Using Paradata to Assess the Quality of an Online Questionnaire

With Focus on Response Times

Magdalena Six	Alexander Kowarik	Marc Plate
Statistics Austria	Statistics Austria	Statistics Austria
Guglgasse 13	Guglgasse 13	Guglgasse 13
1110 Vienna	1110 Vienna	1110 Vienna

May 4, 2016

Abstract

Paradata are process metadata, generated during the process of collecting survey data. The conversion to a new, integrated system for survey data collection (STATsurv) at Statistics Austria makes it possible to collect different kinds of paradata. Our paper presents the first results and insights from evaluating the paradata of the Household Budget Survey 2014/2015. The focus of this descriptive analysis is on the time respondents need to complete a survey question or a set of questions respectively. We examine if these response times differ for different kind of questions and for different groups of respondents. We further give an outlook on how to use these insights to identify for example incomprehensible questions.

1 Introduction

There is no standard definition what exactly constitutes paradata. For the purpose of this paper we define paradata as process metadata, generated during the process of collecting survey data (c.f. Kreuter [2013]). The conversion to a new, integrated system for survey data collection (STATsurv) at Statistics Austria makes it possible to collect different kinds of paradata for the use of measurement and non-response error evaluation. The Austrian Household Budget Survey (HBS) - providing information about expenditure and living standards of private households - is one of the first surveys at Statistics Austria taking use of the new survey software.

This newly collected paradata opens a whole new field of assessing the quality of a questionnaire. How long do respondents need for single questions? Do they need longer for certain categories of questions than for others, or can one even recognize incomprehensible questions by looking at response

times from the paradata? Do the paradata differentiate for different groups of respondents, e.g. for respondents being interviewed vs. self-completed web surveys, or for different age groups? How often are check messages needed and by whom?

We are only at the beginning of exploring the possibilities for the usage of paradata. In this paper we present our first results related to the time respondents needed to complete a question or a set of questions respectively. We present the process how to elicit the response time per question/set of questions from the paradata and give with the help of figures some "anecdotal evidence" about differences in response times for qualitatively different questions and for different categories of respondents.

2   The Household Budget Survey 2014/15

The Household Budget Surveys (HBS) is a sample survey of private households carried out regularly under the responsibility of Statistics Austria. Similar surveys are carried out in each of the members of the European Statistical System. Essentially, the HBS provides information about household consumption expenditures on goods and services, with considerable detail in the categories used; information on income, possession of consumer durable goods and cars; basic information on housing and many demographic and socio-economic characteristics. HBSs are carried out in Austria every five years.

The paradata used in the analysis of this paper refers to the latest HBS at Statistics Austria. The Household Budget Survey 2014/15 was implemented as a probability sample survey and was carried out from October 2015 until November 2016. The primary objective of the survey is to record all expenditures of private households in a detailed way. Therefore the households were asked to register their expenditures in a diary for two weeks. In addition they answered questions about housing, the possession of consumer durables as well as basic demographic questions for each household member (c.f. Statistik Austria [2012]).

The survey consists of several sub-questionnaires, differing in whom they address (the household as a whole or an individual member of the household) and in the point of time, at which the sub-questionnaires have to be completed. The core of the survey is a household account book (a diary), in which the respondents describe in a very detailed manner everything bought in a certain time interval. The paradata analysed in this paper are related to the sub-questionnaire which addresses each member of the household individually and which has to be completed before the start of the household account book.

The questions in this sub-questionnaire cover the age, gender and education of the respondent (the household member), the employment situa-

tion and the occupational status. Further, this sub-questionnaire asks for medical care related expenditure such as expenditures on glasses or hearing aid instruments, expenditures on cosmetic surgery, on stays at a hospital or a health spa and expenditures on dental treatments. Finally, the sub-questionnaire asks about expenditures on mobile phone devices.

It is important to emphasize that this sub-questionnaire includes broadly speaking three kinds of questions: First, questions where the respondent has to choose from several possibilities (e.g. when asked about the employment situation); second, questions which allow only for a "yes" or "no" as answer; and third, questions, where the respondent is asked to name a certain sum of money (the expenditures on certain goods and services). In this overview, we examine the latter two categories.

