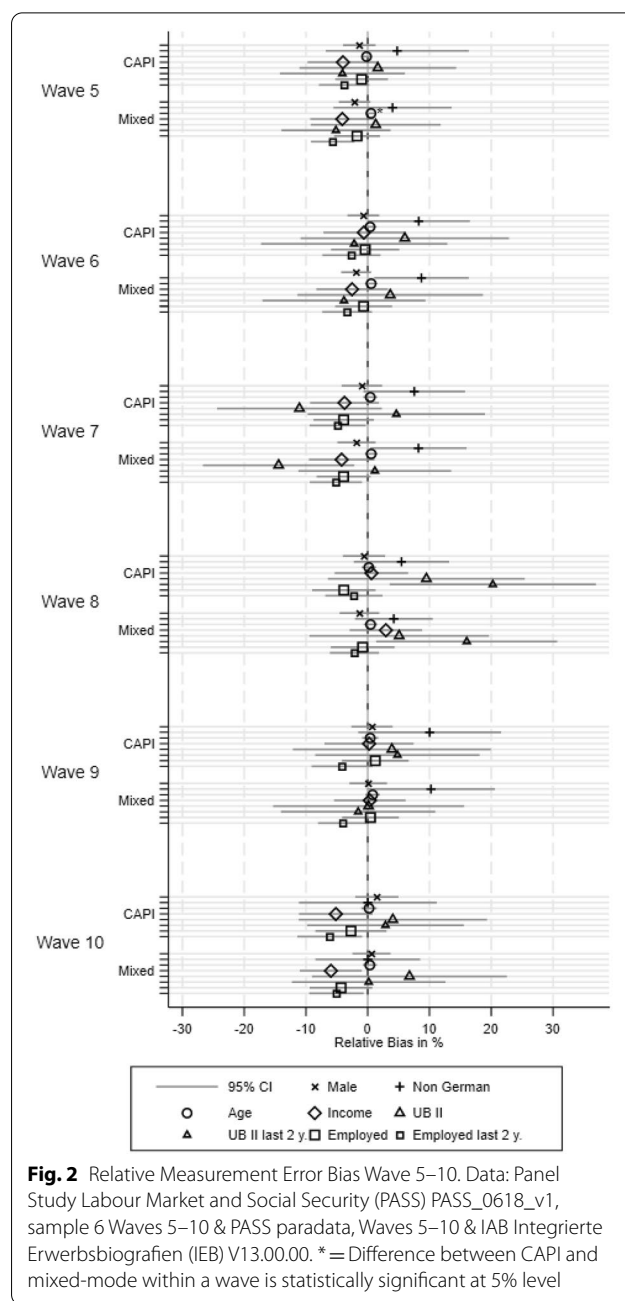
Fig. 2 Relative Measurement Error Bias Wave 5–10. Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10 & IAB Integrierte Erwerbsbiografien (IEB) V13.00.00. * = Difference between CAPI and mixed-mode within a wave is statistically significant at 5% level

Other than for nonresponse bias that cumulates across waves, there is also no clear time-trend for measurement error bias.

---

[6] Since relative biases are reported, with low values in the gross sample (true score) even small absolute biases can lead to large relative biases. For example, with welfare benefit receipt in Wave 7 in the mixed mode scenario, an absolute bias of -1.5 percentage points with a true score of 10.9 percent leads to a relative bias of -13.8 percent.

