2019

IARIW-World Bank

Special IARIW-World Bank Conference “New Approaches to Defining and Measuring Poverty in a Growing World” Washington, DC, November 7-8, 2019

Comprehensive Data Quality Studies as a Component of Poverty Assessments

Mark Brooks

Rattiya S. Lippe

Hermann Waibel

Paper Prepared for the IARIW-World Bank Conference

Washington, DC, November 7-8, 2019

Session 4A: Issues of Poverty Data and the SNA

Time: 13:30 – 15:30, November 8

Comprehensive data quality studies as a component of poverty assessments

By Mark Brooks*, Rattiya S. Lippe, Hermann Waibel

Faculty of Economics and Management, Institute of Development and Agricultural Economics, Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany

Abstract

High quality household data are essential for realistic poverty assessments and provide the basis for designing effective policies to sustainably reduce poverty. Despite of this, income measures from household surveys are often plagued by non-sampling errors such in the form of non-response and measurement error. Current research, while generating important lessons, is often limited in scope and largely examines quantifiable interviewer and respondent characteristics on the prevalence of non-sampling errors. This paper presents a comprehensive study on data quality based on an ongoing long-term panel in Thailand and Vietnam. We show that respondent and interviewer characteristics alongside the interview and survey environment play a major role on the prevalence of non-sampling errors. Moreover, we provide recommendations to survey producers on how to improve the process of identifying and reducing non-sampling errors.

Key words

Non-sampling error; household survey; poverty assessments; data quality; Thailand; Vietnam

---

1 Introduction

Since the implementation of the Millennium Development Goals and Sustainable Development Goals, extreme poverty has declined from 36 percent in 1995 to approximately 9 percent in 2017 (Cuaresma et al., 2018). Following this decline, it has become necessary to improve the process of identifying those who fall into the category of extreme poverty. Hence, improving the reliability and validity of survey data is a necessary step (c.f. Squires et al., 2019). Generally, survey errors refer to deviations of obtained values from true or expected values (Groves et al., 2009). A large proportion of survey error is introduced by non-sampling errors, which occur due to interactions between the interviewer and respondent or weaknesses in the survey design (e.g. Groves, 1989). Non-response at a unit-level can reduce the representativeness of a survey, whereas item non-response (e.g. missing values) and measurement errors (e.g. outlier or implausible values) can reduce the validity of data. Especially data on the value of assets, income and consumption are susceptible to higher shares of both missing information and measurement errors due to their sensitive nature, recall bias, and respondent behavior, resulting in under-/overestimations (Frick & Grabka, 2014; Meyer et al., 2015; Meyer et al., 2018; Moore et al., 2000; Nicoletti et al., 2011; Watson & Li, 2016). Assets, income and consumption data are the foundation for research on poverty – e.g. shock coping strategies, risk mitigation strategies, poverty dynamics and transitions. Shortcomings of data on household wealth can lead to an under-/overestimation of poverty leading to wrong policy conclusions.

While there is an abundance of surveys conducted in some countries, there exist to date relatively few surveys that are suitable for use in calculating reliable poverty estimates (Booth, 2019; Gibson, 2016; Jolliffe et al., 2015). Serajuddin et al. (2015) established that for the period of 2002-2011 there were as many as 57 countries for which at most one yearly poverty estimate could be calculated. In many developing countries, however, the quantity of high-quality survey data sets remains sparse (Booth, 2019; Dang & Carletto, 2018).

In developing countries, advanced survey tools such as "Survey Solutions" by the World Bank (https://mysurvey.solutions) have increased the effectiveness of data collection. Experimental evidence suggests that Computer Assisted Personal Interviews (CAPI) can prevent numerous errors due to the implementation of, for example, automated plausibility checks (e.g. Banks & Laurie, 2000; Caeyers et al., 2012; de Leeuw et al., 1995). Although data access, use, and collection have benefited greatly, such innovations do not automatically solve the issue of low-quality data. Issues such as a lack of interviewer consistency, misinterpretation of questions, and challenging interview and survey conditions remain prevalent in developing countries (Lupu & Michelitch, 2018). Furthermore, to date there are only few studies presenting empirical evidence on the determinants and impact of non-sampling errors in developing countries.

There are at least three shortcomings in the literature on data quality of household surveys in developing countries. Firstly, most studies rely on either cross-section data or experiments and therefore are limited in scope as regards the types of non-sampling errors studied. Secondly, emphasis has been given to quantitative variables such as age, education and other characteristics of interviewer and respondents, whereas qualitative information such as interviewer/respondent behavior, their personality traits, and motivations are not yet considered. Thirdly, most studies focus on individual aspects of data quality such as the effects of the interviewer and/or respondent characteristics, but rarely account for the circumstances of the interview itself, e.g. interview and survey environment (Lupu & Michelitch, 2018). One such paper by Phung et al. (2015) accounts for interview and survey aspects such as the period of day or season during which the interview took place. In summary, studies seldom apply a comprehensive approach that includes quantitative and qualitative information that allows researchers to identify the relative importance of: (a) interview environment, (b) survey environment, (c) interviewer and respondent characteristics, and (d) time.

In this paper, we analyze the determinants of item non-response errors and measurement errors in data items relevant for the measurement of household wealth and poverty. Our study adds to the literature by taking into account a rich set of quantitative and qualitative interviewer and respondent characteristics based on a household panel survey in two countries in South East Asia. We find that interviewer and respondent characteristics are the main drivers of non-sampling errors. Characteristics such as interviewer experience, field of study, status of the respondent within the household and the occupational background of the respondent significantly affect data quality. Furthermore, interview and survey environment play a key role. We find that measurement errors represent the greatest threat to the quality of household survey income data.

The outline of the paper is as follows: In Section 2, we formulate a model to determine factors influencing the prevalence of non-sampling errors in variables pertaining to income. This is followed by a description of the household survey data used in our analysis. Section 4 provides our main results and discusses likely factors influencing the prevalence of non-sampling errors. In the final section, a conclusion is drawn and recommendations for improving data quality in household surveys are proposed.

2 Indicators of data quality

Following the literature, non-sampling errors are widely considered to consist of: (i) coverage errors, (ii) non-response errors, and (iii) measurement errors (e.g. Groves, 1989; Lessler & Kalsbeek, 1992; Weisberg, 2005). Such errors transpire over the course of the collection of survey data and during post-survey data processing.

Coverage errors occur when the sample of respondents used in a survey is not representative of the target population. Accordingly, coverage errors are grounded in decisions made during the survey design phase and are the result of faulty information on the likelihood of the inclusion of a sampling unit in the sample. They occur due to over- or under-coverage of sampling units, such as: (i) incorrect sampling units being included in the sampling frame; (ii) important sampling units not being included in the sampling frame; or (iii) duplication of sampling units in the sampling frame (Groves et al., 2009).

Non-response occurs when measurements on sampled units cannot be obtained (Fowler, 2013). Non-response errors are defined as consisting of unit non-response and item non-response. Unit non-response refers to the sampling unit being unavailable throughout the entirety of the interviewing period. For example, a respondent may not have been available when the survey team was on-site or may have refused to participate in the survey (Lynn & Clarke, 2002). Item non-response occurs when measurements on the sampling unit are only partially collected. Non-response on an item level can result from interviewers erroneously skipping questions, respondents being unable to recall required information, or respondents refusing to answer questions.

The final category of non-sampling errors, measurement errors, occurs when there is a deviation from the true value of a measurement and the value provided by the sample unit. There are three types of measurement errors: response, interviewer and post survey errors (Weisberg, 2005). For example, a measurement error could occur due to a respondent failing to provide an accurate estimate, misinterpreting the question or because of an entry mistake by the interviewer.

The focus of this paper is on non-sampling errors that occur during the interview. Hence, we omit coverage errors as these are mainly attributed to the survey design phase. In our analysis, we differentiate between two types of non-response errors: missing values and refusal values. Missing values occur when a question remains unanswered or is erroneously skipped. Refusals are defined as questions in which a code (e.g. "no answer") has been selected, which indicates that the respondent has actively decided to not answer the question. Measurement errors are split into outlier values and implausible values. We define outlier values as values that are far greater (or lower) than the boundaries determined within the survey guidelines. Implausible values are defined as responses that do not comply with survey plausibility rules, e.g. if an incorrect entry occurs during an interview or if there are inconsistencies between similar questions in varying sections of a questionnaire.

There is a broad scope of explanations for how and why non-sampling errors can occur in surveys, of which we will provide a non-exhaustive overview in the following paragraphs.

The cause of many non-sampling errors is the process of respondents formulating responses (Tourangeau et al., 2000). Several problems can occur ranging from misinterpretations of questions to deliberate misreporting. In addition, recall bias is often accounted for in research and occurs when a respondent's judgement or the method applied to provide an estimated response is flawed. This bias is particularly prevalent for questions referring to events that occurred long before the interview and it is of immense importance to account for such bias in particular in the context of agricultural information (e.g. Beegle et al., 2012). Many studies have dealt with the role of the respondent regarding the prevalence of non-sampling errors and have focused on aspects such as proxy respondents, who must provide in-depth answers on all household members and activities (e.g. Alwin, 2007; Bardasi et al., 2011; Stoop et al., 2010). Household heads are often targeted as respondent in developing countries and have been found to provide interviews with fewer missing values (e.g. Phung et al., 2015) and in some cases to underestimate the income of other household members (e.g. Fisher et al., 2010). The cognitive ability of respondents, which can be captured by their age and level of education, also affects the quality of data. Interviews with respondents with lesser cognitive ability lead to higher share of satisficing (e.g. Knäuper et al., 1997; Knäuper, 1999; Krosnick, 1991). Gender generally significantly affects data quality, although results are inconclusive as to whether male or female interviewers reduce the number of non-sampling errors (e.g. Fowler & Mangione, 1990; Lessler & Kalsbeek, 1992; Phung et al., 2015).

