



UvA-DARE (Digital Academic Repository)

A novel iOS data donation approach

Automatic processing, compliance, and reactivity in a longitudinal study

Baumgartner, S.E.; Sumter, S.R.; Petkevič, V.; Wiradhany, W.

DOI

[10.1177/08944393211071068](https://doi.org/10.1177/08944393211071068)

Publication date

2023

Document Version

Final published version

Published in

Social Science Computer Review

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Baumgartner, S. E., Sumter, S. R., Petkevič, V., & Wiradhany, W. (2023). A novel iOS data donation approach: Automatic processing, compliance, and reactivity in a longitudinal study. *Social Science Computer Review*, 41(4), 1456-1472. <https://doi.org/10.1177/08944393211071068>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

A Novel iOS Data Donation Approach: Automatic Processing, Compliance, and Reactivity in a Longitudinal Study

Social Science Computer Review
2023, Vol. 41(4) 1456–1472
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08944393211071068
journals.sagepub.com/home/ssc



Susanne E. Baumgartner¹, Sindy R. Sumter¹, Vladislav Petkevič¹, and Wisnu Wiradhany²

Abstract

For social scientists who are interested in understanding smartphone-usage patterns, as well as the causes and effects of smartphone use, it is crucial to get accurate estimates of smartphone and app usage. For this reason, we present a novel data donation approach based on screen video recordings of information provided by the iOS Battery Section. In this paper, we describe the Python script that can be used to process the data automatically. Moreover, we tested accuracy of the script, feasibility of the data collection method, as well as compliance and reactivity of the approach among a sample of 93 University students who participated in a 7-day mobile diary study. The findings of the study show that the approach works well, is very feasible and that the script resulted in accurate smartphone use data. Moreover, findings show that compliance rates were rather high, and no substantial signs of reactivity to the method occurred (i.e., participants did not change their smartphone behavior substantially while participating in this study). We therefore hope that the presented approach will encourage future researchers to use this or similar approaches to gain accurate smartphone use data in an ethical way.

Keywords

digital trace data, data donation, automatic processing, compliance, smartphone tracking

Media use, particularly among younger people transfers more and more to mobile phones (e.g., [Anderson & Jiang, 2018](#); [Silver, 2019](#)), with almost half of adolescents stating that they are online almost constantly ([Anderson & Jiang, 2018](#)). For researchers who are interested in understanding smartphone-usage patterns, as well as the causes and effects of smartphone use, it is crucial to get

¹Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

²Psychology Department, Faculty of Humanities, Bina Nusantara University, West Jakarta, Indonesia

Corresponding Author:

Susanne E. Baumgartner, Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands.

Email: s.e.baumgartner@uva.nl

accurate estimates of smartphone and app usage. This is particularly important because self-reports have been shown to be unreliable (e.g., Araujo et al., 2017; Araujo & Neijens, 2020; Naab et al., 2019; Ohme et al., 2020; Parry et al., 2021). For example, in a recent meta-analysis, Parry et al. (2021) found only moderate associations between measures based on self-reports and logged media use data, questioning the validity of using self-reports for assessing media use.

Although, in the past years, a plethora of apps have been developed to track smartphone use unobtrusively, these apps have three shortcomings for use in social scientific research. First, most apps that allow tracking app usage data are limited to Android phones (or provide only limited data for iPhones). This is problematic as approximately 25% of smartphone users worldwide and more than 50% in North America use smartphones which run on iOS rather than Android (statcounter, 2020). Second, these apps are created by third parties which limit the control of researchers regarding measurement, storage, management, and protection of the data (Breuer et al., 2020). Finally, participants in studies employing these apps have limited control over the type and amount of data that is logged from their phones, and have limited knowledge on the exact amount and type of information they are sharing during their research participation. This might infringe on participants' decision rights concerning their own data (Boonstra et al., 2018; Harari et al., 2016; Schneble et al., 2020).

To circumvent these problems, there have been recent attempts to request data donations from participants based on their iOS *Screen Time* feature (Ohme et al., 2020). *Screen Time* was a new feature Apple introduced with iOS 12, which—following activation by the iPhone user—provides detailed accounts of daily and/or weekly app usage (e.g., minutes a specific app was used per day/week). These data donation approaches require participants to provide, for example, screen shots of their *Screen Time* information and share this with the researcher. Although average time spent with apps for a day or longer period can provide useful insights, for some social scientific research, we need an even higher level of granularity about media usage during the course of the day, or for particular time periods during the day. For example, researchers examining the effects of smartphone use on sleep might be specifically interested in app usage in the evening or during the night.

More granular information on app usage is provided by the iOS *Battery Section*. The Battery Section displays which apps are used for how many minutes each hour of the day, including information about on screen as well as background usage of these apps. Thus, collecting Battery Section data can provide researchers with the opportunity to use a more granular approach and assess app usage in those time slots most relevant to their design. For this reason, we will present in this paper a data donation approach based on screen video recordings of information provided by the iOS Battery Section. Moreover, to ensure feasibility of this approach for large samples, and hence a vast quantity of video content, we will present an approach to automatically process the submitted screen recordings. Finally, we will report on compliance and reactivity of this approach during a 7-day mobile diary study.