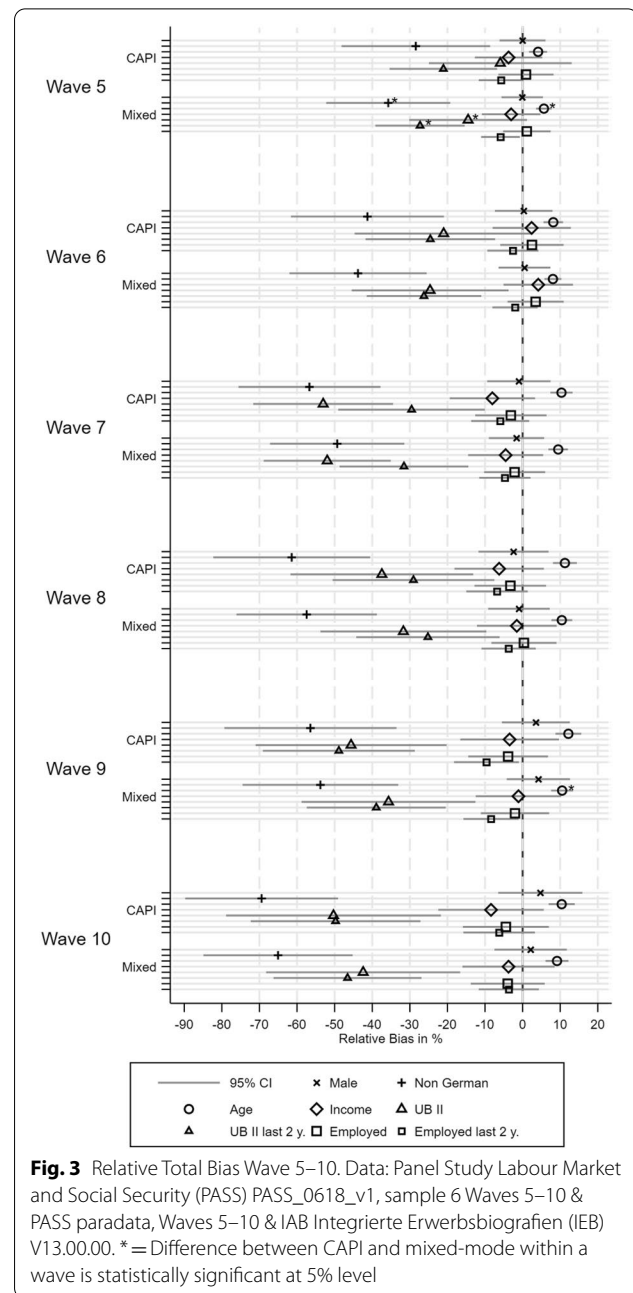### 5.3 Total bias/MSE: CAPI vs. Mixed mode

The total bias, shown in Fig. 3 (see Additional file 1: Table S1 of the online supplemental materials for a tabular version; absolute and absolute relative total biases are shown in Additional file 1: Figures S4 and S7, respectively), is the sum of the nonresponse bias and the measurement error bias. If we compare Fig. 3 to the previous figures, we note that total bias is strongly dominated by nonresponse bias and patterns are similar to those in Fig. 1. In Wave 5 and to a lesser extent in Wave 6 the total biases are smaller for most of the variables in the CAPI mode, while from Wave 7 onward there is a consistent pattern that total bias is smaller for mixed-mode than for CAPI. Differences between mode designs are again not statistically significant in most waves. Only some of the differences in favor of the single-mode CAPI design in Wave 5 (non-German citizenship, past and present UB II receipt and age) and the difference in favor of mixed-mode in Wave 9 are statistically significant at the 5% level. Like nonresponse bias, total bias increases across waves and reaches its maximum in Wave 10. The largest total bias is found for non-German citizenship and for past and current unemployment benefit receipt.

To account for the sampling variance of the variables in both mode scenarios we additionally estimate the mean squared error for each variable and mode scenario for Waves 5 to 10. In Table 6, we report the root-mean-square error (RMSE). The patterns are consistent with the total bias patterns as depicted in Fig. 3. The larger biases of the single mode CAPI design in later waves are even widened when sampling variance is added as numbers of cases would be smaller in a single-mode design due to increased panel attrition.

Here we see that in the start of Wave 5 the RMSEs are larger for male, income, and employment (actual and last 2 years) in the CAPI mode. Since Wave 7, the RMSE is higher or comparable for all variables in the CAPI mode. This is even more evident for all variables in Waves 8 to 10.

6 Discussion

In this article we investigated the total survey error of a CATI-CAPI-mixed-mode design of a longitudinal labour market survey. By focusing on a set of variables for which gold standard measurements were available in administrative data for respondents as well as nonrespondents, we were able to assess total survey error and to split it up into its most important components, nonresponse and measurement error bias. By defining a counterfactual situation in which no CATI-follow-up would have been implemented, we were able to investigate how the mixed-mode design affected biases as compared to a single-mode CAPI design.