The role of the interviewer in obtaining high quality data has been extensively researched and documented in the literature. By deviating from procedures defined in the survey guidelines interviewers can influence the respondent's answers. The interviewer can rephrase questions, disregard interview instructions (e.g. by reading out the codes for an open-ended question), skip certain questions (e.g. often questions the interviewer perceives to be sensitive in nature), or purposefully/accidentally enter a response that does not match with that of the respondent. Such faulty methods of enumeration can be the result of a lack of training or experience in conducting interviews (e.g. Campanelli & O'Muircheartaigh, 1999; Singer et al., 1983; Sinibaldi et al., 2009). Furthermore, the interviewer can bias the response in other ways, for example by assisting the respondent with difficult questions and/or by applying probing techniques in order to obtain answers. However, even if the interviewer were to follow all survey procedures variations in the emphasis or intonation of parts of a question can prompt a response that is not equivalent to the true response of the respondent (Groves, 2009). The role of interviewer experience in surveys is well researched. Prior experience in survey activities provides interviewers with basic survey knowledge regarding interviewing and behaviors to elicit cooperation and accurate response from respondents (Couper & Groves, 1992). However, some conflicting research suggests that interviewers without an extensive survey background may provide data of higher quality (e.g. Fowler & Mangione, 1990; Fowler, 2013; Sinibaldi et al., 2009). While the literature generally suggest that the gender of interviewers affects the quality of data, the results are not consistent. Campanelli & O'Muircheartaigh (1999) find that male interviewers provide data of poorer quality, whereas Phung et al. (2015) observe the opposite effect. Interviewers, who exert friendly or motivating behaviors, often achieve higher cooperation rates and accordingly provide data of higher quality (Jäckle et al., 2013; Olson et al., 2016).

The interview is defined as a structured social interaction by social psychologists (e.g. Kahn & Cannell, 1957), who found that demographic and socio-economic characteristics of both the interviewer and respondent influence the quality of data. Following the assumption that interviewer and respondent interactions will significantly affect the quality of data, research has dealt with the effect of congruent characteristics such as age, gender, and ethnicity (e.g. Baird et al., 2008; Phung et al., 2015). In particular, interviews on sensitive topics have been found to yield more reliable data when respondent and interviewer characteristics match (e.g. Catania et al., 1996). Congruent ethnicity has been found to be highly significant in survey populations that capture a significant share of minority groups that differ from ethnic majority groups, e.g. in terms of culture or language.

The environment, in which the interview takes place, plays an important role regarding data quality. Studies generally control for the duration of interviews. Longer interviews when compared to the survey mean are found to be more prone to non-sampling errors, which can be explained by increasing levels of both interviewer and respondent fatigue and decreasing levels of motivation (e.g. Galesic & Bosnjak, 2009; Phung et al., 2015). The presence of other people during interviews may have an impact on respondent behavior in that they adjust their responses to adhere to social norms, in particular regarding questions on sensitive matters (e.g. Krumpal, 2013; Smith, 1997). Aspects of the survey environment are rarely accounted for in analyzing non-sampling errors. An aspect that should significantly affect the quality of data is the quantity and quality of supervision during surveys. Inadequate scrutiny of data and untimely data processing will lead identify underlying issues in the behavior of interviewers or in the survey instrument too slowly. A lack of monitoring during survey activities may also lead to more issues in terms of data quality (e.g. Groves et al., 2009).

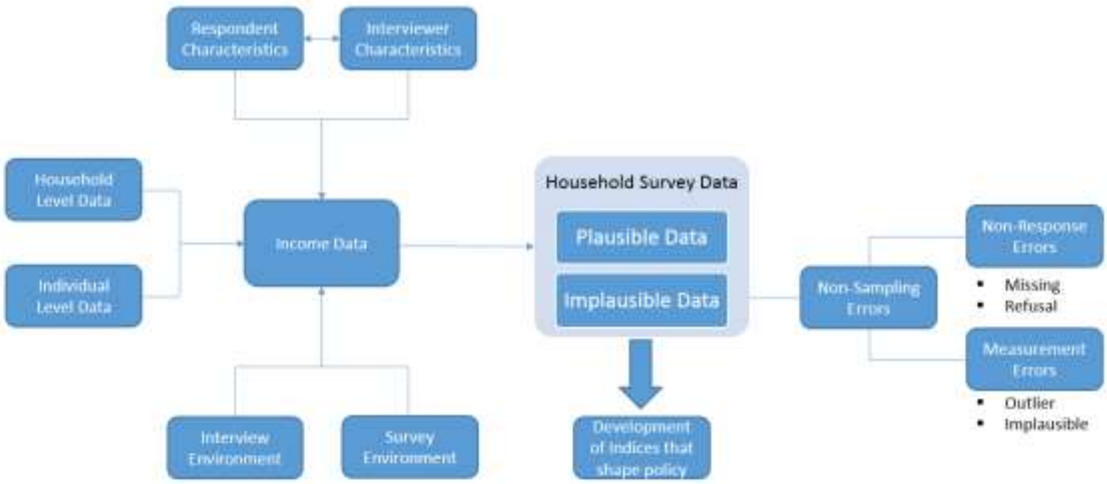
Based on the literature review we hypothesize that five factors influence the prevalence of non-sampling errors, namely (i) interviewer and (ii) respondent characteristics but also for factors from the (iii) interview and (iv) survey environment. Moreover, we assume that while characteristics of interviewers and respondents play the largest role, the role of the interview and survey environment is often underestimated in current literature on non-sampling errors. Lastly, we hypothesize that some factors may have different determinants when examining surveys in different cultural contexts.

3 Conceptual framework and methodology

Household surveys incorporate both household and individual level data in collection of income data. Individual level income can be measured, for example, by capturing income from employment activities or transfer payments. Household level income is drawn from agricultural activities, household dynamics and remittances, and agricultural activities in the context of most surveys. Survey research considers that while survey data items are largely plausible, there are several sources of error, which lead to deviations between an obtained value and a true value that must be considered (Groves et al., 2009; Weisberg, 2005). In our framework, we argue that non-sampling errors in their various forms are influenced by (i) interviewer and (ii) respondent characteristics, (iii) congruent characteristics and factors from the (iii) interview and (iv) survey environment (c.f. Figure 1).

Based on our framework we developed our model, and applied a comprehensive approach that accounts for an extensive range of variables drawn from factors (i)-(v). Our approach provides an extension to Phung et al. (2015) by further differentiating between different types of non-sampling errors. In order to achieve our objective of assessing simultaneously the effect and relative importance of these factors on the share of non-sampling errors, we follow the sub-categories as defined in Section 2.

Figure 1.
Conceptual framework



Source: Own representation

We establish two related regression models with a set of explanatory variables and five dependent variables, which are based on the share of erroneous responses to income relevant variables by interview: (a) missing values; (b) refusal values; (c) outlier values; (d) implausible values; and (e) total erroneous values.

The first model is a probit model that captures the determinants of an interview containing a non-sampling error. As income data, is generally described as being susceptible to a higher share of non-sampling errors (e.g. Frick & Grabka, 2014; Meyer et al., 2015; Moore et al., 2000; Watson & Li, 2016). We then run a second OLS model to examine the determinants of the errors in income relevant variables. By comparing between the two models, we can observe whether select determinants are consistent in terms of their effect on the occurrence and share of non-sampling errors.

Our analysis is carried out separately for two countries in South East Asia, namely Thailand and Vietnam, to determine whether the relative importance and effect of determinants are consistent in different cultural contexts in the long-term household survey data used in our analysis.

In the first step, we apply a probit model in order to measure the impact of interviewer and respondent characteristics, the interview environment and survey environment on the binary dependent variable capturing whether an interview contains one of the five types of non-sampling errors as specified above or not. The model specification is as follows:

$$Y_{in} = \alpha_0 + \beta_k * X_{ki} + \delta_m * Z_{mi} + \rho_n * F_{ni} + \eta_o * I_{oi} + \vartheta_p * S_{pi} + \lambda_{ji} * H_{ji} + \varepsilon_i \quad (1)$$

where Y_{in} represents whether or not an error is prevalent in the context of an interview (= 1 if an interview includes an error of type n); X_{ki} are interviewer characteristics; Z_{mi} are respondent characteristics; F_{oi} are congruent characteristics between the interviewer and respondent; I_{oi} represent characteristics of the interview environment; S_{pi} are the characteristics of the survey environment; and H_{ji} captures household characteristics.

The second model measures the impact of the explanatory variables used in the probit model on the share of erroneous values as captured by the four aforementioned independent variables. In this model, we disregard interviews that were found to be free of errors. The model is specified as follows:

$$Y_{in} = \alpha_0 + \beta_k * X_{ki} + \delta_m * Z_{mi} + \rho_n * F_{ni} + \eta_o * I_{oi} + \vartheta_p * S_{pi} + \lambda_{ji} * H_{ji} + \varepsilon_i \quad (2)$$

where Y_{in} represents the share (in %) of non-sampling errors by error type within variables relevant to income.

