Thailand Vietnam Socio Economic Panel

Comprehensive data quality studies as a component of poverty assessments

Mark Brooks, Rattiya S. Lippe, Hermann Waibel

Leibniz Universität Hannover

2020

TVSEP Working Paper

WP-019

Gefördert durch DFG Deutsche Forschungsgemeinschaft

Leibniz Universität Hannover

Georg-August-Universität Göttingen

The Thailand Vietnam Socio Economic Panel (TVSEP) is a long-term research project for data collection and scientific analysis funded by the Deutsche Forschungsgemeinschaft (DFG). The high quality panel data sets have been collected in rural areas of Thailand and Vietnam since 2007. The panels cover many important aspects of living standard measures, demography, geography, shocks, risks, expectations and behavioural traits of individuals. TVSEP data is available for scientific research purposes free of charge.

The TVSEP Working Papers disseminate the findings of work in progress of researchers who use the TVSEP database. Topics cover a broad range related to different fields of development economics. Papers are available online and can be downloaded at: https://www.tvsep.de/workingpapers

Thailand Vietnam Socio Economic Panel (TVSEP)
TVSEP Database Centre
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover, Germany

Contact: thanh.nguyen@iuw.uni-hannover.de

Comprehensive data quality studies as a component of poverty assessments*

Mark Brooks[a, b], Rattiya S. Lippe[a], Hermann Waibel[a]

June 2020

Abstract

Realistic poverty assessments necessitate high-quality household survey data. Such data provide the foundation for designing sound policies to sustainably reduce poverty. Despite of this, welfare measures from household surveys are often plagued by non-sampling errors in the form of non-response and measurement error. Current research, while generating important lessons, is often limited in scope and the majority of studies on determinants of data quality deal with quantifiable interviewer and respondent characteristics. A comprehensive study on data quality of an ongoing long-term household panel survey in Thailand and Vietnam is presented in this paper. Determinants drawn from respondent and interviewer characteristics, the interview and survey environment and interview paradata are found to have a significant effect on the overall quality of income-related data. We suggest that survey managers utilizing computerized questionnaires further develop and optimize validation and plausibility guidelines in order to minimize non-sampling errors. Furthermore, referring to validation data (e.g. from administrative records) during data processing is likely to be a promising approach in improving the identification of such errors.

Keywords: Non-sampling error, Paradata, Household survey, Poverty assessments, Southeast Asia, TVSEP

JEL: C83, I3

RePEc:tvs:wp-019

* An earlier version of this paper was presented at the Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World" in Washington, DC, November 7-8, 2019 (Brooks et al., 2019).

[a] Faculty of Economics and Management, Institute of Development and Agricultural Economics, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany

[b] Corresponding author, Tel.: +49 511-762-5414, Fax: +49 511 -762-2667, E-mail: brooks@ifgb.uni-hannover.de

1. Introduction

Globally the share of households/individuals living in extreme poverty has been declining steadily with extreme poverty falling from 36 percent in 1995 to approximately 9 percent in 2017 (Cuaresma et al., 2018). This seemingly indicates that poverty reduction policies have been a worldwide success. However, recent World Bank disaggregated statistics report that poverty has decreased significantly in all but one region. In Sub-Saharan Africa extreme poverty rose by almost 50% to 413 million individuals in 2015 when compared to statistics from 1990 (Barne & Wadhwa, 2018). Furthermore, forecasts indicate that progress made toward achieving the 2030 goal of eradicating poverty will not suffice with almost 500 million people being predicted to remain in extreme poverty in 2030 (See appendix: Figure 3).

The identification of those who can be characterized as living in extreme poverty is frequently based on household survey data. In particular, data on household wealth drawn from household surveys are the foundation for research on poverty. Income, consumption and assets being used to analyze shock coping and risk mitigation strategies as well as poverty dynamics and transitions. Accordingly, improving the reliability and validity of survey data is a necessary complementary task (c.f. Squires et al., 2019) – not only to ensure that poverty is correctly measured, but also to ensure that no cases near the poverty line are left behind and erroneously classified as non-poor. However, such data have been found to be susceptible to non-response and measurement error, which stems from their sensitive nature and respondent behavior, which frequently results in under-/overestimation of household wealth (Frick & Grabka, 2014; Meyer et al., 2015; Meyer et al., 2018; Moore et al., 2000; Nicoletti et al., 2011; Parvathi & Nguyen, 2018; Watson & Li, 2016).

Generally, in survey methodology, errors refer to deviations of obtained values from the true value of measurement (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 survey design (e.g. Groves, 1989). One can differentiate between two types of non-sampling error: non-response and measurement error. Non-response can occur at a unit-level (e.g. if a sample case cannot or refuses to be contacted) or at an item level (e.g. missing values of individual data items). These missing data can reduce the representativeness of a survey, whereas measurement errors (e.g. implausible values) can reduce their validity. Accordingly, shortcomings of data on household wealth can significantly affect estimations of poverty and lead to misguided policy conclusions.

A further constraint lies in the suitability of surveys for use in poverty assessments - while there is an abundance of household surveys in some countries, there exist to date relatively few that are suitable for calculating reliable poverty estimates (Booth, 2019; Gibson, 2016; Jolliffe et al., 2015). Furthermore, the quantity of high-quality survey data in developing countries remains sparse (Booth, 2019; Dang & Carletto, 2018). A study by Serajuddin et al. (2015) established that for 57 countries one poverty estimate at most could be calculated for the period of 2002-2011.

Innovations to survey tools utilized in developing countries such as Computer Assisted Personal Interviews (CAPI) have increased the overall effectiveness of data collection. The World Bank, for example, by designing and implementing the android-based software "Survey Solutions" (https://mysurvey.solutions) has made an important step in ensuring that surveys based in developing countries are able to develop the capacities needed to procure high quality data in a timely manner. Experimental evidence suggests that conducting surveys with CAPI over the more traditional Paper and Pencil Interviews can prevent numerous errors due to the implementation of; for example, automated routing and plausibility checks (e.g. Banks & Laurie, 2000; Caeyers et al., 2012; de Leeuw et al., 1995). In spite of advances made in terms of data access, use, and collection, innovations to CAPI and other data collection methods do not automatically solve the problem of low-quality data (Meyer et al., 2015). Issues such as a lack of interviewer consistency, mis-interpretation of questions, and difficult interview and survey conditions remain a challenge especially in developing countries (Lupu & Michelitch, 2018). Furthermore, to date few studies present empirical evidence on the determinants and impact of non-sampling errors in surveys based in developing countries.