Fig. 3 Relative Total Bias Wave 5–10. Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10 & IAB Integrierte Erwerbsbiografien (IEB) V13.00.00. *=Difference between CAPI and mixed-mode within a wave is statistically significant at 5% level

For the eight variables we were able to investigate: past and present welfare benefit receipt, past and present employment, earned income, foreign nationality, age, and gender, we found nonresponse bias to be considerably larger than measurement error bias. More importantly, as disadvantaged groups like foreign nationals and welfare benefit recipients remained hard to recruit over the whole course of the panel, negative nonresponse bias increased from wave to wave and reached substantial levels of -50 percent to -70 percent by the time of Wave 10.

Table 6 Root-Mean-Square Error for CAPI and Mixed mode, Wave 5–10

	Wave 5		Wave 6		Wave 7		Wave 8		Wave 9		Wave 10	
	CAPI	Mixed	CAPI	Mixed	CAPI	Mixed	CAPI	Mixed	CAPI	Mixed	CAPI	Mixed
Sex (male, %)	1.3	1.2	1.6	1.5	2.1	1.9	2.4	1.9	2.8	2.8	3.3	2.2
Non-German (%)	3.0	3.6	4.0	4.3	5.5	4.8	5.9	5.6	5.5	5.2	6.6	6.2
Age (in years)	1.8	2.4	3.6	3.6	4.7	4.3	5.2	4.8	5.7	5.0	5.0	4.4
Income (in Euro)	57.8	49.6	54.1	64.8	116.8	76.4	99.6	63.0	96.7	74.4	146.5	99.8
Unemployment benefit receipt (actual, %)	1.5	1.9	2.5	2.7	5.9	5.7	3.8	3.3	4.2	3.4	4.3	3.6
Unemployment benefit receipt (last 2 years, %)	4.1	5.1	4.3	4.5	4.9	5.2	4.5	3.9	6.8	5.5	6.4	5.9
Employment subject to social security (actual, %)	1.8	1.4	2.1	2.4	2.7	2.2	2.7	2.1	3.1	2.4	3.7	3.4
Employment subject to social security (last 2 years, %)	3.9	3.8	2.5	2.2	4.2	3.4	4.5	3.0	6.5	5.6	4.7	3.3
Number of variables with MSE (CAPI) > MSE (Mixed)	4		2		7		8		7		8	

Data: Panel Study Labour Market and Social Security (PASS) PASS_0618_v1, sample 6 Waves 5–10 & PASS paradata, Waves 5–10 & IAB Integrierte Erwerbsbiografien (IEB) V13.00.00

With respect to nonresponse bias, we found benefits of a mixed-mode design in later waves. In particular, the mixed-mode design was able to reduce negative nonresponse bias for socially disadvantaged groups like benefit recipients and foreign nationals. In addition, we found evidence that younger respondents were easier to recruit after CATI was added and thus, the positive age-bias could be reduced. Measurement error bias was fairly stable over time and not strongly affected by mode design for most variables. We found a puzzling positive measurement error bias for benefit receipt in CAPI in spite of it being a socially undesirable trait. This effect was reduced in the mixed-mode design. Total bias was dominated by nonresponse bias and consequently most results on nonresponse bias carried over to total bias. Introducing the CATI mode as follow-up (contrary to the intention) increased total bias for most variables in the initial waves of the panel. However, in the later waves (from Wave 7 onward) we did see a noticeable reduction in the total bias by introducing the CATI mode, particularly for socially disadvantaged groups (UB II recipients and foreign nationals). While we observe this same pattern for a wide range of different variables, in most cases differences between CAPI and mixed-mode estimates do not differ significantly at the 5% level. The same pattern was observed when looking at the mean-squared errors, which were lower in mixed-mode for the majority of the variables in later waves.

It should be noted that all biases were assessed before weighting adjustments. In a panel context, weighting can mostly eliminate the nonresponse bias we observed, because the lower response propensity of the disadvantaged groups is observed and can be incorporated into response propensity models like those used in the PASS panel. Nevertheless, the perpetual increase in nonresponse bias will inevitably lead to ever increasing weights for respondents from groups with low response propensities and thereby increase the variance of weights and make estimations more inefficient. Thus, it is in the interest of a panel study to decrease bias before weighting adjustments.

We acknowledge some study limitations. First, the study design was based on simulating the counterfactual single-mode (CAPI) design. It is possible that the effects we found would differ if an actual single-mode design had been implemented in an experimental setting. In addition, evaluating measurement error was only possible for a subset of PASS survey variables for which corresponding administrative data were available, though these variables are very important for labour market research. While the variables included in this study include socially undesirable items, such as welfare benefit receipt, for which measurement effects are likely, other types of variables for which mode effects might be even stronger (e.g. attitudes, rating scales) could not be included due to unavailability of gold-standard measures.