In terms of interviewer characteristics, we include quantifiable characteristics such as age, gender, education, locality, and interviewer experience. In addition, our model captures qualitative information such as the level of interviewer professionalism and interviewer personality traits. According to the literature, interviewer experience generally influences non-sampling errors (e.g. Baird et al., 2008; Campanelli & O'Muircheartaigh, 1999; Olson & Bilgen, 2011; Sinibaldi et al., 2009; Townsend et al., 2013). The literature also suggests that local interviewers reduce the frequency of non-sampling errors (e.g. Phung et al., 2015). Sinibaldi et al. (2009) find that extraverted and conscientious interviewers increase respondent co-operation in surveys, whereas more open and agreeable interviewers, contrary to expectations, reduce the probability of co-operation.

Respondent characteristics such as age, gender, education, status, as well as character and personality traits are controlled for. The literature suggests that older respondents give less precise information. For example, Knäuper et al. (1997) use respondent age as a proxy for cognitive ability and find that respondents with lower cognitive ability (e.g. higher age) are less likely to provide accurate information. It is suggested in the literature and a common practice in household surveys in developing countries to interview the household head. Although proxy reporting generally has been found to reduce unit and item non-response and simultaneously increase the amount of measurement errors (e.g. Moore, 1988; Weir et al., 2011) this approach has been found to yield interviews of higher quality when compared with the use of other household members as proxy respondents (Phung et al., 2015). However, conflicting studies have suggested that interviewing male household heads can lead to significant underreporting of household income (Fisher et al., 2010).

Among congruent characteristics of interviewers and respondents, Baird et al. (2008) and Phung et al. (2015) suggest that congruent age can significantly improve the degree of co-operation during an interview and thus increase the quality of data. Further, the authors found pairing interviewers and respondents of the same ethnicity facilitated trust and thus improve co-operation during the interview.

The duration of the interview significantly affects non-sampling errors as longer interview times serves as a proxy for not only respondent, but also interviewer fatigue (e.g. Galesic & Bosnjak, 2009; Phung et al., 2015). Additionally, the presence of other household- and non-household members is likely to influence the interaction quality during the interview as well as the timing of the interview itself.

In terms of the survey environment, the frequency of participation of the respondent in the survey is included, which captures respondent fatigue, motivation and learning effects over the span of their survey participation. Additionally, we control for the week in which the interview takes place. The literature suggests that interviews towards the beginning of the survey activity will be of lower quality and some studies find that the quality also drops significantly towards the end of the survey activity (Baird et al., 2008; Townsend et al., 2013).

In the next section, an overview of the data used is provided and descriptive statistics of the variables used are presented.

4 Data

The long-term panel survey "Thailand Vietnam Socio Economic Panel" (TVSEP) is the source of data for the ensuing analysis of non-sampling errors in income components and measurements. A consortium of German research institutes implemented the project with the goal of advancing the understanding of vulnerability to poverty in the context of emerging economies in Southeast Asia. The panel, which consists of 4,400 households is located in six provinces in Thailand and Vietnam, was first conducted in 2007 and is ongoing. To date eight waves of data have been collected.

The survey instrument follows the standard design of household surveys and includes components such as household characteristics, including education and health modules. Further components deal with household dynamics, assets and resources as well as a detailed section on the household's homestead. Sources of income are captured from the perspective of agriculture, small-scale business self-employment and wage employment. Additional modules contain information on the financial state of the household, namely borrowing, lending, savings and public transfers. An in-depth segment on shocks and perceived risks further benefits the project main goal of analyzing vulnerability to poverty.

The basis for our analysis is the 2017 household survey, which consists of 10 sections. The paper-based questionnaire spans 81 pages and the tablet version of the questionnaire contains over 900 variables. Due to some sections containing multiple rows of data (e.g. on each individual household member), the mean number of data items per questionnaire is 1,524. The 2017 survey consists of 3,812 households, which to date remain representative for the original population of rural households (Liebenehm et al., 2018).

In total, 84 variables from the survey instrument are relevant to the calculation of the income aggregate. In terms of the overall number of items, the majority of income data stems from sections on crop production, livestock and livestock produce, and natural resource extraction activities.

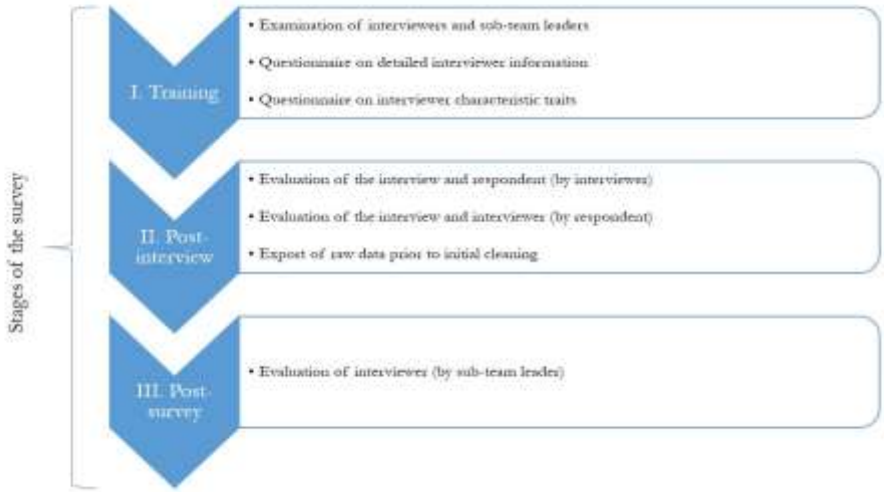
The target population is comprised of rural households that are suitable for the objective of delving into vulnerability studies. The survey area is characterized by low per capita income, inequality in village-level wealth distribution, a high share of agriculture-based household income, poor infrastructure, adjacency to the Laotian or Cambodian border, and high development potential (c.f. Hardeweg et al., 2013).

In order to be able to draw a greater sample of non-sampling errors the survey data structure of TVSEP was adjusted to facilitate research on data quality using the data gathered during the 2017 survey. The standard data checking process was restructured in order to obtain interviews with the least possible level of supervision and on-site adjustments (see Appendix: Figure 2). Hence, a comparison of the raw interview can be made with post-survey data and data that has been subject to the post-survey cleaning process of TVSEP.

In addition, paradata[1] were generated during several stages of the survey process. Firstly, during the interviewer training, paradata consisting of examinations of interviewers, in-depth interviewer information and self-assessed interviewer personality traits were compiled. Secondly, after the completion of each interview, both the interviewer and respondent evaluated the interview and the interaction with their counterpart. Thirdly, after the conclusion of the survey, sub-team leaders evaluated the performance of each of their interviewers (see Figure 3). The implementation of a section on personality traits following the example of the German Socio Economic Panel (SOEP) (www.diw.de/en/soep) and based on the Big Five model developed by Cost and McCrae (1992; 1997) allows for novel insights regarding their influence on data quality. The TVSEP section on personality traits was found to be valid and applicable in the context of both Thailand and Vietnam by Buehler et al. (2019).

1 Paradata refer to data collected that describe the process of survey production that are not part of the interview itself (Couper & Lyberg, 2005; Kreuter et al., 2010)

Figure 3.
Data quality modules added to TVSEP 2017 survey



Source: Own representation

Following survey documentation and guidelines, errors in the variables of interest for the income aggregate were identified and categorized according to our five dependent variables. In our analysis, we use the errors that remain after minimal on-site supervision in order to measure which factors determine the prevalence and share of non-sampling errors during the interview process.

5 Results

In Table 1a/1b, the descriptive statistics of the determinants applied to the two regression models are presented. The mean share of non-sampling errors ranges between 6.43% in Thailand and 2.96% in Vietnam. This can be explained in part by the difference in the count of interviews that were found to be free of non-sampling errors between the two countries. In Thailand, this consists of 170 interview or ~9% of all interviews and 237 interviews (~14.5%) in Vietnam. The main type of non-sampling errors are similar in both countries and mainly consist of missing and implausible values. The share of refusal and outlier values is generally lower. The majority of non-sampling errors stem from income relevant data from the crop section.

The models' independent variables are adjusted based on the differences between the survey population of Thailand and Vietnam and in the selection process of interviewers. Hence, some variables are included in Vietnam that are not accounted for in the case of Thailand. For example, the impact of congruent ethnicity is not controlled for in the Thailand data set because ethnic diversity among respondents is small. The other, originally non-Thai ethnic groups in Northeast Thailand are homogenous in terms of culture and language when compared to the majority ethnic Thais. In Vietnam, however, there is both a higher share of ethnic minority groups and larger differences between ethnic groups (Dang, 2012). The interviewers hired also differed between the two countries. Interviewers in Thailand were mainly hired from local universities and accordingly almost exclusively have a local background. In Vietnam, interviewers are often freelancers with a background in the "survey industry" and varying levels of education. Interviewer age does not vary greatly in both countries with interviewers generally being in their mid-twenties to early thirties.

