

Identifying a Set of Key Paradata Indicators in Web Surveys Based on the Relationship with Response Quality, Respondent Characteristics and Survey Estimates

Vasja Vehovar¹  · Nejc Berzelak²  · Gregor Čehovin¹ 

¹Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana

²National Institute of Public Health

Web survey paradata refer to the digital traces left by respondents, which can provide valuable insights into the response process. An exhaustive literature review revealed that web survey paradata have predominantly been studied in relation to response quality and sometimes also in relation to respondent characteristics and survey estimates. A broad range of paradata indicators was already employed for these purposes in the literature. This paper aimed to identify a minimal set of paradata indicators that can be used as general characteristics of respondents, similar to sociodemographic variables. In the empirical part of the paper, a comprehensive set of raw paradata for a typical web survey was collected ($n = 3,458$). Based on the literature, an initial set of 112 paradata indicators was identified and further reduced through subsequent steps. Finally, the potential paradata indicators were examined for their statistically significant relationship with 9 response quality indicators, 8 respondent characteristics (including the Big Five personality dimensions), and 13 survey estimates related to computers and Internet use. The result of this process was a set of 12 key paradata indicators that could serve as a starting point for establishing a standardised set of paradata indicators, which could potentially augment the web survey response data.

Keywords: web surveys; paradata; smartphones; response quality; data quality; survey estimates; respondent characteristics; big five

1 Introduction

The rise of web surveys (Callegaro et al., 2015) has also increased the importance of web survey *paradata*, which refer to the digital traces left by respondents in web surveys (Couper & Lyberg, 2005). Given their usefulness, there is an evident need for their collection, exploitation and documentation, as indicated in the literature (McClain et al., 2019). This paper focuses on paradata reflecting respondents' direct interactions with a survey questionnaire, which primarily replicate the characteristics of respondents' devices and their navigation through a questionnaire. The corresponding paradata can be captured relatively easily as a by-product of the web survey data collection process (see Callegaro, 2013; Kreuter, 2013). However, the complexity involved in

preparing and processing paradata may require considerable resources (Kunz & Hadler, 2020), posing a notable barrier to their advanced utilisation.

A preliminary literature review (see Vehovar & Čehovin, 2023) indicated that web survey paradata have mainly been studied in relation to three domains: response quality (e.g. speeding), respondent characteristics (e.g. personality traits) and survey estimates (i.e. substantive variables). While different surveys require different sets of paradata, this paper argues that there may exist a common denominator (i.e. a standardised set of *key paradata indicators*) akin to sociodemographic variables (e.g. gender, age and education) that can be useful for all web surveys because they provide general characteristics of respondents. Efforts in this direction may be exemplified by surveys in the CRONOS panel study (Villar et al., 2018) part of the European Social Survey, where a specific set of respondent-level paradata indicators was publicly archived for potential integration with survey response data (European Social Survey, 2018). Additionally, similar endeavours can be observed in the GESIS Panel (Weyandt et al., 2022).

Corresponding author: Gregor Čehovin, Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia (Email: gregor.cehovin@fdv.uni-lj.si)

Given this context, the aim of this paper is thus to identify a set of key paradata indicators that can augment web survey response data. Specifically, the intention is to identify paradata indicators that are robust and straightforward to compute, making them beneficial for general use in survey practices. This can be particularly important for researchers who encounter challenges in determining which paradata to collect. Namely, if paradata are chosen arbitrarily, they might offer little or no value to the corresponding research. Consequently, the effort and resources dedicated to collecting and processing them would be wasted. Conversely, failing to capture paradata that are functional for the corresponding research could be an even more important problem.

This paper first reviews the literature and conceptual issues. Next, it processes and analyses a comprehensive set of raw paradata captured in a typical web survey ($n = 3458$). Various steps were taken to process the raw paradata, compute the paradata indicators and reduce them. This procedure resulted in a final set of 12 key paradata indicators that were found to be statistically significantly related to *response quality indicators* (RQIs), respondent characteristics or survey estimates. The results are discussed as a step towards a standardised set of key paradata indicators that can enrich web survey response data.

2 Background

The literature indicates that web survey paradata can potentially reflect respondents' underlying behaviour when filling out web surveys, which we refer to as *response style*. Some authors also refer to it as *response behaviour* (e.g. Greszki et al., 2015; Höhne & Schlosser, 2019), *response pattern* (e.g. Boulianne et al., 2011; Braekman et al., 2020) or *response optimisation strategy* (e.g. Krosnick, 1991, 2018). There is ample evidence that response styles are associated with the domains of response quality (e.g. Höhne, Schlosser, et al., 2020; Zhang & Conrad, 2014), respondent characteristics (e.g. Bowling et al., 2016; Sturgis et al., 2019) or survey estimates (e.g. Andersen & Mayerl, 2017; Tzafilkou & Nicolaos, 2018). In this context, web survey paradata studies have predominantly addressed the domain of response quality (see Vehovar & Čehovin, 2023). Particularly critical is the suboptimal performance of respondents who seek shortcuts to reduce cognitive effort (Tourangeau et al., 2000), sometimes denoted as *survey satisficing* (Krosnick, 1991) or *insufficient effort responding* (Huang et al., 2012). These behaviours often manifest as extreme or mid-point responses, straightlining, item nonresponse or random responding (Roberts et al., 2019), all of which are aspects that can be inferred from paradata. While these response styles are sometimes observable in the substantive

data, paradata such as response times are often used to assess their presence. This is often done with the assumption that very short or very long response times may indicate lower response quality (e.g. Horwitz et al., 2017; Kumar et al., 2022; Matjašič et al., 2018; Revilla & Ochoa, 2015; Sturgis et al., 2019). In the domain of response quality, some studies have used paradata to analyse nonresponse problems, for example, breakoffs (e.g. Galesic & Bosnjak, 2009; Mittereder, 2019) or attrition (e.g. Roßmann & Gummer, 2016), as well as 'do not know' answers (e.g. Sturgis et al., 2019; Turner et al., 2015). Similarly, mouse movement paradata have been used to address respondent cognitive difficulty (e.g. Fernández-Fontelo et al., 2022; Horwitz et al., 2020; Lenzner et al., 2010) and on-device multi-tasking (e.g. Höhne, Schlosser, et al., 2020). Paradata from smartphone sensors, such as accelerometer paradata, have sometimes been included to study the effects of respondents' movements on response quality in mobile web surveys (e.g. Höhne, Revilla, et al., 2020; Höhne & Schlosser, 2019). Furthermore, focus-out event detection (i.e. when a respondent leaves a browser tab or window containing the web questionnaire) has been applied to detect multi-tasking, which potentially decreases response quality (e.g. Sendelbah et al., 2016).

The second domain of web survey paradata studies focuses on respondent characteristics. In this context, the standard sociodemographic variables (e.g. age, gender and education) are frequently addressed in relation to response styles (e.g. Conrad, Tourangeau, et al., 2017; Yan & Tourangeau, 2008; Zhang & Conrad, 2014). Some paradata studies in the domain of respondent characteristics have also examined the relationship between personality traits (e.g. *Big Five* personality traits) and various aspects of response styles, such as insufficient effort responding (e.g. Bowling et al., 2016), satisficing (e.g. Sturgis & Brunton-Smith, 2023) and panel wave nonresponse patterns (e.g. Cheng et al., 2020).

The third domain of web survey paradata studies addresses the relationship between paradata and survey estimates. Specifically, it explores whether a particular response style, as indicated by paradata, can be linked to certain survey estimates. In this context, paradata, such as keyboard and mouse actions, have been used to study negative emotions (e.g. Hibbeln et al., 2017), subjective ambivalence (e.g. Schneider et al., 2015) and self-efficacy, learning readiness and risk perception (e.g. Tzafilkou & Nicolaos, 2018). Some paradata studies in the domain of survey estimates have used mouse movements to predict emotions (Yamauchi & Xiao, 2018), cognitive impairment (e.g. Seelye et al., 2015) or correctness of responses (Kumar et al., 2022). Studies have also used response time paradata to analyse voting intentions (e.g. Greszki et al., 2015) and

undesirable attitudes or behaviours (e.g. Andersen & Mayerl, 2017).

3 Defining, capturing and structuring paradata

3.1 Web survey paradata definition

Paradata must be clearly separated from *metadata* describing survey characteristics (i.e. general information and features of a survey), *auxiliary data* (i.e. data derived from external sources, such as the sampling frame) (Kunz & Hadler, 2020) and *survey data* (i.e. respondents' answers).

The primary focus of this paper is paradata that can be captured easily from any web survey. As mentioned, the essential objective is to identify standardised paradata indicators suitable for general usage. Following this aim, we focus on passively collected web survey paradata, which are generated automatically with explicit survey-related actions taken by the respondent. We followed Callegaro et al. (2015) and denoted these paradata as *direct paradata*. This differs from a) *indirect paradata*, which require external equipment (e.g. eye-tracking or brainwave-monitoring devices) or external observation (e.g. behavioural coding); b) *prior-survey paradata*, which include longitudinal or panel paradata from previous waves of a survey; c) *contact paradata*, which involve additional intermediate procedures, such as study-specific preprocessing related to different types of invitations (McClain et al., 2019); and also from d) the broader notion of *passive* or *non-reactive data* (e.g. Leiner, 2019), which trace actions or behaviours not only within the survey response window (i.e. direct survey paradata) but also beyond, such as *ambient* or *sensor data*. This includes movement (e.g. acceleration and motion), location (e.g. GPS), light and sound (Hart et al., 2022; Kunz & Hadler, 2020; Struminskaya et al., 2020). Similarly, direct paradata are distinct from *metered data*, another subtype of passive data collection, which are captured through a specific application (i.e. a metre) that participants install on their devices (Bosch & Revilla, 2021).

The typology by Callegaro et al. (2015), based on the *object of description*, additionally clarifies that the empirical study in this paper, which deals with direct paradata, addresses only the (b) device type and (c) questionnaire navigation paradata:

- a. *Contact info paradata*: Contact attempts (e.g. email invitations outcome)
- b. *Device-type paradata*: Respondent device details (e.g. type, operating system and screen size)

- c. *Questionnaire navigation paradata*: Respondent progress through the questionnaire (e.g. response times, mouse movements and prompts)

Similarly, the classification of paradata proposed by McClain et al. (2019), which is based on the *phases of the data collection process*, spells out that direct paradata in this empirical study involve only (d) the response phase:

- a. *Prior survey paradata*: Previous waves in longitudinal studies or earlier stages of multi-stage surveys (e.g. device use, missing data and response speed)
- b. *Recruitment phase paradata*: Behaviour surrounding survey contact attempts and respondent contact with the researcher
- c. *Access phase paradata*: Various attempts by recruited units to access the web survey (e.g. time from first contact to access attempt and number of access attempts)
- d. *Response phase paradata*: Based on timestamps, keystrokes, clicks, mouse movements, device characteristics, and more

The reduction in the scope of this paper, as elaborated by the above discussion of paradata typologies, directly relates to the focus of the empirical study. This reduction is necessary to address the aim of this study, which is to establish a set of robust paradata indicators that can be generally recommended for use in web surveys. As such, the direct paradata specified through the selected typological categories mentioned above are expected to be easily captured in a standardised manner across virtually all web surveys.

3.2 Paradata capturing and processing

Direct paradata, hereafter simply referred to as paradata, can be obtained through the server that hosts the web questionnaire (i.e. *server-side paradata*) or on a respondent's device (i.e. *client-side paradata*) (Heerwegh, 2003). Server-side paradata primarily entail simple or basic paradata (e.g. pages visited, page timestamps and device characteristics), and their capturing is often integrated into the web survey software. These same paradata can also be collected on the client side. In addition, client-side paradata can include more advanced paradata, such as keystrokes, mouse clicks, zooming, scrolling and focus-out events. Their collection typically requires additional scripts or extensions to the web survey software (Callegaro, 2013), as well as considerable data cleaning efforts to prepare the paradata for analysis (McClain et al., 2019). This includes harmonising different web browsers, handling missing values, establishing relational links between different paradata types, dealing with outliers and handling various sorts of noise (see Kunz &

Hadler, 2020). After cleaning the raw paradata, additional processes are needed (e.g. resolving data inconsistencies and reconciling different paradata types) to aggregate and calculate meaningful indicators. All in all, the process of capturing and processing advanced paradata demands considerable resources.

Kaczmirek (2009) described four hierarchical paradata levels of aggregation:

- Level 1: Records of individual respondent actions (e.g. timestamps, clicks/taps, zooming, scrolling and entering answers) related to a given questionnaire element (e.g. item, question or page)
- Level 2: First-level data aggregated across individual respondent actions per questionnaire element (i.e. item, question or page, such as the total number of mouse clicks on a page)
- Level 3: Second-level data aggregated across respondents *per variable* (e.g. item nonresponse per variable) or aggregated across variables *per respondent* (e.g. mean number of answer changes per respondent)
- Level 4: Aggregated across variables *and* respondents, providing a single value per survey (e.g. mean response time)

This empirical study began with the collection of raw paradata related to respondents' actions (Level 1) and then

focused on second-level aggregation at the respondent level (Level 3) because the survey data and RQIs were also organised at the respondent level, which thus represents a crucial level of analysis.