Digital Trace Data in the Social Sciences

Due to the increased digitalization of society, a wealth of data is tracked by and stored on the personal devices people use. The capabilities of data logging and storing on personal devices grow exponentially, with chips growing smaller and multiple sensors being developed. In many cases, mobile phones are the hub of unobtrusive data collection. As a result, for every individual that takes part in social scientific research, there is potentially already an abundance of data available (e.g., Fukazawa et al., 2019; Montag et al., 2020; Trifan et al., 2019). These data are typically referred to as digital trace data, and can be defined as “*records of activity (trace data) undertaken through an online information system (thus digital). A trace is a mark left as a sign of passage; it is*

recorded evidence that something has occurred in the past. For trace data, the system acts as a data collection tool [...]” (Howison et al., 2011, p. 769).

Digital trace data provides researchers in a wide range of fields—including psychology, sociology, and communication science—with incredible opportunities (e.g., Choi, 2020; Rafaei et al., 2019; Stachl et al., 2020; Stier et al., 2020). Digital trace data are celebrated because they offer researchers the opportunity to collect media use data in accurate and unobtrusive ways (Araujo, et al., 2017; Boase, 2016; Jones-Jang et al., 2020; Reeves et al., 2019). However, one of the most important concerns of researchers who are considering the use of digital trace data might be the ethical ramifications of such an approach (Boeschoten et al., 2020). At the moment, ethical guidelines for the use of digital trace data are largely missing and the benefits of measuring something objectively and unobtrusively may not outweigh the infringement on privacy inherent to these approaches (Christen et al., 2017; Stier et al., 2020). To make it possible to use digital trace data in an ethical manner, alternative methods have been devised, including digital data donations (e.g., Boeschoten et al., 2020; Halavais, 2019).

Challenges for Data Collection and Processing

Digital data donations consist of data that are automatically tracked by digital devices, and that users then make available to researchers (e.g., Boeschoten et al., 2020; Halavais, 2019; Ohme et al., 2020). The main advantage of data donations in comparison to automatic tracing data is that participants have some control over the amount and type of data they are sharing with the researchers. At the same time, there are concerns that using a data donation approach in research might be less inclusive and may result in attrition, because data donations require several additional actions and skills from participants (e.g., Elevelt et al., 2019). In the case of smartphone or social media data donations, participants have to follow a number of steps to make their data available to researchers (see Figure 1). All these steps require a basic level of digital literacy from participants.

First, participants need to use the requested media (e.g., social media), or media devices (e.g., smartphones) as they usually would. Second, participants should be able to locate the data requested by the researchers on the mobile device. For this, they need to access their mobile device, including the specific widgets, smartphone features, or social media accounts. Third, they need to either directly download the requested data on their personal device (as is possible, for example, for social media accounts, or video streaming accounts such as Netflix), or they need to capture the

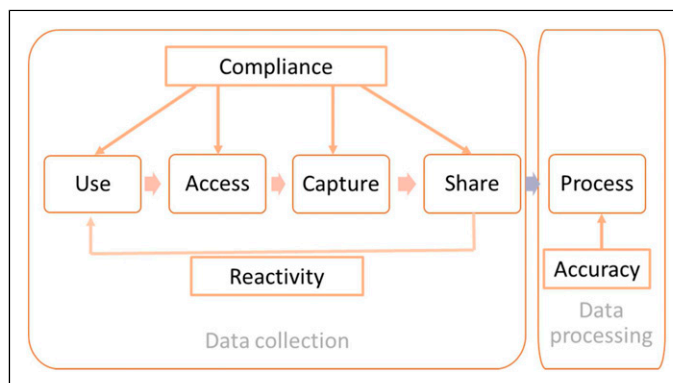


Figure 1. Multi-step data donation process: From data collection to data processing.

requested data with screen shots or video recordings (e.g., if data from smartphone features is used). Fourth, once the data has been downloaded and/or recorded, these files need to be shared with the researchers (e.g., uploaded to a server, or securely mailed to researchers). Specifically, for smartphone data based on the Battery or Screen Time feature, participants have to access the respective feature, capture the information with a screen shot or video recording, and finally share the captured data with the researchers.

Considering this multi-step process of data donations, this approach has at least three challenges which are related to data collection and data processing that need to be considered to fully capitalize on this method. With regards to data collection, following all these steps and sharing data over a longer period of time requires both technical as well as psychological compliance from the participants. Moreover, participating in a data donation study might lead to reactivity to the method. Finally, after participants have shared their data, researchers need to process the acquired information in efficient and accurate ways.

Challenges for Data Collection: Compliance and Reactivity

Compliance in Multiple Day Data Donation Studies. Compliance is typically seen as the willingness to participate in a study, and to follow all requirements related to study participation. Willingness to comply is particularly problematic in intensive longitudinal studies such as experience sampling or diary studies in which participants have to respond over the course of several days and sometimes several times per day (Rintala et al., 2019). In social scientific research, these types of methods are becoming increasingly popular, and combining experience sampling with digital trace data or data donations provides many opportunities for understanding media use, and its effects in ecological valid ways (Choi, 2020; Stier et al., 2020). Typically, compliance decreases over the course of a longitudinal study (e.g., Rintala et al., 2019). The willingness to participate in a longitudinal study including data donations might be even lower as active steps are required from the participants to make data available in comparison to just answering a few questions (e.g., Silber et al., 2021; Skatova & Goulding, 2019). Willingness to comply can be compromised not only by the experienced burden of participation, but also due to forgetting to complete the requested actions.

