FORS GUIDES
to survey methods
and data management

Preparation of survey data

Alexandre Pollien[1], Jessica M. E. Herzing[1,2], and Erika Antal[1]

[1] FORS, Lausanne

[2] LINES, Faculty of Social and Political Sciences, University of Lausanne

FORS Guide No. 13, Version 1.0

December 2020

Abstract:

This guide focuses on the data preparation phase, which starts after data collection and ends before their analysis. This first assessment of the "raw" survey data is crucial since data preparation can affect the quality of the data in a positive or negative way. After an overview of the different types of errors, the guide discusses the remedies and issues related to these editing procedures.

Keywords: plausibilization, valid cases, validation, verification, error correction, quantitative data, databases

How to cite:

Pollien, A., Herzing, J. M. E., & Antal, E. (2020).*Preparation of survey data*. FORS Guide No. 13, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences (FORS).
doi:10.24449/FG-2020-00013

The FORS Guides to survey methods and data management

The FORS Guides offer support to researchers and students in social sciences, who intend to collect data, as well as to teachers at University level, who want to teach their students the basics of survey methods and data management. Written by experts from inside and outside of FORS, the FORS Guides are descriptive papers that summarize practical knowledge concerning survey methods and data management. The FORS Guides go beyond the documentation of specific surveys or data management tools and address general topics of survey methodology. Considering the Swiss context, the FORS Guides can be especially helpful for researchers working in Switzerland or with Swiss data.

Editor:

FORS, Géopolis, CH-1015 Lausanne
www.forscenter.ch/forsguides
Contact:info@forscenter.ch

1. INTRODUCTION

Survey data are generated from measurement in a data collection process in which errors can occur (Biemer & Lyberg, 2003). When looking at the survey life cycle from a data quality perspective (referred to as the Total Survey Error framework), we can identify four error types on the dimension of representation (coverage error, sampling error, nonresponse error, and adjustment error; Groves et al., 2009, p. 48) and three error types on the dimension of individual measurement (validity, measurement error and processing error). Validity lies between the underlying construct and the survey instrument, measurement error is the observational gap between the ideal measurement and the response provided by the respondent, and processing errors occur after data collection and prior to estimation (Groves et al., 2009, p. 53). In this guide, we are looking at the measurement errors and address how to edit them, adding as little processing error as possible.

While editing responses bears the risk to introduce more error to the data, they usually aim at detecting and reducing errors and hence, increase data quality. In the present guide, we will refer to this data processing step between data capture and data analysis as data preparation, which covers *data editing* or *data cleaning*. To try to avoid processing errors, data preparation steps need to be done thoughtfully, as they involve decisions regarding the validity of original responses, and these decisions are in turn made based on the identification and treatment of abnormalities in the data.

Data preparation does not aim to eliminate errors at all costs, as errors are statistically inherent to any measurement. As statistical inference can only account for random errors, the purpose of the cleaning of the raw data is to detect and to reduce *errors which do not occur at random*. Although data preparation aims at fixing errors, it does affect the variance of variables and thus, may induce systematic errors, which can affect statistical estimates (Jones & Hidiroglou, 2013; Osborne, 2010). Therefore, data preparation should be thoroughly documented in technical reports by describing data cleaning methods, error types and rates, error correction rates, and differences in outcomes with and without outliers (extreme values). The documentation should include the flagging of suspected cases, diagnostic information, and information on the type of editing (Van den Broeck, Argeseanu Cunningham, Eeckels, & Herbst, 2005) to communicate assumptions about the editing process and highlight when an assumption is risky (e.g. correlation with the tested hypothesis; Krishnan, Haas, Franklin, & Wu, 2016).

This guide will not discuss within-analysis procedures for measurement error, such as stratification or weighting techniques. This guide is rather targeted at scientists who are using raw survey data or who want to evaluate survey data quality.

As the hunt for errors in survey measures requires prior knowledge on potential error mechanisms, one has to understand what types of errors can occur and what sources potential errors can have. Based on this knowledge, one might then detect errors and can decide for the appropriate approach to correct for potential errors in the measures. In the following, we will discuss these aspects and then give two examples on how data preparation can be handled. Finally, this guide ends with some general recommendations for practice.

