Web Survey Paradata on Response Time Outliers: A Systematic Literature Review

Miha Matjašič[1]   Vasja Vehovar[2]   Katja Lozar Manfreda[3]

Abstract

In the last two decades, survey researchers have intensively used computerised methods for the collection of different types of paradata, such as keystrokes, mouse clicks and response times, to evaluate and improve survey instruments as well as to understand the survey response process. With the growing popularity of web surveys, the importance of paradata has further increased. Within this context, response time measurement is the prevailing paradata approach. Papers typically analyse the time (measured in milliseconds or seconds) a respondent needs to answer a certain item, question, page or questionnaire. One of the key challenges when analysing the response time is to identify and separate units that are answering too quickly or too slowly. These units can have a poor response quality and are typically labelled as response time outliers. This paper focuses on approaches for identifying and processing response time outliers. It presents a systematic overview of scientific papers on response time outliers in web surveys. The key observed characteristics of the papers are the approaches used, the level of time measurement, the processing of response time outliers and the relationship between response time and response quality. The results show that knowledge on response time outliers is scattered, inconsistent and lacking systematic comparisons of approaches. Consequently, there is a need to improve and upgrade the knowledge on this issue and to develop new approaches that will overcome existing deficiencies and inconsistencies in identifying and dealing with response time outliers.

1   Introduction

Survey researchers are not only intensifying the use of computerised methods for collecting responses to survey questions but also increasing the collection of different types of survey paradata (Couper, 1998) such as keystrokes, mouse clicks and response times. Paradata refers to data on the process of answering the survey questionnaire. The main purposes of using paradata are (1) to evaluate and improve survey instruments; (2) to provide insight into the respondent's behaviour (e.g. to measure the amount of information processing necessary to answer a question; see Mayerl, 2013); (3) to systematically detect, measure and analyse response quality; and (4) to potentially intervene in the answering process (e.g. when a respondent proceeds through the questionnaire too quickly).

---

[1] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; miha.matjasic@fdv.uni-lj.si
[2] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; vasja.vehovar@fdv.uni-lj.si
[3] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; katja.lozar@fdv.uni-lj.si

Within this context, response time measurement has been the most frequently used paradata approach. Response time is usually measured in milliseconds and refers to the time the respondent spends on answering an item or a set of items (e.g. question, page, block of questions or entire questionnaire). With the growing popularity of web surveys, where extensive paradata can be easily collected (see e.g. Heerwegh, 2003), the analysis of response time is increasingly related to the evaluation of response quality (e.g. satisficing and item nonresponse).

From a technical point of view, collecting response time paradata can be done on the client side or on the server side. The server-side paradata are collected at the server where the web survey resides and include details of respondents' visits to a particular web questionnaire. On the other hand, the client-side paradata are collected at the level of the respondent's device (with the help of JavaScript) and can include keystrokes, mouse clicks and response time at the item, page or survey level (see Callegaro et al., 2015).

Researchers have demonstrated that response time analysis is a highly useful approach particularly because it allows empirical insight into the respondent's behaviour (Mayerl, 2013; Kreuter, 2013; Gummer and Roßmann, 2015). Among other things, researchers have proven that respondents with less stable attitudes require more time to answer survey questions (Bassili and Scott, 1996; Heerwegh, 2003). Similarly, poorly designed survey questions (e.g. too long sentences and unique response categories) increase respondents' cognitive effort and therefore increase their response time (Bassili and Scott, 1996). Longer response time has been associated with deeper cognitive processing due to respondents' increased motivation or engagement (Callegaro et al., 2009), interruptions or multitasking (Stieger and Reips, 2010), increased complexity of questions (Yan and Tourangeau, 2008), as well as ambivalent attitudes in the case of opinion questions and a lack of knowledge in the case of knowledge questions (Heerwegh, 2003). On the other hand, very short response time, also called speeding (i.e. giving answers very quickly), has been associated with responding too quickly to give much thought to answers and is highly likely to arise when respondents are motivated primarily to finish the questionnaire rather than to provide careful and accurate responses (Greszki et al., 2015; Barge and Gehlbach, 2012; Zhang, 2013; Callegaro et al., 2015).

