Article

Social Science Computer Review
2021, Vol. 39(6) 1253–1271

DOI: 10.1177/0894439320971233
journals.sagepub.com/home/ssc


Completion Conditions and Response Behavior in Smartphone Surveys: A Prediction Approach Using Acceleration Data

Christoph Kern[1], Jan Karem Höhne[2,4], Stephan Schlosser[3], and Melanie Revilla[4]

Abstract

This study utilizes acceleration data from smartphone sensors to predict motion conditions of smartphone respondents. Specifically, we predict whether respondents are moving or nonmoving on a survey page level to learn about distractions and the situational conditions under which respondents complete smartphone surveys. The predicted motion conditions allow us to (1) estimate the proportion of smartphone respondents who are moving during survey completion and (2) compare the response behavior of moving and nonmoving respondents. Our analytical strategy consists of two steps. First, we use data from a lab experiment that systematically varied motion conditions of smartphone respondents and train a prediction model that is able to accurately infer respondents' motion conditions based on acceleration data. Second, we use the trained model to predict motion conditions of respondents in two cross-sectional surveys in order to compare response behavior of respondents with different motion conditions in a field setting. Our results indicate that active movement during survey completion is a relatively rare phenomenon, as only about 3%–4% of respondents were predicted as moving in both cross-sectional surveys. When comparing respondents based on their predicted motion conditions, we observe longer completion times of moving respondents. However, we observe little differences when comparing moving and nonmoving respondents with respect to indicators of superficial responding, indicating that moving during survey completion does not pose a severe threat to data quality.

Keywords

acceleration data, data quality, machine learning, multitasking, smartphone surveys, survey motion

[1] School of Social Sciences, University of Mannheim, Germany
[2] Collaborative Research Center "Political Economy of Reforms," University of Mannheim, Germany
[3] Center of Methods in Social Sciences, University of Göttingen, Germany
[4] Research and Expertise Center for Survey Methodology, Universitat Pompeu Fabra, Barcelona, Spain

Corresponding Author:
Christoph Kern, School of Social Science, University of Mannheim, A 5, 6, 68131 Mannheim, Germany.
Email: c.kern@uni-mannheim.de

Introduction and Background

The use of smartphones in web surveys has continuously increased during the last decade. Gummer et al. (2019), for example, show across 18 web surveys that were conducted in Germany that the proportion of smartphone respondents has successively increased from about 5% to 20% between 2012 and 2016. Revilla et al. (2016) observe a similar trend for Spain, Portugal, and several Latin American countries (see also Bosch et al., 2019b). The reasons for this development in web surveys seem to be twofold. On the one hand, the number of people who own a smartphone has increased. On the other hand, high-speed mobile Internet is more widely accessible. This applies to most countries around the globe, but especially to developing countries (Pew Research Center, 2018a, 2018b).

Smartphones offer many opportunities for innovative data collection that can be used for studying social science research questions as well as for methodological research. Examples include collecting expenditure data with receipt scanning apps (Jäckle et al., 2019), offering voice input options for answering open-ended questions (Revilla et al., 2020; Revilla & Couper, 2019), and asking respondents to take photographs for supplementing survey responses (Bosch et al., 2019a). Particularly, smartphone sensors, such as accelerometer, gyroscope, and GPS, allow the collection of detailed information that can be used to enrich and augment web survey data from smartphones. Data from built-in sensors allow, for instance, to unobtrusively record respondents' location in time use surveys (Elevelt et al., 2019) and to measure objective physical activity for studying health (Weiss et al., 2018), happiness (Lathia et al., 2017), or (side-)effects of long-term unemployment (Kreuter et al., 2018).

At the same time, however, the increasing usage of smartphones as survey devices might also introduce challenges for survey data quality. In principle, smartphones enable respondents to participate in web surveys whenever and wherever they want with almost no locality, situation, or time restrictions (Mavletova, 2013). Consequently, smartphone respondents are frequently surrounded by third parties during survey participation (Toninelli & Revilla, 2016b), which might have a negative impact on response behavior because respondents can get distracted due to their environment. Lynn and Kaminska (2012) distinguish three types of distractions during survey participation: (1) distractions that demand aural attention, such as music playing in the background (see also Wenz, 2019), (2) distractions that demand visual attention, such as looking after children, and (3) multitasking, such as having a conversation. Indeed, Toninelli and Revilla (2016a) show that smartphone respondents report more distractions and/or multitasking behavior, such as watching TV and talking with other people, than PC respondents. However, one problem associated with these findings on distractions and multitasking is that they are based on self-reports, implying that respondents must admit that they might not pay constant attention. Thus, self-reports on distractions and multitasking are subject to social desirability bias and recall error, which might cause imprecise measures (Wenz, 2019).

This study aims to bridge the challenges and opportunities of smartphones as survey devices by utilizing passively collected acceleration data from smartphone sensors as an avenue for studying distractions and multitasking of smartphone respondents. More specifically, this study has two objectives: First, we demonstrate how acceleration data can be used to infer motion conditions of respondents (i.e., identifying moving vs. nonmoving respondents during survey completion). Second, we link the predicted motion conditions to response behavior and indicators of superficial responding. This twofold approach attempts to exemplify how methodological advances in data collection with smartphones and analysis of sensor data can be utilized in survey research to study motion conditions and response behavior. While we estimate the prevalence of different motion conditions and their effects on (superficial) responding in two concrete smartphone surveys, the proposed prediction approach allows future research to identify moving and nonmoving respondents and to analyze their response behavior in (future) smartphone surveys that collect acceleration data.

Previous studies on distractions and multitasking proposed to overcome the drawbacks of self-reports by passively collecting paradata[1] during web survey completion. For instance, Sendelbah et al. (2016) suggest to use so-called JavaScript "OnBlur" functions that measure the in-/activity of web survey pages—whether the browser tab or window that hosts the web survey is the active or processed one (see Callegaro, 2013; Couper & Peterson, 2017; Diedenhofen & Musch, 2017; Höhne et al., 2017, 2018; Höhne & Schlosser, 2018, 2019; Höhne, Schlosser, et al., 2020; Revilla & Couper, 2018a, 2018b; Sendelbah et al., 2016)—to infer on-device multitasking, such as switching browser tab or window to check incoming emails or social media notifications. Höhne, Schlosser, et al. (2020) show that on-device multitasking gathered by means of OnBlur functions occurs among 10% of all smartphone respondents in a web survey.

Beyond the use of paradata to detect multitasking, Höhne and Schlosser (2019) and Toepoel and Lugtig (2015) suggest to passively collect sensor data in smartphone surveys to draw conclusions about respondents' completion conditions. Accelerometers measure the rate of change of velocity of an object over time, allowing researchers to unobtrusively record physiological states, such as movements (see Elhoushi et al., 2017; Harari et al., 2016; He et al., 2016; Höhne, Revilla, et al., 2020; Höhne & Schlosser, 2019; Toepoel & Lugtig, 2015). If a person moves or walks, they are creating acceleration (He et al., 2016), which is detected by the smartphone that is commonly worn on the body (e.g., in the pocket). This situation can be applied to respondents that have the smartphone in their hands and complete a web survey, referring to a "respondent-device" link (Höhne, Revilla, et al., 2020; Höhne & Schlosser, 2019). To put it differently, respondents' motions are detectable by the acceleration sensor of smartphones, allowing researchers to classify smartphone respondents on the basis of their motion levels (i.e., acceleration).

In contrast to paradata, which can be used to study on-device multitasking, the usage of acceleration data provides an avenue to learn about the (situational) conditions in which respondents complete web surveys. For instance, Höhne, Revilla, et al. (2020) show that completing surveys outdoors, moving around during survey completion, and the presence of third parties during survey completion significantly increase the acceleration of smartphones. Thus, acceleration data may be used to study distractions that go beyond respondent-initiated multitasking activities, without relying on self-reports. More specifically, respondents who move around (e.g., outdoors or at home) are likely to encounter aural and/or visual distractions, which might divert their attention away from the web survey. In relation to this, Höhne and Schlosser (2019) report significantly longer completion times and stronger primacy effects for respondents who were experimentally assigned to a "moving" condition while completing a smartphone survey, compared to respondents who were assigned to a "nonmoving" condition.[2] While the detection of such completion conditions is limited when using self-reports or paradata in the form of OnBlur functions, acceleration data from smartphone sensors represent a new gateway for mobile web survey research to infer distractions and multitasking.

In what follows, we outline the conceptual twofold approach of this study for predicting motion conditions and studying their relationship with superficial responding and present the associated research questions. Then, we describe the data sources and study designs including the underlying samples, the survey questions used, and the analytical strategies. We then present the results of the study, and finally, we discuss practical implications associated with the feasibility of predicting motion conditions in smartphone surveys and address future research perspectives.

Conceptual Approach and Research Questions

Although smartphone sensors allow researchers to unobtrusively collect information about respondents' physiological states, there is little research that utilizes acceleration data to study completion conditions in web surveys or classifies respondents based on predicted motion conditions. However, as previously outlined, preliminary evidence suggests that respondents with different motion

conditions may differ in terms of response behavior. We expect that moving is associated with more (situational) distractions and multitasking than nonmoving and thus fosters superficial responding. More specifically, we assume that performing multiple tasks at the same time (e.g., answering survey questions and walking around) implies a relocation of cognitive resources (Zwarun & Hall, 2014), which may increase completion times because respondents must devote time for the additional task (e.g., environmental orientation) and must (partially) restart the cognitive response process when refocusing on the survey. The same mechanism may increase low (or high) intraindividual response variability (IRV) and primacy effects for moving respondents who are engaging in multiple tasks (e.g., safely reaching their destination and completing the survey). In line with this reasoning, we address the following research questions: (1) How accurately can we predict respondents' motion conditions in smartphone surveys using acceleration data? (2) How common are different types of motion conditions in smartphone surveys? and (3) How do respondents with different predicted motion conditions differ with respect to response behavior?

To study motion conditions of smartphone respondents and their effects on response behavior by using acceleration data, we employ a two-step approach. In a first step, we train machine learning models using high-frequency acceleration data to learn the link between different motion conditions and acceleration profiles. The acceleration data were collected in a lab experiment that systematically varied the motion conditions (i.e., sitting, standing, walking, and stair climbing) of respondents while they completed a web survey on their smartphone. More specifically, we train and evaluate regularized regression and tree-based models using grouped cross-validation, reflecting the hierarchical structure of the acceleration data with survey pages nested in respondents. In building our prediction models, we focus on a binary outcome variable with the labels moving (i.e., walking and stair climbing) and nonmoving (i.e., sitting and standing). Note that the data from the lab experiment allow the prediction models to learn patterns from the acceleration data that distinguish moving from nonmoving respondents given the "ground truth" (i.e., the true outcome is observed). Without this knowledge, ad hoc decisions on, for instance, a fixed (and arbitrary) acceleration threshold that differentiates both conditions would need to be specified in order to infer motion conditions in new data sets. Furthermore, such an approach would most likely focus on average acceleration (over time) and would therefore not utilize the full level of granularity that acceleration data offer.

In a second step, the trained models are used to predict respondents' motion conditions in two (new) self-administered smartphone surveys that also collected high-frequency acceleration data of respondents' smartphones during web survey completion. This allows us to study the prevalence of different predicted motion conditions and the effects of motion conditions on response behavior in a field setting. In this context, we compare completion times, IRV, primacy effects, and compliance with an instructional manipulation check (IMC) between respondents with different predicted motion conditions.

Method

Data Sources and Study Designs

In this study, we use three different data sources: data from a lab experiment (*Data Source 1*) and data from two cross-sectional surveys (*Data Sources 2 and 3*). All data sets contain high-frequency acceleration data that were collected using the open-source JavaScript-based tool "SurveyMotion (SMotion)" developed by Höhne and Schlosser (2019).[3] SMotion collects the total acceleration (TA) of mobile devices, such as smartphones, on a survey page level. TA is defined as follows (see, for instance, He et al., 2016):

$$TA = \sqrt{a_x^2 + a_y^2 + a_z^2}.$$

Equation 1: Calculating TA.
Note. Accelerations (a) along the x-, y-, and z-axis are defined as a_x, a_y, and a_z, respectively.

The sampling rate of the TA primarily depends on the device and/or on frequency restrictions set in the JavaScript code. In this study, the TA of smartphones was measured without any frequency restrictions set in the JavaScript code to register it as precisely as possible. On average, the TA was measured every 17.9 ms.

In addition, we collected several kinds of paradata, such as completion times and User-Agent-Strings, that were also collected on a survey page level. For this purpose, we used the open-source JavaScript-based tool "Embedded Client Side Paradata" developed by Schlosser and Höhne (2018). We obtained informed consent for the collection of acceleration data and paradata in the lab experiment and in the two cross-sectional surveys.

While the first data set (Data Source 1) serves for training the machine learning models, the last two data sets (Data Sources 2 and 3) serve for predicting respondents' motion conditions in a field setting and to compare the response behavior of respondents with different predicted motion conditions.

Data Source 1: Lab Experiment. A total of $N = 89$ students from the University of Göttingen in Germany took part in the lab experiment in 2017. These participants were between 18 and 42 years with a mean age of 24.5 ($SD = 4.4$). Overall, 55% were female, and at least 85% had participated previously in a web survey. Furthermore, 99% used their smartphone and 96% used the Internet on a daily basis.

Participants were randomly assigned to one of four experimental groups. The first group ($n = 22$) was seated in front of a desk with the smartphone lying on the desk during survey completion (sitting condition). The second group ($n = 22$) stood at a fixed point with the smartphone in their hands during survey completion (standing condition). The third group ($n = 23$) walked along an aisle with the smartphone in their hands during survey completion (walking condition). The fourth group ($n = 22$) climbed stairs with the smartphone in their hands during survey completion (stair-climbing condition). The entire study was supervised by an experimenter to ensure proper compliance with the assigned motion conditions.

To evaluate the effectiveness of random assignment and the sample composition between the four experimental groups, we conducted χ^2 tests. The results showed no statistically significant differences regarding age, gender, survey participation, smartphone usage, and Internet usage.[4]

Data Source 2: Cross-Sectional Survey 1. The data were collected at the University of Göttingen in Germany in December 2017. Students were invited by email stating the topic of the study and including a link that directed them to the survey. A total of 1,711 students started the survey using a smartphone,[5] which took about 9 min. Among these, 466 were dropped because they only visited the title page ($n = 48$), broke off before being asked any study-related questions ($n = 54$), had no accurately measured acceleration data because of deactivated JavaScript ($n = 8$), or their smartphones were not equipped with accelerometers that allow for the measurement of pure TA without gravity ($n = 356$). This leaves us with 1,245 respondents available for statistical analyses.

These respondents were aged between 17 and 74 years, with a mean age of 23.3 ($SD = 3.9$), and 67% of them were female. All respondents had completed a college preparatory secondary school or university-level education. The distribution of sociodemographic characteristics is largely similar for the cases that were dropped from the analyses (see Online Appendix A in the Online Supplement of this article).

Data Source 3: Cross-Sectional Survey 2. The data were collected by the nonprobability access panel Respondi in Germany in July and August 2019. A total of 1,726 respondents started the smartphone survey, which took about 20 min.[6] Among these, 579 were dropped because they only visited the title page ($n = 57$), broke off before being asked any study-related questions ($n = 24$), had no accurately measured acceleration data because of deactivated JavaScript ($n = 5$), or their smartphones were not equipped with accelerometers that allow for the measurement of pure TA without gravity ($n = 493$). This leaves us with 1,147 respondents.

These respondents were aged between 18 and 70 years, with a mean age of 43.9 ($SD = 14.8$), and 63% of them were female. In terms of education, 10% had completed lower secondary school, 38% intermediate secondary school, and 52% college preparatory secondary school or university-level education. We again observe little differences between the sample with complete information and the dropped cases (see Online Appendix A in the Online Supplement of this article).

Survey Questions

Data Source 1: Lab experiment. The lab data are based on three single questions that were presented individually on the survey page and eight questions with an item-by-item approach that were presented together on one survey page (see Online Appendix B in the Online Supplement of this article). The three single questions and the eight item-by-item questions dealt with achievement and job motivation. These questions were adopted from the Cross Cultural Survey for Work and Gender Attitudes (Hanson Frieze, 2010) and the German General Social Survey (Terwey, 2000). The presentation order of the questions was randomized to limit order effects. The questions were in German, which was the mother tongue of 93% of the participants.

Data Sources 2 and 3: Cross-Sectional Surveys 1 and 2. The data are based on five single questions that were presented individually on the survey page and 16 questions with an item-by-item approach that were presented on two survey pages (see Online Appendix B in the Online Supplement of this article). The single and item-by-item questions dealt with achievement and job motivation and were also adopted from the Cross Cultural Survey for Work and Gender Attitudes (Hanson Frieze, 2010) and the German General Social Survey (Terwey, 2000). The response scales of these questions slightly differed between the two cross-sectional surveys (i.e., endpoint labeled and fully labeled). In addition, both surveys included an IMC that was adopted from Berinsky et al. (2014; see Online Appendix B in the Online Supplement of this article). All questions and the IMC were in German, which was the mother tongue of 93% and 96% of the participants, respectively.

Analytical Strategies

Step 1: Training prediction models based on acceleration data. To infer completion conditions of respondents in smartphone surveys, we use the acceleration data that were collected in the lab experiment (Data Source 1) to train and evaluate several prediction models. In this step, we extract features (i.e., predictor variables) from the acceleration data of the lab experiment by aggregating over the repeated acceleration measurements that were collected for each survey page. More specifically, we compute the following summary statistics for the TA values of each respondent and survey page (i.e., respondent-page): mean, median, variance, median absolute deviation, interquartile range, minimum, maximum, range, and percentiles (i.e., 5%, 10%, 25%, 75%, 90%, and 95%). This data preparation strategy allows us to sidestep the fact that the number of acceleration values per respondent and survey page depends on the sampling rate of respondents' smartphone and the time to complete a given survey page (i.e., the number of observed acceleration values varies widely

Table 1. Distribution of Acceleration Features: Lab Experiment.

	Training Data (Moving)				Training Data (Nonmoving)			
Feature	Minimum	Mean	Maximum	*n*	Minimum	Mean	Maximum	*n*
TA mean	0.307	1.776	3.907	358	0.015	0.269	1.793	348
TA median	0.250	1.659	3.960	358	0.010	0.212	1.240	348
TA variance	0.049	0.793	3.165	358	0	0.102	3.638	348
TA MAD	0.178	0.762	1.423	358	0	0.090	0.801	348
TA IQR	0.250	1.057	2.040	358	0.010	0.152	3.648	348
TA minimum	0	0.188	0.838	358	0	0.015	0.180	348
TA maximum	1.550	5.665	18.830	358	0.060	1.890	16.800	348
TA range	1.550	5.477	18.740	358	0.060	1.874	16.780	348
TA q5	0.070	0.614	2.550	358	0	0.107	0.950	348
TA q10	0.090	0.790	2.740	358	0	0.124	0.950	348
TA q25	0.150	1.162	3.340	358	0	0.158	0.960	348
TA q75	0.400	2.233	4.530	358	0.020	0.313	3.860	348
TA q90	0.592	2.892	5.520	358	0.026	0.466	4.623	348
TA q95	0.730	3.376	7.338	358	0.030	0.622	5.232	348

Note. TA = total acceleration; MAD = median absolute deviation; IQR = interquartile range.

Table 2. Distribution of Acceleration Features: Cross-Sectional Surveys 1 and 2.

	Cross-Sectional Survey 1				Cross-Sectional Survey 2			
Feature	Minimum	Mean	Maximum	*N*	Minimum	Mean	Maximum	*N*
TA mean	0.006	0.308	13.949	12,483	0.010	0.312	19.504	9,992
TA median	0	0.240	13.950	12,483	0.005	0.240	19.500	9,992
TA variance	0	0.135	43.466	12,483	0	0.133	20.339	9,992
TA MAD	0	0.121	5.308	12,483	0	0.120	3.173	9,992
TA IQR	0	0.194	7.547	12,483	0	0.189	4.330	9,992
TA minimum	0	0.018	1.500	12,483	0	0.018	0.997	9,992
TA maximum	0.040	2.284	79.120	12,483	0.030	2.728	72.170	9,992
TA range	0.040	2.266	79.090	12,483	0.030	2.710	72.170	9,992
TA q5	0	0.104	13.850	12,483	0	0.106	19.370	9,992
TA q10	0	0.124	13.870	12,483	0	0.126	19.410	9,992
TA q25	0	0.167	13.900	12,483	0	0.168	19.460	9,992
TA q75	0.010	0.363	14.010	12,483	0.010	0.360	19.580	9,992
TA q90	0.010	0.568	14.663	12,483	0.020	0.573	19.680	9,992
TA q95	0.020	0.756	15.012	12,483	0.020	0.777	19.758	9,992

Note. TA = total acceleration; MAD = median absolute deviation; IQR = interquartile range.

between respondent-pages). Tables 1 and 2 present descriptive statistics of the acceleration features (i.e., independent variables) for the lab experiment and both cross-sectional surveys.

In terms of the outcome variable, we focus on a binary variable that condenses the four conditions of the lab experiment into the categories moving (i.e., walking and stair climbing) and nonmoving (i.e., sitting and standing). In line with previous research, we expect that these physiological states are most informative when studying completion conditions in smartphone surveys (see Höhne, Revilla, et al., 2020; Höhne & Schlosser, 2019).

In building our prediction models, we consider regularized regression methods (including standard logistic regression as a special case), decision trees, and tree-based ensemble methods to

include both parametric and nonparametric (i.e., more flexible) techniques in our set of models. A key feature of the tree-based methods is their ability to automatically adapt to complex functional forms of the relationship between the predictors and the outcome, which is particularly valuable when little theoretical guidelines for a priori model specification exist (Kern et al., 2019). Specifically, we train prediction models using the following methods and tuning grids[7] (see, for instance, Hastie et al., 2009; Kuhn & Johnson, 2013):

- Elastic net (GLMnet; Friedman et al., 2010)
  - $\alpha = [0, 0.5, 1]$, $\lambda = [0.1, \ldots, 0]$
- Conditional Inference Trees (CTREE; Hothorn & Zeileis, 2015)
  - mincriterion = [.99, .95, .90, .85, .75]
- Random forests (RF) and extremely randomized trees (Wright & Ziegler, 2017)[8]
  - splitrule = [gini, extratrees], mtry = [3, 4], min.node.size = [15]
- Extreme Gradient Boosting (XGBoost; Chen & Guestrin, 2016)
  - max_depth = [3, 5, 7, 9], nrounds = [500, 1,000, 1,500], $\eta = [.01, .05]$.

The models are tuned and evaluated using nested cross-validation with 10 outer folds and 10 inner folds. In the inner cross-validation loop, the best tuning parameter setup for each method is selected based on cross-validated binomial log-loss. In the outer cross-validation loop, the respective best models are evaluated and compared based on cross-validated accuracy, log loss, and area under the receiver operating characteristic (ROC) curve (ROC–AUC). We employ grouped cross-validation in the outer cross-validation loop to account for the hierarchical structure of the training data by ensuring that all observations (survey pages) from a given respondent are either in the training set or in the holdout set. By (repeatedly) using different subsets of the data for model training and model evaluation, the cross-validation results allow us to learn about the out-of-sample performance of the prediction models (i.e., their ability to generalize to new data). Nesting two cross-validation loops cleanly separates model selection (tuning) and model assessment such that honest evaluation of the selected models with new (outer) test sets is possible (Hastie et al., 2009).

Step 2: Predicting motion conditions and analyzing response behavior. In the next step, the selected final model is employed to predict motion conditions in the two cross-sectional surveys (Data Sources 2 and 3). Again, aggregated acceleration variables are used as features. We compare the response behavior of moving and nonmoving respondents using (generalized) mixed-effects regression models (with survey pages nested in respondents; see Raudenbush & Bryk, 2002) and logistic regressions, with the predicted motion condition being the predictor of main interest. We further include age, gender, and mother tongue (German) for Data Sources 2 and 3 as control variables, assuming that particularly younger respondents might be more likely to complete surveys "on the go" and have been shown (along with male respondents and immigrants) to be more prone to superficial responding (Schonlau & Toepoel, 2015; van Vaerenbergh & Thomas, 2013). The outcome variables (i.e., response behavior indicators) of the (mixed-effects) regression models are defined as follows (see Online Appendix A in the Online Supplement of this article for descriptive statistics)[9]:

- Completion time: Time in seconds that respondents needed to complete the web survey page (i.e., the time elapsing between question presentation on the screen and submitting the response by clicking next).
  - We applied a distribution-based outlier definition: We excluded all respondents with completion times below the 5% percentile or above the 95% percentile.
  - Calculated for survey pages with single and item-by-item questions.

- Low (high) IRV: *SD* of responses on a survey page lower (higher) than the 20% (80%) percentile (see Dunn et al., 2018; Marjanovic et al., 2015).
  - Binary (low IRV): 1 = "IRV $\leq$ 20% percentile," 0 = "IRV > 20% percentile."
  - Binary (high IRV): 1 = "IRV $\geq$ 80% percentile," 0 = "IRV < 80% percentile."
  - Calculated for survey pages with item-by-item questions only.
- Primacy effects: selecting the first category of a response scale (van Vaerenbergh & Thomas, 2013).
  - Binary: 1 = "first category," 0 = "not first category."
  - Calculated for survey pages with single and item-by-item questions.
- Compliance with an IMC: correctly answering an IMC (Berinsky et al., 2014).
  - Binary: 1 = "compliance," 0 = "noncompliance."

The code and the final (trained) RF model that is used in this article are available at the following GitHub repository: https://github.com/chkern/survey-motion.

Results

Research Question 1: How accurately can we predict respondents' motion conditions in smartphone surveys using acceleration data?

The cross-validation results of the machine learning models predicting motion conditions with the training data (Data Source 1) are reported in Table 3. More precisely, we report summary statistics for the distribution of accuracy, ROC–AUC, and log loss over all outer holdout sets (i.e., samples were not used for model tuning) for the best model of each method. For both accuracy (i.e., the proportion of correct classifications, range in [0, 1]) and ROC–AUC (i.e., range in [0, 1] with 0.5 representing a noninformative model), higher values indicate better prediction performance. In contrast, for binomial log-loss (i.e., the distances between the predicted probabilities and the true classes, no upper bound with the log loss of a [useless] model which predicts 0.5 for all cases being $\ln(0.5) \approx 0.693$), lower values indicate better performance.

Table 3. Prediction Performance.

Model	Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Accuracy						
GLMnet	.889	0.986	0.994	.976	1	1
CTREE	.884	0.964	0.985	.973	0.997	1
RF	.913	0.976	0.986	.980	1	1
XGBoost	.875	0.964	0.986	.971	1	1
ROC–AUC						
GLMnet	.963	0.994	1	.993	1	1
CTREE	.909	0.994	1	.988	1	1
RF	.970	1	1	.995	1	1
XGBoost	.971	0.996	1	.994	1	1
Log loss						
GLMnet	0	0.004	0.024	.078	0.114	0.350
CTREE	.008	0.015	0.044	.334	0.086	2.573
RF	.004	0.016	0.041	.106	0.077	0.492
XGBoost	.010	0.013	0.046	.094	0.128	0.317

Note. GLMnet = Elastic net (Friedman et al., 2010); CTREE = Conditional Inference Trees (Hothorn & Zeileis, 2015); RF = Random forests (Wright & Ziegler, 2017); XGBoost = Extreme Gradient Boosting (Chen & Guestrin, 2016).

The results show a strong prediction performance for all four machine learning models, irrespective of the performance measure used. Thus, the derived acceleration features can be used to build prediction models that sharply discriminate between the two motion conditions (moving and nonmoving) with mean ROC–AUCs between 0.988 and 0.995. As even the (simpler) CTREE model achieves a mean ROC–AUC of 0.988, this result underlines the predictive power of the acceleration data for the prediction task at hand. In order to utilize the best possible model for predicting motion conditions in the two cross-sectional surveys (Data Sources 2 and 3), we selected the RF as the final prediction model, as it achieves the highest mean accuracy (0.980) and the highest mean ROC–AUC (0.995).

Research Question 2: How common are different types of motion conditions in smartphone surveys?

We calculated proportions of respondents that are predicted as moving and nonmoving in the two cross-sectional surveys (Data Sources 2 and 3) based on the RF model that was trained with data of the lab experiment as outlined above (Data Source 1). Overall, a small proportion of respondents is predicted as moving while answering the survey questions. This is the case for both cross-sectional surveys (Data Sources 2 and 3) and all web survey pages under investigation in this study. On average, we predict that 3.7% (Data Source 2) and 2.7% (Data Source 3) of the respondents were moving, indicating only minor differences between the two cross-sectional surveys. Our predictions indicate that active movement, such as walking around, is a relatively rare completion behavior in both studies at hand.[10] Proportions of respondents that are predicted as moving and nonmoving for each survey page of both cross-sectional surveys are presented in Online Appendix D in the Online Supplement of this article.

Research Question 3: How do respondents with different predicted motion conditions differ with respect to response behavior?

The predicted motion conditions allow us to further investigate whether and to what extent moving respondents differ from nonmoving respondents in terms of response behavior. In this context, we assume that moving, compared to nonmoving, is associated with more (situational) distractions and, thus, longer completion times and lower response quality.

Completion Time

As a first step, we compared completion times between respondents that were predicted as moving and respondents that were predicted as nonmoving by using mixed-effects regression models. The results are presented in Table 4. Note that all regressions represent pooled models that are based on data from Cross-Sectional Surveys 1 and 2. Model 1 reveals that completion times are higher for moving respondents than for nonmoving respondents. Specifically, we observe 0.772 s longer completion times for predicted moving respondents on average. Model 2 shows that these relationships are amplified when including demographic controls (i.e., age, gender, and mother tongue). In contrast, we observe shorter completion times in Cross-Sectional Survey 2 compared to Cross-Sectional Survey 1 (Model 2). With little surprise, we observe that completion times are higher for web survey pages with an item-by-item approach (Model 3). Model 4 indicates that longer completion times of moving respondents can specifically be observed on survey pages with item-by-item questions, as shown by the positive interaction effect of Moving $\times$ Item-by-item. By contrast, there is little variation in the positive effect of moving between the two cross-sectional surveys (Model 5).

Table 4. Mixed-Effects Regression Parameters (Estimated Coefficients and *SE*) on the Dependent Variable Completion Time (in Seconds).

Predictor Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Moving	0.772 (.327)	0.959 (.334)	0.958 (.334)	0.377 (.386)	0.844 (.420)
p	.019	.005	.005	.330	.045
Survey 2	−1.204 (.115)	−3.196 (.176)	−3.197 (.176)	−3.186 (.176)	−3.208 (.178)
p	.000	.000	.000	.000	.000
Item-by-item			28.161 (.487)	28.103 (.488)	28.163 (.487)
p			.000	.000	.000
Moving × Item-by-item				1.883 (.629)	
p				.003	
Moving × Survey 2					0.301 (.675)
p					.656
Constant	14.928 (5.208)	15.780 (5.200)	7.734 (0.296)	7.745 (0.297)	7.738 (0.296)
Demographic controls		Yes	Yes	Yes	Yes
Respondents	1,858	1,795	1,795	1,795	1,795
Pages	7	7	7	7	7
Observations	15,239	14,698	14,698	14,698	14,698
r^2 (fixed effects)	.002	.006	.802	.802	.801
r^2 (fixed and random effects)	.853	.855	.836	.836	.836
AIC	98,923.260	95,062.610	95,025.260	95,017.390	95,026.010
BIC	98,969.040	95,130.970	95,101.210	95,100.940	95,109.560

Note. *SE* in parentheses. Coding of the variables: moving (1 = *yes*), Cross-Sectional Survey 2 (1 = *yes*), and item-by-item (1 = *yes*). We controlled for the following demographics: age, gender, and mother tongue (German) AIC: Akaike information criterion; BIC: Bayesian information criterion. r^2 was calculated following Nakagawa and Schielzeth (2013).

Low and High IRV

We now focus on web survey pages with item-by-item questions only and investigate low (e.g., nondifferentiation) and high (e.g., random responses) IRV, as indicators of superficial responding. Table 5 reports the results. While Models 1–3 focus on low IRV, Models 4– 6 use high IRV as the dependent variable.

Low IRV does not seem to be affected by the predicted motion condition (Model 1). This similarly applies when controlling for demographic characteristics (Model 2). The interaction Moving × Cross-Sectional Survey 2 in Model 3, however, indicates a stronger (positive) effect of moving on low IRV in Cross-Sectional Survey 2.

Interestingly, for high IRV, we find a negative effect of moving that increases from Model 4 to Model 5. This indicates that moving respondents are less likely to exhibit highly variable response patterns on web survey pages with item-by-item questions. The conditional main effect of moving (while including the interaction term Moving × Cross-Sectional Survey 2) shows that this is specifically the case for Cross-Sectional Survey 1 (Model 6).

Primacy Effects

In the generalized mixed-effects regressions on primacy effects, we again consider web survey pages with single and item-by-item questions. The results are presented in Table 6. Overall, Models 1 and 2 show only little differences between moving and nonmoving respondents with respect to their tendency to select the first category of a response scale. Primacy effects are

Table 5. Generalized Mixed-Effects Regression Parameters (Estimated Coefficients and *SE*) on the Dependent Variables Low IRV (1 = *Yes*) and High IRV (1 = *Yes*).

	Low IRV			High IRV		
Predictor Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Moving	0.030 (.213)	0.114 (.214)	−0.197 (.273)	−0.336 (.265)	−0.489 (.279)	−0.676 (.348)
p	.888	.595	.472	.206	.080	.053
Survey 2	−0.156 (.075)	−0.109 (.108)	−0.147 (.109)	0.004 (.085)	−0.265 (.126)	−0.283 (.127)
p	.039	.314	.179	.965	.035	.026
Moving × Survey 2			0.892 (.445)			0.554 (.574)
p			.046			.335
Constant	−1.230 (.060)	−1.334 (.090)	−1.317 (.091)	−1.704 (.080)	−1.348 (.104)	−1.340 (.104)
Demographic controls		Yes	Yes		Yes	Yes
Respondents	1,851	1,808	1,808	1,851	1,808	1,808
Observations	4,794	4,653	4,653	4,794	4,653	4,653
Pseudo-r^2 (fixed effects)	.002	.002	.004	.001	.011	.011
Pseudo-r^2 (fixed and random effects)	.156	.149	.148	.272	.263	.264
AIC	5,257.212	5,079.390	5,077.416	4,738.738	4,606.210	4,607.294
BIC	5,283.113	5,124.507	5,128.978	4,764.639	4,651.327	4,658.856

Note. *SE* in parentheses. Coding of the variables: moving (1 = *yes*) and Cross-Sectional Survey 2 (1 = *yes*). We controlled for the following demographics: age, gender, and mother tongue (German). Pseudo-r^2 was calculated following Nakagawa and Schielzeth (2013). AIC = Akaike information criterion; BIC = Bayesian information criterion; IRV = intraindividual response variability.

more common on survey pages with item-by-item questions (Model 3). In addition, Model 4 indicates a stronger (negative) effect of moving on primacy effects for web survey pages with item-by-item questions (interaction of Moving × Item-by-item), compared to the weak (positive) effect of moving for web survey pages with single questions (conditional main effect of moving). Furthermore, the conditional main effect of moving (while including the interaction term Moving × Cross-Sectional Survey 2) in Model 5 indicates that moving respondents in Cross-Sectional Survey 1 are less likely to exhibit primacy effects, compared to nonmoving respondents.

IMC

Finally, we conducted a logistic regression investigating the compliance with an IMC of moving and nonmoving respondents. We now focus on one single web survey page. Table 7 reports the results. Models 1 and 2 show that both moving and nonmoving respondents are equally likely to comply with the IMC. However, respondents in Cross-Sectional Survey 2 generally have a lower probability of complying with the IMC. Finally, Model 3 shows no substantial interaction between the predicted motion condition and the data source (Moving × Cross-Sectional Survey 2).

Table 6. Generalized Mixed-Effects Regression Parameters (Estimated Coefficients and *SE*) on the Dependent Variable Primacy (1 = *Yes*).

Predictor Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Moving	−0.106 (.082)	−0.115 (.084)	−0.115 (.084)	0.266 (.164)	−0.183 (.102)
p	.195	.171	.170	.106	.075
Survey 2	−0.050 (.028)	−0.058 (.044)	−0.058 (.044)	−0.060 (.044)	−0.067 (.044)
p	.073	.186	.186	.174	.134
Item-by-item			1.356 (.297)	1.371 (.298)	1.357 (.298)
p			.00001	.00001	.00001
Moving × Item-by-item				−0.460 (.175)	
p				.009	
Moving × Survey 2					0.198 (.171)
p					.246
Constant	−1.955 (.266)	−1.933 (.272)	−2.320 (.166)	−2.333 (.166)	−2.318 (.166)
Demographic controls		Yes	Yes	Yes	Yes
Respondents	1,873	1,809	1,809	1,809	1,809
Pages	7	7	7	7	7
Observations	50,497	48,864	48,864	48,864	48,864
Pseudo-r^2 (fixed effects)	.0002	.001	.075	.076	.075
Pseudo-r^2 (fixed and random effects)	.275	.276	.268	.268	.268
AIC	52,281.600	50,725.290	50,717.790	50,713.540	50,718.480
BIC	52,325.740	50,795.660	50,796.960	50,801.510	50,806.450

Note. *SE* in parentheses. Coding of the variables: moving (1 = *yes*), Cross-Sectional Survey 2 (1 = *yes*), and item-by-item (1 = *yes*). We controlled for the following demographics: age, gender, and mother tongue (German). Pseudo-r^2 was calculated following Nakagawa and Schielzeth (2013); AIC = Akaike information criterion; BIC = Bayesian information criterion.

Discussion and Conclusion

This study aimed at utilizing acceleration data from smartphone sensors to predict motion conditions of smartphone respondents and to study the effects of different motion conditions on superficial responding. Specifically, we built prediction models to classify respondents into two motion conditions (moving and nonmoving) on the survey page level using data from a lab experiment. On this basis, we predicted the motion conditions of respondents in two "real-world" cross-sectional surveys and compared moving and nonmoving respondents with respect to their response behavior.

Regarding our first research question (i.e., how accurately can we predict respondents' motion conditions in smartphone surveys?), the cross-validation results showed that acceleration data from smartphone sensors can indeed be used to precisely predict motion conditions. We observed strong prediction performance for all methods that were considered (GLMnet, CTREE, RF, and XGBoost), with RF showing the best results over several metrics (accuracy, log loss, and ROC–AUC). The trained RF model is released along with this article and can be used by researchers at their own discretion to predict motion conditions in (new) smartphone surveys.

In order to answer our second research question (i.e., how common are different types of motion conditions in smartphone surveys?), we utilized the RF model to predict motion conditions in two cross-sectional surveys. The results indicated that active movement, such as walking around, during survey completion is a relatively rare phenomenon. On average, we predicted about 4% (Cross-Sectional Survey 1) and about 3% (Cross-Sectional Survey 2) of the respondents were moving while answering the survey questions under investigation, with little variation across survey

Table 7. Logistic Regression Parameters (Estimated Coefficients and *SE*) on the Dependent Variable Compliance With IMC (1 = *Yes*).

Predictor Variable	Model 1	Model 2	Model 3
Moving	.284 (.261)	.225 (.264)	.566 (.351)
p	.277	.395	.107
Survey 2	−.897 (.085)	−.969 (.121)	−.939 (.122)
p	.000	.000	.000
Moving × Survey 2			−.945 (.595)
p			.113
Constant	.231 (.058)	.174 (.092)	.161 (.092)
Demographic controls		Yes	Yes
Observations	2,392	2,329	2,329
Pseudo-r^2	.058	.068	.069
AIC	3,184.754	3,090.659	3,089.964
BIC	3,202.093	3,125.178	3,130.237

Note. *SE* in parentheses. Coding of the variables: moving (1 = *yes*) and Cross-Sectional Survey 2 (1 = *yes*). We controlled for the following demographics: age, gender, and mother tongue (German). Pseudo-r^2 was calculated following McKelvey and Zavoina (1975). IMC = instructional manipulation check; AIC = Akaike information criterion; BIC = Bayesian information criterion.

pages. An important point to note in this context is that the acceleration data from smartphones' accelerometer are not affected by (steady) movement of a car or train. That is, respondents who are sitting still in a cab or train are likely classified as nonmoving, while entering and exiting a vehicle induces acceleration that would classify respondents as moving. One could argue that this definition of moving is in line with our theoretical arguments since particularly active movement (as opposed to sitting in a train) requires additional cognitive resources and coordinating multiple tasks. Nevertheless, we acknowledge that also passive movement (e.g., riding in a train) can cause distractions that may divert the attention of respondents away from the survey task. Thus, future research could aim at differentiating between these different types of movement, potentially by combining acceleration and GPS data.

To answer our third research question (i.e., how do respondents with different predicted motion conditions differ with respect to response behavior?), we compared respondents with respect to a variety of aspects. First, we observed that completion times are higher for moving respondents than for nonmoving respondents. This was particularly the case for survey pages with item-by-item questions, mirroring the findings of Höhne and Schlosser (2019). This result provides supporting evidence that our lab-trained prediction model indeed identified respondents with different completion conditions in two smartphone surveys that were conducted in a field setting.

Second, in contrast to our expectation, we found only little differences when comparing predicted moving and nonmoving respondents with respect to indicators of superficial responding (i.e., low IRV, high IRV, primacy effects, and compliance with an IMC). Exceptions include a negative effect of moving on high IRV and somewhat weaker primacy effects for moving respondents in Cross-Sectional Survey 1. In combination with the low proportion of predicted movers, our findings indicate that the portability of smartphones does not necessarily seduce respondents to carelessly complete surveys on the move. However, considering the overall low proportion of predicted moving respondents, these findings should be interpreted with caution. Also, we advocate for future research on the relationship between motion conditions and response behavior in smartphone surveys employing further indicators of superficial responding, such as break-offs and item nonresponse (the occurrence of these response behaviors was negligible in both cross-sectional surveys of

this study). Ideally, this research would study motion conditions over many web survey pages and with samples that observe a sufficient number of examples for such outcomes.

This study can be extended with respect to the following limitations. First, it is important to note that our prediction models were trained with data from a lab experiment (i.e., motion conditions were varied in a controlled setting) with a rather small sample of university students and by using a between-subject design. This poses the question of generalizability, that is, whether the lab-trained (and cross-validated) RF generalizes to data that is collected in the field. However, while the summary statistics of our acceleration features do show a higher maximum acceleration in the two cross-sectional surveys, we do not observe severe shifts in the distributions of our features between the different data sources (see the descriptive statistics in Tables 1 and 2). Nonetheless, it would be worthwhile to collect further training data using a more heterogeneous sample and a less artificial setting. Second, we only considered a binary outcome as the prediction objective, whereas studying multiple levels of movements (e.g., nonmoving, moving slow, moving fast) might be instructive when comparing response behaviors between motion groups. Note, however, that in our case, refined motion conditions would have resulted in subgroups with even smaller numbers of respondents. Third, predicted motion conditions could be studied from a longitudinal perspective (e.g., by exploring sequences and/or changes of motion conditions during survey completion on the respondent level over time). Relatedly, time dependencies between web survey pages could be considered by, for instance, adding lagged variables that include information about the predicted motion conditions of previous survey pages. Fourth, the cross-sectional surveys in this study measured only a limited number of sociodemographic background variables that could be included when modeling superficial responding. Detailed background information, ideally combined with larger sample sizes, might not only be valuable for extending the set of control variables but also for studying the prevalence of different motion conditions in different sociodemographic subgroups.

This study contributes to web survey research by proposing a prediction approach that enables to study respondents' motion conditions in smartphone surveys. In addition, it provided insights on the prevalence of different motion conditions in two smartphone surveys and on the effects of motion conditions on response behavior. Our prediction model for identifying moving and nonmoving respondents—in combination with the SMotion tool (Höhne & Schlosser, 2019) for collecting acceleration data—allows future research to analyze respondents' completion conditions and to extend this study in various ways. This includes, for instance, studying the effects of (various forms and patterns of) active movement on break-offs and superficial responding in larger smartphone surveys. In addition, major health-related surveys, such as the Health and Retirement Study and the Survey of Health, Ageing, and Retirement in Europe, conduct fitness tests, such as balance and walk tests. These tests are overseen by interviewers that monitor respondents' compliance and performance. The use of smartphone surveys, coupled with the collection of acceleration data and our prediction approach, may allow validating compliance in a self-administered survey setting instead. Documentation and code examples on how the prediction approach and the trained model from this study may be used by future research are provided in the GitHub repository that accompanies this article (https://github.com/chkern/survey-motion). So far, the usage of sensor data, in general, and acceleration data, in particular, in smartphone surveys to explore respondent behavior is still in its infancy. Thus, various research directions exist that could be addressed in the future to shed light on the importance and implications of different completion conditions in smartphone surveys.

Data Availability

Data may be accessed on-site at the University of Mannheim by contacting the first author of this article (c.kern@uni-mannheim.de).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We acknowledge financial support by the German Science Foundation through the Collaborative Research Center 884 "Political Economy of Reforms" at the University of Mannheim in Germany for conducting Cross-Sectional Survey 2 (*Data Source 3*).

Software Information

All data preparations and modeling steps were conducted using R (Version 3.5.1; R Core Team, 2018). The training, tuning, and evaluation infrastructure for building the prediction models was provided by the caret package (Version 6.0-81; Kuhn, 2018), using the glmnet (Version 4.0-2), partykit (Version 1.2-9), ranger (Version 0.12.1) and xgboost (Version 1.2.0.1) packages. The mixed-effects models were fitted with the lme4 package (Version 1.1-19; Bates et al., 2015).

Supplemental Material

The supplemental material is available in the online version of the article.

Notes

1. Even though paradata and sensor data share many similarities (e.g., they can be measured via JavaScript and inform about the survey completion process), it makes sense to distinguish both data types for the purposes of this study. We therefore define sensor data as data that are specifically collected via built-in sensors, such as accelerometers, of smartphones during the survey completion process.
2. Note that this was only observed for web survey pages with multiple questions but not for web survey pages with one single question, indicating that question presentation (single or item-by-item) matters.
3. The International System unit for acceleration is meter per second squared (m/s^2).
4. For a more detailed description of the setup of the lab experiment, see Höhne and Schlosser (2019).
5. The survey was conceptualized as smartphone survey, but some respondents ($n = 937$) took part using a different device. These respondents were identified using the User-Agent-Strings that inform about device properties, such as device type and model (see Callegaro, 2013).
6. The device type was detected at the beginning of the survey, and respondents who attempted to access the survey using a different device than a smartphone were prevented from continuing the survey and asked to use a smartphone.
7. Default settings are used for hyperparameters not listed.
8. Random forests and extremely randomized trees are implemented in caret (Version 6.0-81; Kuhn, 2018) in one function (with the method being a tuning parameter).
9. We also considered item nonresponse as an additional indicator for superficial responding. Given the low prevalence of item nonresponse, we report the corresponding results in Online Appendix C in the Online Supplement of this article.
10. These results hardly change when using a different prediction method.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*, 739–753.

Bosch, O. J., Revilla, M., & Paura, E. (2019a). Answering mobile surveys with images: An exploration using a computer API. *Social Science Computer Review*, *37*, 669–683.

Bosch, O. J., Revilla, M., & Paura, E. (2019b). Do millennials differ in terms of survey participation? *International Journal of Market Research*, *61*, 359–365.

Callegaro, M. (2013). Paradata in web surveys. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic uses of process information* (pp. 261–280). John Wiley.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Retrieved March 20, 2020, from https://arxiv.org/abs/1603.02754

Couper, M. P., & Peterson, G. J. (2017). Why do web surveys take longer on smartphones? *Social Science Computer Review*, *35*, 357–377.

Diedenhofen, B., & Musch, J. (2017). PageFocus: Using paradata to detect and prevent cheating on online achievement tests. *Behavior Research Methods*, *49*, 1444–1459.

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, *33*, 105–121.

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., & Toepoel, V. (2019). Where you at? Using GPS locations in an electronic time use diary study to derive functional locations. *Social Science Computer Review*. https://doi.org/10.1177/0894439319877872

Elhoushi, M., Georgy, J., Noureldin, A., & Korenberg, M. J. (2017). A survey on approaches of motion mode recognition using sensors. *IEEE Transactions on Intelligent Transportation Systems*, *18*, 1662–1686.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22.

Gummer, T., Quoß, F., & Roßmann, J. (2019). Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones? *Social Science Computer Review*, *37*, 371–384.

Hanson Frieze, I. (2010). Cross cultural survey for work and gender attitudes. Retrieved November 5, 2020, from https://sites.google.com/site/friezewebsite/cross-cultural-survey-of-work-and-gender-attitudes

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*, 838–854.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

He, J., Hu, C., & Wang, X. (2016). A smart device enabled system for autonomous fall detection and alert. *International Journal of Distributed Sensor Networks*, *12*, 1–10.

Höhne, J. K., Revilla, M., & Lenzner, T. (2018). Comparing the performance of agree/disagree and item-specific questions across PCs and smartphones. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *14*, 109–118.

Höhne, J. K., Revilla, M., & Schlosser, S. (2020). Motion instructions in surveys: Compliance, acceleration, and response quality. *International Journal of Market Research*, *62*, 43–57.

Höhne, J. K., & Schlosser, S. (2018). Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata Survey focus. *Social Science Computer Review*, *36*, 369–378.

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*, 379–391.

Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. (2020). Switching away: On-device media multitasking in web surveys. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2020.106417

Höhne, J. K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of question formats in web surveys using paradata. *Field Methods*, *29*, 365–382.

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. *Survey Research Methods*, *13*, 23–44.

Kern, C., Klausch, T., & Kreuter, F. (2019). Tree-based machine learning methods for survey research. *Survey Research Methods*, *13*, 73–93.

Kreuter, F., Haas, G.C., Keusch, F., Bähr, S., & Trappmann, M. (2018). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, *38*, 533–549.

Kuhn, M. (2018). *Caret: Classification and regression training. R package version 6.0-81*. Retrieved March 20, 2020, from https://CRAN.R-project.org/package=caret

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lathia, N., Sandstrom, M. G., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLoS One*, *12*. https://doi.org/10.1371/journal.pone.0160589

Lynn, P., & Kaminska, O. (2012). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, *77*, 586–605.

Marjanović, Z., Holden, R. R., Struthers, W., Cribbie, R. A., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, *84*, 79–83.

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, *31*, 725–743.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*, 103–120.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.

Pew Research Center. (2018a). Across 39 countries, three-quarters say they use the internet. Retrieved March 20, 2020, from http://www.pewglobal.org/2018/06/19/across-39-countries-three-quarters-say-they-use-the-internet/

Pew Research Center. (2018b). Smartphone ownership on the rise in emerging economies. Retrieved March 20, 2020, from http://www.pewglobal.org/2018/06/19/2-smartphone-ownership-on-the-rise-in-emerging-economies/

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Revilla, M., & Couper, M. P. (2018a). Comparing grids with vertical and horizontal item-by-item formats for PCs and smartphones. *Social Science Computer Review*, *36*, 349–368.

Revilla, M., & Couper, M. P. (2018b). Testing different order-by-click question layouts for PC and smartphone respondents. *International Journal of Social Research Methodology*, *21*, 695–712.

Revilla, M., & Couper, M. P. (2019). Improving the use of voice recording in a smartphone survey. *Social Science Computer Review*. https://doi.org/10.1177/0894439319888708

Revilla, M., Couper, M. P., Bosch, O. J., & Asensio, M. (2020). Testing the use of voice input in a smartphone web survey. *Social Science Computer Review*, *38*, 207–224.

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels need to adapt surveys for mobile devices? *Internet Research*, *26*, 1209–1227.

Schlosser, S., & Höhne, J. K. (2018). *ECSP—Embedded Client Side Paradata*. Zenodo. https://doi.org/10.5281/zenodo.1218941

Schonlau, M., & Toepoel, V. (2015). Straightlining in web survey panels over time. *Survey Research Methods*, *9*, 125–137.

Sendelbah, A., Vehovar, V., Slavec, A., & Petrovčič, A. (2016). Investigating respondent multitasking in web surveys using paradata. *Computers in Human Behavior*, *55*, 777–787.

Terwey, M. (2000). ALLBUS: A German general social survey. *Schmollers Jahrbuch*, *120*, 151–158.

Toepoel, V., & Lugtig, P. (2015). Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods, Data, Analyses*, *9*, 155–162.

Toninelli, D., & Revilla, M. (2016a). Is the smartphone participation affecting the web survey experience? In *Proceedings of the 48th scientific meeting of the Italian statistical society*, Salerno, Italy. ISBN: 9788861970618.

Toninelli, D., & Revilla, M. (2016b). Smartphones vs PCs: Does the device affect the web survey experience and the measurement error for sensitive topics? A replication of the Mavletova & Couper's 2013 experiment. *Survey Research Methods*, *10*, 153–169.

van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*, 195–217.

Weiss, L., Scherpenzeel, A., & Angleys, N. (2018). *Testing the logistics of the accelerometer project in share. Paper presented at the BigSurv conference, Barcelona, Spain*. https://www.bigsurv18.org/program2018?sess=13#185

Wenz, A. (2019). Do distractions during web survey completion affect data quality? Findings from a laboratory experiment. *Social Science Computer Review*. https://doi.org/10.1177/0894439319851503

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17.

Zwarun, L., & Hall, A. (2014). What's going on? Age, distraction, and multitasking during online survey taking. *Computers in Human Behavior*, *41*, 236–244.

Author Biographies

Christoph Kern is a postdoctoral researcher at the Professorship for Statistics and Methodology at the University of Mannheim. His current research focuses on the usage of machine learning methods in survey research, particularly in the context of nonresponse prediction (email: c.kern@uni-mannheim.de).

Jan Karem Höhne is a postdoctoral researcher at the Collaborative Research Center 884 "Political Economy of Reforms" at the University of Mannheim, Germany, and research fellow at the "Research and Expertise Centre for Survey Methodology" at the Universitat Pompeu Fabra in Barcelona, Spain. His research combines survey methodology, psychology, computer science, and data science (email: hoehne@uni-mannheim.de).

Stephan Schlosser is a doctoral candidate and research associate at the Center of Methods in Social Sciences at the University of Göttingen (Germany). His main research interests focus on survey methodology, data science, and passive data collection (email: stephan.schlosser@sowi.uni-goettingen.de).

Melanie Revilla is Deputy Director of the Research and Expertise Centre for Survey Methodology (RECSM) and an adjunct professor at the Universitat Pompeu Fabra, Spain. Her main research interests include survey methodology and passive data collection (email: melanie.revilla@upf.edu).