2. ERRORS IN SURVEY MEASURES

2.1 PHENOMENOLOGY OF ERRORS IN SURVEY MEASURES

Errors in survey measures are represented by abnormalities in the data. The researcher defines data abnormalities when reviewing the initial measures provided by respondents. Hence, data abnormalities reflect a researcher's point of view on assumptions about everyday life (rules of communication, human biology, culture, etc.). This assertion about errors in survey measures can be assessed logically by the comparison with alternative information (e.g. external data sources or a comparison with other survey measures) or assessed statistically by measuring the deviation of the survey measure with the help of an estimate (standard deviation, standard error). By defining a data value/point as a likely error, the researcher states that the observation "deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism" (Hawkins, 1980). There is, thus, a swaying back and forth between a theory about the underlying response mechanisms (assessment of an error with respect to the measured concept) and a phenomenology of errors (detection of anomalies). The challenge of going back and forth is to circumscribe the errors to identify their cause, their prevalence in the data, and their influence on the estimates.

2.2 TYPES OF ERRORS IN SURVEY MEASURES

Systematic errors or *errors not at random* are a consistent bias caused by a measurement that has a flawed design (e.g. scale is incorrect). If the measurement is repeated, the bias should be the same. Errors not at random may have unpredictable effects on the correlations among items. They can substantially alter the likelihood of making both type I (false positive) and type II errors (false negative; Osborne, 2010). In contrast, *errors at random* have probability distributions with their own variance and expectation (Jones & Hidiroglou, 2013). Errors at random are expected not to be correlated with other variables in the survey. Theoretically, errors at random can be controlled by statistical techniques, although the increase of variance can lead to spurious within-group variability and a lower reliability, which in turn will attenuate correlations and potentially create type II errors in hypothesis testing (Meade & Craig, 2012).

The challenge of data preparation is to deal with errors in the data without adding systematic bias, i.e. errors that are not at random and that occur among subgroups. When the "true" value is not accessible, data cleaning should aim at transforming as much as possible systematic errors into "random errors". As this valuation needs expert knowledge and can result in harmful correction if the researcher only detects a certain proportion of errors, it is sometimes better to leave the decision of what to do to the final data user. Indeed, the aims of a study determine the level of data cleaning. It will determine whether a correction could correlate with the phenomena studied and, thus, introduce a risk of systematic bias in estimates (Besselaar & Sandström, 2016; Van den Broeck et al., 2005). Therefore, it is not the errors in the data themselves that are harmful but the potential bias to estimates.

2.3 SOURCES OF SYSTEMATIC ERRORS IN SURVEY MEASURES

It is difficult to detect an error in survey measures with certainty. A first step to detect errors in survey measures is gaining knowledge on potential error sources of survey measures, which requires insight into all stages of the data collection process and all actors (e.g. researcher, survey practitioner, respondents, interviewers or the interview itself). One source of error in survey measures can be the *survey design*, which is developed by the researcher and the survey practitioner. Besides many other error types, errors in the survey measure, due to the survey design, can occur because of poorly formulated questions, ambiguous definitions, badly designed filters or faulty design of response choices (e.g. not covering “don’t know” as a valid answer to attitudinal questions). For example, questions about the number of persons in the household often overlook the ambiguous nature of cohabitation situations: should a child in shared parenting or a student who lives elsewhere during the week be mentioned?

Another source of errors in survey measures are the *respondents* themselves. In case respondents mishandle one of the four steps of the response process (question and instruction comprehension, retrieval of information, rendering of the judgement, and reporting of an answer; Tourangeau, Rips, & Rasinski, 2000), the probability of error increases. For example, when respondents do not read the question conscientiously, they might report yearly income rather than monthly income, children living elsewhere instead of children present in the household, etc. In addition, respondents can intentionally misreport an information (motivated misreporting), in order to avoid, for example, additional survey questions (e.g. Eckman et al., 2014). Furthermore, respondents can develop strategies to decrease the burden to answer survey questions, which is often associated with low quality of the responses given. This response behavior is referred to as satisficing (as opposite to optimizing, see Krosnick, 1991).