4 Inventory of paradata indicators

We systematically reviewed the literature on the use of direct paradata. For this purpose, we updated the preliminary literature review by Vehovar and Čehovin (2023). In total, 57 overlapping references were found, 51 were related to the response quality domain (44 of them involved response times), 11 were related to respondent characteristics and 5 were related to survey estimates (Table 1).

The literature review thus provided the basis for the identified set of *initial paradata indicators* elaborated on in Sect. 6.1.

After addressing the considerations above, the main *research question* can be formulated as follows: What paradata indicators can comprise a minimal set of key paradata indicators associated with response quality, respondent characteristics or survey estimates? It is worth repeating that the research question is addressed within the context of paradata indicators that, on the one hand, are relatively easy to capture and process, while on the other hand, serve in

Table 1

Direct paradata used in the literature to examine response quality, respondent characteristics and estimates

Paradata domain	References
Response quality indicators	
Response time	Andersen & Mayerl, 2017; Bowling et al., 2016; Callegaro et al., 2009; Cepeda et al., 2021; Conrad et al., 2006, 2007, 2017; Crawford et al., 2001; Fernández-Fontelo et al., 2023; Funke et al., 2011; Galesic & Bosnjak, 2009; Greszki et al., 2015; Gummer et al., 2021; Gummer & Roßmann, 2015; Gutierrez et al., 2011; Haraldsen et al., 2005; Healey, 2007; Heerwegh, 2003, 2002; Heerwegh & Loosveldt, 2002; Höhne, Revilla, et al., 2020; Höhne & Schlosser, 2019; Horwitz et al., 2013, 2017; Huang et al., 2012, 2015; Jenkins et al., 2015; Kaczmirek, 2009; Lenzner et al., 2010; Malhotra, 2008; Maniaci & Rogge, 2014; Matjašič et al., 2021; Meade & Craig, 2012; Paas & Morren, 2018; Revilla & Ochoa, 2015; Roßmann & Gummer, 2016; Schneider et al., 2015; Schroeders et al., 2022; Sendelbah et al., 2016; Smyth et al., 2006; Stern, 2008; Stieger & Reips, 2010; Sturgis et al., 2019; Tourangeau et al., 2004; Wells et al., 2010; Wise & Kong, 2005; Yamauchi & Xiao, 2018; Yan & Tourangeau, 2008; Zhang & Conrad, 2014
Mouse actions	Cepeda et al., 2021; Fernández-Fontelo et al., 2022, 2023; Healey, 2007; Hibbeln et al., 2017; Horwitz et al., 2017, 2020; Jenkins et al., 2015; Kaczmirek, 2009; Kühne & Kroh, 2018; Schneider et al., 2015; Seelye et al., 2015; Stieger & Reips, 2010; Tzafilkou & Nicolaos, 2018; Yamauchi & Xiao, 2018
Keyboard actions	Stieger & Reips, 2010; Tzafilkou & Nicolaos, 2018
Device characteristics	Horwitz et al., 2013; Kaczmirek, 2009; Matjašič et al., 2021; Roßmann & Gummer, 2016; Stieger & Reips, 2010
Multitasking	Höhne, Schlosser, et al., 2020; Sendelbah et al., 2016
Respondent characteristics	Bowling et al., 2016; Cepeda et al., 2021; Cheng et al., 2020; Conrad et al., 2017; Gummer & Roßmann, 2015; Hibbeln et al., 2017; Seelye et al., 2015; Sturgis et al., 2019; Yamauchi & Xiao, 2018; Yan & Tourangeau, 2008; Zhang & Conrad, 2014
Survey estimates	Andersen & Mayerl, 2017; Greszki et al., 2015; Gutierrez et al., 2011; Schneider et al., 2015; Tzafilkou & Nicolaos, 2018

a manner similar to sociodemographic variables, i.e. as general characteristics of respondents.

5 Data and method

5.1 Study design

Respondents were recruited from the largest Slovenian access panel (Valicon, 2022) in January–February 2020. The data collection process was carried out at the University of Ljubljana using 1KA software (1KA, 2023) that was additionally adapted for paradata collection. A total of 11,169 panellists were invited (initial email plus one reminder), and 4771 clicked on the web questionnaire (participation rate of 43%). Respondents used their preferred device, with 2516 (54%) responding on *personal computers* (PCs) and 2128 (46%) on *smartphones* (SPs). The survey was adapted for SP completion. Tablet respondents ($n = 127$) were excluded because they behaved as a very inconsistent mix of PC and SP respondents (Peterson et al., 2017), which blurred the analysis, while their share was far too small for standalone analysis. Of the remaining 4644 units, 1102 were screened out because they reported very few online activities, so they were not eligible for questions about online behaviour, which represented the bulk of the questionnaire content. The remaining 3542 respondents reported regular Internet usage (specifically defined as having shopped online within the past 12 months), 3309 of whom finished the questionnaire and 233 were breakoffs. Once the respondent started the survey, the device could not be changed, and the survey had to be completed in a single session. Soft reminders were used to prompt respondents to answer all items, but they were not required to answer all questions (i.e. no hard reminders). The survey data were weighted for gender, age, education and region. References to the data, questionnaire and scripts are cited in the relevant sections of the paper; they are also summarised in the Online Appendix, Sect. 1.

5.2 Questionnaire

There were 42 questions (240 items), including 15 grid questions (158 items), but due to skips (i.e. branching), the respondents may not have received all questions. Nevertheless, the four grids of attitudinal items, which were used to calculate certain RQIs, were delivered to all respondents. Attitudes were measured using five-point ordinal scales and covered opinions towards online shopping, patterns in Internet use, trust in computers and the Big Five personality traits (20-item short form; see Donnellan et al., 2006). The

remaining 11 grids addressed the frequency of various online behaviours. The exact wording of all items is provided in the complete questionnaire (see Centre for Social Informatics & The Samuel Neaman Institute for National Policy Research, 2021). The median duration of questionnaire completion was 20.6 min.

5.3 Procedures for capturing paradata

Device characteristics and detailed respondent actions were recorded using client-side JavaScript code (see Berzelak et al., 2022) integrated into the web survey software (1KA). The recorded actions appeared as direct output from the web survey software and were stored in five *raw paradata* datasets listed below (see Berzelak et al., 2023). The rows in these five datasets represent the specific actions taken by the respondents. Each row included ID variables—i.e. respondent, page, page session number (in cases where a respondent returns to a certain page in multiple sessions), question item and response ID—and the following comma-separated values (see Online Appendix, Sect. 2, for technical details):

- a. **Page sessions:** 18 variables, including respondent sequence, timestamps and device details (e.g. browser, operation system and screen size);
- b. **Events:** 10 variables, including event timestamp, element type (e.g. typing, clicking, zooming and scrolling), input value, coordinates, element ID and CSS class (e.g. radio button and checkbox);
- c. **Responses:** 7 variables, including timestamp, response type, and response value;
- d. **Mouse actions:** 9 variables, including start and end timestamps, coordinates and distance travelled;
- e. **Alert prompts:** 10 variables, including alert display and close timestamp, alert type, trigger, ability for respondents to ignore the alert, alert text and respondent action.

These five raw paradata datasets exhaustively documented all the essential digital traces resulting from respondents' actions while answering the web survey. They were the basis for calculating the paradata indicators.

5.4 Response quality indicators

The RQIs in this study are based on the work of Alwin (2007), Ganassali (2008) and Callegaro et al. (2015), encompassing measurement errors arising from cognitive problems (Tourangeau et al., 2000) and selected non-response errors. The paper, including the RQIs, does not encompass measurement errors arising from questionnaire characteristics, respondent characteristics, socially desirable responding, and falsification (e.g. Biemer & Lyberg,

2003; Groves, 2005). Although these aspects can contribute to a broader understanding of response quality, they are not inherent to the regular response process, where respondents are expected to answer survey questions honestly and accurately without deliberate misrepresentation. No further conceptual elaboration of response quality is provided here, but the most typical RQIs were selected from the literature (e.g. Mittereder, 2019; Roberts et al., 2019; Vehovar et al., 2022). The empirical study thus included nine frequently used RQIs calculated at the *respondent level* (i.e. the same aggregation level as response data and paradata). See Online Appendix, Sect. 7, Table A.13 for descriptive statistics of RQIs.

The RQIs can be grouped into two sets. The first set comprises the six *direct* RQIs, which reflect actual response quality problems and are therefore of greater importance:

1. *Breakoff* is a dichotomous characteristic used to describe respondents who quit (anywhere in the questionnaire) before finishing it completely (Callegaro et al., 2015).
2. *Item nonresponse* is expressed as the number of unanswered items to which the respondent was exposed divided by the number of all items presented to the respondent.
3. *Straightlining*, a form of satisficing behaviour (Kim et al., 2019; Roberts et al., 2019), was calculated as the number of grids (out of four attitudinal grids) where a respondent's answers had a standard deviation of zero (i.e. the respondent selected the same response for all items). About half of the items in the grid on the Big Five personality dimensions were reverse worded (see Donnellan et al., 2006). Missing values were ignored, but only grids with a minimum of two items answered were considered.
4. *Extreme and midpoint responses* are also forms of satisficing (Roberts et al., 2019). The shares of extreme negative, extreme positive (i.e. left-most and right-most response options) and midpoint responses were identified for each item in the four attitudinal grids. The corresponding means were then calculated for each respondent.
5. *Instructional manipulation check (IMC) failure* indicates the number of attention failures (e.g. Morren & Paas, 2020; Revilla & Ochoa, 2015). Two fictitious online stores were included in a grid for online shopping. Respondents failed the IMC if they stated that they had visited a fictitious online store. They could fail the IMC one (5%) or two times (4%). Item nonresponse was not counted as an IMC failure.
6. *Outliers* were based on Mahalanobis distance (De Maesschalck et al., 2000; Peck & Devore, 2012), which detected respondents with very unusual response patterns (Curran, 2016; Hong et al., 2020) that likely reflected inconsistent (or even random or blind) responses. The

metrics used variables from the four attitudinal grids; a higher score indicated less consistent responses. Respondents were identified as outliers if they had statistically significant distance values relative to the corresponding centroid in multivariate space ($p < 0.01$).

The second set encompasses three *indirect* RQIs associated with undesirable response styles that have potentially negative effects on response quality:

7. Self-reported multitasking can negatively affect response quality (Sendelbah et al., 2016). *Concurrent multitasking* included activities that could be done in parallel with the responding process (e.g. listening to music or watching TV). *Sequential multitasking* meant pausing the response process due to alternative activities (e.g. visiting other websites and doing household chores). The number of reported multitasking activities was calculated for each respondent.
8. *Duration* comprised the time spent by the respondents on all survey questions. A natural log transformation was applied to the response time values to compensate for skewness.
9. *Effort* and *burden* were based on self-reported scores (five-point scale) to two questions: 1) 'How much did you work at providing the most accurate answers you can to the questions in this survey?' and 2) 'How burdensome was it to complete this survey?' Effort and burden, though related, are two distinct measures of the conceptual dimension related to the perception of questionnaire difficulty. They are thus treated separately in the analysis, although they are addressing the same concept of questionnaire difficulty.

5.5 Respondent characteristics

Respondent characteristics included three standard sociodemographic variables—age, gender and education—and five variables representing the Big Five personality traits (i.e. extraversion, agreeableness, conscientiousness, neuroticism and openness [also called imagination]) calculated from 20 items (see Donnellan et al., 2006). Age and gender are used here for illustration rather than primary focus. When precise age and gender data are available from the survey or auxiliary sources, corresponding estimates from paradata are redundant. Nonetheless, exploring the associations with paradata serves multiple potential purposes. For example, understanding this correlation is valuable when respondent information is missing or inaccurately recorded. Such analysis also contributes to assessing the predictive power of paradata.

5.6 Survey estimates

Regarding survey estimates, it is important to acknowledge that the empirical study specifically focused on respondents' activities on the Internet. However, it should be noted that the pattern of Internet activities is often associated with various other substantive and methodological issues, including survey participation (Bottoni & Fitzgerald, 2021). Thirteen typical survey estimates from this survey were chosen for further analysis, seven of which were related to Internet use (see Online Appendix, Table A.15) and six to general trust in computers (see Online Appendix, Sect. 9, Table A.15). Due to the different contexts addressed by these two sets of estimates, they were observed, analysed and interpreted separately.

6 Results

6.1 Selection, computation and reduction of paradata indicators

The literature (Sect. 4) served as a starting point for the identification of paradata indicators. The exceptions—which were already excluded when capturing the corresponding raw paradata (Sect. 5.3)—were a few highly specific paradata indicators that appeared only in single research studies and were also extremely complex to capture and process. Examples include the angular velocity of the mouse pointer (e.g. Cepeda et al., 2021), time elapsed between key press and key release (e.g. Tzafilkou & Nicolaos, 2018) and detailed mouse movement trajectories (e.g. Fernández-Fontelo et al., 2022).

In addition, the screen resolution paradata (i.e. width, height, pixel ratio), which were captured and included among the raw paradata (Sect. 5.3), served exclusively for describing device characteristics and identifying device type. These attributes (along with associated events such as zoom changes and window resizes) underwent no separate processing due to intricate technical challenges. Specifically, the integration of device type, screen size, browser type, scaling settings, and other specifics of respondent's device settings proves extremely complex. To our knowledge, no evidence exists to support the value of such an endeavour, and existing literature presents no solution. Earlier research addressed only screen resolution challenges related to varying questionnaire appearance across devices (Horwitz et al., 2013), differences in survey code presentation across browsers (Kaczmirek, 2009), capturing mouse data to account for resolution differences (Jenkins et al., 2015), and detecting respondents' browser window maximisation (Stieger & Reips, 2010). To our knowledge,

no study connected standalone screen resolution indicators to response quality, respondent characteristics, or survey estimates. Nonetheless, as mentioned, the screen resolution paradata remain useful (as used here) for device identification, distinguishing between PCs and SPs.

After the literature review (Sect. 4) and above step of *preliminary* considerations (i.e. omission of some paradata), the *first* step in determining the paradata indicators involved establishing a set of 112 *initial paradata indicators*. For their calculation, a Python script (Berzelak et al., 2022) was used to process the raw paradata (Sect. 5.3). The 112 initial paradata indicators were defined at different levels of aggregation (i.e. 8 at the item, 7 at the question, 29 at the page and 68 at the respondent levels) and could be structured into nine categories: questionnaire length, device, page navigation, responses, window focus, inactivity, clicks and pointer actions, page display and validation prompts (see Online Appendix, Sect. 5, Table A.9).

In the *second* step, the 112 initial paradata indicators were subject to careful inspection (i.e. expert evaluation) and correspondingly reduced based on seven potential exclusion criteria: redundancy (i.e. not providing substantial added value compared to another indicator), data quality or availability issues (i.e. high noise levels), very low predictive value (i.e. based on the literature), availability of a more accurate or relevant measure (i.e. in another indicator), not being relevant as a predictor (i.e. according to literature) and aggregation of an indicator to another level (e.g. the page-level indicator *total number of page number pageviews* was aggregated into a respondent-level indicator *total number of pageviews*). Three co-authors of the paper independently proposed and iteratively evaluated the initial 112 paradata indicators; the outcome is described in the last two columns of Table A.9 (Online Appendix, Sect. 5). This reduction process resulted in 29 paradata indicators, which were all aggregated to the respondent level (Table 2). It is worth noting that the 83 excluded paradata indicators (out of 112) were also highly specific and outside the scope of general usage in web surveys; they were rarely found in the literature and mainly appeared in single research studies with an extremely narrow focus.

The purpose of the *third* step was to further refine the selection process by closely analysing the remaining 29 paradata indicators for any overlap and eliminating any redundant indicators that may have been present. In cases where an indicator was conceptually very similar, highly correlated and substantially less relevant than another existing indicator, it was removed and replaced. Three co-authors implemented the above criteria and conducted the evaluations independently. Subsequently, they reached a consensus on the outcomes (see Online Appendix, Sect. 3). As a result, 15 of the 29 paradata indicators were eliminated because they could be substituted by a similar but more

Table 2*The 29 respondent-level paradata indicators and the reduced set of 14 paradata indicators*

#	Paradata indicator name	Selected for further processing	Selected for key paradata indicators
1	Total number of pageviews	No; replaced by #3	–
2	Total number of page visits	No; replaced by #3	–
3	Total number of pages visited	Yes	Yes
4	Number of repeatedly visited pages	Yes	Yes
5	Duration	Yes	Yes
6	Duration adjusted for focus-out	No; replaced by #5, #29	–
7	Total number of branching items omitted	Yes	Yes
8	Type of device	Yes	Yes
9	Device brand	No; replaced by #8	–
10	Device model	No; replaced by #8	–
11	Device touch capability	No; replaced by #8	–
12	Browser	No; replaced by #8	–
13	Browser version	No; replaced by #8	–
14	Operating system ^a	Yes	Yes
15	Operating system version	No; replaced by #8, #14	–
16	Total number of responses provided in the questionnaire	Yes	No; high multicollinearity
17	Total number of answer changes	Yes	Yes
18	Total number of items with answer changes	No; replaced by #17	–
19	Total number of validation prompts	No; replaced by #20	–
20	Total number of item nonresponse prompts	Yes	Yes
21	Total number of clicks	Yes	No; high multicollinearity
22	Total number of excess clicks (i.e. total clicks minus clicks needed to complete certain action)	Yes	Yes
23	Mouse pointer movement duration	No; replaced by #24	–
24	Mouse pointer movement distance ^b	Yes	Yes
25	Mouse pointer movement speed	No; replaced by #22	–
26	Total number of pages with orientation change	No; replaced by #8	–
27	Total number of focus-out events	No; replaced by #28	–
28	Total number of focus-out events (longer than five seconds)	Yes	Yes
29	Total focus-out duration	Yes	Yes

Some manuscripts are associated with multiple paradata domains.

^a Effects found for SP, not for PC

^b Measured only for PC, not for SP

relevant indicator, as indicated in the second column of Table 2. This led to a *reduced set* of 14 paradata indicators.

In the *fourth* step, the 14 paradata indicators were used as predictors (i.e. independent variables) in the 34 multiple regression analyses, where the dependent variables were each from the three domains: 13 RQIs, 8 respondent characteristics and 13 survey estimates (see Sect. 5). All 14 indicators showed at least some correlation with dependent variables from the three domains; however, multicollinearity with dependent sets of variables was detected for the *total number*

of responses (#16, Table 2) and *total number of clicks* (#21, Table 2), which were therefore removed. Multicollinearity was considered problematic if the variance inflation factor (VIF) exceeded 5, indicating that highly correlated predictors would not be suitable for inclusion in the model. We thus ended up with 12 *key paradata indicators* listed in Table 2, which were included in the freely available dataset (Vehovar et al., 2023b) and code (Vehovar et al., 2023a).

The key paradata indicators were otherwise interlinked with complex multivariate correlation patterns (see Online

Table 3

Associations between key paradata indicators and response quality indicators (second line shows standard errors)

	Break-offs ^a	Outliers ^a	Item non-resp. ^{b,c}	Straightening ^b	Extreme Positive ^{b,c}	Extreme Negative ^{b,c}	Mid-point ^{b,c}	Concurrent Multitask ^b	Sequential Multitask ^b	Duration ^b	Effort ^b	Burden ^b	IMC ^b
Device (Ref.: PC)	3.054	0.967	0.014	0.002	0.057**	0.042*	-0.025	-0.070***	-0.076***	0.104***	0.030	-0.019	-0.013
Focus-out events	0.596	0.179	0.039	0.026	0.300	0.456	0.682	0.020	0.022	0.034	0.035	0.036	0.014
Focus-out duration	1.095	0.986	-0.005	-0.007	0.047	-0.005	0.047	0.097***	0.062**	0.076***	0.007	-0.003	-0.024
	0.081	0.053	0.010	0.007	0.076	0.116	0.173	0.005	0.006	0.009	0.009	0.009	0.004
	1.319*	0.984	0.021	-0.009	-0.023	0.005	-0.043	0.049	0.130***	0.347***	-0.004	0.054*	-0.031
	0.128	0.047	0.009	0.006	0.071	0.107	0.160	0.005	0.005	0.008	0.008	0.009	0.003
Duration	0.580	0.812	0.008	-0.010	-0.053**	-0.031	-0.001	0.031	0.069**	^d	0.024	0.007	-0.036*
	0.450	0.132	0.021	0.014	0.159	0.241	0.360	0.010	0.012	^d	0.019	0.019	0.008
Item nonresponse prompts	0.742	0.931	0.524***	0.060**	-0.001	-0.002	0.028	-0.006	-0.033	-0.012	-0.006	-0.023	0.119***
	0.258	0.067	0.012	0.008	0.096	0.145	0.217	0.006	0.007	0.011	0.011	0.011	0.005
Excess clicks	1.007	1.000	-0.033	-0.031	0.004	-0.031	-0.019	0.057**	0.030	0.021	-0.040*	0.017	0.017
	0.004	0.001	0.000	0.000	0.002	0.003	0.005	0.000	0.000	0.000	0.000	0.000	0.000
Branching items omitted	0.995	1.027*	-0.031	-0.055**	0.009	0.161***	-0.029	-0.039*	-0.072***	0.014	0.031	-0.024	-0.518***
	0.035	0.011	0.002	0.002	0.018	0.027	0.041	0.001	0.001	0.002	0.002	0.002	0.001
Pages visited	1.006	1.000	-0.081***	-0.053**	-0.030	0.012	-0.018	0.019	-0.002	0.144***	0.039*	0.155***	-0.027
	0.005	0.002	0.000	0.000	0.003	0.005	0.007	0.000	0.000	0.000	0.000	0.000	0.000
Repeatedly visited pages	1.09	0.918	0.071***	-0.027	0.039*	-0.028	0.02	0.009	0.013	0.168***	0.046*	0.02	-0.038*
	0.088	0.047	0.008	0.005	0.061	0.093	0.139	0.004	0.005	0.007	0.007	0.007	0.003
Answer changes	0.989	1.020***	-0.051**	0.015	0.085***	0.071***	-0.03	0.018	0.007	0.066***	0.022	0.062**	0.02
	0.027	0.004	0.001	0.001	0.011	0.016	0.024	0.001	0.001	0.001	0.001	0.001	0.001
Dem. R-sq. (in %) ^{e,f}	8.1	2.6	1.1	2.4	4.5	8.3	6.1	2.4	3.2	0.1	10.4	7.2	0.9
Part. R-sq. (in %) ^{e,g}	11.6	3.8	4.9	2.9	5.0	10.4	6.2	3.7	6.4	9.7	11.2	10.5	24.9
Tot. R-sq. (in %) ^{e,h}	17.5	5.3	29.2	3.2	5.6	10.8	6.3	5.4	8.8	25.6	11.1	10.9	26.3
Observations	3222	3168	3131	3131	3131	3131	3131	3116	3116	3131	3123	3123	3113

Each model included controls for the sociodemographic characteristics. Sampling weights applied.

^a Odds ratios are reported (binary logistic regression).^b Standardised beta coefficients are reported (multiple linear regression).^c This variable was scaled by 100 when reporting standard errors so that its share values correspond to percentages, thus avoiding zero values.^d Duration was not included here because it was present in this study as both an RQ (i.e. dependent variable) and a paradata indicator (i.e. independent variable).^e Nagelkerke's pseudo R² coefficient is shown for binary logistic regression, and the adjusted R² coefficient is shown for linear regression.^f Dem. R² shows the R² coefficient when only controlling for the sociodemographic characteristics.^g Part. R² shows the R² coefficient when including six paradata indicators, which can be calculated also on server-side (see Fig. 1 and Sect. 7.2), and controlling for the sociodemographic characteristics.^h Total R² shows the R² coefficient when including 10 key paradata indicators and controlling for the sociodemographic characteristics.* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Appendix, Sect. 6, Tables A.10–A.12). Interestingly, the *total number of pages visited* (#3, Table 2) and the *total number of branching items omitted* (#7) did not show notable multicollinearity. This is perhaps because engaging in fewer online activities resulted in the respondent being exposed to fewer items, although not necessarily to fewer pages (i.e. non-displayed items were concealed within the corresponding grid, which was located on a page still visited by the respondent). For instance, if the respondent did not indicate visiting the website of a specific store in a previous question, the item about online shopping in that store would not have been shown, although the corresponding page (with other items) would have been shown. It should be added that the *total mouse pointer movement distance* (#24, Table 2) data were only available for PCs, as SPs do not use a pointer, while the analysis revealed that *operating system* (#14) was relevant only on SPs but not on PCs, where the effects were negligible. Therefore, these two paradata indicators were excluded from the main analysis of key paradata indicators (Sects. 6.2–6.4), where regressions required a complete set of paradata indicators. They are analysed and discussed separately in Sect. 6.5.

6.2 Key paradata indicators and response quality indicators (RQIs)

Sects. 6.2–6.4 utilized a narrow set of 10 key paradata indicators. Sect. 6.5 included additional analyses for the *total mouse pointer movement distance* (available only for PCs) and *operating system* paradata indicators (relevant only for SPs). In this section, a set of 10 key paradata indicators was used to analyse their association with RQIs. Therefore, each RQI was included as a dependent variable in a series of 13 regression analyses where the paradata served as predictors; each model also controlled for all respondent characteristics (i.e. age, gender, education and the Big Five personality traits) to improve the generalisability of the results and minimise confounding effects. The results (in Table 3) show that all paradata indicators were statistically significantly associated with at least one RQI. Statistical analyses, including binary logistic regression and linear regression, were conducted using IBM SPSS software version 28. Due to the sufficiently large sample size and the limited number of paradata indicators, there was no need for additional data reduction techniques (e.g. Sharma, 2019).

The *device type* (i.e. SP) was associated with less concurrent multitasking (i.e. standardised beta coefficient of -0.070 , $p < 0.001$), less sequential multitasking (i.e. standardised beta coefficient of -0.076 , $p < 0.001$), longer duration (as expected; standardised beta coefficient of 0.104 , $p < 0.001$), and additional extreme responses (i.e. standardised

beta coefficients of 0.057 and 0.042 , $p < 0.01$ and $p < 0.05$). The value of -0.070 related to concurrent multitasking represents a standardised beta coefficient derived from multiple linear regression. It signifies the expected change in the number of concurrent multitasking activities for a one-unit shift in the independent variable (specifically, transitioning from a PC to an SP), while accounting for other variables in the model. The negative coefficient thus indicates that switching from a PC to an SP is associated with a decrease in the number of concurrent multitasking activities by 0.070 . Additionally, descriptive statistics (Online Appendix, Sect. 7, Table A.13) show that the mean number of concurrent multitasking activities in the study was 0.22 . A decrease by 0.070 thus signifies that the number of concurrent multitasking activities drops by approximately 32% on average when transitioning from a PC to an SP. For sequential multitasking, the coefficient of -0.076 indicates that switching from a PC to an SP is associated with a decrease in the number of multitasking activities by 0.076 , or by 36% on average. For duration (refer to Sect. 5.4), the coefficient of 0.104 suggests that shifting from a PC to an SP raises the natural log-transformed survey duration by 0.104 . If the median duration of 20.6 min were natural log-transformed and increased by 0.104 (i.e. $e^{0.104}$), this results in a duration of approximately 22.9 min, indicating an increase of about 2.3 min or 11% when transitioning from a PC to an SP. Regarding additional extreme responses, the coefficients of 0.057 (for extreme positive responses) and 0.042 (for extreme negative responses) indicate that transitioning from a PC to an SP approximately doubles the average number of extreme positive responses across the four attitudinal grids, while increasing the average number of extreme negative responses by about a third across these grids.

A greater *number of focus-out events* was associated with more concurrent and sequential multitasking and longer duration. A longer *focus-out duration* was associated with longer duration and more sequential multitasking, but also a higher level of respondent burden and breakoffs. Longer *duration* was associated with a greater amount of sequential multitasking and fewer IMC failures, thus reflecting more attentive (and slower) respondents. In addition, an increase in the duration was also associated with fewer extreme positive responses. An increase in the *number of item nonresponse prompts* was associated with more straightlining, more IMC failures and a higher item nonresponse rate. A greater number of *excess clicks* was associated with more concurrent multitasking and a lower score for self-reported effort, which may reflect less attentive respondents.

A greater *number of branching items omitted* was associated with less straightlining, less concurrent and sequential multitasking and fewer IMC failures, as well as additional outliers and extreme negative responses. A greater *number*

of pages visited was expectedly associated with longer duration and a higher level of self-reported effort and burden. In addition, this was related to a lower item nonresponse rate and less straightlining. Because the analysis revealed no adverse effects, such as outliers or breakoffs, associated with increased page visits, it suggests that more visits signify greater attentiveness and motivation among respondents. This aligns with the survey's branching, as ICT-oriented respondents were expected to encounter more pages.

While a greater *number of repeatedly visited pages* was associated with a higher item nonresponse rate and additional extreme positive responses, it was also related to longer duration (as expected), a higher level of self-reported effort and fewer IMC failures. A greater *number of answer changes* was associated with additional outliers, additional extreme responses, longer duration and a greater level of self-reported burden but less item nonresponse.

6.3 Key paradata indicators and respondent characteristics

Each of the eight respondent characteristics (Table 4) was used in the regression analysis as a dependent variable, while 10 key paradata indicators served as predictors. Each model was also controlled for the remaining respondent characteristics (i.e. seven out of eight characteristics, except for the characteristic used as the dependent variable) to avoid confounding effects and provide generalisability and control. Having respondent characteristics (e.g. gender) as dependent variables may challenge the conventional principles of causality, as gender, for example, can potentially influence the response style reflected in paradata, while the reverse is not possible. However, in this context, we estimate the likelihood that a respondent had certain personal characteristics (e.g. being female) based on available paradata indicators. Therefore, the corresponding model needed to be oriented in the opposite direction. The results (Table 4) showed that, except for *number of excess clicks* and *number of repeatedly visited pages*, all key paradata indicators had statistically significant associations with at least one respondent characteristic. However, there were somewhat fewer effects observed with the Big Five personality traits.

The *device type* (i.e. SP) was associated with gender, with SP users more likely to be female than male (i.e. odds ratio of 2.22, $p < 0.001$), as well as with lower age, a lower education level—which is not surprising due to the lower age of SP respondents—and greater extraversion. A greater *number of focus-out events* was also associated with lower age. A longer *focus-out duration* was associated with lower age, a higher education level and higher scores for agreeableness, neuroticism and openness. Longer *duration* was associated with higher age. An increase in the

number of item nonresponse prompts was also associated with higher age and a lower education level. A greater *number of branching items omitted* was associated with higher age, a lower education level and lower scores for extraversion and openness. A greater *number of pages visited* was associated with a higher conscientiousness score. A greater *number of answer changes* was associated with lower age, a higher education level, and a lower conscientiousness score.

6.4 Key paradata indicators and survey estimates

Each of the 13 survey estimates (i.e. seven estimates on Internet use and six estimates related to trust in computers) served as dependent variables in the regression analyses, where 10 key paradata indicators were used as predictors. Again, all respondent characteristics were controlled for.

The analysis of seven estimates, which addressed Internet use (Table 5), revealed that the *device type* (i.e. SP) was associated with increased Internet usage frequency, a higher utilization of SPs and smart TVs for web browsing, reduced reliance on PCs for web browsing and greater use of SPs for personal purposes. A greater *number of focus-out events* was associated with a lower utilization of smart TVs for web browsing. A longer *focus-out duration* was associated with a higher utilisation of tablets and smart TVs for web browsing. Longer *duration* was associated with a greater use of PCs for web browsing. An increase in the *number of item nonresponse prompts* was associated with a lower frequency of Internet usage and reduced reliance on PCs and smart TVs for web browsing. A greater number of *excess clicks* was associated with a lower utilization of smart TVs for web browsing. A greater *number of branching items omitted* was associated with a lower frequency of Internet usage, reduced utilization of PCs, SPs, tablets and smart TVs for web browsing, as well as decreased use of SPs for personal purposes. A greater *number of pages visited* was associated with reduced utilization of PCs for web browsing. Moreover, a greater *number of repeatedly visited pages* was also associated with reduced reliance on PCs for web browsing; additionally, it was associated with an increased frequency of Internet usage. A greater *number of answer changes* was associated with a lower frequency of Internet usage.

The analysis of six estimates focusing on trust in computers (Table 6) showed an association between the *device type* (i.e. SP) and reduced trust in spelling and grammar check functions. In contrast, a greater *number of focus-out events* was associated with higher trust in spelling and grammar checks. A longer *focus-out duration* was associated with higher trust in playlist selection. Longer *duration* was associated with reduced trust in spelling and grammar checks.

An increase in the *number of item nonresponse prompts* was associated with reduced trust in best route selection in a GPS navigation app while driving and higher trust in the diagnosis of medical status by an AI system. A greater *number of branching items omitted* was associated with reduced trust in functions related to spelling and grammar checks, playlist selection, best route selection in a GPS navigation app while driving, autonomous driving of a motor vehicle and diagnosis of medical status by an AI system. A greater *number of pages visited* was associated with reduced trust in functions related to autocompletion of text, spelling and grammar checks and autonomous driving of a motor vehicle and higher trust in best route selection in a GPS navigation app while driving. A greater *number of repeatedly visited pages* was associated with increased trust in autocompletion of text and playlist selection functions. A greater *number of answer changes* was associated with increased trust in autocompletion of text.

It could be observed that *a greater number of branching items omitted* was statistically significant across nearly all models presented in Tables 5 and 6. Furthermore, these associations showed diminished trust ratings across all six estimates in Table 6. This is not surprising, as respondents subjected to a reduced item count due to branching participated in fewer activities related to Internet use, potentially exhibiting in decreased confidence in technology. This issue is addressed in the discussion.

6.5 Specifics of PCs and SPs

In addition to the above 10 indicators, which apply to both PCs and SPs, separate analyses were performed for the total *mouse pointer movement distance* (available only for PCs) and *operating system* paradata indicators (relevant only for SPs). An increase in the *mouse pointer movement distance* (Online Appendix, Sect. 4.1, Tables A.1–A.4) was significantly associated with lower straightlining (standardised beta coefficient of -0.07 , $p < 0.01$) and longer duration (0.11 , $p < 0.001$), suggesting the increased attentiveness of the respondents. However, as a statistically significant impact was only detected for 2 out of 13 RQI variables, the associations between *mouse pointer movement distance* and RQIs were relatively weak. Regarding respondent characteristics, an increase in the *mouse pointer movement distance* was significantly associated only with increased age (0.16 , $p < 0.001$), signalling that older respondent covered a longer total distance with the pointer, which might reflect specific patterns characteristic of older respondents (e.g. moving the pointer while reading, more hesitation and less impulsive responding). However, this might also reflect more frequent use of larger screens or some other screen resolution specifics. In addition, *mouse pointer movement*

distance had no statistically significant association with any of the 13 survey estimates. All in all, the *mouse pointer movement distance* acted as a relatively weak paradata indicator.

The *operating system* was analysed only for SPs, while its effects on PCs were negligible. In terms of the RQIs, the *iOS operating system* (Online Appendix, Sect. 4.2, Tables A.5–A.8) was associated with fewer extreme positive responses (standardised beta coefficient of -0.07 , $p < 0.05$) and higher perceived burden (0.09 , $p < 0.01$), suggesting that iOS respondents were slightly more attentive. For respondent characteristics, the *iOS operating system* was expectedly associated with younger age (beta of -0.18 , $p < 0.001$) and the female gender (odds ratio of 1.52 , $p < 0.05$), as well as lower agreeableness (beta of -0.07 , $p < 0.05$). This is consistent with some studies that showed similar differences in personality between users of different SP operating systems (e.g. Ang et al., 2018; Götz et al., 2017). Regarding the estimates, the *iOS operating system* was significantly associated only with increased odds of using a computer to browse the web (odds ratio of 3.13 , $p < 0.01$) and lower trust scores for selecting a playlist function (beta of -0.07 , $p < 0.05$).

7 Discussion

The main research question (Sect. 4) was related to the search for the minimal set of key paradata indicators associated with response quality, respondent characteristics or survey estimates while remaining practical enough for general usage. For this purpose, a list of 112 initial paradata indicators was established based on relevant literature and was subsequently subjected to reduction process to finally yield 12 key paradata indicators, 10 of which were related to both PCs and SPs. Regression analysis was used to systematically investigate the association between the paradata indicators and 13 RQIs, 8 respondent characteristics and 13 survey estimates related to Internet usage and trust in computers. Fig. 1 summarises the results.

All 10 paradata indicators were statistically significantly associated with at least some dependent variables; however, there were notable differences (see Fig. 1). On the one hand, the *number of branching items omitted* was significantly associated with almost all dependent variables in both sets of survey estimates (i.e. 100% of 7 variables and 83% of 6 variables), as well as 6 of 13 RQIs (i.e. 46%) and 4 of 8 respondent characteristics (i.e. 50%). A greater number of associations were also observed for *device type* and *number of pages visited*. Conversely, the *number of excess clicks* was significantly associated with only 2 out of 13 (i.e. 15%) RQIs and almost none of the respondent characteristics or survey estimates.

Table 4

Associations between key paradata indicators and respondent characteristics (second line shows standard errors)

	Gender (ref.: male) ^a	Age ^b	Education (ref.: lower) ^a	Extraversion ^b	Agreeableness ^b	Conscientiousness ^b	Neuroticism ^b	Openness ^b
Device type (Ref.: PC)	2.216***	-0.278***	0.694***	0.045*	-0.018	-0.009	0.004	0.000
Focus-out events	0.088	0.437	0.087	0.028	0.023	0.024	0.026	0.026
Focus-out duration	1.007	-0.045*	0.994	-0.013	0.006	-0.002	-0.038	-0.026
Duration	0.023	0.116	0.023	0.007	0.006	0.006	0.007	0.007
Item nonresponse prompts	0.984	-0.159***	1.056**	-0.022	-0.048*	0.036	0.060**	0.059**
Excess clicks	0.021	0.107	0.021	0.007	0.005	0.006	0.006	0.006
Branching items omitted	1.019	0.150***	0.949	0.019	0.032	-0.004	-0.017	-0.012
Pages visited	0.047	0.239	0.047	0.015	0.012	0.013	0.014	0.014
Repeatedly visited pages	1.014	0.045**	0.946*	0.017	-0.014	-0.031	0.002	-0.024
Answer changes	0.028	0.146	0.028	0.009	0.007	0.008	0.008	0.008
Dem. R-sq. (in %) ^{c,d}	1.000	0.023	1.000	0.015	0.013	-0.006	0.010	-0.009
Part. R-sq. (in %) ^e	0.001	0.003	0.001	0.000	0.000	0.000	0.000	0.000
Tot. R-sq. (in %) ^{c,f}	1.010	0.201***	0.986**	-0.110***	-0.033	0.018	-0.021	-0.096***
Observations	0.005	0.027	0.005	0.002	0.001	0.001	0.002	0.002
	1.000	-0.003	1.001	0.005	-0.023	0.056**	-0.016	-0.024
	0.001	0.005	0.001	0.000	0.000	0.000	0.000	0.000
	1.028	0.011	0.996	-0.017	0.016	0.014	-0.007	0.013
	0.018	0.093	0.018	0.006	0.005	0.005	0.005	0.005
	0.997	-0.186***	1.010**	0.011	0.011	-0.069***	-0.032	-0.008
	0.003	0.016	0.004	0.001	0.001	0.001	0.001	0.001
	12.2	10.3	11.6	14.8	19.6	10.0	11.6	13.7
	16.3	26.3	13.2	16.1	19.6	10.2	11.5	14.3
	16.3	31.6	13.8	16.1	19.7	10.8	11.6	14.4
	3222	3131	3222	3131	3131	3131	3131	3131

Each model included controls for the sociodemographic characteristics. Sampling weights applied.

^a Odds ratios are reported (binary logistic regression).

^b Standardised beta coefficients are reported (multiple linear regression).

^c Nagelkerke's pseudo R² coefficient is shown for binary logistic regression, and the adjusted R² coefficient is shown for linear regression.

^d Dem. R² shows the R² coefficient when only controlling for the sociodemographic characteristics.

^e Partial R² shows the R² coefficient when including six paradata indicators, which can be calculated also on server-side (see Fig. 1 and Sect. 7.2), and controlling for the sociodemographic characteristics.

^f Total R² shows the R² coefficient when including 10 key paradata indicators and controlling for the sociodemographic characteristics.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 5

Associations between key paradata indicators and estimates about Internet use (second line shows standard errors)

Device type (Ref.: PC)	Used any of the following devices to browse the web in the last 12 months										Smartphone for personal purposes ^b
	Internet use freq.: last 12 months ^b	Desktop or laptop computer ^a	Tablet computer ^a	Mobile phone or smartphone ^a	Smart TV ^a	Other devices ^a					
	0.089***	0.130***	1.057	6.192***	1.349**	1.376					5.759***
Focus-out events	0.018	0.246	0.089	0.245	0.106	0.168					0.357
	0.038	1.059	0.976	1.027	0.944*	1.025					1.073
Focus-out duration	0.005	0.091	0.023	0.060	0.028	0.034					0.131
	0.011	1.055	1.046*	1.010	1.068**	0.994					1.117
Duration	0.004	0.058	0.021	0.047	0.025	0.037					0.082
	-0.017	1.247*	0.929	1.088	0.948	1.050					0.895
Item nonresponse prompts	0.009	0.111	0.049	0.133	0.059	0.082					0.169
	-0.041*	0.790***	1.019	0.949	0.925*	0.944					0.895
Excess clicks	0.006	0.044	0.028	0.052	0.038	0.053					0.069
	-0.035	0.999	1.001	0.999	0.998*	1.001					0.997
Branching items omitted	0.000	0.002	0.001	0.001	0.001	0.001					0.002
	-0.103***	0.965**	0.929***	0.947***	0.916***	0.907***					0.919***
Pages visited	0.001	0.012	0.006	0.010	0.007	0.010					0.014
	-0.017	0.998*	1.000	1.000	1.001	1.000					0.995
Repeatedly visited pages	0.000	0.002	0.001	0.002	0.001	0.002					0.002
	0.048*	0.929**	1.023	1.049	0.986	1.015					1.053
Answer changes	0.004	0.032	0.018	0.043	0.024	0.034					0.056
	-0.050*	1.030	0.998	1.002	1.007	1.008					0.996
	0.001	0.010	0.003	0.007	0.003	0.005					0.008
Dem. R-sq. (in %) ^d	3.8	5.2	4.0	12.9	8.4	9.4					13.1
Part. R-sq. (in %) ^e	4.9	18.2	11.9	21.6	17.3	18.8					21.5
Tot. R-sq. (in %) ^f	5.5	21.0	12.1	21.7	18.2	18.8					23.1
Observations	3126	3223	3217	3217	3217	3217					3217

Each model included controls for the sociodemographic characteristics. Sampling weights applied.

^a Odds ratios are reported (binary logistic regression).
^b Standardised beta coefficients are reported (multiple linear regression).
^c Nagelkerke's pseudo R² coefficient is shown for binary logistic regression, and the adjusted R² coefficient is shown for linear regression.
^d Dem. R² shows the R² coefficient when only controlling for the sociodemographic characteristics.
^e Part. R² shows the R² coefficient when including six paradata indicators, which can be calculated also on server-side (see Fig. 1 and Sect. 7.2), and controlling for the sociodemographic characteristics.
^f Total R² shows the R² coefficient when including 10 key paradata indicators and controlling for the sociodemographic characteristics.
 * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6

Associations between key paradata indicators and estimates about trust in computers for task performance (second line shows standard errors)

	Text autocompletion	Spelling and check	Grammar and preferences	Playlist selection based on music preferences	Optimizing navigation route while driving	Autonomous vehicle driving	Medical diagnosis by AI
Device type (Ref.: PC)	-0.020	-0.051**	0.026	0.030	-0.020	-0.009	-0.009
Focus-out events	0.041	0.040	0.040	0.037	0.043	0.040	0.040
Focus-out duration	0.042	0.062**	0.074	0.019	-0.014	-0.010	-0.010
Duration	0.011	0.010	0.010	0.009	0.011	0.010	0.010
Item nonresponse prompts	-0.006	-0.005	-0.043*	-0.005	0.003	-0.003	-0.003
Excess clicks	0.010	0.009	0.009	0.009	0.010	0.009	0.009
Branching items omitted	-0.026	-0.054**	-0.017	-0.025	-0.010	0.032	0.032
Pages visited	0.022	0.021	0.021	0.020	0.023	0.021	0.021
Repeatedly visited pages	0.003	-0.026	-0.024	-0.046*	0.024	0.046*	0.046*
Answer changes	0.013	0.013	0.013	0.012	0.014	0.013	0.013
Dem. R-sq. (in %) ^{a,b}	0.028	0.003	0.030	-0.019	-0.013	0.015	0.015
Part. R-sq. (in %) ^{a,c}	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tot. R-sq. (in %) ^{a,d}	-0.126	-0.083***	-0.119***	-0.100***	-0.16***	-0.168***	-0.168***
Observations	0.002	0.002	0.002	0.002	0.003	0.002	0.002
	-0.043***	-0.057**	-0.012	0.042*	-0.05**	-0.028	-0.028
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.039*	0.032	0.043*	0.034	0.023	0.021	0.021
	0.008	0.008	0.008	0.008	0.009	0.008	0.008
	0.004*	0.015	-0.033	-0.012	0.000	0.010	0.010
	0.001	0.001	0.001	0.001	0.002	0.001	0.001
	1.5	1.3	1.6	1.2	2.2	2.7	2.7
	0.6	2.8	3.4	2.6	1.9	2.9	2.9
	2.1	4.1	5.0	3.8	4.1	5.6	5.6
	3130	3120	3123	3123	3121	3122	3122

Each model included controls for the sociodemographic characteristics. Sampling weights applied.

^a Nagelkerke's pseudo R² coefficient is shown for binary logistic regression, and the adjusted R² coefficient is shown for linear regression.

^b Dem. R² shows the R² coefficient when only controlling for the sociodemographic characteristics.

^c Partial R² shows the R² coefficient when including six paradata indicators, which can be calculated also on server-side (see Fig. 1 and Sect. 7.2), and controlling for the sociodemographic characteristics.

^d Total R² shows the R² coefficient when including 10 key paradata indicators and controlling for the sociodemographic characteristics.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

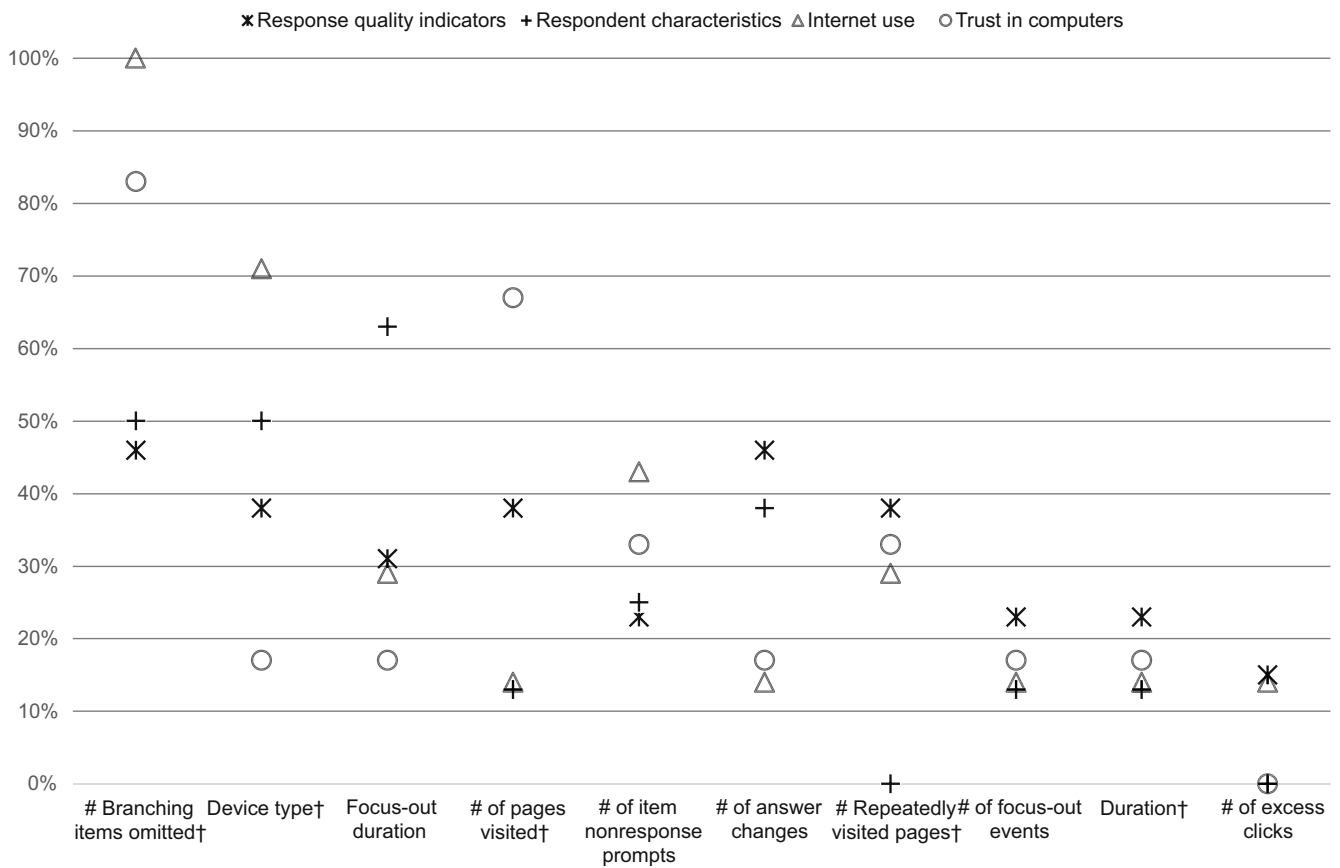


Fig. 1

The share (%) of the number of variables—among the total number of 13 RQI variables, 8 respondent characteristics, 7 estimates about Internet use and 6 estimates about trust in computers—that were statistically significantly ($p < 0.05$) associated with the corresponding paradata indicator († denotes possibility of server-side paradata capturing)

7.1 Patterns of interpretation

Based on the interpretations in Tables 3, 4, 5 and 6, four general patterns can be observed. The *first* general pattern is that the *number of branching items omitted* had significant associations with a large number of dependent variables (see also the peak shares in Fig. 1). Namely, the higher values indicated lower engagement in online activities, which is directly related to lower Internet usage and indirectly to lower trust in computers and sociodemographic variables (i.e. higher age and lower formal education) (cf. Alzahrani et al., 2017). Of course, this effect could only appear when questionnaire branching was related to computer skills and Internet usage (i.e. less intensive Internet users were exposed to fewer questions). Therefore, while this indicator was formally based on paradata, it did not reflect the respondents' response styles but rather the questionnaire content, structure and logic. If the aim is to include only the paradata indicators that reflect response style, then this indicator

should be moved to the set of covariates. This was done in Figure A.1 (see Online Appendix, Sect. 4.3) as a replication of Fig. 1 but without the *number of branching items omitted* as a paradata indicator, which was instead treated as a covariate (in a similar fashion to sociodemographic variables). However, presenting Figure A.1 without this paradata indicator does not affect the relationships and findings related to other key paradata indicators. Nevertheless, the presence of the *number of branching items omitted* in Fig. 1 illustrates the confounding role of the paradata indicators in situations where the questionnaire logic is correlated with the substantive content of the survey.

The *second* general pattern is related to the different levels of associations between the paradata indicators and the sets of dependent variables. Namely, all 10 paradata indicators were associated with RQIs moderately and evenly (i.e. between 15 and 45%) as indicated by dot chart markers with star symbols in Fig. 1. Conversely, the other three sets (i.e. respondent characteristics, estimates on Internet use and trust in computers) showed a much more unbalanced

pattern. This is particularly true for the two sets of survey estimates, where the *number of branching items omitted* stood out notably for both sets (i.e. exceeding 80% share in Fig. 1). Additionally, *device type* stood out within Internet use and the *number of pages visited* stood out within Trust in computers (the corresponding shares in Fig. 1 approximating 70% or higher). Regarding respondent characteristics, besides the *number of branching items omitted*, only three paradata indicators showed somehow higher shares (i.e. close to 40% or above) of statistically significant associations: *focus-out duration*, *device type* and *number of answer changes*.

The *third* general pattern is related to the respondent characteristics. The three sociodemographic variables (gender, age and education) showed the expected associations with the paradata indicators, particularly age (e.g. longer duration). Regarding the Big Five personality traits, some respondents were prone to faster and sometimes less careful responses, leading to lower response quality (Table 3), which is sometimes positively related to conscientiousness (see Table 4). These results confirm some other findings regarding the relationship between satisficing and conscientiousness (e.g. Sturgis et al., 2019). It is worth noting that the observed relationship between paradata indicators and age (e.g. *the number of answer changes* in Table 4) and known correlations between age and personality, as detected by prior psychology research (e.g. Donnellan & Lucas, 2008; Wortman et al., 2012), may have also resulted in the relationship between some paradata indicators and personality dimensions. Specifically, extraversion, neuroticism and openness tend to decline with age (e.g. Donnellan & Lucas, 2008; Wortman et al., 2012). However, it is important to note that in this paper, we did not directly investigate the relationship between sociodemographic variables (including personality traits) and RQIs because this is beyond the scope of the study. For reference, prior studies have been conducted on the relationship between personality dimensions and response style in surveys (e.g. Hibbing et al., 2019)—and more broadly between personality dimensions and other respondent characteristics (e.g. Donnellan & Lucas, 2008; Marsh et al., 2013; Roehrick et al., 2023)—but direct paradata were not included in those studies.

The *fourth* pattern is related specifically to response quality, which is by far the most studied domain in web survey paradata research. Our results were mostly consistent with the literature, particularly with respect to the negative impact on response quality arising from SP *device type* (e.g. de Leeuw & Toepoel, 2017; Fisher & Bernet, 2014; Mittereder, 2019; Vehovar et al., 2022), *number of focus-out events* (e.g. Höhne, Schlosser, et al., 2020; Sendelbah et al., 2016) and longer *duration* (e.g. Gummer & Roßmann, 2015; Matjašič et al., 2018; Vehovar et al., 2022). Within the response quality context, it is surprising that certain paradata indica-

tors—particularly the *number of item nonresponse prompts*, the *number of excess clicks* and *focus-out duration*—had relatively weak associations with RQIs. For PC users, this was true also for *mouse pointer movement distance*.

It is important to reiterate that the first pattern was specific to the study, as more intensive Internet users (due to branching) were exposed to a greater number of pages and questionnaire items. Consequently, a greater number of associations emerged between the study-specific estimates (i.e. Internet use and trust in computers) and the corresponding paradata indicators (i.e. *number of pages visited* and *number of branching items omitted*). However, this pattern can manifest in any study where questionnaire branching relies on Internet use, and survey estimates are (indirectly) associated with Internet use. Yet, the remaining three patterns discussed above are more prevalent and can generally be anticipated in studies.

7.2 Practical considerations

The 12 identified paradata indicators can enhance response data and be archived alongside response data, providing valuable insights into response quality, respondent characteristics and survey estimates. Furthermore, the paradata indicators can be aggregated for additional analysis. For example, clustering analysis could be performed on the paradata indicators to improve the identification and utilisation of sociodemographic segments.

The aim of this study was not only to determine the minimal set of key paradata indicators but also to identify robust and easily captured indicators potentially suitable for general usage. Achieving the latter in survey practice can be relatively straightforward only for paradata indicators that can be potentially captured at the server-side, which include the *number of pages visited*, *number of repeatedly visited pages*, *duration*, *number of branching items omitted*, *device type* and *operating system* (i.e. the paradata indicators #3, #4, #5, #7, #8, and #14 in Table 2). These paradata indicators are also denoted in Fig. 1. Still, the overview of 77 *web survey software* tools that provide a readily available trial version in the English language (see Vehovar et al., 2021) showed that, by default, the majority of software provided very few server-side paradata indicators. At most, they provided the following: (i) an indicator of the completeness level of the questionnaire, (ii) device paradata string (e.g. device type, browser, operating system and screen resolution), (iii) timestamps at page level and (iv) overall duration. The lack of more systematic and extensive server-side paradata integration into the web survey software perhaps also reflects the perception that—from the perspective of users and software providers—the anticipated usefulness of paradata indicators is generally low.

Integrating the collection of client-side paradata into web survey software is not typical, as this requires the use of specialised client-side scripts. The client-side paradata indicators, namely the *number of answer changes*, *number of item nonresponse prompts*, *number of excess clicks*, *mouse pointer movement distance*, *number of focus-out events* and *focus-out duration* (i.e. paradata indicators #17, #20, #22, #24, #28, and #29 in Table 2), all require complex paradata capturing and processing. Various procedures exist, ranging from more general (e.g. Berzelak et al., 2022; Heerwegh, 2003; Höhne, Schlosser, et al., 2020; Kaczmirek & Neubarth, 2007) to highly specialised, such as those addressing very detailed mouse movement (e.g. Peng & Ostergren, 2016). While some advanced web survey software tools capture and integrate some types of client-side paradata, such as the time stamp for when an item was answered within a survey page, they still fail to capture and integrate all the necessary paradata needed to calculate key paradata indicators, such as focus-out events (e.g. Höhne & Schlosser, 2018).

In any case, the existing state of paradata integration in web survey software presents an opportunity to improve paradata collection and facilitate the establishment of a standardised set of paradata indicators. Due to potential challenges in obtaining client-side paradata indicators, it is worth noting that the utilisation of only the six server-side paradata indicators highlighted in this study can already be highly beneficial. Specifically, these indicators accounted for most of the explained variance in 31 out of the 34 regression models examined (see total R^2 and partial R^2 in Tables 3, 4, 5 and 6).

Furthermore, it is also worthwhile to use standard RQIs based on survey response data to complement the paradata indicators, particularly the following *key RQIs*: *share of non-substantive responses*, *item nonresponse*, *breakoffs* and *completeness level of the questionnaire*, as well as various *satisficing indicators* (e.g. straightlining). These key RQIs should also be routinely calculated by online software tools (in a standardised way) similar to the above proposed set of key paradata indicators.

As this paper is a feasibility study that aimed to identify a standard set of paradata indicators, it is useful to provide a comparison with the paradata indicators from the CRONOS panel (see European Social Survey, 2018), one of the rare studies in which paradata are publicly archived alongside survey response data. The comparisons in Table A.14 summarise the differences and illustrate the challenges in developing a standardised set of paradata indicators. Although the indicator sets in Table A.14 do not match perfectly, they nevertheless mostly address similar underlying concepts. Despite differences in indicator sets and in corresponding technical computations, all CRONOS paradata indicators were included in the proposed key set

of paradata indicators, except for screen resolution details and number of sessions (see Online Appendix, Sect. 8, Table A.14), provided that the above-mentioned key RQIs are also integrated with paradata indicators. While we had previously justified the omission of screen resolution as a standalone indicator (Sect. 6.1), the exclusion of the number of sessions was due to a limitation of the empirical study in which respondents were required to complete the survey in a single session. We may add that the GESIS Panel set of paradata indicators (Weyandt et al., 2022), which includes the Universal Client Side Paradata Script (Kaczmirek, 2014), is similar to the CRONOS set but much narrower, focusing only on response times, page visits and navigation, item nonresponse prompts, focus-out events, mouse clicks, survey window size and browser version.

7.3 Limitations

Some limitations of this research are linked to the specifics of the case study. While the selection of RQIs and sociodemographic variables was standardised so they are also relevant for other surveys, a different survey topic might show different effects. Even so, the Internet-related behaviours and attitudes addressed in this study—covering online shopping, Internet use and trust in computers—are substantively a very important area with a profound impact on numerous areas (e.g. Bottoni & Fitzgerald, 2021).

Another specific aspect of this study is the structure of the questionnaire, with its branching pattern, which exposed more active Internet users to a greater number of items. Nevertheless, nearly every survey includes specific branching, and if appropriately handled (as in our case), its impact is incorporated into the paradata analyses. The *number of branching items omitted* must be thus included in any set of key paradata indicators or added as an adjacent covariate.

Another specific aspect of this study has to do with the nature of the data from the access panel, where the respondents were already familiar with pre-existing panel-specific procedures, including incentives. This could lead to higher survey participation (e.g. Bosnjak et al., 2005; Keusch et al., 2014) and fewer breakoffs. The respondents were also accustomed to hard reminders (which did not allow them to continue without providing answers); the soft reminders in this study were an exception for them. Therefore, changing the reminders and incentives might have revealed additional patterns in response quality. Nevertheless, it is very unlikely that the above specifics would compromise the internal validity of the results. In terms of external validity, it should be noted that probability-based and non-probability panels produce similar effects with respect to response quality (Cornesse & Blom, 2023). Furthermore, general population surveys are increasingly being conducted via access panels.

The self-selection of the device (i.e. PC or SP) used to complete the surveys is also a characteristic of this study. The related device effects could have been also the result of uncontrolled factors, such as the higher technical skills of mobile device respondents (e.g. Conrad, Schober, et al., 2017). However, initial oversampling and weighting considerably compensated for these effects. It is also true that the effects on response quality found in quasi-experimental designs are generally comparable to those of experimental studies, so strong circumstantial evidence exists that they would also remain under full experimental conditions. Another important argument supporting the relevance of the results is the fact that in survey practice, respondents use their preferred devices anyway, which the researcher cannot control, so minimising these effects is desirable. The disadvantages of experimentally pre-selecting devices should also be noted, as they force respondents to use the device they might not prefer, creating additional nonresponse and other response quality effects (e.g. Peterson et al., 2017).

One notable aspect of this study was the requirement for respondents to respond in a single session. Of course, this constraint, unfortunately, precluded the use and analysis of additional paradata indicators associated with the number of sessions, which could be related to response styles. It is highly probable that the *number of sessions* would have been added to the 12 key paradata indicators proposed in this study.

Besides the above study specifics, which, importantly, do not interfere with the internal consistency of the results, two seemingly arbitrary parts of the research process should be addressed. The first deals with the selection of 112 initial paradata indicators. This selection was based on an exhaustive literature review; however, an arbitrary cutting point eliminated certain very complex (e.g. mouse movement velocity) or technically highly problematic paradata indicators (e.g. screen resolution as a standalone indicator). Although these restrictions were fully elaborated and justified, it is still true that without them, the set of initial paradata indicators might have been broader. The other seemingly arbitrary part of the research has to do with the reduction from the 112 initial paradata indicators to 29 and then to 14 paradata indicators. Although these two steps were based on clear and reproducible criteria and were evaluated independently by three experts, it is possible that some paradata indicators would have been additionally included or excluded if the process had been more formalised (i.e. data-driven). However, more elaborate reduction procedures would have disproportionately increased the complexity of the research process, which might have gone beyond the aim of this paper. We should recall that this paper is about a feasibility study that aimed to provide initial insight into the potential of creating a standardised set of key paradata indicators

suitable for general usage in practice, so further iterations of the study may verify and modify the proposed solution.

7.4 Future research

All of the limitations discussed above present opportunities for future studies, particularly in the context of a more formal process for reducing paradata indicators. Replicating the analysis in different substantive or methodological contexts would also be extremely valuable. An important extension of this research would be the identification, standardisation and integration of the key RQIs, which are already closely related to the set of key paradata indicators. This would further contribute to efforts to detect standardised segments of respondents according to their response styles and response quality.

In addition, systematic studies could examine the relationships between RQIs and respondent characteristics, between RQIs and survey estimates and between survey estimates and respondent characteristics. Of course, future research could also expand beyond the direct paradata examined in this study to include indirect paradata and passive paradata (e.g. ambient or sensor paradata). Although such extensions would be intriguing, it is worth noting that the technical complexity involved would considerably limit the breadth and applicability of the findings for general use in survey practice.

8 Conclusion

The literature on web survey paradata has primarily focused on their relationship to the domain of response quality, while explorations into their relationships with respondent characteristics and survey estimates have occurred less frequently. Nevertheless, the objective of this paper was to identify a set of key paradata indicators related to all three domains while remaining easy to capture and calculate so that they could be used for general purposes to enhance respondent data with respondent-level paradata indicators.

Following the literature review and conceptual elaboration, 112 initial paradata indicators were identified. In the empirical section, a typical web survey that captured corresponding raw paradata was carried out. The reduction processes resulted in a final set of 12 key paradata indicators that were statistically significantly related to variables from any of the three domains. Certain paradata from this set could be captured also on the server side (i.e. *total number of pages visited*, *total number of repeatedly visited pages*, *duration*, *total number of branching items omitted*, *device type* and *operating system*), while others required a client-based script (i.e. *total number of answer changes*,

total number of item nonresponse prompts, total number of excess clicks, total mouse pointer movement distance, total number of focus-out events and total focus-out duration).

The 12 key paradata indicators identified in this study can serve as a starting point for establishing a standardised set of paradata indicators. Such standardisation would enhance comparability, reproducibility and knowledge discovery in web surveys. Future research should replicate and verify the procedures used in this study, overcome their limitations and apply the approach to other substantive areas.

The results also highlight the challenges facing web survey software providers. First, there is a need to incorporate more server-side paradata, which already comprise the bulk of the paradata indicators and are relatively easy to capture and process. Second, with respect to client-side paradata, there is a need to standardise the corresponding scripts. Finally, web survey software providers might consider offering additional guidance to facilitate the use of paradata indicators by researchers.

Funding This work was supported by the Slovenian Research Agency [grant numbers P5-0399, J5-9334, J5-8233, NI-0004, J5-3100, and V5-2157].

References

- 1KA (2023). *OneClick Survey*. 1KA Web Surveys. <https://www.1ka.si/d/en>
- Alwin, D.F. (2007). *Margins of error: a study of reliability in survey measurement*. Wiley.
- Alzahrani, L., Al-Karaghoul, W., & Weerakkody, V. (2017). Analysing the critical factors influencing trust in e-government adoption from citizens' perspective: A systematic review and a conceptual framework. *International Business Review*, 26(1), 164–175. <https://doi.org/10.1016/j.ibusrev.2016.06.004>.
- Andersen, H., & Mayerl, J. (2017). Social desirability and undesirability effects on survey response latencies. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 135(1), 68–89. <https://doi.org/10.1177/0759106317710858>.
- Ang, C.C., Chow, D.K., Goh, T.W., & Quah, W.G. (2018). *A study on the characteristics of iOS and android phone users in penang [final year project]*. Tunku Abdul Rahman University College. <https://eprints.tarc.edu.my/1687/>
- Berzelak, N., Hrvatin, P., & Vehovar, V. (2022). *Javascript scripts for capturing and python scripts for processing client-based paradata in web surveys*. <https://doi.org/10.5281/zenodo.6806131>.
- Berzelak, N., Hrvatin, P., & Vehovar, V. (2023). Paradata datasets for: identifying a set of key paradata indicators in web surveys [dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.8154489>.
- Biemer, P.P., & Lyberg, L.E. (2003). *Introduction to survey quality*. John Wiley & Sons.
- Bosch, O., & Revilla, M. (2021). *When survey science met online tracking: Presenting an error framework for metered data*. Universitat Pompeu Fabra. <https://doi.org/10.13140/RG.2.2.36032.66569>.
- Bosnjak, M., Tuten, T.L., & Wittmann, W.W. (2005). Unit (non)response in web-based access panel surveys: an extended planned-behavior approach. *Psychology and Marketing*, 22(6), 489–505.
- Bottoni, G., & Fitzgerald, R. (2021). Establishing a baseline: Bringing innovation to the evaluation of cross-national probability-based online panels. *Survey Research Methods*, 15(2), Article 2. <https://doi.org/10.18148/srm/2021.v15i2.7457>.
- Boulianne, S., Klofstad, C.A., & Basson, D. (2011). Sponsor prominence and responses patterns to an online survey. *International Journal of Public Opinion Research*, 23(1), 79–87. <https://doi.org/10.1093/ijpor/edq026>.
- Bowling, N.A., Huang, J.L., Bragg, C.B., Khazon, S., Liu, M., & Blackmore, C.E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, 111(2), 218–229. <https://doi.org/10.1037/pspp0000085>.
- Braekman, E., Demarest, S., Charafeddine, R., Berete, F., Drieskens, S., Van der Heyden, J., & Van Hal, G. (2020). Response patterns in the Belgian health interview survey: web versus face-to-face mode. *European Journal of Public Health*. <https://doi.org/10.1093/eurpub/ckaa166.1295>.
- Callegaro, M. (2013). Paradata in web surveys. In F. Kreuter (Ed.), *Improving surveys with paradata: analytic use of process information* (pp. 261–279). John Wiley & Sons. <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/41148.pdf>.
- Callegaro, M., Yang, Y., Bhola, D.S., Dillman, D.A., & Chin, T.-Y. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 103(1), 5–25. <https://doi.org/10.1177/075910630910300103>.
- Callegaro, M., Lozar, M.K., & Vehovar, V. (2015). *Web survey methodology*. SAGE.
- Centre for Social Informatics & The Samuel Neaman Institute for National Policy Research. (2021). Supplementary materials for: Digital transformation of quantitative data collection in social science re-

- search: Integrating survey data collection in social science research: Integrating survey data collection with big data and paradata for identifying social behaviour. Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana; The Samuel Neaman Institute for National Policy Research, Technion-Israel Institute of Technology. <https://doi.org/10.23668/psycharchives.5106>
- Cepeda, C., Dias, M. C., Rindlisbacher, D., Gamboa, H., & Cheetham, M. (2021). Knowledge extraction from pointer movements and its application to detect uncertainty. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2020.e05873>.
- Cheng, A., Zamorro, G., & Orriens, B. (2020). Personality as a predictor of unit nonresponse in an Internet panel. *Sociological Methods & Research*, 49(3), 672–698. <https://doi.org/10.1177/0049124117747305>.
- Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and non-use of clarification features in web surveys. *Journal of Official Statistics*, 22(2), 245–269. <http://www.websm.org/db/12/919>.
- Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing features of human dialogue to web surveys. *Applied Cognitive Psychology*, 21(2), 165–187. <https://doi.org/10.1002/acp.1335>.
- Conrad, F. G., Schober, M. F., Antoun, C., Yan, H. Y., Hupp, A. L., Johnston, M., Ehlen, P., Vickers, L., & Zhang, C. (2017). Respondent mode choice in a smartphone survey. *Public Opinion Quarterly*, 81(S1), 307–337. <https://doi.org/10.1093/poq/nfw097>.
- Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, 11(1), Article 1. <https://doi.org/10.18148/srm/2017.v11i1.6304>.
- Cornesse, C., & Blom, A. G. (2023). Response quality in nonprobability and probability-based online panels. *Sociological Methods & Research*, 52(2), 561–1102. <https://doi.org/10.1177/0049124120914940>.
- Couper, M. P., & Lyberg, L. (2005). *The use of paradata in survey research*. Proceedings of the 55th Session of the International Statistical Institute.
- Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: perceptions of burden. *Social Science Computer Review*, 19(2), 146–162. <https://doi.org/10.1177/089443930101900202>.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19. <https://doi.org/10.1016/j.jesp.2015.07.006>.
- De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1), 1–18. [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7).
- Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging*, 23(3), 558–566. <https://doi.org/10.1037/a0012897>.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192–203. <https://doi.org/10.1037/1040-3590.18.2.192>.
- European Social Survey (2018). CRONOS codebook: paradata. http://www.1ka.si/uploadi/editor/doc/1711463472CRONOS_Paradata_e01_Codebook.pdf
- Fernández-Fontelo, A., Henninger, F., Kieslich, P. J., Kreuter, F., & Greven, S. (2022). *Classification ensembles for multivariate functional data with application to mouse movements in web surveys*. arXiv:2205.13380. <https://doi.org/10.48550/arXiv.2205.13380>.
- Fernández-Fontelo, A., Kieslich, P. J., Henninger, F., Kreuter, F., & Greven, S. (2023). Predicting question difficulty in web surveys: a machine learning approach based on mouse movement features. *Social Science Computer Review*, 41(1), 141–162. <https://doi.org/10.1177/08944393211032950>.
- Fisher, B., & Berner, F. (2014). *Device effects: how different screen sizes affect answer quality in online questionnaires*. General Online Research Conference (GOR). <http://www.websm.org/db/12/17232/>
- Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the smart: type of rating scale on the web interacts with educational level. *Social Science Computer Review*, 29(2), 221–231. <https://doi.org/10.1177/0894439310376896>.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>.
- Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, 2(1), 21–32. <https://doi.org/10.18148/srm/2008.v2i1.598>.
- Götz, F. M., Stieger, S., & Reips, U.-D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLOS ONE*, 12(5), e176921. <https://doi.org/10.1371/journal.pone.0176921>.
- Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing “too fast” responses and respondents from web surveys. *Public Opinion Quarterly*, 79(2), 471–503. <https://doi.org/10.1093/poq/nfu058>.
- Groves, R. M. (2005). *Survey errors and survey costs* (2nd edn.). John Wiley & Sons.

- Gummer, T., & Roßmann, J. (2015). Explaining interview duration in web surveys: a multilevel approach. *Social Science Computer Review*, 33(2), 217–234. <https://doi.org/10.1177/0894439314533479>.
- Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: properties and implementation. *Sociological Methods & Research*, 50(1), 238–264. <https://doi.org/10.1177/0049124118769083>.
- Gutierrez, C., Wells, T., Rao, K., & Kurzynski, D. (2011). *Catch them when you Can: speeders and their role in online data quality*. Midwest Association for Public Opinion Research (MAPOR). <http://www.websm.org/db/12/16145/>
- Haraldsen, G., Kleven, Ø., & Sundvoll, A. (2005). Big scale observations gathered with the help of client side paradata. Quest Workshop. <http://www.websm.org/db/12/15969/>
- Hart, A., Reis, D., Prestele, E., & Jacobson, N.C. (2022). Using Smartphone sensor paradata and personalized machine learning models to infer participants' well-being: ecological momentary assessment. *Journal of Medical Internet Research*, 24(4), e34015. <https://doi.org/10.2196/34015>.
- 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(1), 111–128. <https://doi.org/10.1177/0894439306293888>.
- Heerwegh, D. (2002). *Describing response behavior in websurveys using client side paradata*. Web Survey Workshop and Symposium. <http://www.websm.org/db/12/345/>
- Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, 21(3), 360–373. <https://doi.org/10.1177/0894439303253985>.
- Heerwegh, D., & Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review*, 20(4), 471–484. <https://doi.org/10.1177/089443902237323>.
- Hibbeln, M. T., Jenkins, J.L., Schneider, C., Valacich, J.S., & Weinmann, M. (2017). How is your user feeling? Inferring emotion through human-computer interaction devices. *MIS Quarterly*, 41(1), 1–21. <https://doi.org/10.25300/MISQ/2017/41.1.01>.
- Hibbing, M.V., Cawvey, M., Deol, R., Bloeser, A.J., & Mondak, J.J. (2019). The relationship between personality and response patterns on public opinion surveys: the big five, extreme response style, and acquiescence response style. *International Journal of Public Opinion Research*, 31(1), 161–177. <https://doi.org/10.1093/ijpor/edx005>.
- Höhne, J.K., & Schlosser, S. (2018). Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata surveyfocus. *Social Science Computer Review*, 36(3), 369–378. <https://doi.org/10.1177/0894439317710450>.
- Höhne, J.K., & Schlosser, S. (2019). SurveyMotion: what can we learn from sensor data about respondents' completion and response behavior in mobile web surveys? *International Journal of Social Research Methodology*, 22(4), 379–391. <https://doi.org/10.1080/13645579.2018.1550279>.
- Höhne, J.K., Revilla, M., & Schlosser, S. (2020). Motion instructions in surveys: compliance, acceleration, and response quality. *International Journal of Market Research*, 62(1), 43–57. <https://doi.org/10.1177/1470785319858587>.
- Höhne, J.K., Schlosser, S., Couper, M.P., & Blom, A.G. (2020). Switching away: exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, 111, 106417. <https://doi.org/10.1016/j.chb.2020.106417>.
- Hong, M., Steedle, J.T., & Cheng, Y. (2020). Methods of detecting insufficient effort responding: comparisons and practical recommendations. *Educational and Psychological Measurement*, 80(2), 312–345. <https://doi.org/10.1177/0013164419865316>.
- Horwitz, R., Tancreto, J.G., Zelenak, M.F., & Davis, M. (2013). *Use of Paradata to assess the quality and functionality of the American community survey Internet instrument* (p. 79). United States Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_01.pdf
- Horwitz, R., Kreuter, F., & Conrad, F.G. (2017). Using mouse movements to predict web survey response difficulty. *Social Science Computer Review*, 35(3), 388–405. <https://doi.org/10.1177/0894439315626360>.
- Horwitz, R., Brockhaus, S., Henninger, F., Kieslich, P.J., Schierholz, M., Keusch, F., & Kreuter, F. (2020). Learning from mouse movements: improving questionnaires and respondents' user experience through passive data collection. In *Advances in questionnaire design, development, evaluation and testing* (pp. 403–425). John Wiley & Sons. <https://doi.org/10.1002/9781119263685.ch16>.
- Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., & DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114. <https://doi.org/10.1007/s10869-011-9231-8>.

- Huang, J.L., Bowling, N.A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: evaluating validity and participant reactions. *Journal of Business and Psychology*, 30(2), 299–311. <https://doi.org/10.1007/s10869-014-9357-6>.
- Jenkins, J.L., Larsen, R., Bodily, R., Sandberg, D., Williams, P., Stokes, S., Harris, S., & Valacich, J.S. (2015). *A multi-experimental examination of analyzing mouse cursor trajectories to gauge subject uncertainty*. 2015 Americas Conference on Information Systems, AMCIS 2015. <http://www.scopus.com/inward/record.url?scp=84963625990&partnerID=8YFLogxK>
- Kaczmirek, L. (2009). *Human survey-interaction: Usability and nonresponse in online surveys*. University of Mannheim. <https://ub-madoc.bib.uni-mannheim.de/2150/1/kaczmirek2008.pdf>
- Kaczmirek, L. (2014). UCSP: universal client side paradata. <http://kaczmirek.de/ucsp/ucsp.html>
- Kaczmirek, L., & Neubarth, W. (2007). Nicht-reaktive datenerhebung: Teinahmeverhalten bei befragungen mit paradata evaluieren. [Non reactive data collection. Evaluating response behavior with paradata in surveys]. In M. Welker & O. Wenzel (Eds.), *Online-forschung 2007. Grundlagen und fallstudien* (pp. 293–311). Herbert von Halem Verlag.
- Keusch, F., Batinic, B., & Mayerhofer, W. (2014). Motives for joining nonprobability online panels and their association with survey participation behavior. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick & P. J. Lavrakas (Eds.), *Online panel research: a data quality perspective* (pp. 171–191). John Wiley & Sons.
- Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D.P. (2019). Straightlining: overview of measurement, comparison of indicators, and effects in mail-web mixed-mode surveys. *Social Science Computer Review*, 37(2), 214–233. <https://doi.org/10.1177/0894439317752406>.
- Kreuter, F. (2013). Improving surveys with paradata: Introduction. In F. Kreuter (Ed.), *Improving surveys with paradata: analytic use of process information* (pp. 1–11). Wiley.
- Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.
- Krosnick, J.A. (2018). Improving question design to maximize reliability and validity. In D.L. Vannette & J. A. Krosnick (Eds.), *The Palgrave handbook of survey research* (pp. 95–102). Palgrave.
- Kühne, S., & Kroh, M. (2018). Personalized feedback in web surveys: does it affect respondents' motivation and data quality? *Social Science Computer Review*, 36(6), 744–755. <https://doi.org/10.1177/0894439316673604>.
- Kumar, M., Valacich, J., Jenkins, J., & Kim, D. (2022). *Too fast? Too slow? A novel approach for identifying extreme response behavior in online surveys*. SIGHCI 2022 Proceedings. <https://aisel.aisnet.org/sighci2022/4>
- Kunz, T., & Hadler, P. (2020). *Web paradata in survey research*. GESIS—Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_037.
- de Leeuw, E., & Toepoel, V. (2017). Mixed-mode and mixed-device surveys. In D.L. Vannette & J. A. Krosnick (Eds.), *The Palgrave handbook of survey research* (pp. 51–61). Springer. https://doi.org/10.1007/978-3-319-54395-6_8.
- Leiner, D.J. (2019). Too fast, too straight, too weird: non-reactive indicators for meaningless data in Internet surveys. *Survey Research Methods*, 13(3), Article 3. <https://doi.org/10.18148/srm/2019.v13i3.7403>.
- Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: a psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7), 1003–1020. <https://doi.org/10.1002/acp.1602>.
- Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5), 914–934. <https://doi.org/10.1093/poq/nfn050>.
- Maniaci, M.R., & Rogge, R.D. (2014). Caring about carelessness: participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83. <https://doi.org/10.1016/j.jrp.2013.09.008>.
- Marsh, H.W., Nagengast, B., & Morin, A.J.S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, 49(6), 1194–1218. <https://doi.org/10.1037/a0026913>.
- Matjašič, M., Vehovar, V., & Lozar Manfreda, K. (2018). Web survey paradata on response time outliers: a systematic literature review. *Metodološki Zvezki*, 15(1), 23–41. <http://ibmi.mf.uni-lj.si/mz/2018/no-1/Matjasic2018.pdf>.
- Matjašič, M., Vehovar, V., & Sendelbah, A. (2021). *Combining response times and response quality indicators to identify speeders with low response quality in web surveys [dataset]*. <https://doi.org/10.23668/psycharchives.4718>.
- McClain, C.A., Couper, M.P., Hupp, A.L., Keusch, F., Peterson, G., Piskorowski, A.D., & West, B.T. (2019). A typology of web survey paradata for assessing total survey error. *Social Science Computer*

- Review*, 37(2), 196–213. <https://doi.org/10.1177/0894439318759670>.
- Meade, A., & Craig, B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>.
- Mittereder, F.K. (2019). *Predicting and preventing breakoff in web surveys*. Vol. 2027. Ann Arbor. Doctoral dissertation. <https://deepblue.lib.umich.edu/handle/2027.42/149963>
- Morren, M., & Paas, L.J. (2020). Short and long instructional manipulation checks: what do they measure? *International Journal of Public Opinion Research*, 32(4), 790–800. <https://doi.org/10.1093/ijpor/edz046>.
- Paas, L.J., & Morren, M. (2018). Please do not answer if you are reading this: respondent attention in online panels. *Marketing Letters*, 29(1), 13–21. <https://doi.org/10.1007/s11002-018-9448-7>.
- Peck, R., & Devore, J.L. (2012). *Statistics: the exploration & analysis of data*. Cengage Learning.
- Peng, H., & Ostergren, J. (2016). Capturing survey client-side paradata. http://www.blaiseusers.org/2016/papers/4_3.pdf
- Peterson, G., Griffin, J., LaFrance, J., & Li, J. (2017). Smartphone participation in web surveys. In P.P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker & B.T. West (Eds.), *Total survey error in practice* (pp. 203–233). Wiley.
- Revilla, M., & 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(1), 97–114. <https://doi.org/10.1177/0894439314531214>.
- Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Research synthesis: Satisficing in surveys: a systematic review of the literature. *Public Opinion Quarterly*, 83(3), 598–626. <https://doi.org/10.1093/poq/nfz035>.
- Roehrick, K., Vaid, S.S., & Harari, G.M. (2023). Situating Smartphones in daily life: big five traits and contexts associated with young adults' Smartphone use. *PsyArXiv*. <https://doi.org/10.31234/osf.io/v2jgk>.
- Roßmann, J., & Gummer, T. (2016). Using paradata to predict and correct for panel attrition. *Social Science Computer Review*, 34(3), 312–332. <https://doi.org/10.1177/0894439315587258>.
- Schneider, I.K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., & Koole, S.L. (2015). The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology*, 6, 996. <https://doi.org/10.3389/fpsyg.2015.00996>.
- Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting careless responding in survey data using stochastic gradient boosting. *Educational and Psychological Measurement*, 82(1), 29–56. <https://doi.org/10.1177/00131644211004708>.
- Seelye, A., Hagler, S., Mattek, N., Howieson, D.B., Wild, K., Dodge, H.H., & Kaye, J.A. (2015). Computer mouse movement patterns: a potential marker of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(4), 472–480. <https://doi.org/10.1016/j.dadm.2015.09.006>.
- Sendelbah, A., Vehovar, V., Slavec, A., & Petrovčič, A. (2016). Investigating respondent multitasking in web surveys using paradata. *Computers in Human Behavior*, 55, 777–787. <https://doi.org/10.1016/j.chb.2015.10.028>.
- Sharma, S. (2019). Paradata, interviewing quality, and Interviewer effects [Thesis, University of Michigan]. <http://deepblue.lib.umich.edu/handle/2027.42/150047>
- Smyth, J.D., Dillman, D.A., Christian, L.M., & Stern, M.J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70(1), 66–77. <https://doi.org/10.1093/poq/nfj007>.
- Stern, M.J. (2008). The use of client-side paradata in analyzing the effects of visual layout on changing responses in web surveys. *Field Methods*, 20(4), 377–398. <https://doi.org/10.1177/1525822X08320421>.
- Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: a paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6), 1488–1495. <https://doi.org/10.1016/j.chb.2010.05.013>.
- Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J.K. (2020). Augmenting surveys with data from sensors and Apps: opportunities and challenges. *Social Science Computer Review*. <https://doi.org/10.1177/0894439320979951>.
- Sturgis, P., & Brunton-Smith, I. (2023). Personality and survey satisficing. *Public Opinion Quarterly*, 87(3), 689–718. <https://doi.org/10.1093/poq/nfad036>.
- Sturgis, P., Schober, M.F., & Brunton-Smith, I. (2019). *Am I being neurotic? Personality as a predictor of survey response styles*. 8th European Survey Research Association Conference. <https://www.europeansurveyresearch.org/conf2019/prog.php?sess=8#490>
- Tourangeau, R., Rips, L.J., & Rasinski, K.A. (2000). *The psychology of survey response*. Cambridge University Press.

- Tourangeau, R., Couper, M.P., & Conrad, F.G. (2004). Spacing, position, and order: interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68(3), 368–393. <https://doi.org/10.1093/poq/nfh035>.
- Turner, G., Sturgis, P., & Martin, D. (2015). Can response latencies be used to detect survey satisficing on cognitively demanding questions? *Journal of Survey Statistics and Methodology*, 3(1), 89–108. <https://doi.org/10.1093/jssam/smu022>.
- Tzafilkou, K., & Nicolaos, P. (2018). Mouse behavioral patterns and keystroke dynamics in end-user development: what can they tell us about users' behavioral attributes? *Computers in Human Behavior*, 83, 288–305. <https://doi.org/10.1016/j.chb.2018.02.012>.
- Valicon (2022). JazVem. <https://www.jazvem.si>
- Vehovar, V., & Čehovin, G. (2023). *Direct Paradata usage for analysis of response quality, respondent characteristics, and survey estimates: state-of-the-art review and typology of paradata*. Centre for Social Informatics Working Paper Series. University of Ljubljana. <https://www.fdv.uni-lj.si/docs/default-source/cdi-doc/direct-paradata-usage-for-analysis-of-response-quality.pdf>
- Vehovar, V., Bevec, D., & Matjaž, U. (2021). *Web survey software: layouts for grid questions on pc and mobile web surveys*. University of Ljubljana, Faculty of Social Sciences, Centre for Social Informatics. <http://papersseries.cdi.si/>
- Vehovar, V., Couper, M.P., & Čehovin, G. (2022). Alternative layouts for grid questions in PC and mobile web surveys: an experimental evaluation using response quality indicators and survey estimates. *Social Science Computer Review*. <https://doi.org/10.1177/08944393221132644>.
- Vehovar, V., Berzelak, N., & Čehovin, G. (2023a). *Code for: identifying a set of key paradata indicators in web surveys [dataset]*. <https://doi.org/10.23668/psycharchives.12982>.
- Vehovar, V., Berzelak, N., & Čehovin, G. (2023b). *Dataset for: identifying a set of key paradata indicators in web surveys [dataset]*. <https://doi.org/10.23668/psycharchives.12981>.
- Villar, A., Sommer, E., Finnøy, D., Gaia, A., Berzelak, N., & Bottoni, G. (2018). Cross-national online survey (CRONOS) panel: data and documentation user guide. https://www.europeansocialsurvey.org/docs/cronos/CRONOS_user_guide_e01_1.pdf
- Wells, T., Vidalon, M., & DiSogra, C. (2010). Differences in length of survey administration between Spanish-language and English-language survey respondents. 6. <http://www.asasrms.org/Proceedings/y2010/Files/400137.pdf>
- Weyandt, K., Struminskaya, B., & Schaurer, I. (2022). *GESIS panel online paradata related to study zs in ZA5664 and ZA5665*. GESIS – Leibniz Institute for the Social Sciences. <https://dbk.gesis.org/dbksearch/download.asp?id=65301>
- Wise, S.L., & Kong, X. (2005). Response time effort: a new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163–183. https://doi.org/10.1207/s15324818ame1802_2.
- Wortman, J., Lucas, R.E., & Donnellan, B.M. (2012). Stability and change in the big five personality domains: evidence from a longitudinal study of australians. *Psychology and Aging*, 27(4), 867–874. <https://doi.org/10.1037/a0029322>.
- Yamauchi, T., & Xiao, K. (2018). Reading emotion from mouse cursor motions: affective computing approach. *Cognitive Science*, 42(3), 771–819. <https://doi.org/10.1111/cogs.12557>.
- 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. <https://doi.org/10.1002/acp.1331>.
- Zhang, C., & Conrad, F.G. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127–135. <https://doi.org/10.18148/srm/2014.v8i2.5453>.