In addition to the psychological willingness to comply, in research that involves data donations, compliance could also be affected by the technological skills of participants and technical difficulties encountered at each of the aforementioned steps (e.g., Ohme et al., 2020). Although technology dependent compliance is partly outside the control of researcher (e.g., someone's phone is broken), the researcher is able to scaffold the knowledge of participants with limited technical skills. This can be done by making the data donation approach as simple and parsimonious as possible, limiting the number of actions a participant has to take, and making sure that all steps of the process are well explained to the participants. These instructions provided in real time by the researchers can be supplemented with recorded instructions which the participant should be able to access during later stages of the research. A technology helpdesk could also be a way to limit the impact of variance in respondents' technical skills.

There are only a few first studies investigating compliance for data donations (Gower & Moreno, 2018; Ohme et al., 2020). Studies indicate that the willingness to participate in mobile data donation studies are rather low and partly depend on individual characteristics of the participants, such as age and conscientiousness (e.g., Elevelt et al., 2019). However, these studies used cross-sectional designs or a maximum of two data collection moments. In the present study, we therefore test compliance rates and changes in compliance over the course of a 7-day study. We expect compliance to decrease over the course of the study.

Reactivity in Multiple Day Data Donation Studies.

Participating in a longitudinal study may result in a change in respondents' behavior or attitudes, which is typically called reactivity or panel conditioning (Halpern-Manners et al., 2017; Van der Zouwen & Van Tilburg, 2001). Reactivity has been observed in a variety of contexts and might be particularly problematic in studies using smartphone data donations for two reasons. First, by taking part in a study in which smartphone use is tracked, participants might believe that they need to adapt their smartphone behavior in order to demonstrate smartphone use patterns similar to the majority of users. Thus, they may adapt their use in order to comply with social norms. Second, by making screen shots or videos of their own smartphone use statistics, participants get very clear insights into their daily app usage and screen time. This knowledge might make them more aware about the amount of time they spent with their smartphones, and may motivate them to change their behavior.

Both sources of reactivity, that is, 1) reactivity due to mere study participation and 2) due to daily exposure to personal media use insights, can lead participants in data donation studies to either change the overall time they spend with their smartphones or to change the use of specific apps during the course of the study. For example, when participants see that Instagram is the app they use most often, and feel this is more often than they had expected, they might decide to use this app less the next day. Whether reactivity plays a role in data donation studies has not yet been investigated. This is surprising because reactivity might pose a major challenge to data donation approaches. We will, therefore, investigate whether participants change their smartphone behavior, in terms of overall screen time and the types of apps used over the course of the study. This will allow us to get insights into the magnitude and pattern of reactivity for longitudinal data donation approaches.

Challenges for Data Processing: Accuracy

Although the collection of data donations can be easily scaled up, the processing of the collected data requires additional steps. Particularly for larger samples, or if more time points are assessed, an automatic processing of the collected data is necessary in order to make it feasible for usage in larger studies. An automatic approach to process the data will allow researchers to collect more data and to use similar approaches in a variety of studies. To our knowledge, no automated approach that processes the data of such data donations has been published yet. We will, therefore, present in this study a Python script that automatically traces all data in the screen recording videos (for an impression of a screenshot of these videos see [Figure 2a](#) and [b](#)).

The Current Approach

The data donations used in this study are based on information provided by the iOS Battery Section. The Battery Section provides detailed information about app usage in minutes for all apps used (on screen or in background) during each hour of the day. To receive this information, participants were asked to make screen videos of this information by capturing the information provided by the Battery Section. For this, participants had to start the screen video function on their phones, then enter the Battery Section, and finally tap on each hour of the day to show which apps have been used during each hour. Creating these videos took approximately 1 minute per day.

Script for Automated Processing of Battery Section Recordings

In order to allow quick, easy and accurate processing of the video recordings, a Python script was developed. This script processes iPhone screen recordings of the Battery Section pages and

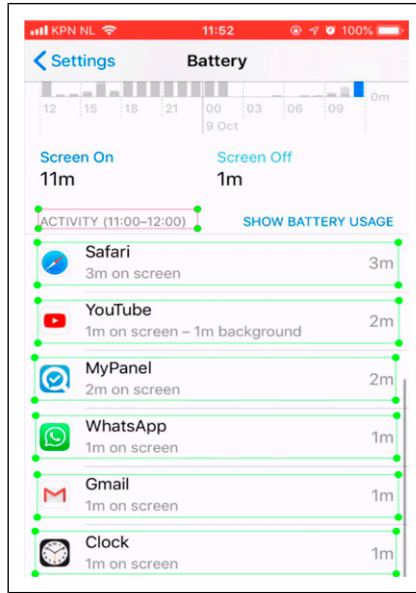


Figure 2a. Example of a manually annotated image used for model training showing various detection fields

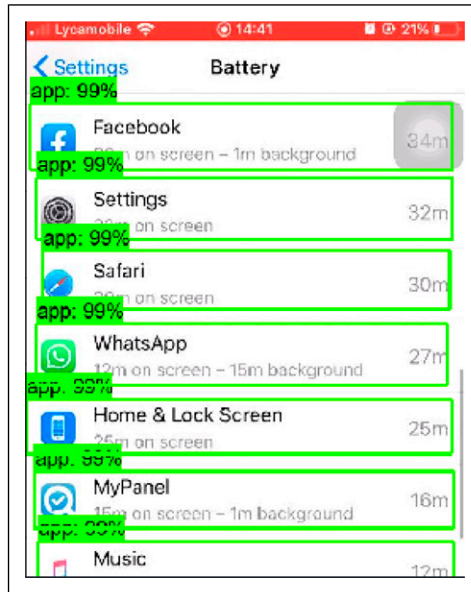


Figure 2b. Visual representation of the object detection model output showing several recognized app fields.

provides a CSV file with the following information: which apps were used at each hour of the day, time in minutes that each app was open on screen per hour, and time in minutes an app was open in the background per hour. This is accomplished by (1) transforming the input video into a series of frames (as numeric arrays), (2) performing object detection to extract the relevant fields (app

name, time indication, etc.) in each frame, (3) sorting and removing irrelevant fields, (4) running optical character recognition (OCR) to transform the detected fields into text data, and (5) processing the text data to re-combine and sort the relevant fields for the CSV output. The script currently processes videos based on the English or Dutch version of the Battery Section.