Van den Broeck and colleagues (2005) note that data cleaning can never be a cure for poor study design or study conduct. To minimize the potential for errors, the survey design for the data collection stage needs to be carefully chosen to avoid errors in the first place (see Dillman, 2007). For example, by implementing interviewer training, proper questions wording, distinct and exhaustive, and correctly specified filters in the questionnaire routing. In web surveys, there is a concern that respondents use external sources (e.g. search for answers online) to validate or answer performance questions (see Clifford & Jerit, 2016). This response behavior is often a result of optimizing the question-answer process (putting additional response efforts into finding the correct answer). As optimizing of knowledge questions is an undesired response behavior (due to the decrease of the validity), which is associated with respondent characteristics (e.g. educational level), it needs to be carefully decided how to correct for this type of response behavior during post-survey adjustments.

There are cases when error is due to *survey practitioners*, for example when erroneous invitation mailings are sent to people who are not in the target group. Address problems result in incorrect coverage or duplicates (multiple records of the same sample unit). Duplicate detection is the process of identifying different or multiple records that refer to one unique real-world entity or object (Elmagarmid, Ipeirotis, & Verykios, 2007). In case of duplicates one case should be deleted. However, there is no convention which one should be deleted (e.g. first time responded, the interview with less missings, etc.) and thus, the rules used in any specific case need to be documented.

A special case of error in survey measures can be introduced by *interviewers* in interviewer-administered surveys. There is a chance for intentional interviewer misreporting, such as the fabrication of entire interviews (interviewer falsification), a partial falsification of interviews, deviating interviewer behavior when selecting respondents (e.g. misclassification of non-cooperative target units), or intentional miscoding of a given response to avoid filter questions (American Association for Public Opinion Research, 2003).

Finally, errors in survey measures can occur due to disturbance in the *survey interview* itself. Whereas this issue is difficult to detect in self-administered mode, face-to-face interviewers sometimes witness a disturbance of third parties during the interview. For example, in the

European Social Survey driven in Switzerland (ESS) 2018, about 5% of the interviews were declared to have been interfered by a partner.

3. TREATMENT OF ERRORS IN SURVEY MEASUREMENT

The management of anomalies is always based on an assessment of the plausibility that the response cannot possibly be correct or that there remains a possibility that it may be correct (Jones & Hidiroglou, 2013). The alternatives are limited to correcting, deleting, or leaving potential errors unchanged. Errors can be corrected by inferring the correct response or putting it in missing value (exclusion of whole cases or setting single observation to missing). From this point of view, validity, logical and format edits can be distinguished.

3.1 VALIDITY EDITS

When the goal is to check the validity of the data, it can be treated at two different levels: at the case level or at the variable level. At case level, one must verify if the actual person (observed case) is identical with the one chosen for the study (i.e. if the person is the selected person for the study or if there might be a substitution, e.g. the partner responding instead). This verification can be performed by comparing the collected data with external data or using information given in the study itself. The external data are mostly the sampling frame or administrative register data. For coverage checks, the researcher should verify each predefined criterion concerning the target population of the study (e.g. nationality, citizenship or age range). In case of evidence for a respondent not being part of the target population, the case can be flagged in the dataset. On the variable level, one might validate single variables and edit the value of the variable in case the external data seems to be more plausible.

External data such as register data is not guaranteed to be error-free either (see Oberski, Kirchner, Eckman, & Kreuter, 2017). Since there is often no evidence whether the mistake lies in the register data or in the survey data, the researcher must assert which source of information is more plausible. Consequently, at least two respondent characteristics are necessary to decide upon the non-validity of a case, such as sex and age. This amount of information is sometimes insufficient, and requires the use of more characteristics: civil status, nationality or composition of the household. The more the survey response of the respondent deviates from the characteristics given in the sampling frame, the more likely it is that the observation is erroneous. In a self-administered survey with a postal mail invitation, substitution by another household member occurs quite often (2 to 5% of sample units in surveys as ESS or MOSAiCH; see survey reports). In face-to-face surveys, substitutions are found to be mainly coming from persons external to the household (neighbor or new resident who arrived after the target person moved).