In terms of respondent age, the majority are older than 50 years, with the mean age of respondents being slightly lower in Vietnam. It is also noticeable that a high share of respondents are heads of their household in both countries (~60%). Considering that the majority of income data stems from agriculture related activities, we include a dummy that captures whether or not the respondent's main or secondary occupation is related to agriculture. Thai respondents primarily work in agriculture in 60% of the cases, whereas 70% do so in Vietnam. In terms of respondent continuity almost 10% of Thai respondents participated in every wave of TVSEP, whereas this was only the case for around 4% of Vietnamese respondents. Accordingly, these respondents have participated in all eight waves of TVSEP since its implementation in 2007. Respondents are predominantly female in both countries, with a similar share of male and female household heads in Thailand (~50%). In Vietnam, the majority of household heads are male (~70%). Interviewers tend to be homogeneous regarding characteristics such as age and level of education due to the selection process of interviewers in the TVSEP. Accordingly, interviewers have a background in university level education, although some 15% of Thai interviewers have not yet completed a bachelor degree. In terms of survey experience, slightly more than half of the Thai interviewers have a background as an interviewer Almost 95% of Vietnamese interviewers have prior experience working in household surveys and all interviewers have attained their bachelor degree.

The average interview time during the 2017 survey was slightly under three hours in Thailand and almost four and a half hours in Vietnam. The majority of interviews followed the standard procedures of the survey, with only 1% of interviews being conducted during the more inconvenient evening interview sessions. Data collection lasted five weeks in Thailand, whereas in Vietnam the survey lasted 6 weeks.

Table 1a. Summary statistics of determinants (Thailand)

Variables	**Obs.**	**Mean**	**Std. dev.**
*Dependent variables **			
Missing values (in %)	1,818 (336)	0.94 (5.09)	4.69 (9.91)
Refusal values (in %)	1,818 (587)	0.65 (2.01)	2.25 (3.60)
Outlier values (in %)	1,818 (539)	0.43 (1.45)	0.95 (1.25)
Implausible values (in %)	1,818 (1,489)	3.81 (4.65)	4.38 (4.42)
Total erroneous values (in %)	1,818 (1,648)	5.83 (6.43)	7.04 (7.13)
Independent variables			
Respondent age (1 = ≥ 60 ; 0 = < 60)	1,818	0.42	0.49
Respondent gender (1 = male, 0 = female)	1,818	0.34	0.47
Respondent education – secondary and higher (1 = yes, 0 = no)	1,818	0.15	0.36
Respondent is head of household (1 = yes, 0 = no)	1,818	0.57	0.49
Respondent agreeableness	1,817	5.75	0.97
Respondent has agricultural background (1 = yes, 0 = no)	1,818	0.70	0.46
Interviewer gender (1 = male, 0 = female)	1,818	0.28	0.45
Interviewer has completed degree (1 = yes, 0 = no)	1,818	0.85	0.36
Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)	1,818	0.22	0.42
Interviewer exam score (in %)	1,818	25.78	15.06
Interviewer household survey experience (1 = yes. 0 = no)	1,818	0.61	0.48
Interviewer other survey experience (1 = yes, 0 = no)	1,818	0.26	0.44
Interviewer TVSEP survey experience (1 = yes, 0 = no)	1,818	0.13	0.33
Interviewer openness	1,818	4.59	1.19
Interviewer extraversion	1,818	3.96	0.79

Log of interview duration (minutes)	1,817	5.05	0.35
Morning Interview (1 = yes, 0 = no)	1,818	0.53	0.49
Respondent participated in all waves (1 = yes, 0 = no)	1,818	0.10	0.29
Survey week	1,818	2.65	1.14
Household size (persons)	1,818	4.57	1.91

Source: Own calculations based on TVSEP survey 2017.Note: *Of the 1,818 questionnaires available for the analysis, 170 were free of errors and hence not included in the second regression model. Number in brackets refers to error descriptives for the OLS model.

Table 1b. Summary statistics of determinants (Vietnam)

Variables	Obs.	Mean	Std. dev.
*Dependent variables **			
Missing values (in %)	1,629 (182)	0.27 (2.44)	1.41 (3.55)
Refusal values (in %)	1,629 (134)	0.08 (0.95)	0.38 (0.96)
Outlier values (in %)	1,629 (484)	0.23 (0.76)	0.50 (0.66)
Implausible values (in %)	1,629 (1,345)	2.38 (2.89)	3.09 (3.18)
Total erroneous values (in %)	1,629 (1,392)	2.96 (3.47)	3.71 (3.79)
Independent variables			
Respondent age (1 = ≥ 60 ; 0 = < 60)	1,629	0.26	0.44
Respondent gender (1 = male, 0 = female)	1,629	0.44	0.50
Respondent education – secondary and higher (1 = yes, 0 = no)	1,629	0.67	0.47
Respondent is head of household (1 = yes, 0 = no)	1,629	0.58	0.49
Respondent agreeableness	1,629	5.88	0.90
Respondent has agricultural background (1 = yes, 0 = no)	1,629	0.82	0.38
Interviewer gender (1 = male, 0 = female)	1,629	0.40	0.49
Interviewer from suitable field of study (1 = economics/agriculture, 0 = other)	1,629	0.48	0.50
Interviewer exam score (in %)	1,629	37.45	19.11
Interviewer household survey experience (1 = yes. 0 = no)	1,629	0.06	0.24
Interviewer other survey experience (1 = yes, 0 = no)	1,629	0.89	0.32
Interviewer TVSEP survey experience (1 = yes, 0 = no)	1,629	0.05	0.22
Interviewer openness	1,629	5.49	0.91
Interviewer extraversion	1,629	4.14	0.67

Congruent ethnicity (1 = yes, 0 = no)	1,629	0.76	0.43
Log of interview duration (minutes)	1,629	5.58	0.32
Morning Interview (1 = yes, 0 = no)	1,629	0.59	0.49
Respondent participated in all waves (1 = yes, 0 = no)	1,629	0.04	0.20
Survey week	1,629	3.78	1.35
Household size (persons)	1,629	4.44	1.79

Source: Own calculations based on TVSEP survey 2017.

Note: *Of the 1,629 questionnaires available for the analysis, 237 were free of errors and hence not included in the second regression model. Number in brackets refers to error descriptives for the OLS model.

Determinants of non-sampling errors are identified by applying a two regression models. The results of the probit model are shown in Table 2a/2b and the results of the OLS are shown in Table 3a/3b.

The results of the probit model show that estimated coefficients generally have the expected signs and these can then be used to determine the consistency of our results in the OLS regression. The major results can be summarized as follows: First, respondents whose main occupation is based in the agriculture sector seems to lead to higher likelihood of non-sampling errors in the survey instrument. This is likely due to such respondents being members of agriculture intensive households, who will have a larger number of data items in the agriculture section, which contains the highest number of income related questions.

In Thailand, interviewers from fields of study that extensively deal with the topics in the TVSEP survey (e.g. economists and agriculturalists) are less likely to produce non-sampling errors in their interviews. The opposite is the case for Vietnam. When examining the impact of survey experience similar results can be found. This suggests that interviewers in Vietnam, who can be characterized as professional, full-time interviewers, are less likely to conform to the survey procedures and guidelines and instead prefer to apply their extensive experience in other surveys to their interviews. This is in line with the findings of Fowler & Mangione (1990); Fowler (2013); and Sinibaldi et al. (2009).

Interviewer personality traits, namely openness and extraversion, which are generally thought to positively influence cooperation in interviews, signal that they decrease the probability of such errors occurring (e.g. Jäckle et al., 2011; West & Blom, 2017). The aspect of improved cooperation seems to be further solidified as errors that are significantly affected by personality traits consist of refusal and implausible values.

Congruent ethnicity plays a significant role as we hypothesized. Matching interviewers and respondents with the same ethnic background reduced the likelihood of implausible responses in interviews. Social norms in countries with very different ethnic groups can affect the way in which respondents interact with interviewers from outside of their own communities and reduce the level of cooperation as found by Adida et al. (2016).

Longer interview durations increase the likelihood of non-sampling errors, as hypothesized. This is presumably due to interviewer and respondent fatigue. Longer interviews will generally also be more complex in nature, which will also significantly affect the likelihood of errors occurring.

The survey shows mixed results in the model, which suggests that the probability of non-sampling errors is not necessarily consistently influenced by interviewer experience in the field with the survey instrument.