There are at least four shortcomings in the literature on data quality of household surveys in developing countries. Firstly, most studies rely on either cross-sectional or experimental data and are therefore limited in scope as regards the types of non-sampling errors studied. Secondly, emphasis has been placed on quantitative interviewer and respondent characteristics such as age, gender and education, whereas qualitative information such as interviewer/respondent behavior, their personality traits, and motivations are seldom considered. Thirdly, most studies focus on individual determinants of data quality such as the effects of the interviewer and/or respondent characteristics, but rarely account for the circumstances during the interview or survey itself (Lupu & Michelitch, 2018). The paper by Phung et al. (2015) accounts for circumstances such as the period of day or season during which the interview took place. Finally, survey paradata[1] are infrequently available to researchers as survey providers seldom collect data about the survey process such as interview interruptions or other unexpected events. 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 and measurement errors prevalent in data items on household wealth and poverty. Our study adds to the literature by accounting for a comprehensive set of quantitative and qualitative interviewer and respondent characteristics based on household panel survey data in two countries in Southeast Asia. We find that interviewer and respondent characteristics are the main drivers of non-sampling errors. Results suggest that characteristics such as interviewer experience, personality traits and the status of the respondent within the household significantly affect data quality. Furthermore, the progression of data collection activities and the presence of others significantly influenced the quality of interviews. Measurement errors were identified as representing the greatest threat to high-quality household surveys on income utilizing CAPI.

The outline of the paper is as follows: In Section 2, we provide a brief overview of indicators of data quality before we formulate a model to determine factors influencing the prevalence of non-sampling errors in variables pertaining to income in Section 3. The household survey data used in our analysis is described in the following section. Section 5 summarizes our main results and discusses how individual determinants may influence the prevalence of non-sampling errors. In the final section, conclusions are drawn and recommendations of methods to improve household survey data quality are offered.

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 error (e.g. Groves, 1989; Lessler & Kalsbeek, 1992; Weisberg, 2005). Such errors transpire throughout the course of survey design, data collection and post-survey data processing.

Coverage errors occur should the survey sample not be representative of the target population. Accordingly, coverage errors are grounded in the survey design phase and are the result of, for example, faulty information on the likelihood of the inclusion of a sampling unit in the sample. Such errors are the result of over- or under-coverage of sampling units, such as: (i) inclusion of incorrect sampling units in the sampling frame; (ii) exclusion of important sampling from the sampling frame; or (iii) duplication of sampling units within the frame (Groves et al., 2009).

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

Non-response occurs when measurements cannot be obtained for a sampling unit or for a specific item of the survey instrument (Fowler, 2013). Unit non-response refers to sampling units that are unavailable for interviewing during survey data collection. For example, a respondent may not be available when the survey team requests an interview or may refuse to participate (Lynn & Clarke, 2002). Item non-response, in contrast to unit non-response, occurs when responses to the sampling unit are only available partially. Erroneous skipped questions, the inability of respondents to recall necessary information of past events, or refusals on behalf on the respondent are examples for causes of item non-response.

Measurement error occurs if the value provided by the sample unit deviates from the true value of measurement. There are three types of measurement error: response, interviewer and post-survey errors (Weisberg, 2005). A respondent being unable to recall accurately the amount of fertilizer input used during the reference period throughout the response process can lead to measurement error. Further, misinterpretation of the intent behind a question or inaccurate reproduction of question text represents a source of such errors. Post-survey errors arise, for example in PAPI surveys, from entry errors when the paper questionnaire is processed and input to a database by survey staff.

In the context of this paper, the research objective is to examine the determinants of non-sampling errors that occur during the interview procedure in order to provide practical recommendations to survey providers in the context of developing countries. Hence, we omit coverage errors as these are attributed to the survey design phase. Furthermore, we omit unit non-response, which commonly occurs prior to the selection of a respondent and the initiation of the interview process. Unit non-response frequently materializes from households refusing to participate before a respondent can be identified, or due to the household in question being unavailable for an interview (e.g. interviewer cannot contact or locate household), or due to a household no longer being eligible for interviewing (Weisberg, 2005).

A non-exhaustive overview of the diverse sources of non-sampling errors in surveys is provided in the following paragraphs.

Many non-sampling errors transpire during the process of respondents formulating responses to survey items (Tourangeau et al., 2000). Several problems can occur ranging from misinterpretation of the contents or intent of survey items to deliberate misreporting. Furthermore, recall bias is a prevalent source of non-sampling errors, which occurs when a respondent's judgement or the method applied to provide an estimated response is flawed. Long periods of recall in questionnaire design frequently lead to such errors and they must be accounted for in particular in the context of agricultural information (e.g. Beegle et al., 2012). Respondents answering in proxy and hence being probed in detail about the characteristics and activities of other household members have been found to introduce non-sampling errors (e.g. Alwin, 2007; Bardasi et al., 2011; Stoop et al., 2010). In spite of these findings, surveys in developing countries generally target household heads as proxy respondents and studies on data quality provide mixed results. On the one hand, household heads were found to provide interviews with fewer missing values (e.g. Phung et al., 2015). On the other hand, household heads have been found to underestimate the income of other household members (e.g. Fisher et al., 2010). Further, the cognitive ability of respondents, as proxied by their age and level of education, has been found to significantly affect the quality of data (e.g. Nguyen & Nguyen, 2020). Accordingly, respondents with lesser cognitive ability, e.g. with a low level of education or elderly household heads, have been found to satisfice more frequently during interviews (e.g. Knäuper et al., 1997; Knäuper, 1999; Krosnick, 1991). Satisficing refers to the decision of a respondent to utilize short-cuts in order to minimize effort required to formulate a response. This stems from the notion that merely providing satisfactory rather than optimal answers to survey items is sufficient. Examples for such behavior are straight-lining (e.g. always choosing the first response option), answering at random, or selecting "don't know" rather than exerting effort to answer survey items (Barge & Gehlbach, 2012).

The role of the interviewer in obtaining high quality data is extensively researched and well documented in the literature. Deviations from survey procedures, as defined and enforced by survey organizers, by the interviewer can influence the respondent's answers. Rephrasing questions, neglecting to follow interview instructions, skipping certain questions to reduce workload or due to perceived sensitivity of a question topic, and recording a response that does not match with the respondent's (e.g. mechanical or response errors) are prominent examples of the influence of the interviewer. Faulty methods of enumeration or recording errors can be the result of insufficient training or a lack of experience in conducting interviews (e.g. Campanelli & O'Muircheartaigh, 1999; Singer et al., 1983; Sinibaldi et al., 2009). Providing direct assistance to the respondent regarding difficult questions and/or applying incorrect probing techniques in order to obtain responses can lead to further errors. Minute variations in the emphasis or intonation of parts of a question can further influence the respondent (Groves, 2009). Prior experience of interviewers in survey activities can provide basic knowledge on interviewing behaviors that elicit cooperation, which is a prerequisite for obtaining accurate responses (Couper & Groves, 1992). However, some studies suggest that interviewers without an extensive survey background may provide data of higher quality as they are less likely to be biased due to prior experience in other surveys and hence more likely to follow closely existing survey guidelines (e.g. Fowler & Mangione, 1990; Fowler, 2013; Sinibaldi et al., 2009). The findings of the literature on the effect of interviewer gender on data quality are inconsistent. For example, Campanelli & O'Muircheartaigh (1999) find that male interviewers collect data of poorer quality, whereas Phung et al. (2015) observe the opposite. Character traits, such as displaying friendly or motivating behaviors, often lead to higher rates of cooperation and accordingly provide data of higher quality (Jäckle et al., 2013; Olson et al., 2016). Interviewers, who are well trained are generally not hypothesized and have not been found to be the main source of non-sampling errors. In practice, however, it is not always possible to give sufficient training to interviewers and exercise enough supervision, due to lack of budget and time (Weisberg, 2005).