Both types of respondents' behaviour, either longer response time or speeding, may be viewed through the prism of cognitive models describing the mental processing of respondents' behaviour (for overviews, see Sudman et al., 1996; Callegaro et al., 2015). The most common reference model is based on four cognitive steps (Tourangeau et al., 2000): understanding the question, obtaining relevant information from memory, using this information to make decisions (judgment) on choice of the correct answer and the selection and delivery of the answer to the question. Here, we should note that in this case response times cannot distinguish between these cognitive steps. The model of four cognitive steps assumes the respondent is sufficiently motivated to answer the question to the best of their ability. When motivated respondents carefully carry out all cognitive steps we say they are using the optimising strategy (Krosnick, 1991). However, in survey research it is unlikely that all respondents choose the optimal strategy for responding. It is likely that certain respondents adjust their cognitive steps depending on the question difficulty and level of their motivation. Thus, some respondents do not take the required cognitive steps, but still provide responses, which are however not optimal. We call this satisficing (Krosnick, 1991), which can range from weak to strong. The strategy of weak satisficing

happens when a respondent still performs all four cognitive steps, but is less thorough and less motivated in doing so, while strong satisficing occurs when a respondent skips certain cognitive steps (e.g. they read the survey questions very quickly, and then give a random answer). Satisficing takes various forms (Krosnick, 1991; Callegaro et al., 2015). Typical forms include: (1) the selection of the first possible answer (primacy effect); (2) agreement with questions irrespective of their content (acquiescence); (3) non-differentiation in answering closed questions that contain measurement scales (straightlining); (4) responding with the answer "do not know" etc. It is generally considered that satisficing reflects respondents' insufficient effort and lack of motivation in completing the questionnaire and therefore may represent poorer quality responses (Callegaro et al., 2009; Zhang, 2013; Conrad et al., 2017). In relation to response times, we may assume that respondents following the optimising strategies would have longer response times than those following the satisficing strategies.

Response time in web surveys is particularly important because, compared to personal interviews, the survey process is conducted solely by means of computer programs and without the presence of interviewers; it therefore presents an essential instrument for insight into (the quality of) the process of answering a web questionnaire.

Of course, measuring response time alone is not enough; we also need an approach, the level of response time measurement and time limits to identify respondents who proceed through the survey questionnaire in an undesired way (e.g. too quickly, too slowly or inconsistently) and are related to certain response time anomalies. We refer to such respondents as response time outliers (or simply as outliers). Researchers have used (or developed) various approaches, levels of time measurement and time limits to detect (see e.g. Ratcliff, 1993; Cousineau and Chartier, 2010; Kreuter, 2013), process (e.g. remove or transform) or analyse response time outliers (e.g. Ratcliff, 1993; Kreuter, 2013).

The main difficulty in analysing the response time is related to the type of approach (i.e. statistical or cognitive), level of response time measurement (item, page or questionnaire) and corresponding time limits (i.e. time thresholds where we declare units as response time outliers). Statistical approaches are denoted here as those that are based solely on the statistical properties of the collected response times, such as central tendency measures (e.g. mean and median), dispersion measures (e.g. standard deviation) or distribution measures (e.g. percentile). On the other hand, cognitive approaches are denoted here as those based on external cognitive criteria (i.e. psychological properties), which are independent of the statistical properties of the collected response times. These approaches are typically based on the respondent's reading speed expressed in words that are expected to be read per second.

In addition to differences related to the type of approach (statistics vs. psychology), approaches can also differ in the level of measurement—namely, the response time can be analysed at the level of a survey item, survey question, survey page or survey questionnaire.

When analysing response time, the following issues arise: "Which approach and which level of response time measurement should be used to address response time outliers?", "What are the lower and upper time limits?", "What should be done with respondents who have a response time outside the time limits?" and "How can respondents who more or less persistently deviate from these limits be identified?" The current research does not provide answers to these questions; rather, we are interested if and how survey

researchers address this issue. We believe it is very important to answer such questions because units that represent poor response quality (e.g. respondents who satisfice) can reduce the internal consistency and reliability of the answers and thus impair the results of statistical analyses (see e.g. Ratcliff, 1993; Meade and Craig, 2012).

To demonstrate the prevalence of the issues stated above, we conducted a systematic literature review of these issues. The scope was restricted to web surveys because they (1) are commonly used for data collection in empirical social research, particularly in internet panels; (2) enable interactive feedback on web questionnaires, which allows survey designers to intervene and thus improve the quality of responses; (3) enable researchers to capture paradata easily (for examples of capturing paradata, see Stieger and Reips, 2010); and (4) prevail (among other computer-assisted survey information collection methods) in response time analysis. Among other things, this will help us indicate the directions of further research in order for web surveys to achieve higher quality.

The structure of this paper is as follows: In section 2, we outline the research questions. In section 3, we present the methodological approach for conducting systematic literature reviews. In section 4, we provide the results and answers to the research questions, while in section 5, we summarise the main findings and place them in a broader context.

2 Research Questions

Due to the growing awareness of the importance of detecting response time outliers in web survey methodology (Malhotra, 2008; Kreuter, 2013; Callegaro et al., 2015), we conducted a systematic literature review of existing papers on response time outliers in web surveys.

Through the systematic literature review, we aimed to answer the following research questions:

Q1. How do researchers measure response time (i.e. approaches and level of response time measurement)?

Q2. How do researchers detect response time outliers (i.e. approaches and time limits used)?

Q3. What is the typical percentage of detected response time outliers?

Q4. How do researchers process the response time outliers?

Q5. How do researchers analyse the relationship between response time and response quality?

The first and second research questions will be addressed in section 4.1. The third and fourth research questions will be addressed in section 4.2, and the fifth research question will be addressed in section 4.3.

Table 1: Systematisation of global and local context effects versus intra-survey and extra-survey stimuli

Inclusion criteria	Exclusion criteria
• Papers reporting primary or secondary research using a web survey as a data collection method	• Papers not reporting primary or secondary research using a web survey as a data collection method
• Papers reporting a measure of response time, an approach used to detect response time outliers or treatment of response time outliers	• Papers not reporting a measure of response time, an approach used to detect response time outliers or treatment of response time outliers
• Full text available in English	• Full text unavailable in English

3 Methodology

3.1 Search Strategy

The international standard for reporting results of systematic reviews and meta-analyses (PRISMA) (see e.g. Moher et al., 2009) was used to guide the methodology of our systematic review. A systematic literature search was conducted using the Digital Library of the University of Ljubljana,[2] which enables access to more than 20 000 paid electronic journals and more than 170 000 paid electronic books. The key search terms were as follows: "paradata", "response time", "reaction time", "response latency", "outliers", "response time outliers", "reaction time outliers", "response latency outliers", "data quality", "response quality", "response time in web survey", "response time outliers and quality of responses", "response time data quality", "cognitive processes in web survey", "satisficing in web survey" and "speeding in web surveys".

3.2 Eligibility Criteria

We reviewed all scientific papers in English that were published on or before 25 June 2017 (with no lower limit) and met the inclusion criteria shown in Table 1. It should also be noted that the focus of this paper is only on scientific papers because, to the best of our knowledge, no book or book chapters rely on empirical work related to the approaches and time limits for identifying response time outliers. Moreover, conference presentations and reports were also excluded because we wanted to focus on the most carefully evaluated material.

We identified the potentially relevant scientific papers by examining the abstracts or the scientific papers as a whole.

---

[2]http://dikul.uni-lj.si

4   Results

Through database searching based on the key search terms, we identified 45 papers deal-
ing with response time in web surveys. Available abstracts of these papers were screened
using the inclusion criteria. In total, 17 abstracts were rejected because they failed to
meet at least one of the criteria. For the 28 abstracts that met the inclusion criteria, full
manuscripts were retrieved for screening, and all of them were eligible for systematic
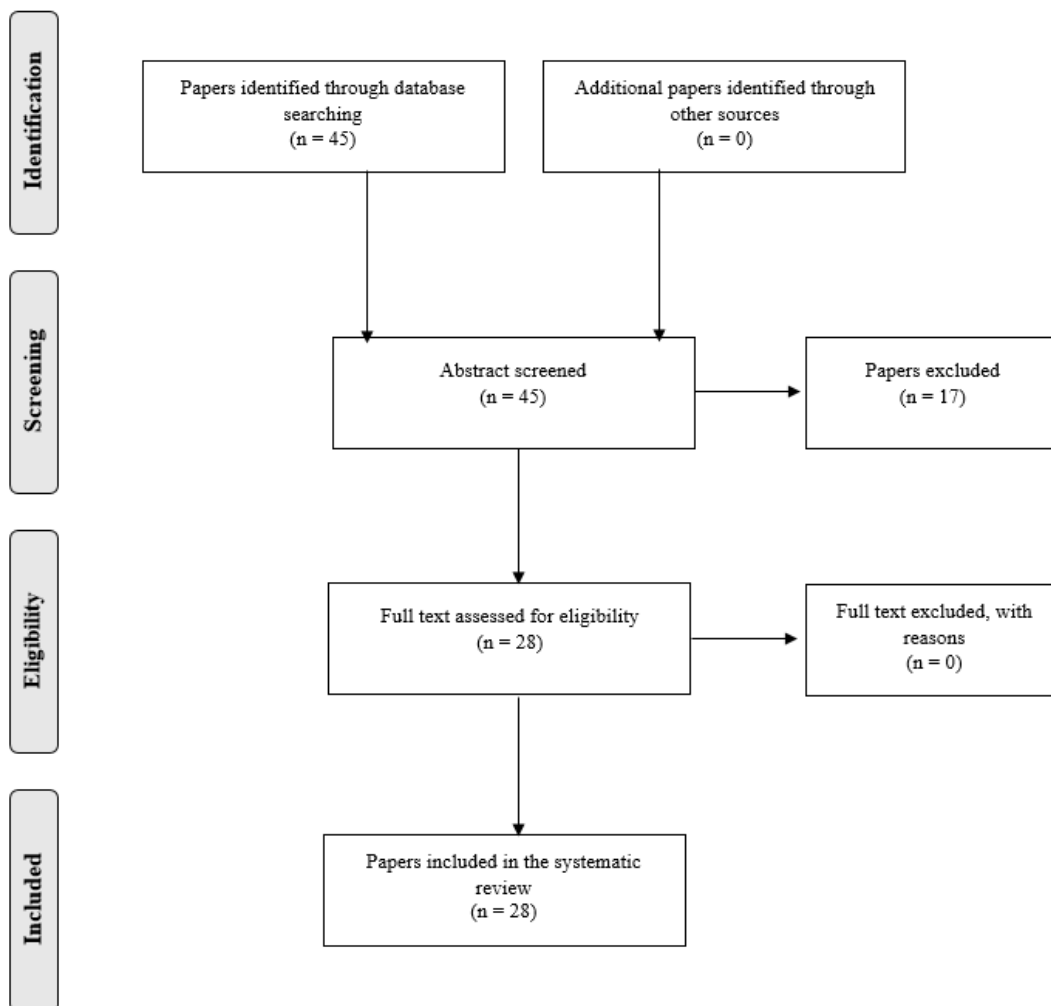review (see Figure 1).

Figure 1: PRISMA flow diagram for included scientific papers

The systematic review thus included 28 scientific papers published between 1 January
2003 and 25 June 2017. The full references and characteristics of the included papers are
presented in Table A1 (Appendix). The sample size of the included papers varied from
132 to 24 273 (median = 1547). The sampling frame also varied from probability samples
to nonprobability samples. Out of 28 papers, 13 were primarily interested in response
times (i.e. the paper titles pointed to analyses of response times).

4.1 Approaches, Levels of Time Measurement and Time Limits

All 28 papers reported on response time measurement and the approach used. Two papers were only interested in too fast response times and 11 were only interested in detecting too slow response times, while 15 papers analysed for both types of response time outliers.

The levels of response time measurement also differed among the papers. The response time per item was measured in 11 papers (either because each item was presented on one page or because the advanced paradata collection was used), the questionnaire completion time was measured in 11 papers and the response time per page was measured in 10 papers.[3] There were also differences in the approaches and time limits used. Twenty one papers used statistical approaches, six papers used a cognitive approach and one paper used a combination of statistical and cognitive approaches. The type of approach, time limits and level of response time measurement used in the papers are summarised in Table 2.

Table 2: Papers according to approach, time limits and level of time measurement

		Level of response time measurement		
Approach	Time limits	Item	Page	Questionnaire
Percentiles	Bottom limit at 1st percentile/top limit at 99th percentile Bottom limit at 5th percentile/top limit at 95th percentile	Yan and Tourangeau (2008)		Gummer and Roßmann (2015); Lenzner et al. (2010) Harms et al. (2017)
	Top limit above 90th percentile		Yan et al. (2010)	Revilla and Ochoa (2015)
Standard deviation	±2 standard deviations from the mean response time	Christian et al. (2009)		Heerwegh (2003); Heerwegh and Loosveldt (2006)

---

[3]In case of level of response time measurement the number of studies is greater than 28 because some studies measured response time on multiple levels.

Approach	Time limits	Level of response time measurement		
		Item	Page	Questionnaire
	2 or 3 standard deviations above the mean response time		Smyth et al. (2006); Smyth et al. (2009); Mahon-Haft and Dillman (2010)	Naemi et al. (2009)
Interquartile range (IQR)	±1.5 IQR			Funke et al. (2011)
	±2.5 IQR	Funke (2016)		
Speeder index (see Roßmann, 2010)	Speeder index value of ±2 standard deviations from the mean response time	Roßmann and Gummer (2016)	Roßmann and Gummer (2016)	
Average time	6 times more than the average question time	Couper et al. (2006)		
Median absolute deviation	5 times the median absolute deviation			Sendelbah et al. (2016)
Combined approach (two statistical approaches)	Logarithmic transformation and 1 standard deviation above the mean response time			Malhotra (2008)
	Logarithmic transformation and response time above 99.9th percentile		Tijdens (2014)	

Approach	Time limits	Level of response time measurement		
		Item	Page	Questionnaire
Reading rate (cognitive approach)	Top limit at 99th percentile and 2 standard deviations above the mean		Sauer et al. (2011)	
	Bottom limit at 1 or 3 seconds/top limit at 5 or 15 minutes Top time limit at 300 or 350 milliseconds	Healey (2007); Callegaro et al. (2009); Stieger and Reips (2010) Zhang and Conrad (2014); Conrad et al. (2017)	Healey (2007); Meade and Craig (2012)	Healey (2007)
Combined approach (statistical and cognitive approach)	Speeder index: bottom limits at 30%, 40% or 50% faster than the median response time per page and reading rate of 7.5 words read per second	Greszki et al. (2015)	Greszki et al. (2015)	

Based on Table 2, there seems to be no association between the level of response time measurement and the approach used. It thus seems that researchers selected their approaches arbitrarily, but it is also possible they were following other researchers (see e.g. Yan and Tourangeau, 2008; Greszki et al., 2015). They might be also using other approaches, but they presented only one of them.

4.2 Treatment of Response Time Outliers

Among the 28 papers, 20 excluded response time outliers entirely from further analysis (i.e. unit nonresponse), one treated response time outliers as missing values (i.e. item nonresponse), five analysed response time outliers and two used substitutions (e.g. substituted times of response time outliers with the 90th percentile value of response times). Table 3 summarises the treatment and proportion of response time outliers for the 28 reviewed papers.

Table 3: Papers according to treatment and proportion of response time outliers

Paper	Number of papers	Average % of response time outliers	Median % of response time outliers
Excluded from the analysis			
Heerwegh (2003); Gummer and Roßmann (2015); Harms et al. (2017); Tijdens (2014); Yan and Tourangeau (2008); Healey (2007); Couper et al. (2006); Lenzner et al. (2010); Malhotra (2008); Funke et al. (2011); Naemi et al. (2009); Funke (2016)	12	4	3
Callegaro et al. (2009); Smyth et al. (2009); Mahon-Haft and Dillman (2010); Sauer et al. (2011); Smyth et al. (2006); Christian et al. (2009); Heerwegh and Loosveldt (2006); Sendelbah et al. (2016)	8	Not reported	Not reported
Analysed			
Greszki et al. (2015); Conrad et al. (2017); Zhang and Conrad (2014); Stieger and Reips (2010)	4	Not possible to calculate	Not possible to calculate
Roßmann and Gummer (2016)	1	Not reported	Not reported
Substituted (replaced)			
Revilla and Ochoa (2015); Yan et al. (2010)	2	7	7
Set as missing values			
Meade and Craig (2012)	1	4	4

Among the 20 papers that excluded the response time outliers from the analysis (i.e. unit nonresponse), 12 papers reported the percentage of excluded respondents, which, on average, amounted to 4% (median = 3%). Among the five papers that analysed the answers of the response time outliers, the proportion of response time outliers varied widely because some papers detected response time outliers according to the respondents'

age (e.g. percentage of speeders by age group), education or level (e.g. respondents speeding on at least one item) or used multiple time limits to detect response time outliers (e.g. 50%, 40% and 30% faster than the median completion time). Therefore, we did not calculate the average and median proportions of the detected response time outliers. Next, both the average and the median proportion of the detected response time outliers of the two papers that used substitutions amounted to 7%. In the paper where all responses of the response time outliers were set as missing (i.e. item nonresponse), the percentage amounted to 4%.

According to Table 3, there also seems to be no association between the approach used and the treatment and proportion of the response time outliers.

4.3 Relationship between Response Time and Response Quality

Thirteen papers addressed the relationship between response time and response quality (see Table 4), of which eight papers found some evidence of a correlation. Within this context, papers typically observed the following aspects of response quality: primacy effects, order effects, dropout, straightlining and response accuracy.

Table 4: Relationship between response time and response quality

Correlation between response time and response quality	Research focus	Conclusions
Smyth et al. (2006)	Investigated check-all and forced-choice question formats in web surveys	Respondents who answered check-all questions quickly marked significantly fewer options and appear to have employed a weak satisficing response strategy (as evidenced by patterns of primacy), more so than their counterparts who answered these questions more slowly.
Malhotra (2008)	Investigated if respondents who complete web surveys more quickly also produce data of lower quality	Respondents with relatively lower cognitive skills who take less time to complete web surveys satisfice and produce lower quality data in the form of order effects.

Correlation between response time and response quality	Research focus	Conclusions
Callegaro et al. (2009)	Investigated the link between response time and optimising/satisficing strategies	Optimisers invest more time than satisficers when answering questions. This supports the perspective that deeper cognitive processing requires greater effort and takes more time. Moreover, response times can be another tool for studying and identifying optimising/satisficing strategies and for assessing the quality of data collected with surveys and questionnaires.
Tijdens (2014)	Investigated dropout rates and response times of an occupation search tree in a web survey	Response time in each step of the search tree is related to the search tree item length or to the respondent's valid self-identification and dropout in the next step. This means that the response time increases with search tree item length, next-step dropout, invalid self-identification, higher age and lower education, but it is not affected by employment status.
Zhang and Conrad (2014)	Investigated impact of speeding on response quality in terms of straightlining	Positive correlation between speeding and straightlining (i.e. non-differentiation in answering closed questions that contain measurement scales).

Correlation between response time and response quality	Research focus	Conclusions
Revilla and Ochoa (2015)	Investigated the links among response time, quality (satisficing) and auto-evaluation of the efforts made	Weak link between response time and quality: worse quality of answers is directly related to shorter response time—that is, to more speeding
Greszki et al. (2015)	Investigated the effects of removing "too fast" responses and respondents from web surveys on substantive findings	Small effects of speeder corrections on substantive findings, which suggest that speeding adds some kind of random noise to the data
Conrad et al. (2017)	Investigated the impact of providing immediate feedback on reducing speeding	Reduction in speeding was associated with some evidence of improved response quality, namely increased response accuracy.

Table 4 shows that there seems to be no association between the type of approach used and the evidence of correlation between response time and response quality. This is also true for the level of response time measurement and the evidence of correlation between response time and response quality.

However, among the eight papers that found some evidence of correlation, seven papers measured the response time per item or per page and only one paper measured the completion time per survey. Furthermore, among the five papers that did not find evidence of correlation, three papers measured the response time per questionnaire. This could indicate that measuring the response time per questionnaire is not a suitable approach for analysing the association between response time and response quality.

5   Summary and Conclusions

In recent years, response time has been frequently discussed in survey methodology, since it can offer insight into different aspects of the interaction between the respondent and the questionnaire (Kreuter, 2013). This is particularly true for computer-assisted modes such as web surveys.

The goal of the systematic literature review in this paper was to determine the contribution of past analytical papers to the knowledge on response time outliers (i.e. how to measure them, how problematic they are and how to deal with them) in web surveys, as well as to identify gaps in this knowledge. For this purpose, we identified and analysed

28 scientific papers dealing with response time in web surveys. The methodology of our systematic review was based on PRISMA.

We can conclude that a series of methodological issues exist. First, despite the various approaches used in previous studies (see Table 2), our analysis showed no consensus in the survey research literature regarding the right conceptual approach (statistical vs. cognitive) and level of response time measurement (item, page or questionnaire). There is also no consensus in defining time limits for detecting response time outliers. We identified 21 papers where researchers used a statistical approach (or statistical approaches) with various time limits (e.g. 2 or 3 standard deviations, 1st or 99th percentile and 5th or 95th percentile), six papers where researchers used a cognitive approach with various time limits (e.g. 1-3 seconds and 300 or 350 milliseconds per word) and one paper where researchers combined statistical and cognitive approaches.

Second, we also found no consensus on how to deal with response time outliers (see Table 3). Out of the 28 papers, 20 removed all response time outliers (i.e. respondents whose response time was determined as outside the time limits) prior to data analysis. Among the remaining eight papers, five preserved the answers of respondents whose response times were determined as outside the time limits, two papers substituted these respondents' response times and one paper set all the answers of these respondents as missing values.

Third, the question about whether response time outliers are associated with poorer response quality (see Table 4) often relies on the implicit assumption that such a link exists. However, this link does not necessarily exist. We identified 13 papers that addressed the relationship between response time and response quality, of which eight papers found some evidence of correlation—that is, response times can be used to identify optimising/satisficing strategies of respondents when answering web survey questionnaires. Based on this limited evidence, the existence of an association between response time outliers and poorer response quality is still not fully conclusive.

Fourth, it seems that survey researchers used approaches arbitrarily, following other researchers or according to their unspecified circumstances. We also found no association between the approach used and the treatment of the response time outliers. There was also no association between the type of approach used and the response quality or between the level of response time measurement and response quality. Within this context, we also warn that measuring the response time per questionnaire may not be appropriate when analysing the association between response time and response quality, although more research is needed to verify this.

The methodological issues addressed in this paper concern not only academic researchers but also the business sector, particularly the internet panels. To assess the prevalence of these issues, we conducted an investigation among leading Slovenian non-probability internet panels and received responses from the Valicon, GfK and Marketagent.com. We found that all three internet panels measure the response time per survey page (by default) or per questionnaire (by request). All three also developed their own approaches (statistical or cognitive), set specific time limits for detecting response time outliers and developed specific strategies to deal with response time outliers. Due to response quality concerns, all three internet panels are looking for ways to improve their approaches for detecting response time outliers. Despite the differences among them, all three internet panels combine cognitive and statistical approaches for detecting and

evaluating response time outliers, which was in fact very rarely the case in our analysed papers.

The inquiry with internet panels basically confirmed the findings from the literature review-namely, issues related to response quality, the level of response time measurement, the optimal combination of approaches and the percentage of detected response time outliers deserve to become the subject of further research to overcome the lack of knowledge on response time outliers and response quality.

It is thus somewhat surprising that these important issues are so weakly researched in the literature. One possible explanation is that the effect of response time outliers on response quality is actually weak (Greszki et al., 2015) and that analyses of the responses containing response time outliers do not necessarily lead to a wrong conclusion (i.e. differences in substantive findings). Accordingly, these issues would not be worth researching and the differences between approaches would also be negligible.

Nevertheless, our review showed that there is a potential relationship between response time outliers and response quality. The elaboration of this link is an important step in promoting greater response quality in survey research, particularly in light of the trend towards continuous measurement using panels in everyday survey practice (Couper, 2005), growth of internet usage through portable mobile devices and increasing difficulties in recruiting respondents for survey research. Within this context, the frequency of data collection, as well as the number of waves and type of respondents being recruited, may influence the prevalence of response time outliers, as well as response quality (Roßmann and Gummer, 2016). Also of particular concern is the emergence of professional respondents who frequently participate in surveys and mainly do so for the incentives (Matthijsse et al., 2015).

There are two other specific issues we have not addressed sufficiently due to space limitations as well as lack of previous research, but which definitely deserve more attention. One issue is the unclear conclusions related to the problem of long response times. While the problem of overly short response times is straightforward, for overly long response times the situation is less clear. As noted in the introduction, reasons for long response times due to respondents' processing can be both "positive" (e.g. engagement, motivation) or "negative" (e.g. multitasking). However, papers rarely address this issue which is also very difficult to research, due to the limited conceptual background of this problem as well as complex methodological issues. For example, even the dedicated elaboration of multitasking in web surveys (Sendelbah et al., 2016) was not able to distinguish multitasking and overly long response times.

Another very specific issue deserving greater elaboration relates to the impact of mobile devices, such as smartphones and tablet computers, on response times in web surveys. The use of mobile devices to fill in web surveys might affect respondents' response time in various ways (e.g. smaller screens might limit the amount of questions visible at any one time so respondents may have to scroll more to view all questions what may result in taking more time to answer). However, addressing these questions can be more difficult and require much bigger research efforts (see Gummer and Roßmann, 2015). Correspondingly, almost no study included in our analysis addressed this issue. In part, this also reflects the low share of respondents who participated via a mobile device, particularly in studies from 2016 or before while we can expect more studies focusing separately on mobile devices in the future.

In any case, we may conclude there is a definite need for more research that will: (1) systematically compare the efficiency of approaches for detecting response time outliers; (2) harmonise the existing approaches and develop new ones; (3) analyse the detected response time outliers; (4) compare how the correlation and substantive results would be affected if the data from the response time outliers were removed; and (5) show if (and which) approaches, levels of time measurement and time limits are associated with response quality.

Acknowledgment

References

[1] Barge, S. and Gehlbach, H. (2012): Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, **53**, 182–200.

[2] Bassili, J. N. and Scott, B. S. (1996): Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, **60**, 390–399.

[3] Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A. and Chin, T. Y. (2009): Response latency as an indicator of optimizing in online questionnaires. *Bulletin de Méthodologie Sociologique*, **103**, 5–25.

[4] Callegaro, M., Lozar Manfreda, K. and Vehovar, V. (2015): *Web Survey Methodology*. Los Angeles: Sage.

[5] Christian, L. M., Parsons, N. L. and Dillman, D. A. (2009): Designing scalar questions for Web surveys. *Sociological Methods & Research*, **37**, 393–425.

[6] Conrad, F., Tourangeau, R., Couper, M. and Zhang, C. (2017): Reducing speeding in Web surveys by providing immediate feedback. *Survey Research Methods*, **11**, 45-61.

[7] Couper, M.P. (1998): Measuring Survey Quality in a CASIC Environment. In: *Proceedings of the Survey Research Methods Section*, 41–49. American Statistical Association.

[8] Couper, M. P. (2005): Technology trends in survey data collection. *Social Science Computer Review*, **23**, 486–501.

[9] Couper, M. P., Tourangeau, R., Conrad, F. G. and Singer, E. (2006): Evaluating the effectiveness of visual analog scales: A Web experiment. *Social Science Computer Review*, **24**, 227–245.

[10] Cousineau, D. and Chartier, S. (2010): Outliers detection and treatment: A review. *International Journal of Psychology Research*, **3**, 58–67.

[11] Funke, F., Reips, U. D. and Randall K. T. (2011): Sliders for the smart: Type of rating scale on the Web interacts with educational level. *Social Science Computer Review*, **29**, 221–231.

[12] Funke, F. (2016): A Web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, **34**, 244–254.

[13] Greszki, R., Meyer, M. and Schoen, H. (2015): Exploring the effects of removing "too fast" responses and respondents from Web surveys. *Public Opinion Quarterly*, **79**, 471–503.

[14] Gummer, T. and Roßmann, J. (2015): Explaining interview duration in Web surveys: A multilevel approach. *Social Science Computer Review*, **33**, 217–234.

[15] Healey, B. (2007): Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in Web surveys. *Social Science Computer Review*, **25**, 111–128.

[16] Harms, C., Jackel, L. and Montag, C. (2017): Reliability and completion speed in online questionnaires under consideration of personality. *Personality and Individual Differences*, **111**, 281–290.

[17] Heerwegh, D. (2003): Explaining response latencies and changing answers using client-side paradata from a Web survey. *Social Science Computer Review*, **21**, 360–373.

[18] Heerwegh, D. and Loosveldt, G. (2006): An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in Web surveys. *Journal of Official Statistics*, **22**, 191–210.

[19] Kreuter, F. (2013): *Improving Surveys with Paradata: Analytic Uses of Process Information*. New York: John Wiley & Sons.

[20] Krosnick, J. A. (1991): Response strategies for coping with cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, **5**, 213–236.

[21] Lenzner, T., Kaczmirek, L. and Lenzner, A. (2010): Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, **24**, 1003–1020.

[22] Mahon-Haft, T. A. and Dillman, D. A. (2010): Does visual appeal matter? Effects of Web survey aesthetics on survey quality. *Survey Research Methods*, **4**, 43–59.

[23] Matthijsse, S. M., De Leeuw, E. D. and Hox, J. J. (2015): Internet panels, professional respondents, and data quality. *Methodology*, **11**, 81–88.

[24] Malhotra, N. (2008): Completion time and response order effects in Web surveys. *Public Opinion Quarterly*, **72**, 914–934.

[25] Mayerl, J. (2013): Response latency measurement in surveys: Detecting strong attitudes and response effects. *Survey Methods: Insights from the Field*. Available at: http://surveyinsights.org/?p=1063

[26] Meade, A. W. and Craig, S. B. (2012): Identifying careless responses in survey data. *Psychological Methods*, **17**, 437–455.

[27] Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009): Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, **62**, 1006–1012.

[28] Naemi, B. D., Beal, D. J. and Payne, S. C. (2009): Personality predictors of extreme response style. *Journal of Personality*, **77**, 261–286.

[29] Ratcliff, R. (1993): Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510–532.

[30] Revilla, M. and Ochoa, C. (2015): What are the links in a Web survey among response time, quality, and auto-evaluation of the efforts done? *Social Science Computer Review*, **33**, 97–114.

[31] Roßmann, J. (2010): Data quality in Web surveys of the German longitudinal election study 2009. In: *3rd ECPR Graduate Conference, Dublin, Ireland*.

[32] Roßmann, J. and Gummer, T. (2016): Using paradata to predict and correct for panel attrition. *Social Science Computer Review*, **34**, 312–332.

[33] Sauer, C., Auspurg, K., Hinz, T. and Liebig, S. (2011): The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods*, **5**, 89–102.

[34] Sendelbah, A., Vehovar, V., Slavec, A. and Petrovčič, A. (2016): Investigating respondent multitasking in Web surveys using paradata. *Computers in Human Behavior*, **55**, 777–787.

[35] Stieger, S. and Reips, U. (2010): What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behaviour*, **26**, 1488–1495.

[36] Smyth, J. D., Dillman, D. A., Christian, L. M. and Stern, M. J. (2006): Comparing check-all and forced-choice question formats in Web surveys. *Public Opinion Quarterly*, **70**, 66–77.

[37] Smyth, J. D., Dillman, D. A., Christian, L. M. and Mcbride, M. (2009): Open-ended questions in Web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, **73**, 325–337.

[38] Sudman, S., Brandburn, N. M. and Schwarz, N. (1996): *Thinking about Answers: The Application of Cognitive Process to Survey Methodology*. San Francisco: Jossey-Bass Publishers.

[39] Tijdens, K. (2014): Dropout rates and response times of an occupation search tree in a Web survey. *Journal of Official Statistics*, **30**, 23–43.

[40] Tourangeau, R., Rips L. J. and Rasinski K. A. (2000): *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

[41] Yan, T., Conrad, F. G., Tourangeau, R. and Couper, M. P. (2010): Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing Web surveys. *International Journal of Public Opinion Research*, **23**, 131–147.

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

[43] Zhang, C. (2013): *Satisficing in Web Surveys: Implications for Data Quality and Strategies for Reduction*. Ph.D thesis of University of Michigan.

[44] Zhang, C. and Conrad, F. G. (2014): Speeding in Web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Method*, **8**, 127–135.