Table 2a. Probit regression results: Dependent variable, non-sampling error dummy (Thailand)

	(1) Missing value	(2) Refusal value	(3) Outlier value	(4) Implausible value	(5) Error value
Respondent age ≥ 60 (1 = yes, 0 = no)	0.155 (1.21)	-0.004 (-0.03)	0.095 (0.82)	0.092 (0.67)	0.018 (0.11)
Respondent is head of household (1 = yes, 0 = no)	0.165 (1.67)	-0.042 (-0.46)	0.019 (0.21)	-0.072 (-0.69)	0.064 (0.50)
Respondent age (= 1) * head of household (Interaction effect)	-0.179 (-1.17)	-0.111 (-0.77)	-0.114 (-0.80)	-0.133 (-0.81)	-0.137 (-0.70)
Respondent gender (1 = male, 0 = female)	0.046 (0.57)	0.115 (1.48)	0.086 (1.11)	0.034 (0.39)	0.042 (0.42)
Respondent education – secondary and higher (1 = yes, 0 = no)	0.057 (0.57)	0.075 (0.84)	0.236** (2.63)	-0.009 (-0.09)	0.101 (0.75)
Respondent agreeableness	0.039 (1.02)	0.015 (0.45)	0.026 (0.76)	0.035 (0.91)	0.041 (0.95)
Respondent has agricultural background (1 = yes, 0 = no)	-0.036 (-0.45)	0.051 (0.67)	0.512*** (6.60)	0.570*** (7.24)	0.514*** (5.61)
Interviewer gender (1 = male, 0 = female)	0.009 (0.10)	-0.038 (-0.46)	-0.374*** (-4.49)	-0.222* (-2.41)	-0.266* (-2.52)
Interviewer has completed degree (1 = yes, 0 = no)	-0.121 (-1.07)	-0.455*** (-4.49)	-0.189 (-1.77)	0.210 (1.84)	-0.023 (-0.18)
Interviewer from suitable field of study (1 = economics / agriculture, 0 = other)	-0.091 (-0.95)	-0.307*** (-3.40)	-0.301*** (-3.34)	-0.088 (-0.92)	-0.238* (-2.24)
Interviewer exam score (in %)	-0.002 (-0.52)	0.000 (0.00)	-0.007* (-2.48)	-0.009** (-2.85)	-0.006 (-1.67)
Interviewer other household survey experience (1 = yes, 0 = no) (No experience is base)	-0.050 (-0.55)	0.084 (1.01)	-0.071 (-0.86)	0.017 (0.18)	0.013 (0.11)
Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)	-0.123 (-1.00)	0.111 (1.03)	-0.120 (-1.13)	0.103 (0.85)	-0.134 (-0.96)
Interviewer openness	0.019 (0.54)	-0.064 (-1.95)	0.000 (0.00)	0.029 (0.80)	0.015 (0.32)
Interviewer extraversion	-0.018 (-0.38)	-0.022 (-0.47)	0.041 (0.91)	-0.114* (-2.15)	-0.061 (-1.00)
Log of interview duration (minutes)	0.090 (0.84)	0.060 (0.61)	0.343*** (3.33)	0.308** (2.74)	0.424** (3.21)
Morning interview (1 = yes, 0 = no)	-0.013 (-0.19)	-0.056 (-0.87)	0.097 (1.49)	0.131 (1.83)	0.036 (0.42)
Respondent participated in all waves (1 = yes, 0 = no)	-0.297* (-2.17)	-0.054 (-0.46)	-0.051 (-0.42)	-0.020 (-0.15)	-0.105 (-0.69)

Survey week	-0.163*** (-4.91)	0.281*** (9.45)	-0.124*** (-4.24)	-0.054 (-1.66)	0.012 (0.30)
Buriram province (1 = yes, 0 = no) (Nakhon Phanom is base)	-0.182 (-1.58)	-0.259* (-2.39)	0.067 (0.60)	-0.058 (-0.46)	-0.020 (-0.14)
Ubon province (1 = yes, 0 = no) (Nakhon Phanom is base)	-0.024 (-0.21)	-0.004 (-0.04)	-0.004 (-0.03)	-0.046 (-0.38)	0.068 (0.46)
Household size (persons)	0.025 (1.34)	0.036* (2.03)	0.048** (2.74)	0.088*** (4.01)	0.089*** (3.31)
Constant	-1.106 (-1.73)	-0.883 (-1.53)	-2.480*** (-4.04)	-0.980 (-1.43)	-1.298 (-1.65)
Observations	1,816	1,816	1,816	1,816	1,816
Pseudo R^2	0.027	0.074	0.071	0.094	0.101

* Significant at 10%.; ** Significant at 5%.; *** Significant at 1%.

Notes: Absolute value of z-statistics in parentheses

Table 2b. Probit regression results: Dependent variable, non-sampling error dummy (Vietnam)

	(1) Missing value	(2) Refusal value	(3) Outlier value	(4) Implausible value	(5) Error value
Respondent age ≥ 60 (1 = yes, 0 = no)	0.275 (1.42)	0.184 (0.87)	-0.012 (-0.07)	-0.192 (-1.13)	0.089 (0.48)
Respondent is head of household (1 = yes, 0 = no)	-0.009 (-0.07)	0.0952 (0.69)	-0.005 (-0.05)	0.074 (0.65)	0.112 (0.94)
Respondent age (= 1) * head of household (Interaction effect)	-0.509* (-2.31)	-0.136 (-0.56)	0.080 (0.42)	0.391* (1.97)	0.091 (0.42)
Respondent gender (1 = male, 0 = female)	0.091 (0.77)	-0.031 (-0.25)	0.161 (1.80)	-0.129 (-1.26)	-0.142 (-1.33)
Respondent education – secondary and higher (1 = yes, 0 = no)	0.094 (0.79)	-0.031 (-0.25)	0.174* (1.78)	0.273* (3.05)	0.268* (2.85)
Respondent agreeableness	0.031 (0.61)	0.022 (0.46)	-0.034 (-0.90)	0.080 (1.77)	0.077 (1.62)
Respondent has agricultural background (1 = yes, 0 = no)	0.295* (2.28)	-0.024 (-0.19)	0.372*** (3.79)	0.796*** (8.30)	0.791*** (8.02)
Interviewer gender (1 = male, 0 = female)	-0.018 (-0.19)	-0.085 (-0.83)	0.015 (0.19)	0.153 (1.75)	0.123 (1.35)
Interviewer from suitable field of study (1 = economics / agriculture, 0 = other)	0.038 (0.40)	0.413*** (4.14)	0.161* (2.07)	0.293*** (3.36)	0.274** (3.01)
Interviewer exam score (in %)	-0.000 (-0.15)	0.006* (2.15)	0.002 (1.11)	0.000 (0.04)	0.002 (0.67)

Interviewer other household survey experience (1 = yes, 0 = no) (No experience is base)	-0.105 (-0.58)	0.796** (2.99)	0.250 (1.63)	0.439* (2.70)	0.524** (3.17)
Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)	-0.647* (-1.99)	1.068** (3.33)	0.119 (0.52)	0.306 (1.25)	0.423 (1.67)
Interviewer openness	0.053 (0.99)	-0.190*** (-3.63)	0.074 (1.75)	-0.085 (-1.58)	-0.068 (-1.19)
Interviewer extraversion	-0.052 (-0.78)	0.066 (0.74)	-0.004 (-0.07)	-0.039 (-0.62)	-0.049 (-0.75)
Congruent ethnicity (1 = yes, 0 = no)	-0.076 (-0.67)	-0.130 (-1.06)	0.049 (0.53)	-0.361*** (-3.43)	-0.278* (-2.55)
Log of interview duration (minutes)	0.306* (2.06)	-0.012 (-0.07)	0.615*** (5.22)	0.808*** (5.29)	0.821*** (5.15)
Morning interview (1 = yes, 0 = no)	-0.000 (-0.00)	0.016 (0.17)	0.018 (0.26)	0.066 (0.82)	0.100 (1.18)
Respondent participated in all waves (1 = yes, 0 = no)	0.215 (0.90)	-0.077 (-0.27)	0.035 (0.18)	-0.425* (-1.96)	-0.280 (-1.26)
Survey week	-0.180*** (-4.00)	-0.047 (-1.08)	-0.076* (-2.20)	-0.065 (-1.68)	-0.091* (-2.24)
Ha Tinh province (1 = yes, 0 = no) (Dak Lak is base)	0.016 (0.13)	-0.111 (-0.80)	0.310*** (3.08)	0.350* (2.94)	0.362* (2.87)
Hue province (1 = yes, 0 = no) (Dak Lak is base)	0.099 (0.62)	-0.207 (-1.33)	0.188 (1.56)	0.448** (3.22)	0.386** (2.73)
Household size (persons)	0.018 (0.67)	0.102*** (3.72)	0.047* (2.16)	0.103*** (3.79)	0.119*** (4.09)
Constant	-2.818** (-2.93)	-1.966 (-1.79)	-5.239*** (-6.62)	-4.919*** (-5.04)	-5.070*** (-4.97)
Observations	1,629	1,629	1,629	1,629	1,629
Pseudo R^2	0.056	0.068	0.072	0.169	0.173

* Significant at 10%.; ** Significant at 5%.; *** Significant at 1%.

Notes: Absolute value of z-statistics in parentheses

The results of our second model, also suggests that determinants from all four specified factors (e.g. interviewer and respondent characteristics; interview environment; and survey environment) significantly affect non-sampling errors (see Table 3a/3b).

Interestingly, our results suggest that if the respondent is the household head this seemingly leads to lower data quality. This is because household heads tend to be older than the mean age of respondents suggests. The mean age of non-household head respondents if 52 in Thailand, which is 10 years younger than that of household heads. In Vietnam, there is a similar gap of 11 years. Older heads, in particular, reduce the reliability of income measures gathered in the context of Vietnam. These results are consistent with the findings of Knäuper et al. (1997) and Krosnick (1991). These results do not match those of Phung et al.

(2015), who find in the 2007 and 2008 waves of TVSEP that interviewing household heads yields data of higher quality. Due to the longevity of the TVSEP, the age of household heads has grown with the panel, which in turn may suggest the decline of cognitive ability in household heads may be responsible for lower quality data in the 2017 wave.

A further aspect that captures the suitability of proxy respondents implemented in our model is the occupational background of the respondent. Results suggest that respondents who are primarily engaged in agriculture provide data of significantly higher quality across all categories of non-sampling error. It is to be expected that members of the household active in agriculture will more reliably represent households that are characterized by a high dependency on agricultural income.