3   STATsurv – The New Data Collection Software at Statistics Austria

Requirements for modern survey data collection software include mixed-mode designs (CAPI, PAPI, CAPI, CAWI), centralization of question-, case- and field-management in one system and the possibility to save paradata on the surveys conducted. Also modularization of the questionnaire is an often named and requested characteristic of modern data collection software. Existing data collection software such as Blaise can only partly meet these challenges. For that reason Statistics Austria has begun to develop its own integrated software called STATsurv. The new software tool is implemented gradually in social surveys conducted at Statistics Austria. The Austrian Household Budget Survey 2014/15 is the first survey which has been implemented and tested on the new IT-system and the new mode CAWI.

4   The Paradata of the Household Budget Survey 2014/2015

The "questionnaire navigation paradata" consists of the identification of the household and the person, the identification of the page of the web survey, action types for each page such as *Go to next page, Show help* and interaction with plausibility warnings. A page of the web survey often consists of only one question, but it sometimes consists of more questions. For each action type of each page, there exists a time stamp. Further, the paradata include information whether the questionnaire was completed by a respondent on his or her own (CAWI) or whether the respondent was asked by an interviewer, who navigated through the questionnaire and typed the respondent's answers (CAPI). Further it includes technical information such as device used, screen-resolution or if the application was online/offline while

the questionnaire was completed.

The net-sample size of the HBS was 7162 households. We have complete paradata for the sub-questionnaire for 2263 households, or 3432 persons respectively. There are several reasons for the smaller number of available paradata. First the paradata characteristics considered by us only make sense for CAWI and CAPI mode. Second, we excluded proxy interviews since we analyze the paradata with respect to the age and the gender of the respondent. And third, there also occurred some technical difficulties, which did not allow to save paradata for all interviews conducted.

5 Time Spent on Each Question/Page

The IT-software STATsurv saves the time stamps for every action performed by the respondent/the interviewer for every page of the survey. If there is more than one question on a page of the survey, it is not possible to assign the actions or the time stamp to a single question but only to the survey page.

We order every data row by the household-ID, the person-ID and the time stamp. Thereby we are able to track the navigation of the respondent (the interviewer respectively) through the survey. In the simplest case, the time spent on a page is the time from the moment the respondent clicked "forward" from the last page until the moment the respondent clicked "forward" on the actual page, leading the respondent to the next page. Since there is no time stamp from "the last page" for the first page of the survey, no time difference can be calculated. A simplified version showing the paradata for the first five actions for one single respondent looks as follows. The action *WEITER GEKLICKT* means *Go to next page*.

```
   householdID personID CAWI/CAPI  question/page           action
1:      687659   767559       CAWI PT1.X01.1.1.10 WEITER_GEKLICKT
2:      687659   767559       CAWI PT1.A02.1.1.10 WEITER_GEKLICKT
3:      687659   767559       CAWI PT1.A01.1.1.10 WEITER_GEKLICKT
4:      687659   767559       CAWI PT1.A01.2.1.10 WEITER_GEKLICKT
5:      687659   767559       CAWI PT1.A01.2.1.40 WEITER_GEKLICKT
                time stamp time
1: 2014-10-05 11:27:47.474   NA
2: 2014-10-05 11:28:05.372 17.9
3:  2014-10-05 11:28:09.75  4.4
4: 2014-10-05 11:28:16.383  6.6
5: 2014-10-05 11:28:47.365 31.0
```

Since it is possible that a respondent did "go back" in the survey and thus visits a survey page more than one time, we have to sum up over all

the visits to get the overall time a respondent stayed on a page. Also the time of actions such as *Show help* is added to the respective page.

We further set a timeout: If a respondent stayed longer on a survey page than three minutes, we assume that the respondent has left the survey and we cut off the time and set it to three minutes (180 seconds).