Future research could take-off from here by considering additional modes, as CATI and CAPI are similar in that they are both interviewer administered. Given insights from the literature, measurement effects should be more pronounced when interviewer administered modes are mixed with self-administered modes. Given the increase in push-to-web designs that follow-up web surveys with telephone or face-to-face interviews, it would be important to check whether our results carry over to such designs. It also seems important to include costs in future research. This is not a simple task. While per unit costs of telephone interviews are usually lower than those of personal interviews, there are additional fixed-costs involved in setting up the infrastructure and logistics of a mixed-mode design. Finally, in this first evaluation of total survey error properties of a longitudinal mixed-mode versus single-mode design, we focused on point estimates at different points in time. However, the purpose of panel surveys is to measure change and exploit individual change in statistical analyses. Thus, an important extension of our research would be to investigate how biases in estimates of change are affected by the mixed-mode design.

7 Conclusion

This study found that sequentially mixing CAPI and CATI modes in later waves of an economic panel survey can be beneficial for reducing nonresponse bias (and total bias and mean-squared error) for socially disadvantaged groups and younger people, compared to a single-mode CAPI design, without inducing strong measurement error effects. These findings have important practical implications for longitudinal labour market surveys. First, there seems to be benefits of adopting a mixed-mode design for reducing nonresponse bias in later waves. Nonresponse bias in later waves is a major concern in longitudinal surveys as nonresponse accumulates over time. The fact that nonresponse bias in these later waves can be reduced, especially for socially disadvantaged groups, by using a mixed-mode design is reassuring. Second, it is also reassuring that measurement error bias was largely unaffected by mixing modes, though we caution that this might have been expected given that both modes are interviewer administered. Mixing interviewer- and self-administered modes is likely to lead to larger measurement error effects than those observed here (Klausch et al. 2017). Lastly, when taken together, our results suggest that nonresponse is

a bigger contributor to the total survey error than measurement error, thus addressing nonresponse seems to be the higher priority and may merit greater allocation of resources to minimize it in economic panel surveys.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12651-022-00328-1.

Additional file 1: Table S1. Relative Nonresponse Bias, Measurement Error Bias, and Total Bias, Wave 5–10 in percent. **Table S2.** Relative Nonresponse Bias Components, W5–10 in percent. **Figure S1.** Relative Nonresponse Bias Components Wave 5 -10. **Figure S2.** Absolute Nonresponse Bias Wave 5-10. **Figure S3.** Absolute Measurement Error Bias Wave 5-10. **Figure S4.** Absolute Total Bias Wave 5-10. **Figure S5.** Absolute Relative Nonresponse Bias Wave 5-10. **Figure S6.** Absolute Relative Measurement Error Bias Wave 5-10. **Figure S7.** Absolute Relative Total Bias Wave 5-10.

Declarations

Ethics approval and consent to participate
Ethics approval for the PASS study was obtained from the IAB.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

Abraham, M., Bähr, S., Trappmann, M.: Gender differences in willingness to move for interregional job offers. Demogr. Res. **40**, 1537–1602 (2019)

Aquilino, W.S.: Interview mode effects in surveys of drug and alcohol use: a field experiment. Public Opin. Quart. **58**(2), 210–240 (1994)

Bianchi, A., Biffignandi, S., Lynn, P.: Web-face-to-face mixed-mode design in a longitudinal survey: effects on participation rates, sample composition and costs. J off. Stat. **33**(2), 385–408 (2017)

Biemer, P.B.: Total survey error: design, implementation, and evaluation. Public Opin. Quart. **74**(5), 817–848 (2010)

Calderwood, L., Peycheva, D., Henderson, M., Silverwood, R., Mostafa, T., Rihal, S.: Next steps: sweep 8 – age 25 user guide, 3rd edn. UCL Centre for Longitudinal Studies, London (2021)

Carpenter, H., Burton, J.: Adaptive push-to-web: experiments in a household panel study. Understanding society Working paper 2018-05. University of Essex, Colchester (2018)

Cernat, A.: The impact of mixing modes on reliability in longitudinal studies. Sociol. Method Res. **44**(3), 427–457 (2015)