3.2 CONSISTENCY AND RANGE EDITS

Another type of data edit stands at the variable level. The answer might not match with what was expected by the researcher or the value is not consistent within the measured concept. The respondent's answers can be unlikely themselves or inconsistent between them. A large number of doubtful responses can lead to consider the entire case as an invalid interview (see section 2.3).

Consistency edits compare different answers from the same record to check their logical consistency. Range edits determine whether values reported are outside of bounds (Jones & Hidiroglou, 2013). Data cleaning strategies depend on, among other things, the variable format. When the variable format is categorical, researchers are often not able to check the information without comparing it to additional information. The variable alone is often not sufficient to determine the source of the error and it is necessary to combine the suspected value with other information, either external to the survey or information from other given responses. In the case of variables with numerical formats, the individual answers may differ from what would be expected and can take so-called extreme values. Extreme values (also referred to as outliers) can be legitimate values that are "far" from the normal distribution of the variable and, thus, can skew the distribution, which causes concern depending on the statistical method used. Extreme or non-plausible values can be detected using methods based on robust estimates of the centrality and dispersion parameters (i.e. variance). For example, the standard deviation method involves selecting a certain number of standard deviations that deviate from the mean (e.g. ± 2 standard deviations from the mean). Osborne (2010) notes that, in a normally distributed population, a score at 3 or more standard deviations from the mean has a probability of 0.26% to occur. However, it should be noted that extreme values increase the standard deviation themselves, increasing the threshold of their detection. Therefore, the median absolute deviation method is more robust. The interquartile range method is another robust way for detecting outliers when there are many extreme values.

Outliers can be dealt with regardless of assumptions about the cause of error. Some authors recommend that true extreme values stay in the analysis (Van den Broeck et al., 2005), as they may reflect a dimension of the analyzed phenomenon. However, when outliers are non-plausible, the extreme value is possibly a true error and should be considered as invalid measurement and thus corrected or deleted (Van den Broeck et al., 2005). Sometimes, suspected errors will fall in between the plausible and non-plausible value. In these cases, it is a good practice to apply a combination of diagnostics. If there is no information that could confirm the “true” status of an outlying data point, a procedure is to go to previous stages of the data preparation process to see whether a value is coherent with other variables in a sociological or statistical sense (Van den Broeck et al., 2005).

In computer-assisted questionnaires, the formats of the possible answers can be predetermined by the questionnaire design. For reasons of comparability with paper mode, survey designers sometimes avoid too sophisticated format controls in computer-assisted questionnaires. For example, the age or the number of persons in the household is not always limited to plausible values in order to avoid deciding on an arbitrary threshold. The more open the answer format, the more unexpected answers can be encountered. Open-ended questions (e.g., for “other” category) entail concerns in interpreting the answers when coding. For example, if respondents provide an answer to their educational level by mentioning an occupation in the open-ended “other” category, a researcher could make an assumption on the highest educational level achieved or code this response as missing. If only specific occupation groups use ambiguous answers, and these are coded as missing, a systematic bias is introduced since these individuals will have the same error in educational level.

3.3 INCONSTANCIES ACROSS SEVERAL VARIABLES

The detection of error across variables can be done through an assessment of coherence between them by using checks for logical inconsistencies, violations of filters, or abnormal response behavior (e.g., straightlining). Their detection is based on the observation of the relation of different information: errors of this type are sometimes detected via rules issued from the questionnaire design. For example, the International Social Survey Programme (ISSP) asks for the total number of persons in the household and, separately, for the number of adults in the household, number of children above school entry age and number of children below school age: the total number of persons in the household should thus be equal to the sum of these categories.

When the variable has a categorical response format, the value can be checked by logical criteria: for example, a person cannot declare to be retired and to be full-time employed at the same time. Furthermore, answers can be compared to administrative rules (e.g. can a self-employed individual work in a public organization?). Answers can be compared with known policies (e.g. to verify that marriage does not occur before the legal age in the given country). If the survey has information from more than one member of a household, relations between sample units allow some checks across respondents, e.g. validating the declared relationships between household members. Panel data provides the possibility of logical verification by considering the information across waves. For example, one can check that someone declared to be married cannot be single in further waves (from a legal point of view) but only separated, divorced or widowed. Longitudinal studies allow and necessitate checking the temporal consistency of data. Temporal consistencies can be checked by looking at differences between waves as a result of a reported event, such as new members in the household and new degree of education achieved. Again, when the detection of inconsistencies is done using characteristics provided by external sources, an assessment of plausibilization between the two sources of information should be carried out, as some predefined criteria will point to variables that are more likely to be erroneous than others.

3.4 FURTHER APPROACHES FOR TREATING ERRORS IN SURVEY MEASURES

In case interviewer falsification could not be prevented during the fieldwork, there are various methods available aiming at identifying falsified data (see DeMatteis et al., 2020), which are similar to the aforementioned strategies (see Blasius & Thiessen, 2013; Murphy et al., 2016; Schwanhäuser, Sakshaug, Kosyakova, & Kreuter, 2020; Weinauer, 2019). If the deviant interviewer behavior can be confirmed, the data should be excluded from the data set.

When respondents intentionally misreport in order to reduce their burden to answer a survey question, one can differentiate between weak and strong satisficing (Krosnick, 1991). With regard to data cleaning, weak satisficing results in respondents selecting the first reasonably appropriate response alternative, which then converts into response order effects and a tendency to agree with statements (also called acquiescence). The identification of weak satisficing in survey data is often difficult and requires advanced statistical modeling (e.g. structural equation models, see Kaminska, McCutcheon, & Billiet, 2010). In case of strong satisficing, respondents often endorse the status quo by expressing opinions or attitudes to “keep things as they are”. Furthermore, respondents do not show differentiation in rating scales, hence they assign the same value to all response categories (e.g. straightlining) or excessively use the “don’t know” and similar non-substantial response options as “neither agree nor disagree”. Strong satisficing can be on the edge of respondents providing random answers (such as “mental coin flipping”). Some strong satisficing techniques are relatively easy to identify (status quo, non-differentiation and excessive don’t know answering). However, random answering is hard to identify within the data and often one needs additional auxiliary data, such as response time (identifying speeding, see Greszki, Meyer, & Schoen, 2014; Zhang & Conrad, 2014), consistency or attention checks (e.g. reversed items in scales, instructed response items, see for bogus items Meade & Craig, 2012), to identify this type of response behavior. If a respondent displays signs of strong satisficing through the whole questionnaire, one strategy is to completely delete the case; otherwise only problematic data can be dropped or set to item nonresponse.

4. EXAMPLES FROM SWISS SURVEYS

When using the method of plausibilization to detect potential invalid responses, one needs to consider different potential sources of error sources (dimensions of error), which need different correct strategies. To exemplify the issue of considering multiple error sources and correction strategies, we use questions of the ISSP 2019[1] about the income estimation (see Figure 1).

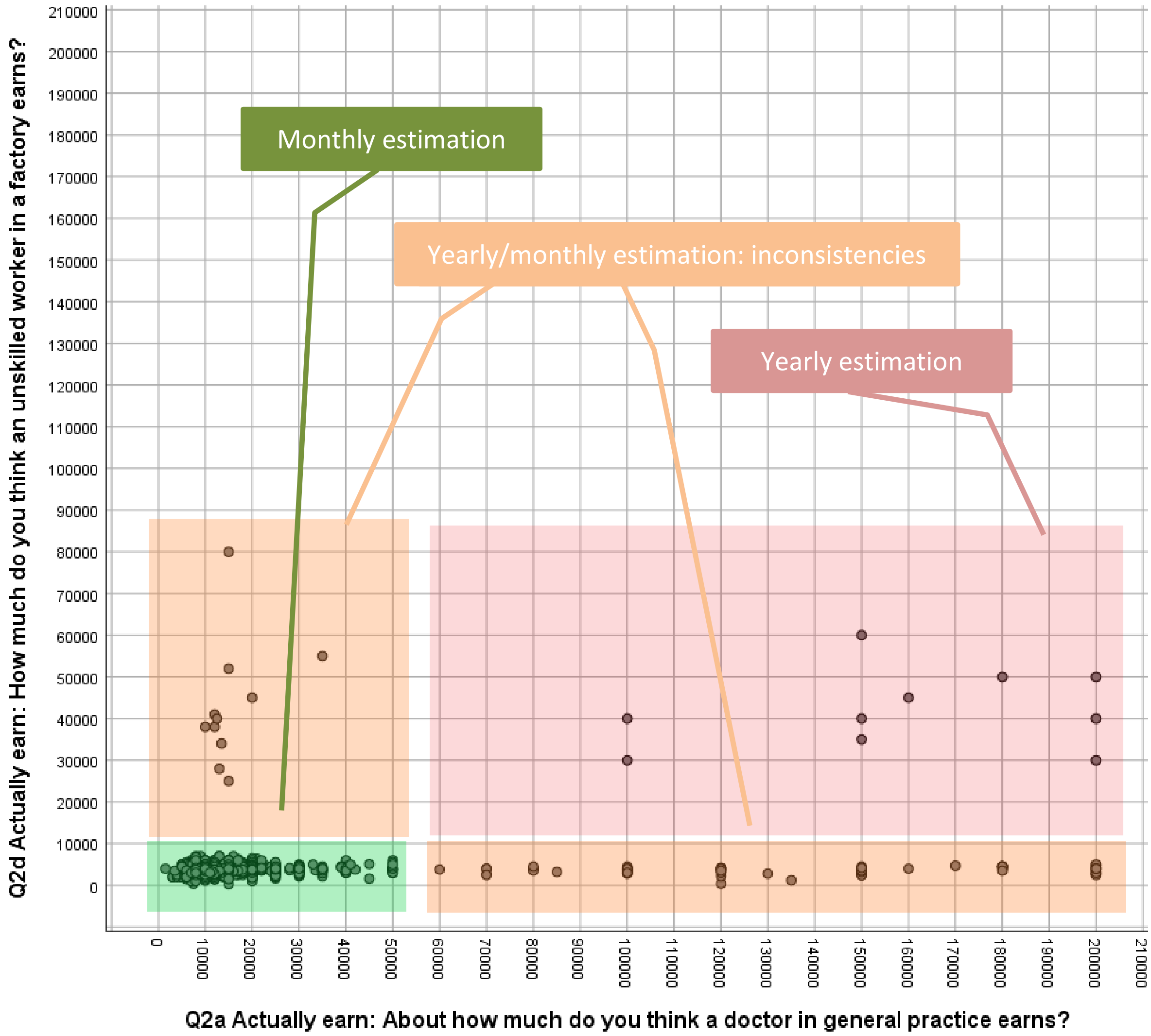
During the survey, respondents were asked to estimate the income for different work occupations. In order to make a comparison between the web mode and the paper mode reliable, the web questionnaire did not perform any validation or format constraints. Thus, many errors appear to be interpreted: if "4k" indicates 4,000 with certainty, when the respondent has written only "4". It is necessary to check the consistency of the different answers of the same respondent; this may indicate a monthly income of 4,000 or an annual income of 40,000. Some answers can be attributed to entry errors, as for example omission of zero (“400”). The ambiguity between a monthly salary and an annual salary also occurs in syntactically correct answers: 80,000 CHF for an unskilled worker is a sociologically non-plausible monthly salary. The difficulty increases when we get closer to plausible values. Here again, the comparison of answers across questions is worthwhile. However, it does not solve all anomalies: many respondents gave inconsistent answers throughout questions. An interesting finding is that high-income occupations are reported more frequently in annual yearly earnings than others: we estimate about 20% of responses for “chairman of a large national corporation”, 4% for “doctor in general practice” and 1% for “unskilled worker in a factory”. Some work occupations are more associated with yearly earnings than other occupations. This shows a strong relation between the error, the design of the question, and the social phenomenon being measured. In summary, this example shows errors of response entry (forgetting or adding zeros), errors due to skipping instructions (monthly earnings), errors of question design (reporting a monthly instead of a yearly salary for a top manager might be less straightforward due to performance-related pay).

As stated earlier, there is the challenge that the correction of errors in the data (e.g. due to mistakes in the questionnaire design) can introduce systematic bias. For example, without a correction of potential errors in the measurement of the number of household members in the ISSP might have systematic bias regarding the number of children reported. The measurement of the composition of household consisted of a set of questions on different categories of household members (e.g. children below or above a certain age). Hence, the total number of people in the household should be equal to the number of children plus the number of adults. Furthermore, a question instruction to put "0" when there is no child in the household was provided. However, many respondents did not indicate anything in the question on children present in the household and left the field blank, either because there is no child in the household, or because they could not answer for one reason or another (don't know, privacy issue). These responses were considered as item missing at first and thus, many

[1] This study was conducted within MOSAiCH, a mixed-mode survey of the Swiss population.

inconsistencies were identified at the data preparation stage, e.g. survey data did not match the sample data. Therefore, variables indicating the number of persons in the household were adjusted using a procedure taking information from this additional data source.

Figure 1. Potential error sources for potential invalid responses for the estimation of salary (based on Swiss ISSP 2019 data).



Since we do not know in which source (survey or register) the true value lies, the procedure of data editing followed a logic of parsimony: make as few and soft changes as possible to achieve a correct total (one change is less than two changes, replacing a nonresponse by a zero is less than by another number). Table 1 reports the result of this procedure.

Table 1 shows that in total 2,756 data edits were performed after the data underwent a review by the researcher. For 69 observations the total number of persons in a household was corrected, notably 12 cases by replacing no answer with the value provided in the sampling frame and 57 cases were replaced with the sampling frame data, because they did not give a coherent total number of household members. Furthermore, there were 195 cases edited regarding the number of adults within a household. These inconsistencies were often based on the fact that there were some people who forgot to count themselves in the number of adults and in the number of persons in the household, despite the instructions. Moreover, 2,427 cases were edited, because the response field for the number of children was left blank. Thus, these item non-responses were replaced by the value 0. We had otherwise the choice either to let the item non-responses, or to replace them by the value 0, which would have created at least 114 errors (we did not consider cases where no answers about category of household members were indicated, considering that the respondent refused to answer). If these data edits would not have been made, 2,756 out of 3,065 cases would have had an incorrect value, which would very possibly have had an effect on survey estimates. This example shows the potential impact of data editing on a later analysis.

Table 1. Change in distribution due to the data editing.

	Missing replaced by zero	Missing replaced by value	Non-missing replaced by value	Total
# of persons in household	0	12	57	69
# of adults	0	75	120	195
# of children above school entry age	1,152	26	21	1,199
# of children below school age	1,275	1	17	1,293
Total	2,427	114	215	2,756

Note. Data source - MOSAiCH 2019.

5. IMPLICATIONS FOR SURVEY PRACTITIONERS

After reading this guide one might have an idea on what issues arise when preparing the data for publication (scientific use file) or for the own analysis. However, one has to keep in mind that one final question remains: “If we analyse the data without removing the invalid cases, what will be the impact on findings and conclusions?” (Andreadis, 2014). To not disconnect parts of the data preparation from the analyses serves as a major safeguard. Further, this separation hinders the ability to experiment and test different data preparation procedures and to tune the parameters to their particular object of study (Krishnan et al., 2016). Keeping this in mind, we recommend the following:

Recommendation 1 – Error prevention within the data collection is always better than treating them in the data preparation.

Recommendation 2 – One needs to be careful to not add systematic bias when correcting errors.

Recommendation 3 – The documentation of potential errors is important, as the correction might have an influence on survey estimates.

Recommendation 4 – One should investigate errors on the case as well as the variable level.

Recommendation 5 – One should use various methods and data sources to identify potential errors.

6. FURTHER READINGS AND USEFUL SOFTWARE PACKAGES

There are several introductory books on the topic of survey quality. Just to name three examples we can recommend: Biemer and Lyberg (2003), Blasius and Thiessen (2012), and Groves et al. (2009). If you are more interested in the sources of error within the response process the book by Tourangeau, Rips, and Rasinski (2000) is very useful. Generally, it can be worthwhile to look at technical reports from datasets that are similar to yours to get ideas of what has been done in other similar cases.

REFERENCES

American Association for Public Opinion Research. (2003). Interviewer falsification in survey research: Current best methods for prevention, detection and repair of its effects. Retrieved October, 24, 2020, from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/falsification.pdf

Andreadis, I. (2014). Data quality and data cleaning. In Garzia, D. & S. Marschall (Eds.), *Matching voters with parties and candidates. Voting advice applications in comparative perspective*. Colchester: ECPR Press, 79–91.

Besselaar, P., & Sandström, U. (2016). What is the required level of data cleaning? A research evaluation case. *Journal of Scientometric Research*, *5*(1), 7-12. doi:10.5530/jscires.5.1.3

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. New York: Wiley.

Blasius, J., & Thiessen, V. (2012). Institutional quality control practices. In Blasius, J. & V. Thiessen (Eds.), *Assessing the Quality of Survey Data*. London: Sage.

Blasius, J., & Thiessen, V. (2013). Detecting poorly conducted interviews. In Winker, P., N. Menold, N., & R. Porst (Eds.), *Interviewers' Deviations in Surveys – Impact, Reasons, Detection and Prevention* (pp. 67-88). Frankfurt: Peter Lang.

Clifford, S., & Jerit, J. (2016). Cheating on political knowledge questions in online surveys: an assessment of the problem and solutions, *Public Opinion Quarterly, 80*(4), 858–887. doi:10.1093/poq/nfw030

DeMatteis, J. M., Young, L. J., Dahlhamer, J., Langley, R. E., Murphy, R. E., Olson, K., & Sharma, S. (2020). *Falsification in surveys*. Report for the American Association for Public Opinion Research. Retrieved October, 24th, 2020, from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Data_Falsification_Task_Force_Report.pdf

Dillman, D. A. (2007). *Mail and internet surveys: the tailored design method* (2nd edition). New York, NY: John Wiley and Sons, Inc.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the Mechanisms of Misreporting to Filter Questions. Surveys, *Public Opinion Quarterly*, *78*(3), 721–733. doi:10.1093/poq/nfu030

Elmagarmid, A. K., Ipeirotis, P. G., & Verykios, V. S. (2007). Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering (TKDE), 19*(1), 1-16.

Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: a data quality perspective* (pp. 238-262). Chichester: Wiley.

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

Hawkins, D. M. (1980). *Identification of outliers*. New York: Chapman and Hall.

Jones, J., & Hidiroglou, M. (2013). Capturing, coding, and cleaning survey data. In G. Snijkers, G. Haraldsen, J. Jones, J., & D. K. Willimack (Eds.), *Designing and conducting business surveys* (pp. 459-504). New York: John Wiley & Sons, Inc.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context, *Public Opinion Quarterly, 74*(5), 956–984. doi:10.2307/40985412

Krishnan, S., Haas, D., Franklin, M. J., & Wu, E. (2016). Towards reliable interactive data cleaning: a user survey and recommendations. *HILDA '16, 9*, 1-5.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. doi:10.1002/acp.2350050305

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455.

Murphy, J., Biemer, P., Stringer, C., Thissen, R., Day, O., & Hsieh, Y. P. (2016). Interviewer falsification: Current and best practices for prevention, detection, and mitigation. *Statistical Journal of the IAOS, 32*, 313-326. doi:10.3233/SJI-161014

Oberski, D. L., Kirchner, A., Eckman, S., & Kreuter, F. (2017). Evaluating the quality of survey and administrative data with generalized multitrait-multimethod models, *Journal of the American Statistical Association, 112*(520), 1477-1489. doi:10.1080/01621459.2017.1302338

Osborne, J. W. (2010). Data cleaning basics: best practices in dealing with extreme scores, *Newborn and Infant Nursing Reviews, 10*(1), 37-43. doi:10.1053/j.nainr.2009.12.009

Schwanhäuser, S., Sakshaug, J., Kosyakova, Y., & Kreuter, F. (2020). Statistical identification of fraudulent interviews in surveys: improving interviewer controls. In Olson K., J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer effects from a total survey error perspective.* Boca Raton: Chapman and Hall/CRC.

The American Association for Public Opinion Research. (2016). Standard definitions: Final dispositions of case codes and outcome rates for surveys. 9th edition. AAPOR.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med, 2*(10), 966-970. doi:10.1371/journal.pmed.0020267

Weinauer, M. (2019). Be a detective for a day: how to detect falsified interviews with statistics. *Statistical Journal of the IAOS, 35*(4), 569 – 575. doi: 10.3233/SJI-190524

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127-135. doi:10.18148/srm/2014.v8i2.5453