While the effect of gender on non-sampling errors remains unclear, the literature, for the greater part, suggests that female interviewers are better suited in ascertaining cooperation and hence provide interviews of higher quality (Campanelli & O'Muircheartaigh, 1999). Our results mirror this for the greater part, although male interviewers in Thailand seem to provide interviews with a lower share of implausible values, whereas interviewer gender does not seem to play a significant role in Vietnam. We are able to match the findings of Phung et al. (2015) regarding the role of gender in Thailand from previous waves of TVSEP, but unable to do so for Vietnam.

Steps taken in determining the knowledge of the interviewer regarding the subject matter of the survey as captured by the interviewer's field of study and their score in an exam towards the end of the survey training, suggest that the share of non-sampling errors is far lower for economists and agriculturalists in Thailand. The score achieved in the exam also seem to be a good predictor for how well an interviewer will fare during the survey itself. While the probit model suggests that the likelihood of a non-sampling error occurring in an interviewer is higher for experienced interviewers, the OLS model suggests that the overall share of errors in questionnaires is far lower. In Thailand, experienced interviewers provide less data items prone to outliers. In Vietnam, interviewers with a background in the "survey industry" have a lower share of implausible errors. Furthermore, interviewers who previously worked in a TVSEP survey provided interviews that are more complete. A possible explanation for this could be that prior experience with the survey instrument assisted interviewers in identifying when questions were not filled in. Interestingly, this suggests that while the overall share of non-sampling errors can be reduced by hiring experienced interviewers, the likelihood of them providing erroneous data sets is higher. Less experienced interviewers lead to a lower likelihood of interviews containing errors, but those that do are far more affected. The share of non-sampling errors decreased throughout the span of the project on a weekly basis. Interviews that took place in the final weeks of the survey had the least share of non-sampling errors. This mirrors the findings of Campanelli & O'Muircheartaigh (1999); Singer et al. (1983); and Sinibaldi et al. (2009).

In terms of personality traits, we observe mixed effects. On the one hand, extraversion seems to improve cooperation and reduce non-sampling errors such as outlier and implausible values. On the other hand, extraverted interviewers have a far greater share of missing values than those who are introverted. A possible explanation could be that interviewers who are more outgoing, while achieving higher levels of cooperation, may be more sociable, thus becoming more be more distracted due to increased interactions with respondents. This may lead to a higher share of erroneously skipped items in the questionnaire.

Contrary to the assumption that respondent fatigue in panel surveys may increasingly be an issue in terms of data quality (e.g. Krosnick et al., 1999) we find that the more often a respondent participates in the survey, the higher the quality of data in Thailand. Respondents who participated in all survey waves yielded data with a lower share of missing values. This suggests that there may be a small form of panel conditioning (Lundmark & Gilljam, 2013), with long-term respondents being able to correct interviewers when data items are erroneously skipped.

Table 3a. OLS regression results: Determinants of the share of non-sampling errors (Thailand)

	(1) Missing values	(2) Refusal values	(3) Outlier values	(4) Implausible values	(5) Total erroneous values
Respondent age ≥ 60 (1 = yes, 0 = no)	0.573 (1.544)	0.728 (0.593)	0.192 (0.169)	0.050 (0.338)	0.681 (0.557)
Respondent is head of household (1 = yes, 0 = no)	0.350 (1.553)	0.688*** (0.234)	0.070 (0.135)	0.492 (0.302)	0.813* (0.431)
Respondent age (= 1) * head of household (Interaction effect)	1.400 (2.311)	-0.678 (0.809)	-0.102 (0.228)	0.131 (0.465)	0.006 (0.769)
Respondent gender (1 = male, 0 = female)	0.356 (1.374)	-0.564 (0.345)	-0.027 (0.127)	-0.055 (0.258)	0.082 (0.408)
Respondent education – secondary and higher (1 = yes, 0 = no)	1.989 (1.832)	0.298 (0.282)	0.373** (0.149)	-0.086 (0.312)	0.506 (0.521)
Respondent agreeableness	-1.274** (0.565)	0.117 (0.208)	-0.018 (0.059)	-0.081 (0.145)	-0.172 (0.255)
Respondent has agricultural background (1 = yes, 0 = no)	-4.305*** (1.389)	-1.086** (0.453)	-0.762*** (0.176)	-1.661*** (0.313)	-2.511*** (0.475)
Interviewer gender (1 = male, 0 = female)	2.511* (1.291)	-0.163 (0.315)	-0.112 (0.172)	-1.072*** (0.290)	-0.568 (0.466)
Interviewer has completed degree (1 = yes, 0 = no)	1.520 (1.188)	-0.017 (0.731)	-0.036 (0.150)	-0.470 (0.358)	-0.464 (0.523)
Interviewer from suitable field of study (1 = economics / agriculture, 0 = other)	-1.850* (1.024)	0.347 (0.696)	-0.284** (0.110)	-0.671** (0.320)	-1.077** (0.447)
Interviewer exam score (in %)	0.096*** (0.035)	-0.008 (0.020)	-0.009** (0.004)	-0.058*** (0.009)	-0.046*** (0.014)
Interviewer other household survey experience (1 = yes, 0 = no) (No experience is base)	-0.882 (1.186)	0.510 (0.486)	-0.235** (0.118)	-0.132 (0.315)	-0.208 (0.465)
Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)	-1.696 (1.476)	0.103 (0.345)	0.227 (0.208)	0.108 (0.329)	0.041 (0.470)
Interviewer openness	-1.119 (0.693)	-0.130 (0.225)	0.001 (0.047)	0.342*** (0.112)	0.077 (0.195)
Interviewer extraversion	2.762*** (0.982)	0.074 (0.239)	-0.178*** (0.065)	-0.258* (0.141)	0.194 (0.267)
Log of interview duration (minutes)	-0.225 (1.410)	-1.544** (0.765)	-0.488*** (0.180)	-0.142 (0.335)	-0.481 (0.596)
Morning interview (1 = yes, 0 = no)	-0.807 (0.983)	0.085 (0.307)	-0.011 (0.104)	0.469** (0.220)	0.253 (0.349)
Respondent participated in all waves (1 = yes, 0 = no)	-3.310** (1.573)	-0.415 (0.471)	0.168 (0.228)	-0.001 (0.396)	-0.839 (0.517)
Survey week	0.024 (0.431)	-0.089 (0.196)	-0.137*** (0.047)	-0.589*** (0.101)	-0.871*** (0.165)

Buriram province (1 = yes, 0 = no) (Nakhon Phanom is base)	-4.013* (2.363)	0.187 (0.468)	-0.253 (0.164)	0.181 (0.363)	-1.382** (0.668)
Ubon province (1 = yes, 0 = no) (Nakhon Phanom is base)	-1.671 (2.329)	0.825*** (0.258)	-0.099 (0.163)	0.025 (0.356)	-0.517 (0.681)
Household size (persons)	0.539 (0.328)	-0.119 (0.074)	-0.084*** (0.029)	-0.329*** (0.064)	-0.125 (0.110)
Constant	5.897 (10.763)	10.308*** (3.385)	6.470*** (1.116)	11.323*** (1.919)	15.301*** (3.354)
Observations	335	585	539	1,487	1,646
Adj. R^2	0.120	0.059	0.143	0.108	0.069

* Significant at 10%.; ** Significant at 5%.; *** Significant at 1%.

Notes: Robust standard errors in parentheses

Table 3b. OLS regression results: Determinants of the share of non-sampling errors (Vietnam)

	(1) Missing values	(2) Refusal values	(3) Outlier values	(4) Implausible values	(5) Total erroneous values
Respondent age ≥ 60 (1 = yes, 0 = no)	-0.293 (0.885)	-0.193 (0.375)	-0.157 (0.135)	-0.652** (0.328)	-0.832** (0.421)
Respondent is head of household (1 = yes, 0 = no)	-0.394 (0.898)	0.168 (0.271)	-0.062 (0.089)	-0.026 (0.235)	-0.114 (0.286)
Respondent age (= 1) * head of household (Interaction effect)	1.912 (1.236)	-0.032 (0.405)	0.306** (0.151)	1.010** (0.422)	1.311** (0.509)
Respondent gender (1 = male, 0 = female)	-0.321 (0.806)	-0.226 (0.198)	0.035 (0.080)	-0.236 (0.263)	-0.207 (0.296)
Respondent education – secondary and higher (1 = yes, 0 = no)	-0.207 (0.858)	0.003 (0.175)	0.045 (0.069)	0.201 (0.203)	0.254 (0.250)
Respondent agreeableness	-0.322 (0.236)	0.091 (0.113)	0.006 (0.033)	-0.076 (0.089)	-0.084 (0.104)
Respondent has agricultural background (1 = yes, 0 = no)	-2.521* (1.282)	-0.895*** (0.327)	-0.470*** (0.139)	-1.266*** (0.388)	-1.395*** (0.420)
Interviewer gender (1 = male, 0 = female)	-0.274 (0.552)	-0.119 (0.141)	0.023 (0.065)	0.810*** (0.198)	0.769*** (0.225)
Interviewer from suitable field of study (1 = economics / agriculture, 0 = other)	-0.096 (0.608)	-0.146 (0.164)	-0.066 (0.065)	0.401* (0.233)	0.457* (0.255)
Interviewer exam score (in %)	0.003 (0.021)	-0.001 (0.004)	0.001 (0.002)	0.009 (0.006)	0.010 (0.007)
Interviewer other household survey experience (1 = yes, 0 = no) (No experience is base)	-2.061 (1.308)	0.186 (0.191)	0.045 (0.099)	-0.753** (0.349)	-1.069** (0.527)
Interviewer TVSEP survey experience (1 = yes, 0 = no) (No experience is base)	-4.347*** (1.537)	0.132 (0.432)	-0.150 (0.133)	-0.384 (0.693)	-1.050 (0.774)
Interviewer openness	0.780**	-0.034	-0.019	-0.004	0.084

	(0.301)	(0.104)	(0.038)	(0.102)	(0.125)
Interviewer extraversion	0.329 (0.387)	-0.116 (0.127)	-0.036 (0.045)	-0.084 (0.112)	-0.080 (0.135)
Congruent ethnicity (1 = yes, 0 = no)	0.038 (0.804)	-0.076 (0.192)	0.150** (0.063)	-0.227 (0.194)	-0.344 (0.256)
Log of interview duration (minutes)	0.487 (0.822)	-0.618** (0.260)	-0.172 (0.109)	-0.369 (0.353)	-0.062 (0.396)
Morning interview (1 = yes, 0 = no)	-0.010 (0.603)	-0.235 (0.153)	0.119* (0.061)	-0.024 (0.161)	-0.039 (0.196)
Respondent participated in all waves (1 = yes, 0 = no)	-0.763 (1.198)	-0.232 (0.308)	-0.140 (0.137)	0.071 (0.364)	-0.012 (0.436)
Survey week	0.136 (0.224)	-0.003 (0.075)	-0.003 (0.035)	-0.351*** (0.116)	-0.458*** (0.120)
Ha Tinh province (1 = yes, 0 = no) (Dak Lak is base)	-1.154 (0.723)	-0.521*** (0.199)	-0.370*** (0.088)	-1.036*** (0.363)	-1.245*** (0.374)
Hue province (1 = yes, 0 = no) (Dak Lak is base)	1.236 (1.017)	-0.365 (0.263)	-0.265** (0.108)	-0.564** (0.248)	-0.437 (0.284)
Household size (persons)	-0.371** (0.154)	-0.006 (0.032)	-0.051*** (0.017)	-0.198*** (0.048)	-0.223*** (0.058)
Constant	2.118 (6.317)	5.841** (2.310)	2.581*** (0.723)	9.575*** (2.908)	9.043*** (3.051)
Observations	182	134	484	1,345	1,392
Adj. R^2	0.086	0.227	0.107	0.070	0.061

* Significant at 10%.; ** Significant at 5%.; *** Significant at 1%.

Notes: Robust standard errors in parentheses

6 Conclusion

In this paper, we addressed the issue of the influence of interviewer and respondent characteristics as well as characteristics of the interview and survey environment on non-sampling errors. We provide insights regarding what factors influence specific types of non-sampling error, namely: missing values, refusal values, outlier values, and implausible values. Furthermore, we measure non-sampling errors by accounting for which variables determine the overall proportion of non-sampling errors.

While our results regarding the influence of interviewer and respondent characteristics mirror the findings of the literature for the greater part, we provide novel insights regarding qualitative characteristics such as personality traits. In order for interviews to provide high-quality data on income variables, both quantitative and qualitative characteristics of interviewers and respondents must be considered. In addition, interviewer knowledge of the survey instrument and topic play an important role. While our results only find a positive effect of congruent ethnicities, matching age and gender are not found to affect data quality. The random allocation of interviewers to respondents and the largely homogenous interviewers in terms of characteristics represent a bottleneck of our analysis. An experimental approach of allocation of interviewers and respondent according to matching characteristics may facilitate future research on the influence of congruent characteristics. Contrary to the findings of Phung et al. (2015), we find that since the introduction of CAPI to TVSEP in 2016, implausible errors have become the biggest issue in the TVSEP data set. Using the 2007 and 2008 waves of panel data from TVSEP, Phung et al. (2015) had previously determined that outlier values had been the most prevalent type of non-sampling error. Plausibility rules and survey guidelines should be extended in order to further reduce non-sampling errors. We hypothesize, while the implementation of plausibility rules and an intensification of supervision in the more recent TVSEP waves has significantly reduced outlier values, data quality may be further improved by extending and optimizing automated plausibility rules.

Following our results, we recommend that household surveys take further steps to ensure that the data they gather is of sufficiently high quality to be used by policy-makers. Possible approaches to improving the quality of survey data could be in the selection of suitable respondents and well-trained interviewers. While we were able to show that the interview and survey environment also significantly affect the prevalence of flawed income data, more detailed research may yield important lessons.

We advise that respondents in household surveys are selected according to their status in the household and their knowledge of household income activities. It may be advisable to interview heads of businesses, household members in charge of financial decisions within the household and household heads in order to extract more reliable income data. However, more research is needed on transitions that may need to take place in long-term panel surveys when suitable respondents age and provide data of lower quality. Furthermore, we suggest that interviewers are selected based on knowledge of the survey subject, previous interviewer experience and on personality traits such as extraversion and agreeableness. For surveys implemented with CAPI it is recommended to develop and implement detailed plausibility ranges and validations in order to identify non-sampling errors that occur during interviews. This will ensure that complex household surveys can produce high-quality income data.

Finally, an interesting trend in the literature in determining the prevalence of non-sampling errors in survey data is to make use of so-called validation data. Such approaches compare, for example, household survey data with employer records, administrative records, previous waves of respondent's reports and similar surveys (e.g. Epland & Kirkland, 2002; Mathiowetz et al., 2002; Meyer et al., 2019). We suggest that a rapid increase of cooperation between surveys in similar contexts will be vital in reliably identifying non-sampling errors, which may otherwise remain hidden. A first important step for household surveys could be to include more a broader scope of validation linkages between individual waves of survey data to further improve the identification of non-sampling errors during data processing.

References

Adida, C. L., Ferree, K. E., Posner, D. N., & Robinson, A. L. (2016). "Who's asking? Interviewer coethnicity effects in African survey data". Comparative Political Studies, 49(12): pp. 1630-1660.

Alwin, D. F. (2007). "Margins of Error: A study of reliability in survey measurement". Hoboken, N. J.: Wiley-Interscience.

Baird, S., Hamory, J., & Miguel, E. (2008). "Tracking, attrition and data quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)". Center for International and Development Economics Research.

Banks, R., & Laurie, H. (2000). "From PAPI to CAPI: the case of the British Household Panel Survey". Social Science Computer Review, 18(4): pp. 397-406.

Bardasi E, Beegle K, Dillon A., & Serneels P. (2011). "Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania". World Bank Economic review, 25(3): pp. 418-447.

Beegle, K., Carletto, C., & Himelein, K. (2012). "Reliability of recall in agricultural data". Journal of Development Economics, Vol. 98 No. 1, pp. 34-41.

Booth, A. (2019). Measuring poverty and income distribution in Southeast Asia. Asian-Pacific Economic Literature.

Buehler, D., Rasadhika S. & Stein, W. (2019). "Personality traits in Southeast Asia - Evidence from rural Thailand and Vietnam". TVSEP Working Paper No. 006. Hannover, Germany: Leibniz Universität

Caeyers, B., Chalmers, N., & De Weerdt, J. (2012). "Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment". Journal of Development Economics, Vol. 98 No. 1, pp. 19-33.

Campanelli, P., & O'Muricheartaigh, C. (1999). "Interviewers, interviewer continuity, and panel survey nonresponse". Quality and quantity, 33(1): pp. 59-76.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). "Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior". Public Opinion Quarterly, 60(3): pp. 345-375.

Costa, J., Paul T., & McCrae, R. R. (1992). "Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual". Odessa, FL, USA: Psychological Assessment Resources

Costa, J., Paul T., & McCrae, R. R. (1997). "Personality trait structure as a human universal". American Psychologist, 52: pp. 587–596.

Couper, M., & Groves, R. M. (1992). "The role of the interviewer in survey participation". Survey Methodology, 18 (2): pp. 263-277.

Couper, M. P., & Lyberg, L. (2005, April). "The use of paradata in survey research". In Proceedings of the 55th session of the International Statistical Institute.

Cuaresma, J. C., Fengler, W., Kharas, H., Bekhtiar, K., Brottrager, M., & Hofer, M. (2018). "Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty". Palgrave Communications, 4(1): p. 29.

Dang, H. A. (2012). "Vietnam: A Widening Poverty Gap for Ethnic Minorities". In G. Hall & H. Patrinos (Eds.). Indigenous Peoples, Poverty and Development (pp. 304-343). New York, NY: Cambridge University Press.

Dang, H. A. H., & Carletto, C. (2018). "The Seemingly Underappreciated Role of Panel Data in Measuring Poverty and Economic Transformation". World Economics, 19(3), 45-60.

De Leeuw, E. D., Hox, J. J., & Snijkers, G. (1995). "The effect of computer-assisted interviewing on data quality. A review". Market Research Society. Journal., 37(4): pp. 1-19.

Epland, J., & Kirkeberg, M. I. (2002), "Comparing Norwegian Income Data in Administrative Registers With Income Data in the Survey of Living Conditions,". In: The International Conference on Improving Surveys (ICIS), Copenhagen, August 25–28, 2002.

Fisher, M., Reimer, J. J., & Carr, E.R. (2010). "Who should be interviewed in surveys of household income?". World Development, 38(7): pp. 966-973.

Fowler Jr, F. J., Fowler, J. F. J., & Mangione, T. W. (1990). "Standardized survey interviewing: Minimizing interviewer-related error" (Vol. 18). Sage.

Fowler Jr, F. J. (2013). "Survey research methods". Sage publications.

Frick, Joachim R.; Grabka, Markus M. (2014). "Missing income data in the German SOEP: Incidence, imputation and its impact on the income distribution". SOEP Survey Papers, No. 225, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.

Galesic M., & Bosnjak, M. (2009). "Effects of questionnaire length on participation and indicators of response quality in a web survey." Public opinion quarterly, 73(2): pp. 349-360.

Gibson, J. (2016). "Poverty measurement: we know less than policy makers realize". Asia & the Pacific Policy Studies, 3(3): pp. 430-442.

Groves, R. M. (1989). "Survey errors and survey costs". New York, NY: Wiley-Interscience.

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009) "Survey methodology (2nd)". Hoboken, NJ: John Wiley and Sons.

Hardeweg, B., Klasen, S., & Waibel, H. (2013). "Establishing a database for vulnerability assessment". In Vulnerability to Poverty (pp. 50-79). Palgrave Macmillan, London.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2011). "The effect of interviewer personality, skills and attitudes on respondent co-operation with face-to-face surveys". (No. 2011-14). ISER Working Paper Series.

Jäckle, A., P. Lynn, J. Sinibaldi and S. Tipping. (2013). "The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys". Survey Research Methods 7(1): pp. 1–15.

Jolliffe, D., Lanjouw, P., Chen, S., Kraay, A., Meyer, C., Negre, M., Prydz, E., Vakis, R., & Wethli, K. (2015). "A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals". The World Bank, Washington.

Kahn, R L., & Cannell C. F. (1957). "The dynamics of interviewing". New York, NY: John Wiley and Sons.

Knäuper, B., Belli, R. F., Hill, D. H., & Herzog, A. R. (1997). "Question difficulty and respondents' cognitive ability: The effect on data quality". Journal of Official Statistics-Stockholm, 13: pp. 181-199.

Knäuper, B. (1999). "The impact of age and education on response order effects in attitude measurement". Public Opinion Quarterly, 63: pp. 347-370.

Kreuter, F., Couper, M., & Lyberg, L. (2010). "The use of paradata to monitor and manage survey data collection". In Proceedings of the joint statistical meetings, American Statistical Association (pp. 282-296). Alexandria, VA: American Statistical Association.

Krosnick, J. A. (1991). "Response strategies for coping with the cognitive demands of attitude measures in surveys". Applied Cognitive Psychology, 5: pp. 213-236.

Krosnick, J. A. (1999). "Survey Research.". Annual Review of Psychology, 50: pp. 537-67.

Krumpal, I. (2013). "Determinants of social desirability bias in sensitive surveys: a literature review". Quality & Quantity, 47(4): pp. 2025-2047.

Liebenehm, S., Degener, N., & Strobl, E. (2018). "Rainfall shocks and risk aversion: Evidence from Southeast Asia". (TVSEP Working Paper No. No. 006). Hannover, Germany: Leibniz Universität

Lessler, J. T., & Kalsbeek, W. (1992). "Nonresponse errors in surveys". New York, NY: John Wiley and Sons.

Lundmark, S., & Gilljam, M. (2013). "Participation effects in panel surveys: evidence from a seven-wave randomized experiment". American Association of Public Opinion Research (AAPOR), Boston.

Lupu, N., & Michelitch, K. (2018). "Advances in Survey Methods for the Developing World". Annual Review of Political Science, Vol. 21, pp. 195-214.

Lynn, P., & Clarke, P. (2002). "Separating refusal bias and non-contact bias: evidence from UK national surveys". Journal of the Royal Statistical Society: Series D (The Statistician), 51(3): pp. 319-333.

Mathiowetz, N. A., Brown, C., & Bound, J. (2002). "Measurement Error in Surveys of the Low-Income Population". In Studies of Welfare Populations: Data collection and research issues, eds. Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro: pp. 157-194. Washington, DC: National Academy Press.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. Journal of Economic Perspectives, 29(4), 199-226.

Meyer, B. D., Mittag, N., & Goerge, R. M. (2018). "Errors in survey reporting and imputation and their effects on estimates of food stamp program participation". (No. w25143). National Bureau of Economic Research.

Meyer, B. D., & Mittag, N. (2019). "Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net". American Economic Journal: Applied Economics, 11(2): pp. 176-204.

Moore, J. C. (1988). "Self/proxy response status and survey response quality". Journal of Official Statistics, 4(2): pp. 155-172.

Moore, J. C., Stinson, L. L., & Welniak, E. J. (2000). "Income Measurement Error in Surveys: A Review". Journal of Official Statistics, Vol. 16 (4): pp. 331-361.

Nicoletti, C., Peracchi, F. & Foliano, F. (2011). "Estimating Income Poverty in the Presence of Missing Data and Measurement Error". Journal of Business & Economic Statistics, 29 (1): pp. 61-72.

Olson, K. & Bilgen, I. (2011). "The role of interviewer experience on acquiescence". Public Opinion Quarterly, 75(1): pp. 99-114.

Olson, K., Kirchner, A., & Smyth, J. (2016). "Do interviewers with high cooperation rates behave differently? Interviewer cooperation rates and interview behaviors". Survey Practice, 9(2).

Phung, T. D., Hardeweg, B., Praneetvatakul, S., & Waibel, H. (2015). "Non-sampling error and data quality: What can we learn from surveys to collect data for vulnerability measurements?". World Development, 71: pp. 25-35.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., Dabalen, A. L. (2015). "Data deprivation: another deprivation to end". Policy Research working paper, no. WPS 7252. Washington, D.C.: World Bank Group

Sinibaldi, J., Jäckle, A., Tipping, S., & Lynn, P. (2009). "Interviewer characteristics, their doorstep behaviour, and survey co-operation". American Association for Public Opinion Research (AAPOR) Conference. pp. 5955-69. Hollywood, Florida: American Statistical Association, Survey Methods Research Section.

Singer, E., Frankel, M., & Glassman, M. (1983). "The effect of interviewer characteristics and expectation on response". Public Opinion Quarterly, 47 (1): pp. 68-83.

Smith T.W. (1997). "The impact of the presence of others on a respondent's answers to questions". International Journal of Public Opinion Resarch, 9(1): pp. 33–47.

Squires, J. E., Hutchinson, A. M., Bostrom, A. M., Deis, K., Norton, P. G., Cummings, G. G., & Estabrooks, C. A. (2012). "A data quality control program for computer-assisted personal interviews". Nursing research and practice, 2012.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). "Improving survey response: Lessons learned from the European Social Survey". Chichester: John Wiley & Sons, Ltd.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). "The Psychology of Survey Response". New York: Cambridge University Press.

Townsend, R. M., Sakunthasathien, S., & Jordan, R. (2013). "Chronicles from the field: The Townsend Thai Project". Cambridge. MA: MIT Press.

Watson, N., & Li, N. (2016). "Evaluating potential improvements to the income imputation methods for HILDA Survey". HILDA Discussion Paper No. 1/16: Melbourne Institute for Applied Economic and Social Research, University of Melbourne.

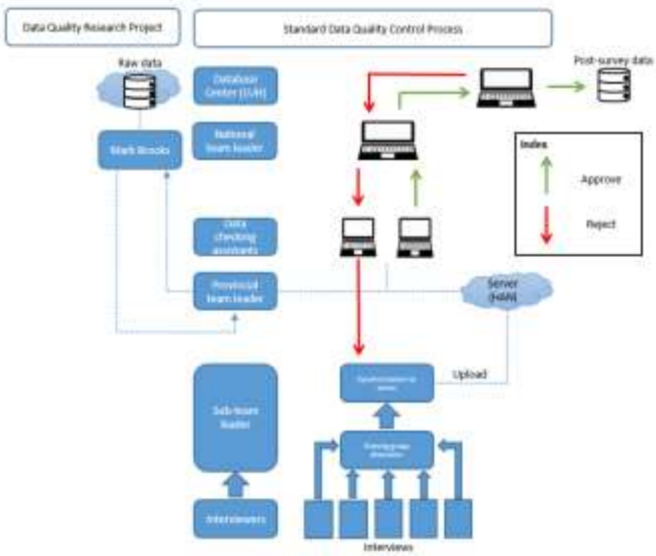
Weir, D., Faul, J., & Langa, K. (2011). "Proxy interviews and bias in the distribution of cognitive abilities due to non-response in longitudinal studies: a comparison of HRS and ELSA". Longitudinal and life course studies, 2(2): pp. 170-184.

Weisberg, H.F. (2005). "The total survey error approach: A guide to the new science of survey research". Chicago, IL: University of Chicago Press.

West, B. T., & Blom, A. G. (2017). "Explaining interviewer effects: A research synthesis". Journal of Survey Statistics and Methodology, 5(2): pp. 175-211.

Appendix

Figure 2.
Survey data structure TVSEP 2017 – Example with one team



Source: Own representation