In general, an interview can be defined as a structured social interaction during which the demographic and socio-economic characteristics of the interviewer and respondent interact and play an important role in determining the quality of data collected (Kahn & Cannell, 1957). Accordingly, the effect of congruent characteristics such as age, gender, and ethnicity are of particular interest (e.g. Baird et al., 2008; Phung et al., 2015). Interviews on sensitive topics yielded reliable data when respondent and interviewer characteristics were congruent in a study by Catania et al. (1996). Furthermore, congruency of ethnicity has been found to significantly affect interactions in survey populations that capture a significant share of minority groups that differ greatly in terms of culture or language from ethnic majority groups.

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 others 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 non-sampling error analysis. 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 to underlying issues in the behavior of interviewers or in the survey instrument being identified too slowly and increase their impact on data quality. A lack of monitoring activities during the survey is also a factor that can reduce the quality of data (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 characteristics, (ii) respondent characteristics, (iii) congruency of aforementioned characteristics, (iv) the interview environment and (v) 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 on income. 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 in the context of most surveys. Survey research considers that while survey data items are largely plausible, there are several potential 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 (iv) interview and (v) survey environment (c.f. Figure 1).

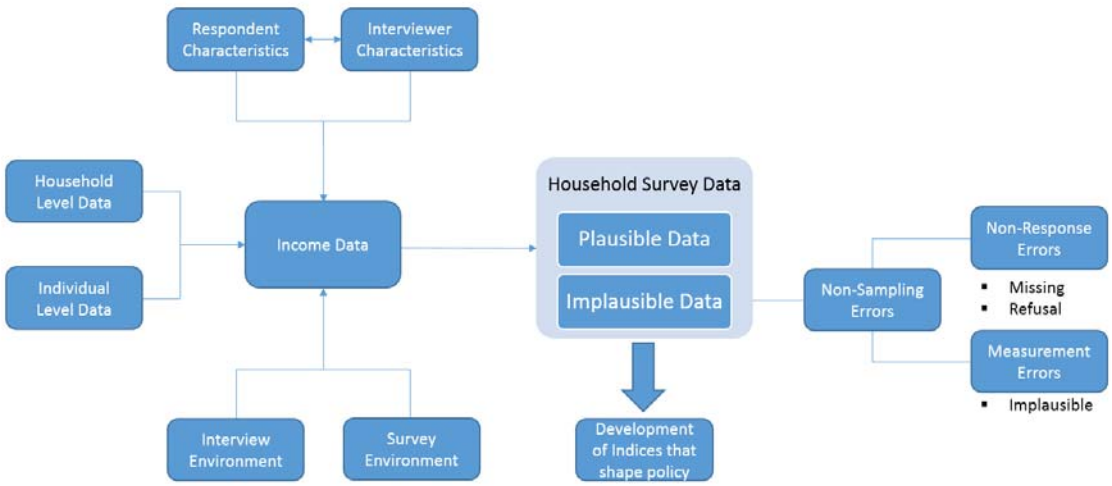

Figure 1. Conceptual framework

In our analysis, we differentiate between two types of item non-response errors: missing values and refusal values. Missing values occur when a question remains unanswered or is erroneously skipped. This generally would occur in the context of computer assisted personal interviews due to mechanical errors either by the interviewer or in the implementation of the computerized questionnaire. 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. For numerical values, interviewers were trained to select a specific number combination when the respondent is identified as being unwilling to provide an answer. The survey that this study is based on, applies a code specified as "98 – no answer" in order to identify potential refusal cases. Interviewers are instructed to make deliberate, but cautious, use of this code for situations in which the respondent seemingly refuses to provide an answer in spite of careful probing. Measurement errors are identified as implausible values, which 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. Mechanical errors such as the mistaken addition or deletion of a digit in terms of a monetary value are also identified as implausible values, as are extreme values that could not plausibly be explained. Such extreme outliers were defined as implausible when taking into account, for example, local market prices, responses from previous survey waves and by means of callback to the household.

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 Figure 1.

To better account for the prevalence of questionnaires not afflicted by non-sampling errors we apply a negative binomial regression, which is an approach generally used for zero-inflated count data, which is applicable for the analysis of errors within survey instruments. We establish a general model with a set of explanatory variables, which is applied to four dependent variables based on the count of non-sampling errors in income relevant variables by interview: (a) missing values; (b) refusal values; (c) implausible values; and (d) total erroneous values. Our individual error regressions are first run as a whole for the combined sample of households and then re-run separately for two countries in South East Asia, namely Thailand and Vietnam. The goal is to determine whether select determinants are consistent in terms of their effects on interview data quality and are applicable to multiple cultural contexts or whether survey providers must themselves identify key determinants of survey error on a case-by-case basis.

The model applies a negative binomial regression model to capture the impact of interviewer and respondent characteristics, the interview environment and survey environment on each of the dependent count variables of non-sampling errors. The model specification is as follows:

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

where Y_{in} represents the overall count of non-sampling errors in income relevant survey items by error type in the context of an interview. 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; and S_{pi} are the characteristics of the survey environment. We then repeat the model specifications on a country-level albeit having slight variations in terms of the independent variables selected. Independent variables differ based on known differences between the survey population, interviewers and cultural differences between Thailand and Vietnam.

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 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, and their status within the household are also captured. In general, older respondents are found to provide 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. The literature suggests that a common practice in household surveys in developing countries is 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 in comparison of alternative 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). We further hypothesis that respondents who are higher in the hierarchy of the household and are either heads or control household finance will be able to provide highly accurate income data.

Among congruent characteristics of interviewers and respondents, Baird et al. (2008) and Phung et al. (2015) suggest that congruent age can significantly improve the extent 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 improved 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). However, the average amount of time taken to answer survey items is likely also a good predictor of interview data quality and may capture fabrications by the interviewer (e.g. Couper & Hansen, 2002; Couper & Kreuter, 2013; Couper et al., 2010). Additionally, the presence of other household- and non- household members is likely to influence the interaction quality during the interview due to concerns of privacy, which is further exacerbated by the sensitive nature of income data (e.g. Grabka & Frick, 2007).

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 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 was first conducted in 6 provinces of Thailand and Vietnam in 2007 and is ongoing. To date eight waves of data have been collected.

The survey instrument follows the typical 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 providing data for 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 influence of supervision and without prior on-site cleaning (see Appendix: Figure 4). 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 were generated throughout several stages of the survey (see Figure 2). 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. 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).

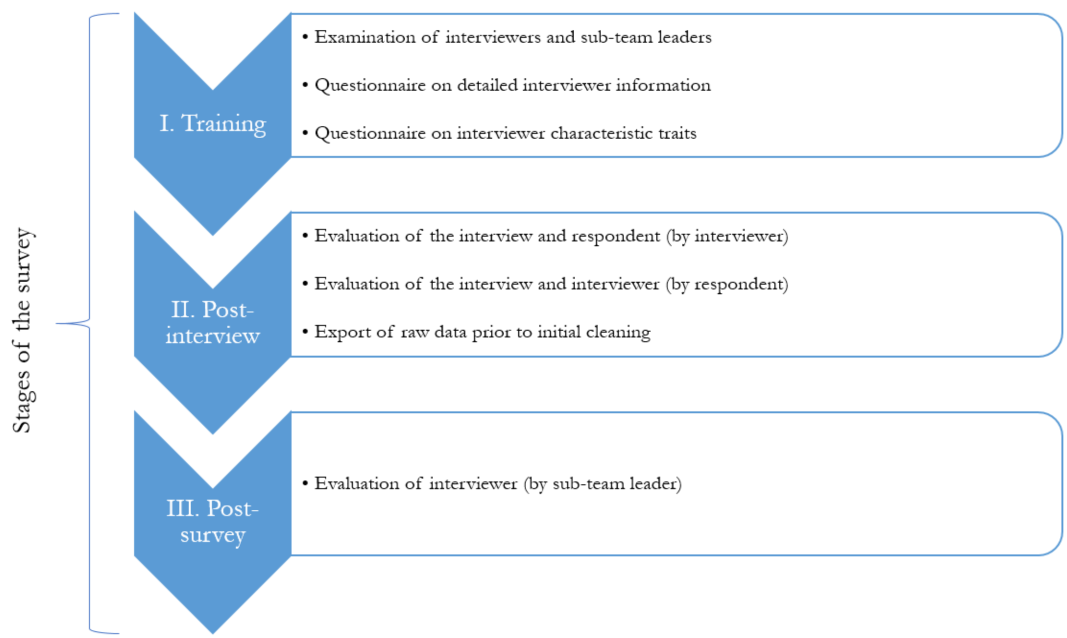

Figure 2. Data quality modules of TVSEP 2017

Following survey documentation and guidelines, errors in the variables of interest for the income aggregate were identified and categorized according to our four 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 of non-sampling errors during the interview process.

5. Results

In Table 1, the descriptive statistics of the determinants applied in our analysis are presented. The mean count of total non-sampling errors ranges between 6.60 in Thailand and 6.58 in Vietnam (or ~6.4% of income variables in Thailand and 3.5% in Vietnam). The standard deviation of total erroneous values is lower in Vietnam and 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 interviews or ~9% of all interviews and 237 interviews or ~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 is generally lower. The majority of flagged issues stem from income data related to the agriculture section.

The models' independent variables are selected based on differences between the survey population of Thailand and Vietnam. Furthermore, variables are included based on differences in the hiring process of interviewers between both countries. Hence, some variables are expected to have an effect in Vietnamese model, but not in the Thai model. In the following, explanations for country-level differences to our model are given. For example, the impact of congruent ethnicity is hypothesized to not play a significant role in the Thailand data set because ethnic diversity among respondents is minimal. The non-Thai ethnic minority groups in Northeast Thailand are generally 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 ethnic groups are less homogenous (Dang, 2012). The characteristics of interviewers hired also differed between the two countries. Interviewers in Thailand were mainly hired from local universities. In Vietnam, interviewers are often freelancers with a background in the "survey industry" and who have all completed at least a bachelor degree. Interestingly, almost all Vietnamese interviewers have previous experience working in other surveys. Hence, we capture interviewer experience based on their participation in prior waves of TVSEP. Interviewer age does not vary greatly in both countries with interviewers generally being in their mid-twenties to early thirties.

The majority of respondents are older than 50 years of age, with the mean age of respondents being slightly lower in Vietnam. It is also noticeable that a high share of respondents are household heads in both countries (~60%). When accounting for respondents who are both heads of their household and in control of the financial affairs TVSEP was able to conduct interviews with almost 40% in Thailand and 50% in Vietnam. The majority of income data stems from agriculture related activities. Accordingly, respondents' main or secondary occupation is often based in agricultural work. 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 previous 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. Overall, respondents are predominantly male 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%).

The average interview time during the 2017 survey was slightly under three hours in Thailand and almost four and a half hours in Vietnam. Accordingly, the number of survey items answered per minute is approximately ten in Thailand and almost seven per minute in Vietnam. The difference in entry speed is driven by thirteen outlier cases in which interviews were completed within one hour. In Vietnam, only one interview was completed within one hour. The majority of interviews followed the standard procedures of the survey (e.g. morning/afternoon sessions), with only 1% of interviews being conducted during the more inconvenient evening interview sessions. Data collection went on for five weeks in Thailand, whereas in Vietnam the survey lasted 6 weeks.

Table 1. Summary statistics of determinants (Thailand/Vietnam)

Variables	Obs.		Mean		Std. dev.	
	TH	VN	TH	VN	TH	VN
Dependent variables *						
Missing values	1,818	1,629	1.14	0.54	7.83	2.72
Refusal values	1,818	1,629	0.68	0.13	2.73	0.54
Implausible values	1,818	1,629	4.78	5.90	5.72	7.20
Total erroneous values	1,818	1,629	6.60	6.58	10.16	8.17
Independent variables						
Respondent age	1,818	1,629	57.85	52.75	12.76	13.89
Respondent gender (1 = male, 0 = female)	1,818	1,629	0.34	0.44	0.47	0.50
Respondent ethnicity (1 = Thai, 0 = other)	1,818	1,629	0.97	0.78	0.18	0.42
Respondent years of schooling	1,818	1,629	5.42	6.90	2.96	3.70
Respondent is head of household and financial affairs (1 = yes, 0 = no)	1,818	1,629	0.39	0.49	0.49	0.50
Interviewer gender (1 = male, 0 = female)	31	42	0.23	0.40	0.43	0.50
Interviewer ethnicity (1 = Thai, 0 = minority)	31	42	0.97	0.98	0.18	0.15
Interviewer has completed degree (1 = yes, 0 = no)	31	42	0.87	1.00	0.34	0.00
Interviewer survey experience (1 = yes. 0 = no)	31	42	0.52	0.93	0.51	0.26
Interviewer TVSEP survey experience (1 = yes, 0 = no)	31	42	0.06	0.05	0.25	0.22
Interviewer from local province (1 = yes, 0 = no)	31	42	0.13	0.17	0.34	0.38
Interviewer extraversion	31	42	3.89	4.13	0.76	0.73
Interviewer agreeableness	31	42	4.19	6.13	1.25	0.71
Log of interview duration (minutes)	1,817	1,629	5.05	5.58	0.35	0.32
Answers per minute	1,817	1,629	10.05	6.89	3.56	2.13
Others present during interview (1 = yes, 0 = no)	1,818	1,629	0.21	0.12	0.41	0.33
Tablet malfunction occurred during interview (1 = yes, 0 = no)	1,818	1,629	0.64	0.71	0.83	0.76
Respondent participated in all waves (1 = yes, 0 = no)	1,818	1,629	0.10	0.04	0.29	0.21
Survey week	1,818	1,629	2.65	3.78	1.14	1.35

Source: Own calculations based on TVSEP survey 2017.

Note: *Of the 1,818/1,629 Thai/Vietnamese questionnaires available for the analysis, 170/237 were free of errors

Determinants of non-sampling errors are identified by applying one regression model for the combined sample and then by repeating the regression on a country-level. The major results of the combined sample model (see Table 2) generally have the expected signs and can be summarized as follows:

Firstly, in examining determinants of missing values we establish that the respondent does not play a significant role in their emergence. Rather, interviewer characteristics such as extraversion and the quality of interactions between respondent and interviewer play a key role. A possible explanation could be that in some cases interviewers who assess themselves as being outgoing do not achieve higher levels of cooperation due to being more sociable, but rather they may become more distracted during their interviews due to increased interactions with respondents. This may lead to a higher count of erroneously skipped items due to mechanical errors in their questionnaires. Further, congruency of ethnicity between respondents and interviewers significantly reduce the number of missing values, as hypothesized. An independent variable implemented in the model to attempt to capture interviewer satisficing and careless behaviors is the number of answers entered to the tablet per minute. The results suggest that the more answers that are provided within the timeframe of one minute, the higher the count of non-sampling errors. This suggests that interviewers will fall into specific patterns in order to speed up the interview process and reduce their workload, which in turn may lead to a higher likelihood of mechanical error (e.g. Olson & Peytchev, 2007). In terms of interviewer experience as captured by the survey week, we find that with increasing experience in the use of the CAPI questionnaire, the number of missing items due to mechanical errors are significantly reduced with each passing week. Hence, interviewer flaws in the form of mechanical errors are the main drivers of missing values in CAPI surveys. In the case of TVSEP, this may be caused by lax routing plausibility rules. A rule was implemented which ascertained that if all questions were answered the final page of the questionnaire would be highlighted green, whereas if a question had not been answered the final page alongside the relevant question would be highlighted red. Furthermore, section labels would follow the same guideline. Uploading, however, was still possible if questions were unanswered. Hence, an implementation of stricter plausibility checks may hinder future missing data in the context of CAPI as generally suggested by the literature (e.g. de Leeuw et al., 1995).

Secondly, refusal to provide a response to survey items by the respondent is significantly impacted by the level of education of the interviewer. Interviewers who have at least completed their bachelor degree are more easily able to elicit answers from respondents, whereas interviewers with less life and study experience have more issues in reliably obtaining information on income. Interviewers with survey experience are also more capable of reducing the likelihood of respondent refusal to answer questions related to household income. Furthermore, local interviewers reduced the frequency of refusal cases. It appears that interviewers with higher levels of experience, who are local to the survey area, are better suited to procuring an answer from unwilling respondents. According to the literature, interviewer extraversion can positively influence the cooperation of respondents during the interview (e.g. Jäckle et al., 2011; West & Blom, 2017). In our model, we find the opposite effect, namely that extraverted interviewers are suggested to provide data more prone to item non-response. Interviewers that self-assess themselves as being agreeable can be characterized as being more sympathetic, cooperative and warm. Such characteristics are found to reduce the cases of refusal in TVSEP. Tablet malfunctions, which resulted in crashes and partial recovery of previous data entry, also have explanatory value for the prevalence of refusal errors. The result of faulty hard-/software during interviews may negatively influence the attitude of the respondent, who may connect such issues with a lack of professionalism on behalf of the survey. Interviewer fatigue likely plays a role in the increasing count of refusals towards the end of the survey as the likelihood of interviewers being able to persuade respondents to respond in full will likely decrease with loss of motivation. Furthermore, the results pertaining to the effect of interviews taking place a day prior to breaks of the survey activity also suggest that interviewer fatigue is an important issues to account for in data quality, albeit the results not being significant in the combined model.

Thirdly, we find that the respondent has the greatest impact on the prevalence of implausible values in surveys on household income. Older, more senior members of households provide more accurate estimates of income, which may be explained partially by such members often being in charge of financial affairs and being household heads. This matches the findings of Phung et al. (2015), who also determine that interviews with household heads yield data of significantly higher quality. Seemingly, in the case of developing counties the argument that with increasing age, the cognitive ability of respondents decreases thus reducing the reliability of income estimates as suggested by Knäuper et al. (1997) and Krosnick (1991) cannot be confirmed in our results. The finding that respondents with higher levels of education are more likely to provide erroneous income data is mirrored by the results. Measurement errors are significantly more prevalent for respondents who have completed at least secondary school, which may be explained by the background of such individuals. Higher levels of education often signal that the individual in question is more well-off, in particular in terms of income. The literature often finds that more wealthy individuals are less likely to dispose accurate information on income. Interviews with household members who are wage- or self-employed yield data that is more accurate. This confirms our hypothesis that interviewing respondents with more knowledge of income in the context of the household will better represent the true extent of household income. We further find that positive character traits such as agreeableness reduce the number of measurement errors. Tablet malfunctions significantly affect the quality of data concerning measurement errors. Such malfunctions are often described as the main disadvantage of CAPI when compared with traditional Pen and Paper Interviews (e.g. Dale & Hagen, 2007).

When regarding non-sampling errors as a whole it seems that characteristics and traits of the respondent and interviewer are the main drivers of error.

In analyzing the individual types of non-sampling error our model suggests that there are significant differences regarding the determinants between Thailand and Vietnam, which warrants further examination by re-running the model on a country level.

Table 2. Combined sample regression results: Dependent variable, non-sampling error count

	(1) Missing values	(2) Refusal values	(3) Implausible values	(4) Erroneous values
Respondent age ≥ 60 (1 = yes, 0 = no)	0.232 (0.190)	-0.0568 (0.112)	-0.209*** (0.0469)	-0.122** (0.0557)
Respondent gender (1 = male, 0 = female)	0.0764 (0.183)	0.0237 (0.0997)	0.145*** (0.0438)	0.137*** (0.0514)
Respondent education – secondary and higher (1 = yes, 0 = no)	0.0702 (0.218)	0.0570 (0.113)	0.196*** (0.0563)	0.211*** (0.0633)
Respondent is head of household and financial affairs (1 = yes, 0 = no)	-0.351* (0.189)	-0.0978 (0.103)	-0.108** (0.0424)	-0.145*** (0.0515)
Respondent participated in all waves (1 = yes, 0 = no)	-0.324 (0.346)	-0.127 (0.160)	-0.0286 (0.0775)	-0.0983 (0.0756)
Respondent main occupation (1 = wage- or self-employed, 0 = other)	-0.0702 (0.211)	-0.0916 (0.118)	-0.258*** (0.0593)	-0.247*** (0.0580)
Interviewer gender (1 = male, 0 = female)	0.387** (0.192)	-0.153 (0.103)	0.00508 (0.0419)	0.0888* (0.0496)
Interviewer survey experience (1 = yes, 0 = no)	0.0146 (0.202)	0.225** (0.112)	-0.146*** (0.0505)	-0.0718 (0.0547)

Interviewer has completed degree (1 = yes, 0 = no)	0.391 (0.278)	-0.432*** (0.141)	0.286*** (0.0745)	0.214*** (0.0774)
Interviewer from local province (1 = yes, 0 = no)	-0.0485 (0.292)	-0.204* (0.123)	0.179*** (0.0520)	0.0756 (0.0546)
Interviewer extraversion	0.222* (0.126)	0.209*** (0.0748)	-0.0196 (0.0261)	0.0611* (0.0332)
Interviewer agreeableness	0.121 (0.111)	-0.313*** (0.0767)	-0.0561* (0.0308)	-0.0497 (0.0340)
Congruent ethnicity (1 = yes, 0 = no)	-0.507** (0.245)	-0.195 (0.153)	-0.0308 (0.0567)	-0.0688 (0.0603)
Answers per minute	0.0408 (0.0254)	0.0683*** (0.0181)	0.0173*** (0.00640)	0.0350*** (0.00836)
Tablet malfunction occurred during interview (1 = yes, 0 = no)	0.167 (0.113)	0.0964* (0.0565)	0.0547* (0.0289)	0.0856*** (0.0300)
Others present during interview (1 = yes, 0 = no)	0.184 (0.231)	0.174 (0.136)	0.113** (0.0491)	0.0760 (0.0560)
Interview took place before break day (1 = yes, 0 = no)	0.241 (0.248)	0.0497 (0.150)	0.0731 (0.0627)	0.106 (0.0686)
Survey Week	-0.155** (0.0631)	0.207*** (0.0465)	-0.112*** (0.0157)	-0.0888*** (0.0185)
Country (1 = Vietnam, 0 = Thailand)	-0.686** (0.322)	-1.164*** (0.182)	0.331*** (0.0805)	0.0910 (0.0968)
Over-dispersion factor	2.726*** (0.0647)	0.973*** (0.115)	-0.105*** (0.0398)	-0.0824** (0.0401)
Constant	-1.630** (0.762)	-0.724 (0.465)	1.864*** (0.197)	1.617*** (0.264)
Observations	3,446	3,446	3,446	3,446
Chi^2	56.98	291.0	236.6	160.7
Degrees of freedom	19	19	19	19
Adjusted R^2	0.02	0.08	0.01	0.01

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

Notes: Robust standard errors in parentheses

The separation of the combined sample at the country-level is applied in order to determine whether there are differences based on different cultural contexts (see Table 3a/b). While the survey instrument applied in both countries is identical, there are small differences in the implementation of the survey. For example, the interviewers hired in Vietnam differ in terms of their characteristics when compared with those hired in Thailand. The Thai interviewers can be characterized as being comprised of younger inexperienced students, whereas the Vietnamese interviewers are from a more professional background and many are experienced full-time interviewers. Accordingly, a comparison at a country-level seems suitable in order to determine if the results of the combined model are consistent at a country-level.

Regarding missing values, the results for Thailand are widely consistent between the combined model and the country-level analysis. In Thailand, an interestingly the results suggest that respondents who have participated in all prior waves of the TVSEP survey may have a positive effect in that they reduce the number of item level missing data. This could be due to such respondents being aware of the general structure of the questionnaire and hence being more aware of when an interviewer implausibly skips a question that must be answered. In Vietnam, missing values seem to be explained mainly by previous interviewer experience in surveys and CAPI-tools as well as by congruency of ethnicity. For both countries, the regression mirrors the finding that interviewers most likely drive missing values and that this is likely due to mechanical errors on their behalf.

The results on refusal to provide data on income at the country-level mirror the combined results in that more positive interview interactions reduce the number of refusal cases and disturbances during the interview such as tablet malfunctions or careless interview behavior lead to an increase of refusal cases. This could be because respondents feel that they are not being taken seriously and become fatigued in light of technical issues during the interview. In Vietnam, the presence of others during the interview has a large impact on the count of refusal cases suggesting that it may be beneficial to further ensure confidentiality for the respondent in particular in Vietnam. Furthermore, in later survey weeks, there seems to be a potential effect of interviewer fatigue, which seemingly manifests in a drop in the capability of eliciting respondent cooperation.

Implausible value determinants are mostly consistent across both countries and the effect of respondent status within the household on the reliability of income data emphasizes the importance of interviewing household heads as found by Phung et al. (2015). While the effect of interviewer gender on non-sampling errors remains unclear in our results, the literature, for the greater part, suggests that female interviewers are better suited in ascertaining cooperation and hence provide interviews of higher quality (e.g. Campanelli & O'Muircheartaigh, 1999). Our results for Vietnam mirror this for the greater part, although male interviewers in Thailand seem to provide interviews with a lower count of implausible values. The latter is consistent with the findings of Phung et al. (2015) regarding the role of gender in Thailand from previous waves of TVSEP. When examining the impact of survey experience, we can only find significant results in Thailand. This suggests that the more professional interviewers in Vietnam 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, which may reduce the impact of interviewer experience in Vietnam. This is in line with the findings of Fowler & Mangione (1990); Fowler (2013); and Sinibaldi et al. (2009). As almost all interviewers in Vietnam have a full-time job in the field of interviewing, a variable capturing prior experience with the TVSEP survey in particular is added to the country-level regression. In Thailand prior experience as an interviewer for the less professional set of interviewers significantly improves the quality of data whereas there is no significant effect for interviewers who have a more professional background in Vietnam. The number of measurement 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 lowest count of implausible values. 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). Regarding the country-level regressions, this suggests and is mirrored in the results that the effect of congruent ethnicity is far greater in Vietnam than in Thailand. This mirrors the findings of Campanelli & O'Muircheartaigh (1999); Singer et al. (1983); and Sinibaldi et al. (2009). Finally, we are neither able to find results that suggest an effect of respondent fatigue (e.g. Krosnick et al., 1999) in the panel, nor of panel conditioning (e.g. Lundmark & Gilljam, 2013).

Table 3a. Regression results (Thailand): Dependent variable, non-sampling error count