Video to Frames Conversion (1)

The videos are read using the OpenCV library (2015) and loaded into memory frame by frame as NumPy (Oliphant, 2006) arrays. Since consecutive frames are likely to contain identical information, the script compares the information in each frame to the information in the subsequent frame and records the absolute difference between the two. This is accomplished by converting copies of the frames into grayscale, applying Gaussian blur (as to minimize the effect of artifacts), and subtracting one frame array from another. Then the baseline in the resulting array of differences is estimated and the values sufficiently distant from the baseline are indexed. These indices are used to select the relevant frames from the initially read NumPy arrays. Currently, the sufficient distance from the baseline is estimated automatically such that the number of relevant frames returned by the function is at least twice the length of the screen recording in seconds.

Object Detection (2)

Once a selection of relevant frames is made, the frames are run through an object detection model to detect the following two fields: indications of which hour is being shown and apps with their logos, names, and durations of activity on screen and in the background. To detect these fields, a TensorFlow (Abadi et al., 2016) model was trained on manually annotated screen captures (screenshots, $N = 247$) from a random sample of videos ($N = 31$). An example of a manually annotated image is presented in Figure 2a. The training was initialized using a pre-trained model (Faster RCNN Inception v2¹) and trained for 170,000 steps on the new data (annotated captures). The script calls the trained model and processes all frames in one TensorFlow session outputting detection boxes (coordinates of the detected fields), types of fields, and detection confidence for each field. A visual representation of the model output is shown in Figure 2b.

Sorting and De-duplication (3)

Then OCR (see below) is performed on the hour indication fields to: (1) detect frames that show app usage not by hour but for the given day overall and remove them, and (2) sort the detected app fields into their corresponding hours. The sorting is done by grouping together all the app fields that appear either in the same frame as the hour indication field or—in case the hour indication field is not in the frame—the app fields in all frames that precede a frame with an hour indication field showing a different hour. Thereafter, an OpenCV (2015) implementation of the scale-invariant feature transform (SIFT; Lowe, 2004) algorithm is used to remove duplicated app fields within each hour to speed up the subsequent steps.

Optical Character Recognition (4)

Another instance of the object detection mode is run on the previously detected app fields to extract the image coordinates of the app name and the time for which it was used in the given hour. As before, to do so, a TensorFlow (Abadi et al., 2016) model was trained on manually annotated app fields ($N = 130$) using a pre-trained model (Faster RCNN Inception v2²). Separating the app name and usage time is done to improve the OCR accuracy. OCR is performed on each detected

field (either the app name or usage time) using PyTesseract³—a Python implementation of the Tesseract OCR engine (Smith, 2007). Prior to supplying the images to Tesseract, they are pre-processed as to improve OCR accuracy: the images are enlarged, transformed into black and white, and blurred. The exact parameters for these pre-processing steps differ depending on the field being processed. An example of a pre-processed image is shown in Figure 3.

Data Sorting (5)

Finally, the script combines all information into a data frame for the CSV output. The combined data frame has four columns: time of the day (by hour), app name, app time on screen, and app time in background. Table 1 shows an example of the Battery Section video output. When the output of all the videos is combined into one dataset, fuzzy string matching is performed to account for minor OCR mistakes and standardize the app names in the dataset. The string matching is done by using simple ratio with the threshold of 90%; that is to say, all app name strings that share more than 90% of characters in the same order are considered to be identical and are standardized. The 90% threshold was chosen by trial and error with lower values resulting in the script erroneously considering different apps to be the same (e.g., ‘email’ and ‘mail’). The remaining steps of the script concern data cleaning: the indications of hours are transformed so that their format is consistent across all processed videos; the time indicators are stripped of all characters but digits (e.g., “1 min” becomes “1”); any duplicated entries are removed; finally, all faulty entries (e.g., NAs in the hour or name fields) are removed.

Method

Sample and Procedure

The data donation approach described in this paper was part of a larger study on smartphone use and sleep. Ethical approval was granted by the ethics board of the Communication Science Department of the University of Amsterdam. The data of 93 participants is presented in this paper⁴. All participants were university students (66% female, $M_{age} = 21.02$, $SD_{age} = 3.08$) who currently used an iPhone. At the beginning of the study, participants were invited in groups of up to five for an introduction meeting at the University’s lab facilities. During the 1-hour introduction meeting, participants received detailed instructions about the study set-up and the data donation approach. Participants also had the opportunity to practice recording their screen and uploading the files.



Figure 3. Example of an app name field before (left) and after (right) pre-processing.

Table 1. Example of the Battery Section Video Output.