Cernat, A., Sakshaug, J.W.: Understanding the patterns of mode switching in longitudinal studies. Surv. Res. Methods **15**(3), 281–298 (2021a)

Cernat, A., Sakshaug, J.W.: Estimating the measurement effects of mixed modes in longitudinal studies: currant practice and issues. In: Lynn, P. (ed.) Advances in longitudinal survey methodology, pp. 227–249. Wiley, Chichester (2021b)

Cernat, A., Sakshaug, J.W.: Measurement error in longitudinal data. Oxford University Press, Oxford (2021c)

Clark, W.A.V., Drever, A.I.: Residential mobility in a constrained housing market: implications for ethnic populations in Germany. Environ. Plann. A **32**(5), 833–846 (2000)

Couper, M.P., Ofstedal, M.B.: Keeping in contact with mobile sample members. In: Lynn, P. (ed.) Methodology of longitudinal surveys, pp. 183–203. Wiley, West Sussex (2009)

de Leeuw, E.D.: Data quality in mail, telephone, and face-to-face surveys. TT-Publicaties, Amsterdam (1992)

de Leeuw, E.: To mix or not to mix data collection modes in surveys. J. off. Stat. **21**, 233–255 (2005)

DeLeeuw, E.D., van der Zouwen, J.: Data quality in face to face interviews: a comparative mea-analysis. In: Groves, R.M., Biemer, P.P., Lyberg, L., Massey, J.T., Nicholls, W.L., II., Waksberg, J. (eds.) Telephone survey methodology, pp. 257–282. Russell Sage Foundation, New York (1988)

Denzer, M., Schank, T., Upward, R.: Does the internet increase the job finding rate? Evidence from a period of expansion in internet use. Inf. Econ. Policy (2021). https://doi.org/10.1016/j.infoecopol.2020.100900

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., Presser, S.: Assessing the mechanisms of misreporting to filter questions in surveys. Public Opin. Quart. **78**(3), 721–733 (2014)

Fessler, P., Kasy, M., Lindner, P.: Survey mode effects on measured income inequality. J Econ Inequal **16**, 487–505 (2018)

Groves, R.M., Fowler, F.J., Jr., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R.: Survey methodology. Wiley, Hoboken (2011)

Gundert, S., Hohendanner, C.: Active labour market policies and social integration in Germany: Do "One-Euro-Jobs" improve individuals' sense of social integration? Eur. Sociol. Rev. **31**(6), 780–797 (2015)

Hetschko, C., Schöb, R., Wolf, T.: Income support, employment transitions and well-being. Labour Econ **66**, 101887 (2020)

Holbrook, A.L., Green, M.C., Krosnick, J.A.: Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. Public Opin Quart **67**(1), 79–125 (2003)

Hope, S, Campanelli, P, Nicolaas, G, Lynn, P, Jäckle, A. The Role of the Interviewer in Producing Mode Effects: Results from a Mixed Modes Experiment Comparing Face-to-Face, Telephone and Web Administration. ISER Working Paper Series, No. 2014-20, University of Essex, Institute for Social and Economic Research (ISER), Colchester (2014)

Jacobebbinghaus, P., Seth, S.: The German integrated employment biographies sample IEBS. Schmollers Jahr. **127**(2), 335–342 (2007)

Jesske, B, Schulz, S: Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS * 5. Erhebungswelle - 2011. (FDZ-Methodenreport, 11/2012 (de)), Nürnberg (2012).

Josten, M., Trappmann, M.: Interviewer effects on a network size filter question. J off. Stat. **32**(2), 349–373 (2016)

Klausch, T., Hox, J., Schouten, B.: Selection error in single- and mixed-mode surveys of the Dutch general population. J. R. Stat. Soc. A **178**, 945–961 (2015)

Klausch, T., Schouten, B., Hox, J.J.: Evaluating bias of sequential mixed-mode designs against benchmark surveys. Sociol. Method Res. **46**(3), 456–489 (2017)

Kogan, I.: New immigrants-old disadvantage patterns? Labour market integration of recent immigrants into Germany. Int. Migr. **49**(1), 91–117 (2011)

Kosyakova, Y., Skopek, J., Eckman, S.: Do interviewers juggle filter questions? Evidence from a multilevel approach. Int. J. Public. Opin. R **27**, 417–431 (2015)

Kreuter, F., Müller, G., Trappmann, M.: Nonresponse and measurement error in employment research: making use of administrative data. Public Opin. Quart. **74**(5), 880–906 (2010)

Kreuter, F., McCulloch, S., Presser, S., Tourangeau, R.: The effects of asking filter questions in interleafed versus grouped format. Sociol. Method Res. **40**, 88–104 (2011)

Krosnick, J.A., Alwin, D.F.: An evaluation of a cognitive theory of response order effects in survey measurement. Public Opin. Quart. **51**, 201–219 (1987)

Krug, G., Prechsl, S.: The role of social integration in the adverse effect of unemployment on mental health – testing the causal pathway and buffering hypotheses using panel data. Soc. Sci. Res. **68**, 102379 (2019)

Lietzmann, T, Frodermann, C: Gender Role Attitudes and Labour Market Behaviours: Do Attitudes Contribute to Gender Differences in Employment in Germany? Work Employ Soc. 1–21 (2021).

Lipps, O.: Non-observation bias in an address-register-based CATI/CAPI mixed mode survey. Methods, Data, Analyses **10**(1), 5–24 (2016)

Lynn, P.: Alternative sequential mixed-mode designs: effects on attrition rates, attrition bias, and costs. J. Surv. Stat. Methodol. **1**(2), 183–205 (2013)

Matschinger, H., Bernert, S., Angermeyer, M.C.: An analysis of interviewer effects on screening questions in a computer assisted personal mental health interview. J. Off. Stat. **21**, 657–674 (2005)

Pohlan, L.: Unemployment and social exclusion. J. Econ. Behav. Organ. **164**, 273–299 (2019)

Revilla, M.: Quality in unimode and mixed-mode designs: a multitrait-multimethod approach. Surv Res Methods **4**(3), 151–164 (2010)

Sakshaug, J.W., Antoni, M., Sauckel, R.: The quality and selectivity of linking federal administrative records to respondents and nonrespondents in a general population survey in Germany. Surv. Res. Methods **11**(1), 63–80 (2017)

Scherpenzeel, A.: Mode effects in panel surveys: a Comparison of CAPI and CATI. Bases statistiques et vues d'ensemble. Bundesamt, für Statistik, Office fédéral de la statistique, Neuchâtel (2001)

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., Klausch, T.: Disentangling mode-specific selection and measurement bias in social surveys. Soc. Sci. Res. **42**, 1555–1570 (2013)

Schwarz, N., Strack, F., Hippler, H.J., Bishop, G.: The impact of administration mode on response effects in survey measurement. Appl. Cognit. Psych. **5**(3), 193–212 (1991)

St-Pierre, M, Beland, Y: Mode Effects in the Canadian Community Health Survey: A Comparison of CAPI and CATI. Proceedings of the American Statistical Association Meeting. Survey Research Methods, Toronto, Canada. American Statistical Association, pp. 1–8 (2004).

Tourangeau, R., Yan, T.: Sensitive questions in surveys. Psychol Bull **133**(5), 859–883 (2007)

Trappmann, M., Bähr, S., Beste, J., Eberl, A., Frodermann, C., Gundert, S., Schwarz, S., Teichler, N., Unger, S., Wenzig, C.: Data resource profile: panel study labour market and social security (PASS). Int. J. Epidemiol. **48**(5), 1411–1411g (2019)

Trappmann, M, Müller, G, Bethmann, A: Design of the Study. In: Bethmann, A, Fuchs, B, Wurdack, A (eds.) User Guide "Panel Study Labour Market and Social Security" (PASS). Wave 6, (FDZ-Datenreport, 07/2013 (en)), Nürnberg, S. 13–22 (2013).

University of Essex, Institute for Social and Economic Research: Understanding Society: Waves 1–11, 2009–2020 and Harmonised BHPS: Waves 1–18, 1991–2009. [data collection]. 14th Edition. UK Data Service. SN: 6614 (2021).

US Census Bureau: Understanding and using American community survey data: what all data users need to know. U.S. Government Publishing Office, Washington, DC (2020)

Vannieuwenhuyze, J.T.A., Loosveldt, P., Molenberghs, G.: Evaluating mode effects in mixed-mode survey data using covariate adjustment models. J. Off. Stat. **30**, 1–21 (2014)

Publisher's Note