	(1) Missing values	(2) Refusal values	(3) Implausible values	(4) Erroneous values
Respondent age ≥ 60 (1 = yes, 0 = no)	0.0133 (0.188)	0.0207 (0.102)	-0.150*** (0.0534)	-0.0561 (0.0513)
Respondent gender (1 = male, 0 = female)	0.189 (0.202)	0.0467 (0.102)	0.193*** (0.0538)	0.174*** (0.0520)
Respondent education – secondary and higher (1 = yes, 0 = no)	0.0213 (0.264)	0.212 (0.135)	0.101 (0.0711)	0.148** (0.0690)
Respondent is head of household and financial affairs (1 = yes, 0 = no)	-0.299 (0.193)	-0.161 (0.101)	-0.123** (0.0533)	-0.175*** (0.0512)
Respondent participated in all waves (1 = yes, 0 = no)	-0.797** (0.347)	-0.169 (0.166)	-0.0935 (0.0911)	-0.203** (0.0884)
Respondent main occupation (1 = wage- or self-employed, 0 = other)	-0.497** (0.231)	-0.113 (0.119)	-0.0753 (0.0609)	-0.122** (0.0593)
Interviewer gender (1 = male, 0 = female)	0.236 (0.248)	0.129 (0.114)	-0.185*** (0.0596)	-0.0709 (0.0583)
Interviewer survey experience (1 = yes, 0 = no)	0.0713 (0.205)	0.174* (0.100)	-0.106** (0.0512)	-0.0329 (0.0504)
Interviewer has completed degree (1 = yes, 0 = no)	0.369 (0.269)	-0.498*** (0.131)	0.265*** (0.0735)	0.172** (0.0703)
Interviewer from local province (1 = yes, 0 = no)	0.467 (0.346)	0.272 (0.178)	0.167* (0.0879)	0.147* (0.0863)
Interviewer extraversion	0.227 (0.153)	0.167** (0.0648)	0.0268 (0.0378)	0.139*** (0.0361)
Interviewer agreeableness	0.0742 (0.133)	-0.334*** (0.0682)	0.00200 (0.0358)	-0.0327 (0.0344)
Congruent ethnicity (1 = yes, 0 = no)	-0.521 (0.411)	-0.120 (0.217)	0.104 (0.107)	0.112 (0.104)
Answers per minute	0.0418* (0.0245)	0.0731*** (0.0137)	0.00252 (0.00723)	0.0257*** (0.00701)
Tablet malfunction occurred during interview (1 = yes, 0 = no)	0.206* (0.122)	-0.00567 (0.0601)	0.0757** (0.0312)	0.116*** (0.0305)
Others present during interview (1 = yes, 0 = no)	-0.0831 (0.224)	-0.0954 (0.121)	0.111* (0.0606)	0.0181 (0.0591)
Interview took place before break day (1 = yes, 0 = no)	-0.213 (0.298)	-0.0985 (0.157)	0.0459 (0.0766)	0.0540 (0.0750)
Survey Week	-0.141* (0.0854)	0.359*** (0.0438)	-0.0777*** (0.0225)	-0.0419* (0.0218)
Ubon province (1 = yes, 0 = no) (Buriram is base)	0.932*** (0.284)	0.440*** (0.140)	-0.0575 (0.0725)	0.0879 (0.0711)
Nakhon Phanom province (1 = yes, 0 = no) (Buriram is base)	1.558*** (0.341)	0.267 (0.187)	0.0309 (0.0903)	0.315*** (0.0881)
Over-dispersion factor	2.455*** (0.0708)	0.709*** (0.0822)	-0.164*** (0.0441)	-0.166*** (0.0392)

Constant	-2.094***	-1.253***	1.355***	0.940***
	(0.719)	(0.446)	(0.224)	(0.203)
Observations	1,817	1,817	1,817	1,817
Chi²	101.6	181.1	111.3	136.9
Degrees of freedom	20	20	20	20
Adjusted R^2	0.03	0.05	0.01	0.01

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

Notes: Robust standard errors in parentheses

Table 3b. Regression results (Vietnam): Dependent variable, non-sampling error count

	(1) Missing values	(2) Refusal values	(3) Implausible values	(4) Erroneous values
Respondent age ≥ 60 (1 = yes, 0 = no)	0.386	-0.0434	-0.401***	-0.323***
	(0.376)	(0.277)	(0.0682)	(0.0690)
Respondent gender (1 = male, 0 = female)	-0.131	0.136	0.122*	0.107
	(0.347)	(0.262)	(0.0666)	(0.0676)
Respondent education – secondary and higher (1 = yes, 0 = no)	0.505	-0.358	0.224***	0.227***
	(0.407)	(0.262)	(0.0670)	(0.0691)
Respondent is head of household and financial affairs (1 = yes, 0 = no)	-0.0172	0.144	-0.0615	-0.0763
	(0.364)	(0.265)	(0.0672)	(0.0685)
Respondent participated in all waves (1 = yes, 0 = no)	-0.213	-0.386	0.0319	0.0106
	(0.774)	(0.594)	(0.141)	(0.144)
Respondent main occupation (1 = wage- or self-employed, 0 = other)	0.705*	0.111	-0.483***	-0.378***
	(0.429)	(0.313)	(0.0786)	(0.0797)
Interviewer gender (1 = male, 0 = female)	0.434	-0.782***	0.128**	0.136**
	(0.328)	(0.256)	(0.0605)	(0.0618)
Interviewer TVSEP survey experience (1 = yes, 0 = no)	-2.076**	0.391	-0.0124	-0.0578
	(1.005)	(0.508)	(0.142)	(0.146)
Interviewer from local province (1 = yes, 0 = no)	0.507	-0.109	0.311***	0.334***
	(0.445)	(0.333)	(0.0817)	(0.0830)
Interviewer extraversion	0.0504	0.238	-0.106**	-0.0925**
	(0.257)	(0.165)	(0.0441)	(0.0452)
Interviewer agreeableness	0.0443	-0.804***	-0.130***	-0.131***
	(0.241)	(0.167)	(0.0436)	(0.0446)
Congruent ethnicity (1 = yes, 0 = no)	-1.402***	-0.488	-0.307***	-0.388***
	(0.440)	(0.309)	(0.0765)	(0.0781)
Answers per minute	0.0793	0.150**	0.0409***	0.0443***
	(0.0751)	(0.0590)	(0.0141)	(0.0143)
Tablet malfunction occurred during interview (1 = yes, 0 = no)	-0.0395	0.628***	0.0738*	0.0807**
	(0.210)	(0.161)	(0.0380)	(0.0388)
Others present during interview (1 = yes, 0 = no)	0.748	1.247***	0.110	0.172*
	(0.569)	(0.310)	(0.0884)	(0.0906)
Interview took place before break day	0.408	0.409	0.0635	0.119*

(1 = yes, 0 = no)	(0.365)	(0.258)	(0.0698)	(0.0712)
Survey Week	-0.0691	-0.0196	-0.108***	-0.103***
	(0.206)	(0.150)	(0.0361)	(0.0369)
Thua Thien Hue province	-0.349	-0.350	-0.185	-0.191
(1 = yes, 0 = no) (Ha Tinh is base)	(0.648)	(0.480)	(0.121)	(0.123)
Dak Lak province	-0.769	0.290	-0.714***	-0.721***
(1 = yes, 0 = no) (Ha Tinh is base)	(0.584)	(0.422)	(0.108)	(0.109)
Over-dispersion factor	3.106***	1.688***	-0.173***	-0.101**
	(0.111)	(0.204)	(0.0499)	(0.0477)
Constant	-1.080	0.648	3.388***	3.420***
	(2.073)	(1.226)	(0.350)	(0.357)
Observations	1,351	1,351	1,351	1,351
Chi^2	29.80	68.56	256.2	232.5
Degrees of freedom	19	19	19	19
Adjusted R^2	0.02	0.07	0.03	0.03

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

Notes: Robust standard errors in parentheses

To conclude the results of our combined and country-level analysis 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). Furthermore, by comparing the determinants of non-sampling errors of two countries that are part of a long-term panel under similar conditions and with an identical survey instrument, we find that there are significant differences at the between determinants of data quality for both countries.

6. Conclusions and Recommendations

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 factors influencing specific types of non-sampling error, namely: missing values, refusal values and implausible values.

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, other congruent characteristics such as matching age and gender were included in previous models and were 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 respondents 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 further developed in order to minimize non-sampling errors. We hypothesize, that while the implementation of plausibility rules and an intensification of supervision in the more recent TVSEP waves has significantly reduced non-sampling errors, data quality may further be improved by extending and optimizing automated plausibility rules.