App	On Screen	In Background	Hour
Gmail	3	NA	16:00–17:00
Settings	23	NA	16:00–17:00
Calculator	3	12	16:00–17:00
Messenger	2	8	16:00–17:00
Weather	4	2	16:00–17:00

Participants downloaded an app (MyPanel) on their phone which was used for daily screen video uploads and additional daily surveys (not reported here). Finally, they filled in an online survey to collect information about demographics, personality traits and their general smartphone use and sleep patterns. After the introduction meeting, they were asked to upload a screen video of their Battery Section every morning (starting the next day) for the next 7 days. A link to an instruction video with a short description of how to take the screen videos was accessible to all participants via the app so that they could check this in case they did not remember how to record the videos. After the 7-day period, participants returned to the lab for an exit meeting, during which they completed a final questionnaire answering questions about their perceived difficulty of sharing the screen recordings, and their perceived reactivity. Participants could choose to be rewarded with 10€ or research credits for participation. After all participants had taken part in the study, the Battery Section videos were processed with the Python script described above.

Results

Descriptives

In total, participants used 773 unique apps during the course of the study. Overall, app use showed a long-tailed distribution, with a few apps being used very frequently, and many apps being used only occasionally. Figure 4 shows the distribution of the top 50 apps. As can be seen the most used app was the “home and lock screen” which is used when unlocking the screen, followed by WhatsApp, Instagram, and Snapchat. Figure 4 includes an overview of the amount of time apps were used per hour (combined for all seven days). It shows that smartphone use is highest in the evening hours, and particularly high between 9:00p.m. and 11:00p.m.

Script Accuracy

The accuracy of the automated video-processing script was assessed by comparing automated script output of 19 videos (1242 app entries) to the manual coding of these same videos (1258 app entries). Both datasets contained the following four columns: app name, time of use on screen, time of use in background, and the hour in which it was used. The accuracy was calculated as the percentage of entries that matched between the manually coded and the automatically processed datasets. Separate scores were calculated for the on screen time, background time, and total time (on screen and background use combined) matches. Forty-two app entries that had neither an indication of time on screen nor of time in background (e.g., some instances of Siri and Torch apps) in the automated output were removed from the manually coded dataset prior to the analysis as such apps were skipped in the automatic coding by design. Similarly, seven apps whose names were not in a Latin script were removed from the automatically generated dataset as these apps were not coded manually. To account for possible minor differences in the app name spelling between the two datasets, fuzzy string matching was used.

The on screen use accuracy (both app name and the on screen time matching between the two datasets) was 86.33%. The background use accuracy was 90.03%, and the total use accuracy (on screen and background use times combined) was 89.60%. In total, 1206 app names (95.86%) present in the manually coded data were correctly identified during the automated processing. Out of these apps, 100 had incorrectly identified screen time, 112 incorrectly identified background time, and 52 had incorrectly identified total time. Finally, 25 apps that were not present in the manual dataset were erroneously included in the automated one and 57 apps were overlooked by the video-processing script; 23 of them due to an entire hour having been missed during the automated processing.

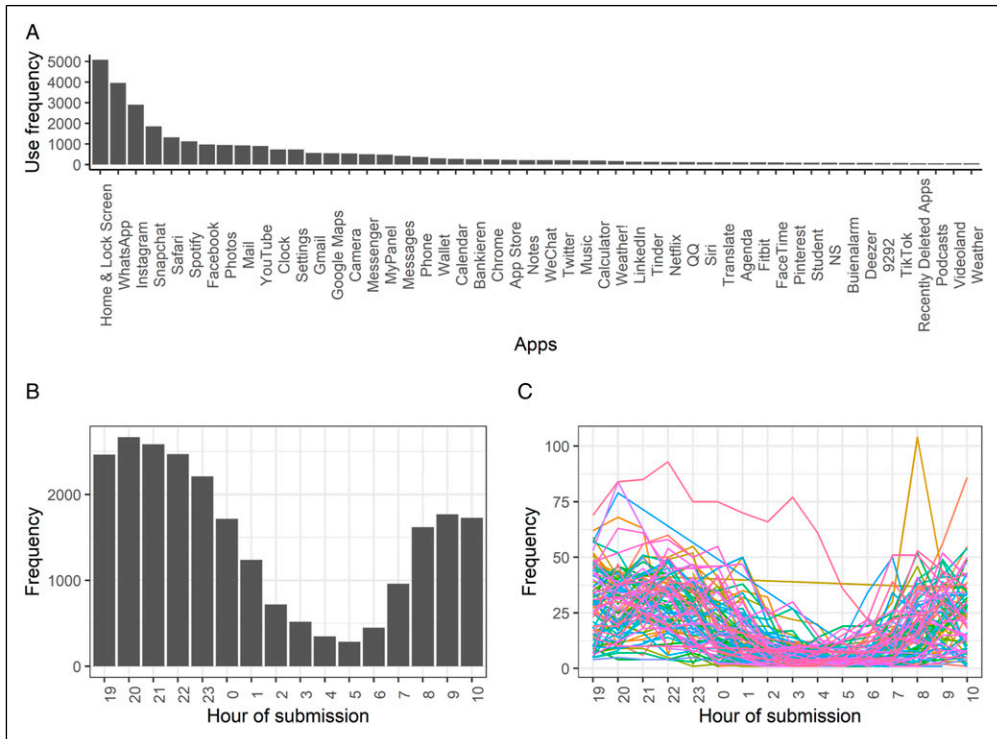


Figure 4. Frequency of the 50 most used apps (A), frequency of apps per hour over all participants (B), and frequency of apps used per hour (per participant) (C).

Compliance

In total, 544 videos were uploaded, which amounted to an average of 5.85 ($SD = 1.76$, median = 6) videos per participant showing overall a high level of compliance. Seven participants uploaded more than the seven requested videos, sharing videos on one or two more days that technically fell outside the scope of the study (these videos were not taken into account for the analyses on compliance and reactivity). When considering long term compliance with daily video uploads, compliance decreased over the course of the study, with a compliance rate of 100% on Day 1, and only 45% on Day 7 (42 videos). See Figure 5 for an overview of how many participants shared videos for how many days. Figure 5 shows that many participants (45%) shared videos on 7 days, and 24% on 6 days. Only 12% uploaded videos for 3 days or less.

The self-report data collected at the end of the study showed that the majority of participants agreed or strongly agreed to the statement that recording the screen videos was easy (56%). However, some somewhat disagreed (19%) or strongly disagreed (6%) with this statement and thus perceived recording the videos as rather difficult indicating that compliance might be partly inhibited due to technical difficulties some participants encountered.

Subjective and Objective Reactivity

Concerning self-reported perceived reactivity, the large majority of participants stated that they believed that participating in this study did not change their smartphone use behavior during the

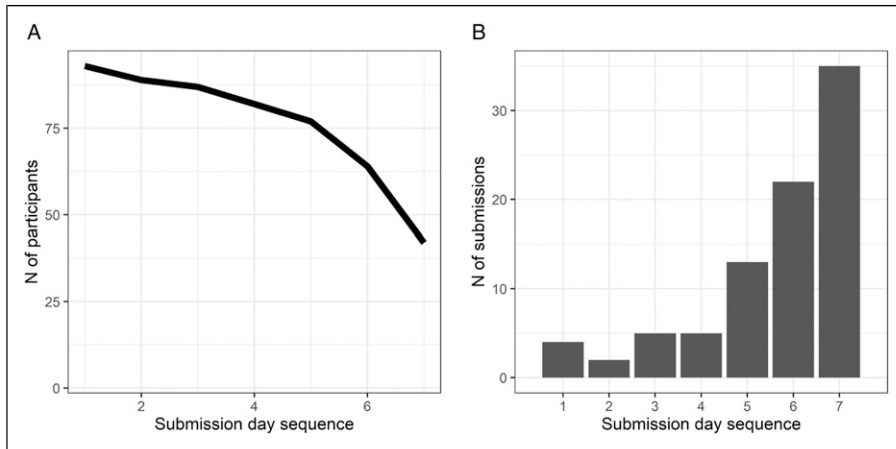


Figure 5. Amount of participants that shared videos on a minimum of 1 to a maximum of 7 days (A), Total submitted videos per day, (B) V.

course of the study (85%). However, 10% indicated they used their phones less, and 3% stated they used their phones more because of study participation.

Objective reactivity was assessed by examining Battery Section data over time. More specifically, we conducted a linear mixed model using the `lme4` package in R (Bates et al., 2015), with time as a predictor, the total minutes spent for apps per day as the outcome measure, and a random slope for participants by time. The random slope takes into account individual differences in app time across days and the unbalanced data (i.e., some participants submitted more data than others). Examining the app use data, it seems that time spent on the phone actually somewhat increased over the 7-day course of the study, $t=2.77$, $p=.005$ (see Figure 6). However, the apps that were used did not change during the course of the study, with the five most used apps being the same on each day (see Figure 6; i.e., Home and lock screen, WhatsApp, Instagram, Snapchat and Safari).

Discussion

The present study introduced a novel approach to collect and process data based on the iOS Battery Section. To our knowledge this is the first study presenting an automated script to process video data from the iOS Battery Section, and the first study reporting compliance and reactivity for a longitudinal data donation study. We hope that the detailed account and the availability of the Python script⁵ will allow other researchers to easily implement this approach in their study, and will enable them to collect accurate smartphone use data.

Concerning compliance, the findings of this study are in line with previous studies showing that compliance typically declines over the course of a longitudinal study (e.g., Rintala et al., 2019). Nevertheless, participants uploaded videos on average on more than five out of the seven days, and more than two-third of the sample uploaded video for six or seven days. Given that we gave participants instructions to upload these daily videos but did not remind them during the study period, the study shows rather high compliance rates and underlines the feasibility of our approach. Several participants noted during the exit meeting that they just forgot to upload videos. We, thus, believe that compliance can be easily heightened through reminders, and are especially important to consider when studies are longer than five days. The study also suffered from diminished compliance due to technological failure. During the exit interviews some participants

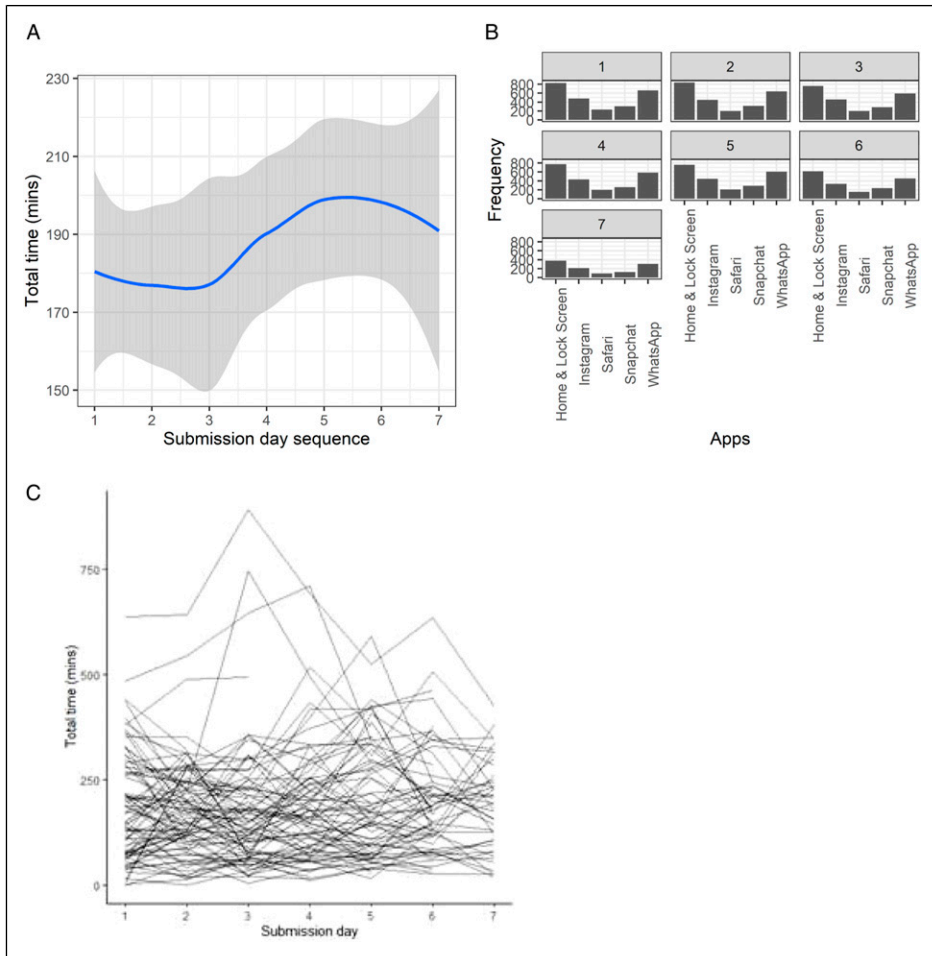


Figure 6. Average time spent on apps per study day (A), the top five most used apps per submission day (B), and the actual fluctuation of app usage per participant (C).

mentioned that they had encountered problems while uploading the videos to the app that was used in this study. These problems can be easily circumvented in future studies by making other uploading options available to participants. Thus, we recommend creating fall back options, for example, allowing participants to upload their files through a research website rather than the research app on their phone.

It is important to note that participants in the current study were all highly educated University students who might have above average technical skills. To use this approach among other age groups and people with lower media and technology literacy might require additional training sessions to improve compliance. Future studies should test this approach and compliance among more diverse samples. However, we believe that at least for the age groups that use smartphones the most in their daily lives ([Pew Research Center, 2019](#)), the current approach is feasible and produces high compliance rates.

A second promising result for future studies is that reactivity seems to play a minor role when using data donations. We expected that awareness and social desirability might lead to a decline in

smartphone use, however, this did not occur. In contrast, we rather found a slight increase in smartphone use over the course of the study. One potential explanation is that participants decreased their smartphone use at the beginning of the study and then returned to normal levels during the course of the study. Importantly, the types of apps that were used in the sample, did not change over the course of the study. Also, only very few participants indicated that they changed their smartphone behavior as a result of taking part in this study. Although this is definitely a good sign, other types of reactivity might still have occurred that are not reflected in the data. For example, participants might have chosen to not use specific apps during the course of the study that they normally would use, or they might have used them on other devices. Future studies are needed to test reactivity in more detail.

The study also showed that the developed Python script worked well and provided accurate app usage data. However, the script can still be improved. The key frame detection algorithm should be optimized in future versions of the script to improve both the app recognition accuracy and the video-processing time. The current approach extracts some superfluous frames (thus unnecessarily increasing the processing time) and can sometimes miss relevant ones (thus reducing the overall accuracy). Better performance can be achieved by using more sophisticated peak detection algorithms and baseline estimation techniques. For instance, an automatic scene change detection algorithm could be introduced and separate baselines estimated for each scene. Another issue that should be addressed in future revisions is the script's incorrect categorization of the apps that were open exclusively in the background in a given hour. Presently such apps are erroneously categorized as being open on screen. This issue arises from the script relying only on the app time indications and their relative positions and can be solved by also detecting the words 'on screen' and 'in background' and employing that information in the categorization.

Implementation Recommendations for Future Research

To enhance the successful use of data donation in combination with the presented Python script, we recommend researchers adhere to the following guidelines when designing their study.

First, it is helpful to organize intake meetings with (small groups of) participants to explain the different actions that are required during the study and include a live demonstration of these actions. This intake meeting can be organized both offline as well as online, where instructions can be provided in a web conference environment with screen sharing features. Second, it is useful to provide instructions for creating data donations as reminders to participants either via video instructions, or other forms of easily comprehensible and accessible information. In the present study, we provided instruction videos implemented in the research app, so that participants could easily access them to refresh their memories. This ensures that participants have access to clear instructions about the most important part of the data donation procedure. When finances allow it, we would also recommend implementing an online helpdesk accessible through the app where people can indicate when they experience any technical difficulties.

Third, the study should be fielded among participants who use the same language, or carefully instruct users to change the language settings of their phones to match the language that the script is developed for, in this case English and Dutch, although additional languages can be easily added to the script. These instructions can be provided during the online or offline intake meeting. Finally, it will be important to develop a method that anonymizes video content before it is stored long term on researcher's cloud space. Until that is possible, to ensure the privacy of participants researchers using this method should instruct respondents to change the name on their phones for the course of the study. If this step is not followed, the uploaded videos are potentially identifiable. In addition, participants should be instructed to re-record their screen recording when a pop-up notification appears including personal information, for example from WhatsApp.

Conclusion

The present study introduced a new, transparent, data collection and processing method to gain accurate app usage data based on the iOS Battery Section. Although the data donation approach used in this study required additional steps from the participants, the compliance rate of participants was high. Moreover, this study provided first evidence that digital data donations do not change the studied behavior of the participants. This highlights the usability of this and similar data donation approaches for research. Moreover, the developed Python script makes the processing of the data donation videos easy and fast, and thus will enable future researchers to use this method without additional work for processing the data manually. We, therefore, hope that the presented approach will encourage more researchers to use this or similar approaches to gain accurate smartphone use data in an ethical way.

Declaration of Conflicting Interests

The author(s) 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: This study was partly funded by the Research Priority Area Communication and its Digital Communication Methods Lab at the University of Amsterdam as well as the University of Amsterdam Data Science Centre

Notes

1. https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tfl_detection_zoo.md
2. https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tfl_detection_zoo.md
3. <https://pypi.org/project/pytesseract/>
4. In total, 125 participants participated in the study. For this paper, only videos were processed that contained English or Dutch versions of the Battery Section ($N = 544$). The remaining videos based on other languages were not processed, but the script can be adapted to detect other languages as well.
5. The Python script can be found in the supplemental materials and on the OSF (<https://osf.io/48dhu/>).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., & Zheng, X. (2016). *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. arXiv preprint arXiv:1603.04467
- Anderson, M., & Jiang, J. (2018). *Teens, social media & technology 2018*. Pew Research Center. <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/>
- Araujo, T., & Neijens, P. (2020). Unobtrusive measures for media research. In J. van den Bulck (Ed.), *The International Encyclopedia of Media Psychology* (Vol. 3). Wiley. <https://doi.org/10.1002/9781119011071.iemp0049>.
- Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, 11(3), 173–190. <https://doi.org/10.1080/19312458.2017.1317337>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Boase, J. (2016). Augmenting survey and experimental designs with digital trace data. *Communication Methods and Measures*, 10(2–3), 165–166. <https://doi.org/10.1080/19312458.2016.1150975>
- Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). *Digital trace data collection through data donation*. ArXiv. <https://arxiv.org/abs/2011.09851>
- Boonstra, T. W., Nicholas, J., Wong, Q. J., Shaw, F., Townsend, S., & Christensen, H. (2018). Using mobile phone sensor technology for mental health research: Integrated analysis to identify hidden challenges and potential solutions. *Journal of Medical Internet Research*, 20(7), Article e10131. <https://doi.org/10.2196/10131>
- Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*, 22(11), 2058–2080. <https://doi.org/10.1177/1461444820924622>
- Choi, S. (2020). When digital trace data meet traditional communication theory: Theoretical/methodological directions. *Social Science Computer Review*, 38(1), 91–107. <https://doi.org/10.1177/0894439318788618>
- Christen, M., Domingo-Ferrer, J., Herrmann, D., & van den Hoven, J. (2017). Beyond informed consent—investigating ethical justifications for disclosing, donating or sharing personal data in research. In *Philosophy and computing* (pp. 193–207). Springer.
- Elevelt, A., Lugtig, P., & Toepoel, V. (2019). Doing a time use survey on smartphones only: What factors predict nonresponse at different stages of the survey process? *Survey Research Methods*, 13(2), 195–213. <https://doi.org/10.18148/srm/2019.v13i2.7385>
- Fukazawa, Y., Ito, T., Okimura, T., Yamashita, Y., Maeda, T., & Ota, J. (2019). Predicting anxiety state using smartphone-based passive sensing. *Journal of Biomedical Informatics*, 93, 103151. <https://doi.org/10.1016/j.jbi.2019.103151>.
- Gower, A. D., & Moreno, M. A. (2018). A novel approach to evaluating mobile smartphone screen time for iPhones: Feasibility and preliminary findings. *JMIR mHealth and uHealth*, 6(11), Article e11012. <https://doi.org/10.2196/11012>
- Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication & Society*, 22(11), 1567–1581. <https://doi.org/10.1080/1369118X.2019.1627386>
- Halpern-Manners, A., Warren, J. R., & Torche, F. (2017). Panel conditioning in the general social survey. *Sociological Methods & Research*, 46(1), 103–124. <https://doi.org/10.1177/0049124114532445>
- 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(6), 838–854. <https://doi.org/10.1177/1745691616650285>
- Howison, J., Wiggins, A., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 767–797. <https://doi.org/10.17705/1jais.00282>
- Jones-Jang, S. M., Heo, Y.-J., McKeever, R., Kim, J.-H., Moscovitz, L., & Moscovitz, D. (2020). Good news! Communication findings may be underestimated: Comparing effect sizes with self-reported and logged smartphone use data. *Journal of Computer-Mediated Communication*, 25(5), 346–363. <https://doi.org/10.1093/jcmc/zmaa009>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant key points. International. *Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Montag, C., Sindermann, C., & Baumeister, H. (2020). Digital phenotyping in psychological and medical sciences: A reflection about necessary prerequisites to reduce harm and increase benefits. *Current Opinion in Psychology*, 36(3), 19–24. <https://doi.org/10.1016/j.copsyc.2020.03.013>
- Naab, T. K., Karnowski, V., & Schlüt, D. (2019). Reporting mobile social media use: How survey and experience sampling measures differ. *Communication Methods and Measures*, 13(2), 126–147. <https://doi.org/10.1080/19312458.2018.1555799>

- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2020). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS screen time function. *Mobile Media & Communication*. <https://doi.org/10.1177/2050157920959106>.
- