



Improving surveys with paradata: analytic uses of response time

A study based on process information of China Family Dynamic Survey

Yi Shi¹ · Jun Feng² · Xiaoqin Luo¹

Received: 31 August 2017 / Accepted: 25 June 2018 / Published online: 18 September 2018
© The Author(s) 2018

Abstract

This study uses paradata and survey data from the China Family Dynamic Survey to analyze the issue of response time in a questionnaire survey. The survey questions were classified into three levels and we compared the inter-group differences of response time between four kinds of questions which included information inquiries, logical calculations, intimate information and logical judgments. We used OLS method to perform a regression analysis to identify the impediments to response time and discussed the roles played by the survey locale, and the characteristics of interviewees and those of interviewers. The major impediments that prolonged response time were factors that impeded comprehension of the survey. The impediments included the difficulties interviewees had understanding questions, questions that did not take into account the different cultural backgrounds of interviewees, the complexity of questionnaires, and the performance of interviewers. Finally, this paper puts forward some suggestions for questionnaire design and investigation processes to reduce impediments to comprehension and improve the ability of interviewers.

Keywords Survey quality · Response time · Paradata · Questionnaire design

✉ Yi Shi
simon6510@163.com

Jun Feng
fengjunhr@126.com

Xiaoqin Luo
xiaoqinl@126.com

¹ China Population and Development Research Center, Beijing, China

² National Academy of Economic Strategy, Chinese Academy of Social Sciences, Beijing, China

1 Introduction

With large-scale surveys being used in many countries, methods of survey quality assessment are getting attention worldwide and are becoming more diverse. However, assessments mainly focus on the internal consistency of a single survey or comparisons between surveys. The development of research methods that can evaluate response behavior and non-sampling errors in social surveys is always a big challenge. Interviewees busy coping with the pressures of daily life often lack the time and energy to complete a survey. In addition to the impacts of the survey locale and characteristics of the interviewers, in some cases interviewees give answers to questions that are not entirely truthful (Feng 2010). More and more researchers have come to realize that response behavior is directly related to the quality of survey data (Fowler 2010). Non-sampling errors are generally believed to result from questionnaire design, the interview process and response behavior (Lessler and Kalsbeek 1992). Research on response behavior to a survey needs to rely on paradata from that survey. Paradata concerns the context and processes surrounding survey data collection (Evans 1996; Lin et al. 1996; Kreuter 2013), and it is difficult to observe and record in traditional surveys. Only a handful of surveys have recorded paradata, but the paradata are rarely published, resulting in slow progress in the development mining techniques for paradata (Zhang et al. 2003; West 2018).

Both paradata and survey data can be used to explain response behavior. Response time, which reflects the understanding interviewees have of survey questions and their willingness to answer questions, provides important information for questionnaire design and is of value for quality evaluations. In addition, data collection methods also affect response rates and quality of survey data (Mac Elroy 2000). “In particular, the interaction between length of survey (both in terms of content and number of questions) and incentive (either total incentive offered as a prize package or the approximate value of the incentive on an individual basis) has been thought to influence the quality of survey” (Johnson et al. 2006). Specifically, during survey design, the factors affecting the questionnaire include whether the questionnaire is acceptable overall, whether survey plans are well-designed, whether questions are concise and whether questions are put in the correct order. When seeking interviewees, the main factor affecting survey quality is whether the interviewees’ contact address and telephone number are obtained. Additionally, special tracking efforts (record of interviewees’ tracking and inquiry information) and well-trained interviewers also influence survey quality (Sun et al. 2011). During the stage of data collection, the quality of the survey depends on the making of appointments with interviewees, conducting interviews at different times, being allowed multiple to access interviewees, the gender of interviewees, establishing trust relationships, and whether there is compensation for participating (He et al. 2008).

Overall, the methodologies used to evaluate survey quality are inadequate and this creates enormous barriers to evaluating the quality of social surveys. The most common approach to evaluations is to compare key survey variables with

data from previous surveys. But to be effective, this approach needs to draw information from many successful investigations. Due to the lack of data support, prior research has focused mainly on the study of survey questions that are not answered for various reasons. A few studies have analyzed response time. The response time reflects the understanding interviewees have of the questions and their willingness to answer. Analyzing the characteristics of response time can make good use of paradata for quality assessments, and can also provide valuable reference points for the future research.

2 Concepts, data and methods

2.1 Concepts

Response time can be defined as the time required for the interviewee to complete the entire questionnaire or answer a single question. The concept can be divided into two types: response time for the questionnaire and response time for each question. Response time for the questionnaire refers to the total time spent on the questionnaire from start to finish. Beginning the questionnaire is marked by clicking the “Start” button on the tablet and the end is marked clicking the “Submit” button. The time between the two clicks is the response time for the questionnaire. The response time for a particular question refers to the time between the beginning of n th question and the beginning of $(n + 1)$ th question. After answering the $(n - 1)$ th question, interviewees need to click the “Next Question” button to move to the n th question. The time between two “Next Question” clicks is the response time for the n th question. The recording of response times is heavily dependent on the use of CAPI (Computer Assisted Personal Interviewing). CAPI can record the response behaviors of people conducting or taking surveys; e.g., the time interviewers spend asking questions, the time interviewees spend reacting, and the time interviewers spend on clicking. Therefore, the notion of response time used in this paper does not simply refer to the reaction time of interviewees, but the time spent on all of the actions needed to complete a questionnaire or a single question.

2.2 Data

The China Family Dynamic Survey, which was conducted in 2014, used a stratified multistage PPS design and covered 31 provinces, autonomous regions and municipalities, targeted 32,500 families from 1624 sample communities in 1560 town or urban districts distributed among 233 prefecture-level cities and 321 county-level cities and districts. A total of 32,494 families were actually interviewed. During the face-to-face interviews, interviewers used the CAPI system on a laptop or PAD to recorded process information such as the response time for the whole questionnaire or a single question. The Survey used three questionnaires: the Family Questionnaire, the Individual Questionnaire and the Community Questionnaire. The Individual Questionnaire was used in interviews with children, teenagers, adults and

the elderly. Adults formed the largest sample of those who completed the Individual Questionnaire and this study uses these adult questionnaires as the object of analysis. After invalid questions were removed, there were 47 valid questions in the adult questionnaire that focused on the health and lifestyle, work and income, marriage and contraception, and migration and social security of interviewees.

2.3 Methods

The purpose of this study is to analyze key determinants of response time in surveys. First, this study used paradata from adult questionnaires that were part of the China Family Dynamic Survey to compare response time length for different questions and subjects. Methods used in other studies to ensure the reliability of measurements were also referred to. Second, because response time is a continuous variable which should follow a normal distribution, OLS regression is an appropriate method to analyze the factors influencing response time.

3 Response time in different groups

3.1 Response differences by variable

Considering the types of question, content and responses, the adult questionnaire can be divided into information inquiries, logical calculations, intimate information and logical judgments. These questions can also be divided into subcategories for personal information, employment information, time calculations, number calculations, sex and reproductive health, subjective judgments, knowledge judgments and conceptual judgments. Both four-classification and eight-classification show response times for the various groups that are significantly different (Table 1).

3.1.1 Response time for information inquiries

Familiarity with the content of questions is one of the major factors influencing response time. While familiarity depends on an individual's life experiences, memory, and cognitive level, information inquiries are generally easy to answer and had an average response time of 9.73 s. Personal information questions about gender, height and weight are especially easy to understand and had average response times of 9.10 s. However, the response process for employment information, such as job experience, occupation, industry type, and work content, is a bit more complicated and the average response time was longer (10.28 s). It is worth noting that the response time for information inquiry questions varied with the individual's memory and cognitive level. Questions about industry type required interviewees to have an understanding of classification standards, so the average response time for this question was much higher than for other information inquiry questions.

Table 1 Outcomes of the clustering based on response time (unit: s)

Classification	Groups	Questions contents	Mean1	SD	Mean2	Mean3
Information inquiries	Personal information	Year of birth	4.77	17.45	9.10	9.73
		Height	3.47	23.99		
		Weight	12.84	23.25		
		Smoker or non-smoker	11.72	10.09		
		Health examination	10.22	23.33		
		Cellphone number	13.76	159.95		
		Time of marriage	6.91	65.91		
		Employment	9.79	39.18	10.28	
		Searching for job or not	10.98	19.90		
		Occupational	10.44	22.76		
		Industry type	16.06	30.72		
		Workload	5.50	11.82		
		Employment status	5.31	15.53		
		Mental/physical work	12.35	13.33		
Social security	11.82	26.81				

Table 1 (continued)

Classification	Groups	Questions contents	Mean1	SD	Mean2	Mean3
Logical calculations	Time calculations	Exercise time per day	16.36	20.59	12.66	12.78
		Study hours per week	14.80	15.95		
		Working weeks per month	14.45	31.79		
		Working hours per week	14.42	26.96		
		Housework time per day	14.16	15.47		
		Time to read news per week	12.73	13.23		
		Working days per year	10.58	15.68		
		Working days per week	6.26	11.17		
		Sleep hours per day	10.17	15.24		
		Number of cigarettes smoked	6.85	8.14	13.05	
		Personal annual income	12.77	73.26		
		Books read	15.45	15.35		
Intimate information	Sexual and reproductive health	Annual education expenditure	17.11	15.54	18.19	18.19
		Change of marriage	34.82	33.76		
		Contraceptive methods	22.42	121.23		
		Types of contraception	10.97	24.13		
		Sexual life	4.30	14.39		
		First sexual behavior	12.32	19.88		
		Whether contraception was used during first sexual behavior	26.38	39.55		
		The source of contraception	20.01	21.22		
		Pregnancy experiences	18.36	18.89		
		Pregnancy results	14.09	50.16		

Table 1 (continued)

Classification	Groups	Questions contents	Mean1	SD	Mean2	Mean3
Logical judgments	Subjective judgments	Work pressure	8.26	8.81		20.82
		Working environment	7.49	45.95	12.26	
		Work-life conflict	9.83	10.69		
Knowledge judgments		Economic pressure	23.47	70.94		
		Reading content	19.64	14.27		
		Health issues of concern	12.85	16.39	14.17	
Conceptual judgments		Chronic disease knowledge	10.02	18.87		
		Migration experience	62.56	95.12		
		Migration distance	30.14	44.84	38.88	
		Length of residence	23.93	34.41		

3.1.2 Response time for logical calculations

Compared with information inquiry questions, the answers to some questions required both basic knowledge of the question's content and a logical calculation. Therefore, the response time for these questions was longer than for information inquiry questions. Unit standardization is an important condition to ensure the efficiency of logical calculations. Compared with complex questions, such as questions concerning reading quantity and annual consumption, the response time for questions requiring time computation is shorter.

3.1.3 Response time for intimate information

Each culture has its own customs and social norms, and as a result, individuals have different attitudes and ways to deal with questions about intimate information. The average response time for intimate information questions was 18.19 s, higher than the response times for information inquiries and logical calculations because personal privacy issues were involved. Moreover, the more intimate the questions were, the longer the response time was. Questions which involved personal memories like the dates of changes in marital status and the first sexual behavior were relatively difficult to answer and the average response time to such questions was more than 20 s. Questions asking whether an individual had had sexual encounters or been pregnant did not require a lengthy thought process to answer, and the average response time was relatively short (generally no more than 10 s).

3.1.4 Response time for logical judgments

Of the four kinds of questions, logical judgment questions had the highest average response time of more than 20 s. This group of questions was more influenced than other kinds of questions by questionnaire design and the investigative process, meaning that interviewees needed more time to figure out the meaning of the concepts. However, the questions based on subjective judgments were relatively easy to answer, and the average response time was 9.73 s. Questions such as "What do you like to read?", "What health problems are you worried about?", and "What chronic diseases do you have?" were harder to answer and the average response time was longer (14.17 s). In general, both interviewers and interviewees had to spend more time understanding questions that touched upon relatively subjective concerns like pressure or relatively abstract and complex concepts like migration experiences, especially in the case of questions that were interconnected with other questions. Response times were longer as a result. In the case of migration, interviewees often had to ask first for clarification of what was meant by the term migration before they could answer questions concerning the frequency, distance and time span of migration. The average response time for these questions approached 40 s and was much longer than the response time for other judgment questions. It is worth noting that once the concept of migration was clarified during the answering of the first question, interviewees had shorter response times answering additional questions about migration.

3.2 Response time differences in different survey locales

The internal factors that affected the total response time for the questionnaire were content and design of the questionnaire. The external factors were related to the survey locale. By comparing total response time for questionnaires completed in the eastern, central and western regions of China, we can see that differences in location had an important influence. The total response time for questionnaires completed in Eastern China was the shortest, with an average time of only 7.82 min. The response time in Central China was in the middle (10.73 min), the response time in Western China was higher (11.17 min), and the response time in Northeastern China was the highest (14.9 min). At the same time, the standard deviation of response time in Eastern China was far below that of other regions (Table 2).

3.3 Response time difference connected to the characteristics of interviewers

In addition to the effects of questionnaire design and the locale where the survey was administered, individual characteristics of different interviewers also influenced the investigative process. Overall, when the interviewers were female, the average response time for completing questionnaires was higher than when interviewers were male. Response times were lower when interviewers had a high education level and higher when their education level was low. It is notable that interviewers who were familiar with CAPI could not complete the questionnaire in a short time, and interviewers whose field was family planning spent more time completing questionnaires than interviewers from other occupations (Table 3).

3.4 Response time differences connected to the characteristics of interviewees

The characteristics of interviewees were directly related to their level of understanding and how they responded to the questionnaires; these are factors that cannot be ignored. In terms of gender, the average response time of males was lower than that of females, and the standard deviations of both groups were high. Interviewees with agricultural household registrations (*hukou*) had a lower average response time than

Table 2 Response time for completing the whole questionnaire in different regions (unit: min)

	Mean	SD	Kurtosis	Skewness	N
Regions					
Eastern	7.82	9.51	1208.82	27.17	7408
Central	10.73	23.49	253.02	12.55	4585
Western	11.17	20.24	82.78	7.78	4669
Northeastern	14.90	27.42	23.45	4.47	1626
Community types					
Rural	9.44	17.34	439.96	15.96	10,360
Urban	10.81	20.39	79.42	7.91	7928
Total	10.03	18.73	231.00	11.60	18,288

Table 3 Response times for completing whole questionnaires given by different interviewers (unit: min)

	Mean	SD	Kurtosis	Skewness	N
Gender					
Female	10.34	18.85	118.92	9.08	11,947
Male	8.71	14.57	1083.03	25.78	6120
Education					
Junior high school	7.30	1.33	0.07	0.94	9
High school	10.29	19.26	101.76	8.88	2494
College	9.52	17.41	520.71	17.42	7898
Undergraduate	9.42	15.71	122.62	9.68	6831
Postgraduate	7.03	2.95	3.23	1.18	113
Proficiency in CAPI					
Incapable	7.06	3.01	1.38	1.13	122
General	8.20	9.30	476.36	16.96	5045
Skilled	10.17	19.34	259.84	12.10	12,178
Occupation					
Other	7.86	7.10	192.02	11.28	847
Family planning	9.66	17.34	310.11	13.21	16,498
Total	9.57	16.99	320.80	13.41	17,345

interviewees with non-agricultural *hukou*, and the kurtosis and skewness of the former were higher than the latter (Table 4).

Education is a reflection of the knowledge, understanding and cognitive abilities of interviewees and greatly affects them. In general, the higher the education level of an interviewee, the shorter the response time. Specifically, graduate students had the lowest response time, using only 7.5 min to complete the whole questionnaire, a time that was far lower than that of other interviewees with less education. Interviewees who had received no formal education had the highest response time (10.73 min), a time far greater than those who had received formal education. However, the average response time of interviewees with a lower level of secondary education was lower than that of interviewees with a higher level of secondary education (Table 5).

Cultural sensitivity also needs to be considered as a factor affecting response time because it has a direct impact on response behavior. This view was confirmed by the difference in response times between ethnic minorities and people of Han

Table 4 Response time for completing the whole questionnaire by interviewees with different characteristics (unit: min)

	Mean	SD	Kurtosis	Skewness	N
Gender					
Male	9.72	17.87	301.42	13.11	15,037
Female	11.51	22.25	72.61	7.47	3248
Hukou					
Agriculture	9.49	17.65	401.77	15.49	11,432
Non-agricultural	10.98	20.81	62.62	7.21	6096

Table 5 Response time for completing the whole questionnaire by interviewees with different education levels (unit: min)

	Mean	SD	Kurtosis	Skewness	N
Unschoolled	10.73	17.22	52.86	6.62	531
Primary school	10.22	18.51	111.29	9.16	3852
Junior high school	9.87	19.81	342.72	14.25	7736
Senior high school	10.27	18.46	107.61	8.97	3460
College degree	10.28	17.43	71.80	7.78	1369
Undergraduate	9.48	15.98	96.75	9.28	1182
Postgraduate	7.50	4.37	12.93	3.06	151

nationality. The response time of many ethnic minorities was much higher than that of people of Han nationality. Among the ethnic minorities, the response time of Tibetan, Dong and Tujia minorities was very high, and this was probably related to the different cultural backgrounds of these interviewees that affected understanding of the questionnaire's content (Table 6).

4 The determinants of response time for the questionnaire survey

4.1 Model design

A descriptive analysis indicates that response time, which is the record of the survey process, is influenced by multiple factors such as questionnaire design, survey locale, and the characteristics of interviewers and interviewees. However, it is difficult to test the diverse ways in which these factors influence response time using descriptive analysis alone. Hence, this research analyzes the impacts of the survey

Table 6 Response time for completing the whole questionnaire by different ethnic groups (unit: min)

	Mean	SD	Kurtosis	Skewness	N
Han	9.51	16.87	336.74	13.77	16,687
Zhuang	6.80	2.81	0.68	0.71	250
Hui	12.22	25.32	35.49	5.80	223
Tujia	19.26	32.57	8.29	3.01	148
Miao	12.91	22.05	22.11	4.43	126
Li	9.06	4.01	-0.15	0.77	122
Manchu	19.25	36.14	16.83	3.77	120
Tibetan	48.35	58.75	7.55	2.17	90
Bai	8.79	3.70	2.14	0.96	79
Yi	10.61	14.09	68.00	8.04	76
Dong	60.16	89.00	2.34	1.65	58
Mongol	10.60	17.29	50.44	6.91	58
Total	10.04	18.73	230.97	11.60	18,285

locale and characteristics of both interviewee and interviewer on total response time and the response time for key questions by using OLS method. The influence of the skip pattern is also controlled in the model because the number of questions that interviewees do not answer has an effect on total response time for the questionnaire (Fig. 1).

4.2 Characteristics of interviewees

Our findings show that total response time for completing the entire questionnaire was significantly lower for male interviewees than it was female interviewees. The response time of Han nationality Chinese for completing the entire questionnaire was significantly lower than that of ethnic minority Chinese, and the response time of interviewees with higher levels of education was significantly lower than that of interviewees with lower levels of education. However, interviewees with non-agricultural *hukou* had remarkably longer response time than interviewees with agricultural *hukou*. Questions about migration are difficult to answer and this is the main reason that the response time for completing the total questionnaire was longer for non-agricultural interviewees. For instance, interviewers asked “Have you ever had to leave your residence (refers to the town/street where you have lived) for more than six months?” Interviewers also asked “When was the last time you left your residence for more than half a year?” These questions about migration were generally relevant for interviewees from rural areas (most migrants in China have agricultural *hukou* and move from rural to urban areas), but were likely to confuse interviewees with non-agricultural *hukou* who had always lived in urban areas. Voice records of the survey process indicate that non-agricultural interviewees often asked interviewers for more details about migration questions and in some cases even terminated the interview. In fact this study found that the extent to which survey questions were universally applicable could change the response time, something that other studies have also found to be the case (Yan and Tourangeau 2008). A study of paradata from the US National Family Development Survey indicates that African-American, Hispanic and other ethnic minority interviewees needed more time to answer questions than white interviewees because the design of survey questions embodied the values of mainstream culture. Other studies have found that the reading habits of

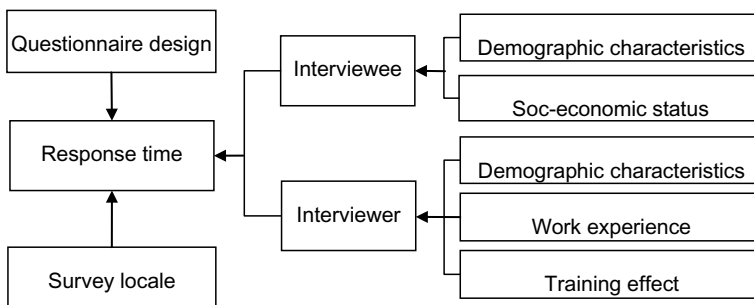


Fig. 1 Model design of determinants of the response time

interviewees have a significant influence on response time. Interviewees who spend more time reading books and news are more likely to be affected by the universality of questions (Couper and Kreuter 2013).

Based on the data for the response time to logical calculation questions, retired and unemployed people answered economic questions more quickly than other interviewees. This is because, first, they had a single source of income and the amount was easy to calculate and, second, they were less likely to be concerned about keeping the amount of their income confidential. On the other hand, interviewees experiencing high levels of economic pressure had longer response times to such questions because they needed more time to calculate their incomes and hesitated to report the amount. With respect to the response time for subjective judgment questions, interviewees participating in labor markets had faster reactions to questions about economic pressure, work stress and work-life conflicts. Farmers, students and retirees had lower response speed because the questions were not relevant to their current life experiences. In the case of knowledge judgment questions, there was a complex relationship between the education levels of interviewees and their responses. For example, interviewees with higher levels of education took less time to answer questions about chronic disease types while interviewees with lower education levels took less time to answer questions about health awareness (refers to health issues of concern in Table 1). A random selection of voice recordings of interviews indicates that interviewees with high education levels paid more attention to their health status and quickly understood questions about chronic diseases (like hypertension, heart disease, diabetes, arthritis, slipped disc, malignant tumors, and cervical spondylosis). However, a number of knowledge judgment questions covering subjects such as smog, water safety, food safety, patient care, medical care, medical insurance, and waste disposal were unclear in intent and confusing. Interviewees usually asked for clarification, for example, asking interviewers whether water shortage problems are part of water safety problems or asking interviewers to explain the difference between “medical care is difficult”, “medical care is expensive”, and “medical insurance”. These questions were raised more frequently by well-educated interviewees who were more likely to ask for clarification than the less educated, who did not ask additional questions and responded more quickly.

With respect to the response times for conceptual judgment questions, interviewees with high education levels took less time than interviewees with low education levels to answer questions about migration experience. Interviewees participating in labor markets needed more time to answer these questions. Moreover, because the concept of migration is multi-faceted and complex, the response time increased rather than decreased for interviewees with good learning and reading habits.

4.3 Characteristics of interviewers

With respect to total response time for completing the questionnaire, male interviewers took less time than female interviewers, and interviewers with high levels of education took less time than those with low levels. It is worth noting that interviewers with higher CAPI proficiency actually took more time to complete the

questionnaire. Interviewers engaged in family planning work took significantly more time than interviewers working in other fields. It is interesting that interviewers who used CAPI proficiently and had family planning work experience could reduce the response time for many questions, but the total response time for completing the questionnaire was, unexpectedly, greater. A random selection of voice recordings of interviews suggests that family planning workers were more familiar with CAPI than other interviewers and had good relationships with interviewees in their area. Interviewees and interviewers tended to interact more during the part of the survey process that covered subjective judgment questions. Conversations often turned to other topics, resulting in increased response time for completing the whole questionnaire. The characteristics of interviewers did not significantly influence interviewees' response time to logical calculation questions. However, interviewers who were family planning workers could significantly reduce the response time to questions about sex and reproductive health. With respect to knowledge judgment questions, the education and CAPI proficiency of interviewers had no significant influence reducing time interviewees needed to think and respond to questions about whether they suffered from chronic diseases or had other health issues. These findings confirm that it was mainly the characteristics of the interviewees, not of the interviewers, that influenced the response time for knowledge judgment questions. With respect to conceptual judgment questions, interviewers with high education levels and CAPI proficiency could give quick, concise explanations of the concept of migration and this helped to reduce the response time of interviewees to these questions.

4.4 Survey locale

Our results show that interviewees in Northeastern China took the longest time to answer the questionnaires. Interviewees in Western and Central China took less time, and response time of interviewees in Eastern China was the least. These findings reveal that reaction speeds differed from region to region, and also reflect the different backgrounds and habits interviewees from different regions have when they are answering questions. The explanation could be that, compared to people living in the relatively more developed Eastern China, people in Central and Western China have a weaker sense of time and more biased understandings which lengthened their response times. In contrast, the sense of time and cultural customs are the major reasons for long response times in Northeastern China where the average education level is high. This conclusion is confirmed by the much shorter response times in the Northeast for questions about migration experiences and other abstract concepts. Furthermore, the response times for subjective judgment, knowledge judgment, and conceptual judgment questions in Northeastern China were far lower than the response times for these questions in other regions. There was also no significant difference in the response times between Northeastern China and other regions for information inquiry, logical calculation, and intimate information questions. This implies that interviewees in Northeastern China were quick to understand the survey questions. However, the time span from clicking "start answer" to answering the first question and from completing the final question to clicking "submit" were

longer for interviewees from the Northeast than they were for interviewees from other regions. This lengthened the response time for completing the entire questionnaire. In other words, the interviewers and interviewees in Northeastern China spent less time interacting about questionnaire content, but more time before starting the survey and after completing the process (Table 7).

5 Conclusions and discussion

This article uses a study of paradata to discuss the relationships between questionnaire design, the characteristics of interviewers and interviewees, the survey locale and the response time for the questionnaire as a whole and for individual questions. First, this study finds that there is a significant correlation between the question type and the response time; for example, interviewees needed more time to answer logical judgment questions than other questions. Second, the characteristics of interviewers and interviewees had a significant impact on response time. Well-educated, male Han interviewees responded more quickly than interviewees with other characteristics, and well-educated male interviewers also played a significant role in decreasing response times. Third, there is a certain similarity in response times for questionnaires completed in the same area. However, response times varied considerably from region to region. The time spent starting and ending the interview process in Northeastern China was remarkably longer than in other regions. These conclusions reflect the fact that variations in survey response times are the result of complex factors, but at the same time, the conclusions also suggest that analyzing response times provides a basis for controlling investigative quality. The conclusions can also support a design process that improves questionnaire quality, especially by reducing non-sampling errors caused by cross-cultural bias and regional differences in large-scale social surveys. The conclusions also provide references in support of innovative decision-making that can help in recruitment and the development of training programs for interviewers, and assist in the implementation of measures to make investigations more robust.

5.1 Understanding limitations and response efficiency

The limitations of questionnaires are a major factor affecting response time. There are three principal aspects to limitations: the ability of interviewees to understand the questionnaire, culture compatibility of the questionnaire and the interviewees, and the complexity of the concepts in questions. First, the empirical analysis results show that the higher the education level of interviewees, the shorter their response times. Results for a variety of different types of questions show that interviewees with higher levels of education understood questions more quickly and responded significantly more quickly than other interviewees. Second, although the survey interviewed ethnic minority people whose cultural background differs from that of Han Chinese, questionnaire design was not sensitive to different cultural backgrounds, and this was not conducive to improving response efficiency. The response time of ethnic minority interviewees was quite long. Questionnaire design did not

Table 7 Determinants of response time for adult questionnaires and the key questions

	Whole questionnaire	Income	Sexual behavior	Economic pressure	Work stress	Work-life conflict	Chronic disease	Health concern	Migration experience
<i>Interviewee's characteristics</i>									
Male	-1.159**	-0.373	0.548	0.738	0.237	-0.224	0.0398	-0.989**	-2.620**
Age	-0.217	0.0884	-0.510	0.105	-0.123+	0.0324	-0.145+	0.114	-0.236
Age-squared	0.003+	-0.001	0.007	0.002	0.002*	0.001	0.003**	-0.001	0.006
Non-agricultural hukou	2.092***	-0.401	-0.551	-1.103*	-0.340	0.305	-0.157	1.845***	2.930**
Han ethnicity	-3.363***	-0.252	0.614	-3.052***	-0.0126	-0.858*	-0.967**	0.282	1.806
Education level (reference group: senior high school)									
No formal education	1.403	0.0537	-0.833	-0.977	-0.236	0.460	-0.160	-3.994***	-0.477
Primary school	0.853+	0.417	1.103	-0.581	0.652*	0.451	0.188	-2.370***	-2.656
Junior high school	0.551	-0.261	-0.073	-0.053	0.580**	0.235	0.144	-1.257***	0.238
College	-0.542	-0.006	0.898	-0.607	-1.292***	-0.697+	-0.369	0.989*	-0.402
Undergraduate	-1.923**	-0.362	-0.619	-0.563	-1.778***	-0.955*	-1.492***	0.652	-3.249*
Postgraduate	-3.701*	-1.662	-0.534	-1.722	-3.308***	-0.988	-2.564**	0.760	-11.43**
Employment status (reference group: farmer)									
Non-agricultural employment	-2.273***	-0.296	-0.0870	-4.835***	-0.286	-0.493*	-0.0810	0.156	5.995***
Retired	-3.448***	-2.932***	0.092	-2.256+	-1.118*	-1.760**	-0.078	-0.385	6.132**
In school	-4.053	-0.166	-3.317	1.442	-2.055	-3.724	0.534	-1.107	2.164
Unemployed	-1.899**	-1.905***	-0.552	-3.994***	-1.130**	-0.892+	0.138	0.606	4.861***
Amount of reading	0.057***	0.016	-0.027	-0.003	-0.001	-0.008	0.004	0.091***	-0.010
Length of study	0.036	0.063*	0.010	0.011	0.024	0.056**	-0.002	0.274***	0.447***
Amount of news read	0.053*	0.035+	-0.016	0.034	0.006	0.041*	0.0194	0.113***	0.329***
Personal annual income	-0.001	-0.001	-0.001	0.001*	-0.001	0.000	0.000	0.002*	0.001***
Economic pressure	-0.099	0.557***	-0.385	0.416*	0.282***	0.346***	0.216**	-0.102	0.880**
<i>Interviewers' characteristics</i>									

Table 7 (continued)

	Whole question-naire	Income	Sexual behavior	Economic pressure	Work stress	Work-life conflict	Chronic disease	Health concern	Migration experience
Male	-1.488***	0.195	-1.329	0.788+	-0.246	-0.226	0.0217	-1.133***	-1.29
Education level (reference group: senior high school)									
Junior high school	-4.350	-0.230	0.001	-6.011	0.721	-0.682	-1.737	-2.902	3.596
College	-1.327***	0.552	1.373	1.683**	0.669**	0.383	0.302	0.098	1.61
Undergraduate	-1.315**	0.498	0.303	0.585	0.397	0.116	0.320	-0.389	-0.494*
Postgraduate	-3.931*	-0.669	-4.206	4.719+	-0.853	-0.877	-0.964	-2.953*	-0.083*
CAPI proficiency (reference group: general)									
Incapable	0.119	0.127	0.0520	-2.992	0.213	-0.142	-1.411	0.622	-0.686
Skillful	2.151***	0.407	0.830	-1.381**	0.506**	0.296	-0.0437	0.319	-1.705*
Family planning worker	2.159***	0.445	-3.448+	1.334	0.396	0.204	0.137	-0.597	1.089
<i>Survey locale (reference group: Eastern China)</i>									
Central China	2.226***	1.863***	1.041	1.450**	0.926***	1.647***	0.925***	0.695*	2.599**
Western China	2.250***	2.121***	1.070	1.585**	0.666**	1.013***	0.981***	0.843**	1.969*
Northeastern China	6.411***	0.502	0.360	-3.330***	-0.369	0.0153	-0.215	-1.237**	-9.368***
Control skip pattern or not	Yes	No	No	No	No	No	No	No	No
Constant	17.200***	4.905*	15.20*	17.49***	7.858***	6.331***	10.40***	8.664***	51.88***
N	15,864	15,834	1220	15,527	13,208	13,210	15,533	15,531	15,825
R ²	0.036	0.015	0.013	0.029	0.022	0.013	0.013	0.061	0.026

†*p* < 0.10; **p* < 0.05; ***p* < 0.01; ****p* < 0.001

take cultural differences into consideration and, as a result, ethnic minority interviewees often had inaccurate understandings of questions. Third, the more complex the question was, the more difficult it was for interviewees to understand the true meaning of the question and for interviewers to explain the question to interviewees. This problem was particularly noticeable in the category for logical judgment questions. The education levels of interviewers and interviewees had no significant influence on the length of the response time to this type of question. The educational levels and cultural backgrounds of interviewees are givens, meaning that questionnaire design must take these factors into account. Using simple, everyday language to ask questions, improving the culture compatibility of the questionnaire, and simplifying the computations needed to answer questions will reduce response time and make the interview process more efficient.

5.2 Development of good interviewers

Research into social investigation methods generally concludes there is a positive correlation between the quality of interviewers and the efficiency of investigations. This conclusion, however, does not apply to all cases, because even well-trained interviewers cannot reduce response times for all types of questions. First, careful recruitment and training of skillful, well-educated interviewers can contribute to reduce the response time for logical judgment questions. In other words, there is a positive correlation between the effects of the training interviewers received and their efficiency in dealing with complicated questions. Second, skillful, well-educated interviewers may increase the response time for subjective judgment questions. This shows that interviewers tend to over interpret and this can have a negative impact on survey efficiency to some extent, especially when interviewees must make subjective judgments to answer questions. Therefore, training is needed to help interviewers find middle ground between explanations that provide insufficient detail and excessive, overly detailed interpretations.

5.3 The relationship between response time and survey quality

It is important to pay attention to the light paradata sheds on items such as non-response behavior and response time. This is crucial because current social sciences research procedures need more and better data to control completely various types of errors in the survey process. We must explore every possible way to reduce such errors. The relevance to survey results of response time, which serves as the time record of an investigation, deserves more attention because it can serve as an important indicator to measure non-sampling errors. First, there is an essential correlation between the length of the response time and the results of an investigation. Some studies have concluded that there is a noticeable negative correlation between response time and response enthusiasm; longer response time for the baseline survey will decrease the probability of follow-up in the second round (Olson 2013). Therefore, starting from the premise of ensuring that a specific amount of data can be collected and then improving questionnaire design to reduce response time should be seen as a method to increase the follow-up rate. Second,

the length of the response time has a direct impact on the answer result. The longer the average response time to a question, the greater the discrepancies among the answers to the question and the higher the standard deviation. With logical calculation and knowledge judgment questions, for example, interviewees whose response times were more than two standard deviations were less objective, and there were more contradictions in the logical associations. The possibility of extreme values in the answers also increased. We should attempt to identify cases of investigations in which there may be problems with response time. Third, there is a significant correlation between the standard errors of the response time and of the answers. When response times have high standard deviation, there is also usually high standard deviation of survey data. Measuring the quality of the questionnaire design by the standard deviation of the response time, therefore, is an effective method to evaluate design quality. This article is limited to discussion of the characteristics of response time and the role it plays in the investigation process and in the relationship between interviewers and interviewees. This study does not offer a detailed look at the relationship between response time and the quality of the investigation. There has been no detailed discussion of the relationship between response duration and survey quality. Further in-depth studies that consider these issue would have crucial practical value and methodological implications for the improvement of social surveys.

Funding This study is funded by the China Social Science Foundation (Grant 16ZDA089)—Construction of China's population data platform (1949–2015).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

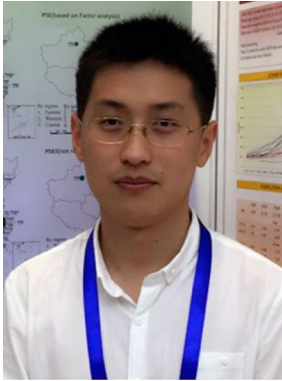
References

- Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society*, 176(1), 271–286.
- Evans, W. (1996). Computer-supported content analysis trends, tools, and techniques. *Social Science Computer Review*, 14(3), 269–279.
- Feng, X. (2010). *Research methods of sociology* (3rd ed.). Beijing: China Renmin University Press.
- Fowler, Floyd. (2010). *Design and evaluation of questionnaires*. Chongqing: Chongqing University Press.
- He, J., Liu, J., & Shu, X. (2008). The study of unanswerable error control in the sampling survey. *Statistics and Decision*, 257(5), 162–163.
- Johnson, T. P., Holbrook, A. L., Cho, Y. I., et al. (2006). Nonresponse error in injury-risk surveys. *American Journal of Preventive Medicine*, 31(5), 427–436.
- Kreuter, F. (2013). Improving surveys with paradata: Analytic uses of process information. *International Journal of Market Research*, 56(5), 691.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York: Wiley-Interscience.
- Lin, J., Ren, R., & Liu, F. (1996). Multi-channel parallel data acquisition technology. *Journal of Changchun College of Geology*, 1, 93–98.
- Mac Elroy B. (2000). Variables influencing dropout rates in Web-based surveys. *Quirks Marketing Research Review*, 583(7), 71.
- Olson, K. (2013). Paradata for nonresponse adjustment. *Annals of the American Academy of Political & Social Science*, 645(1), 142–170.
- Sun, Y., Zou, Y., Ding, H., et al. (2011). The analysis of rejection behavior in the follow-up survey—Taking the Chinese household dynamic tracking survey as an example. *Sociological Studies*, 2, 167–181.

West, B. T. (2018). Paradata in survey research. *Survey Practice*, 4(4), 1–8.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1), 51–68.

Zhang, X., Yun, S., Lu, S., et al. (2003). Parallel data mining research. *Computer Engineering*, 29(17), 58–59.



Yi Shi is an Associate Professor at the Department of Population Study, China Population and Development Research Center.



Jun Feng is a post-doctoral fellow at the National Academy of Economic Strategy (NAES), Chinese Academy of Social Sciences.



Xiaoqin Luo is a Senior Engineer at the Department Information and Network Management, China Population and Development Research Center.