While it is inconclusive whether or not there are clear differences in explanatory variables that significantly affect the quality of income data, our results merit further research into determining to which extent best practice solutions are applicable. More in-depth examination not only related to income but also of other measures of welfare such as consumption is necessary. 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 researchers and policy-makers. Possible approaches to improving the quality of survey data could be in the selection of suitable respondents within households included in survey samples 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 to be learned.

We advise that respondents in household surveys be selected according to their status in the household and their knowledge of household income activities. Interviewing heads of businesses, household members in charge of financial decisions within the household and household heads in order to extract more reliable income data is suggested. 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 or 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 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 obscured, even after thorough cleaning of data. A first important step in this direction for which such cooperation could be to include 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.

Barge, S., & Gehlbach, H. (2012). "Using the theory of satisficing to evaluate the quality of survey data." Research in Higher Education, 53(2): pp. 182-200.

Barne, D., & Wadhwa, D. (2018, December). "Year in Review: 2018 in 14 Charts". Retrieved from: https://www.worldbank.org/en/news/feature/2018/12/21/year-in-review-2018-in-14-charts

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.

Brooks, M., Lippe, R. S., & Waibel, H. (2019). "Comprehensive Data Quality Studies as a Component of Poverty Assessments". Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World". Washington, DC, 7-8 November, 2019.

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., & Hansen, S. E. (2002). "Computer-assisted interviewing". Handbook of interview research: pp. 557-575.

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.

Couper, M. P., & Kreuter, F. (2013). "Using paradata to explore item level response times in surveys". Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(1): pp. 271-286.

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.

Dale, O., & Hagen, K. B. (2007). "Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data". Journal of clinical epidemiology, 60(1): pp. 8-17.

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, J. R., & Grabka, M. (2007). "Item non-response and imputation of annual labor income in panel surveys from a cross-national perspective". IZA Discussion Paper No. 3043. Bonn: IZA.

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.

Nguyen, G., & Nguyen, T. T. (2020). "Exposure to weather shocks: A comparison between self-reported record and extreme weather data". Economic Analysis and Policy, 65: pp. 117-138.

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): pp. 1-11.

Olson, K., & Peytchev, A. (2007). "Effect of interviewer experience on interview pace and interviewer attitudes". Public Opinion Quarterly, 71(2): pp. 273-286.

Parvathi, P., & Nguyen, T. T. (2018). "Is environmental income reporting evasive in household surveys? Evidence from rural poor in Laos". Ecological Economics, 143: pp. 218-226.

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

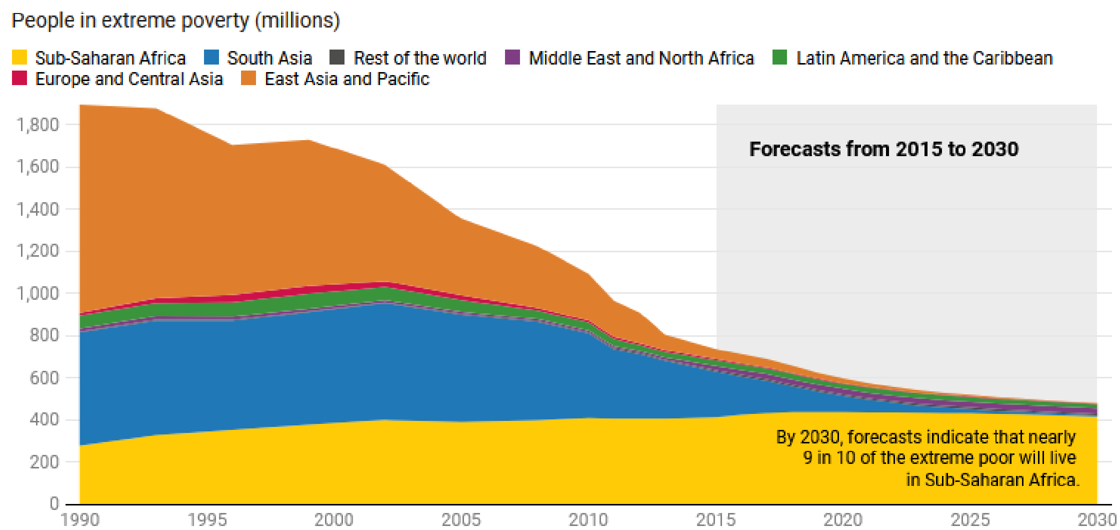

Figure A3. Overview extreme poverty (Source: Barne & Wadhwa, 2018)

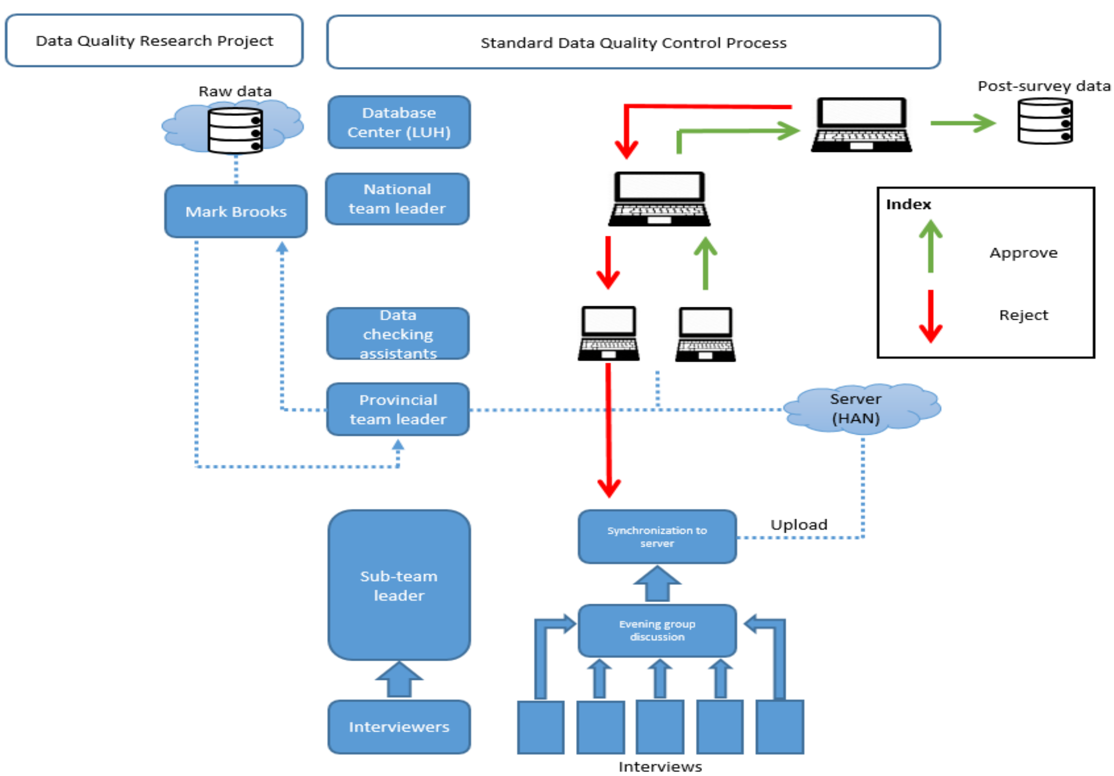

Figure A4. Survey data structure TVSEP 2017 – Example with one survey team