5.1 Yes/No Questions vs. Concrete Numbers Asked

First, we compare questions with "yes" or "no" as possible answers to questions about the same topics, but where the respondents are asked about the expenditures they had if they answered the question before with "yes". Since only those respondents answering with "yes" are asked about the expenditures, the scale of the histograms differs between Figure 1 and Figure 2. Please note that the survey page related to the third subfigure in both Figure 1 and Figure 2 consists of four questions, whereas the other pages only consist of one or two questions. Thus it is natural that respondents need longer to complete this page, which is also reflected in the histograms.

The histograms in Figure 1 and Figure 2 indicate that the variance as well as the medians (the red dotted line) of the time per page/question are smaller for "yes/no"-questions than they are for questions about expenditures. The figures also indicate that the distribution of time of the question about expenditures are "heavy tailed", which might reflect our expectation that respondents do have to think harder or have to look up numbers at invoices so as to be able to answer the question.
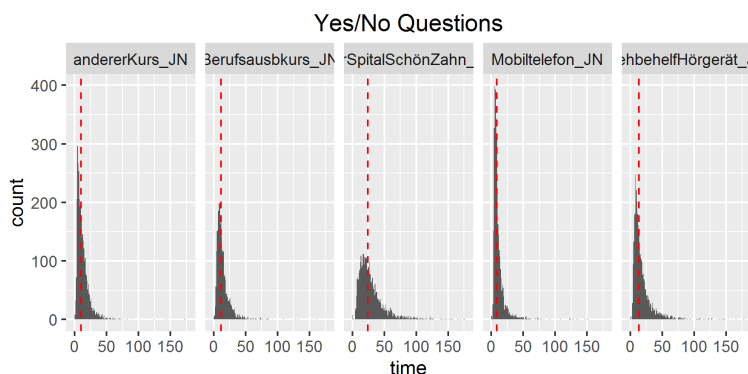

Figure 1: Histograms of five yes/no questions asking if expenditures occurred for (vocational) training programs (subfigure 1 and 2), health expenditures (subfigure 3), mobile phone devices (subfigure 4) and glasses or hearing aid instruments (subfigure 5)
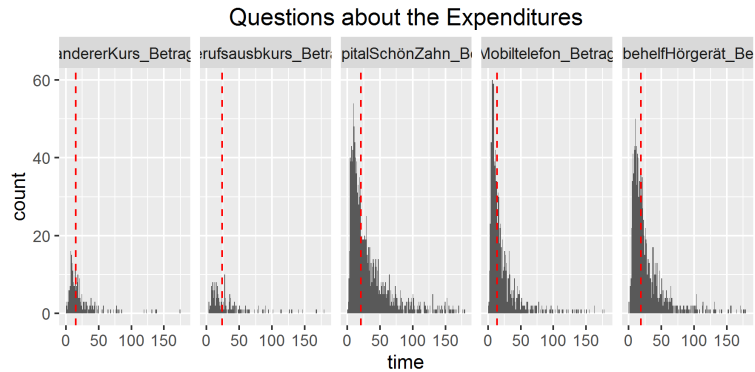
Figure 2: Histograms of five questions asking about the expenditures occurred for (vocational) training programs (subfigure 1 and 2), health expenditures (subfigure 3), mobile phone (subfigure 4) devices and glasses or hearing aid instruments (subfigure 5)

5.2 Male versus Female Respondents

Next we compare the time needed to complete a survey page between men and women. Since we are now more interested in the distribution of the times instead of the absolute numbers, we consider density plots instead of histograms.

We show only the density plots for the questions about expenditures, but the result is the same for all the questions: Neither the form of the distribution nor the median differs significantly between male and female respondents.
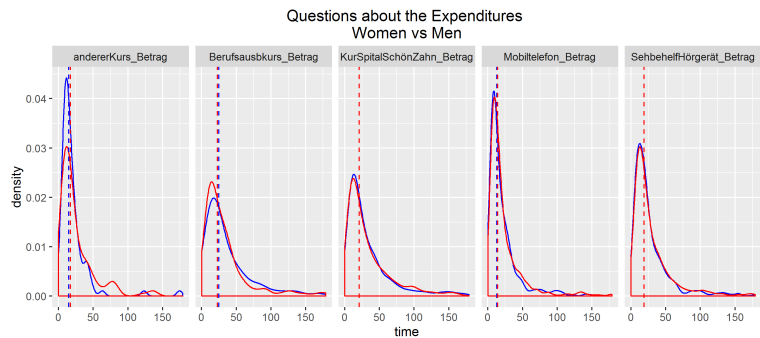
Figure 3: Density plots of the five questions about expenditures as before, differentiated between male (blue) and female (red) respondents.

5.3 Age Groups

Next, we want to examine if the time needed to complete a page/question differs between age groups. Therefore we define three age groups: Respon-

dents between 15 and 40, Respondents between 41 and 60 and Respondents older than 60 years. Since too few elderly respondents attended a vocational program, we excluded that question.

Whereas we see almost no difference among the age groups for the yes/no questions, Figure 5 indicates slightly that elderly people do need longer to complete questions about their expenditures.
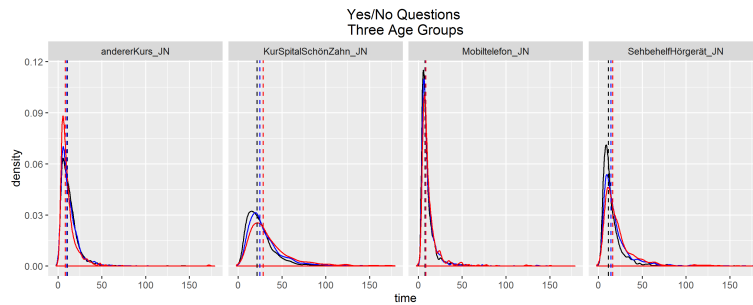

Figure 4: Time needed to complete Yes/NO Questions, for three age categories: $15 - 40$ years (black), $41 - 60$ years (blue) and > 60 years (red).
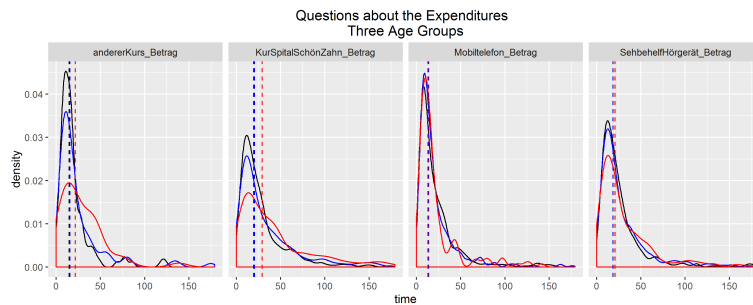
Figure 5: Time needed to complete questions on expenditures, for three age categories: $15 - 40$ years (black), $41 - 60$ years (blue) and > 60 years (red).

5.4 CAWI vs. CAPI

It is especially interesting to be able to compare the times needed to complete a page between respondents who were interviewed (CAPI) and those respondents who completed the survey on their own via a websurvey.

Whereas no difference in the response times is observable for the yes/no questions, respondents interviewed (CAPI) seem to be considerably quicker than respondents completing the survey via CAWI.
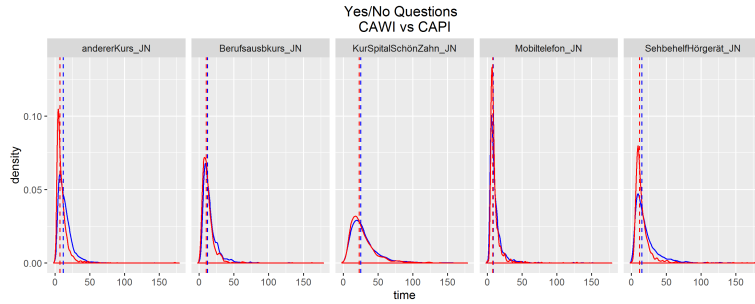
Figure 6: Time needed to complete yes/no questions, for respondents completing the survey themselves via CAWI (blue) and respondents interviewed via CAPI (red)
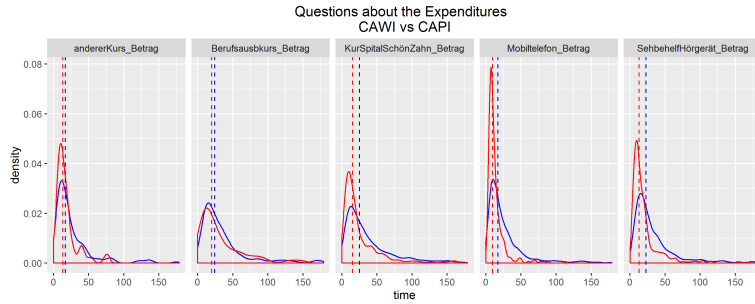

Figure 7: Time needed to complete questions on expenditures, for respondents completing the survey themselves via CAWI (blue) and respondents interviewed via CAPI (red)

5.5 CAWI vs. CAPI and Age Groups

Finally, we want to examine if the age differences observed in Section 5.3 are observable in CAWI as well as CAPI mode.

Surprisingly, the age differences are even magnified among the respondents using CAWI mode, see Figure 8, whereas almost not observable among the respondents being interviewed (CAPI), see Figure 9

We can think of two explanations for this rather surprising result. First, it is possible that elderly people need more time to navigate through the online questionnaire. Second, it is also conceivable that interviewer put respondents under more time pressure than they would take for themselves if they are on their own. In any case, more research is necessary on this question in the future.
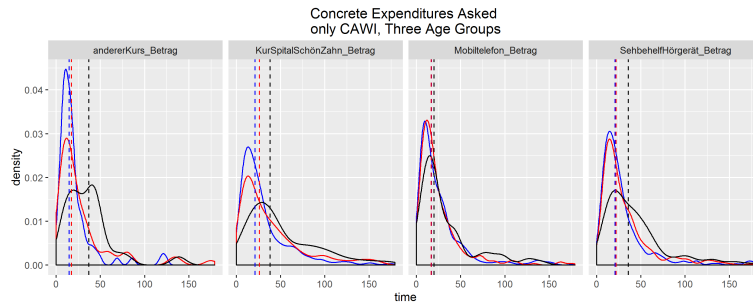
Figure 8: Time needed to complete questions on expenditures, for three age categories: $15 - 40$ years (black), $41 - 60$ years (blue) and > 60 years (red), only CAWI mode considered
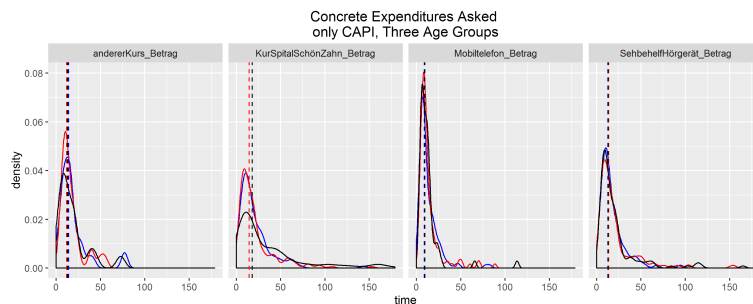

Figure 9: Time needed to complete questions on expenditures, for three age categories: $15 - 40$ years (black), $41 - 60$ years (blue) and > 60 years (red), only CAPI mode considered

6 Summary and Outlook

Through our in-depth analysis of response times, we have detected several interesting relationships between categories of respondents and response times. We have come a big step closer to our goal to identify incomprehensive questions by looking at the paradata, but there is still a lot to do.

Aside from the response times, we also want to elicit other measures such as the number of times clicked on a question/page, the number of explanatory messages per question/page or the number of plausibility warnings from the paradata.

Of course our feedback about the possibilities and difficulties in using the paradata is a valuable input for the IT department of Statistics Austria, who can develop the software STATsurv further for upcoming surveys.

References

F. Kreuter. *Improving Surveys with Paradata*. Wiley, New Joursey, 2013.

Statistik Austria. Verbrauchsausgaben, Sozialstatistische Ergebnisse der Konsumerhebung. Homepage Statistik Austria, 2012.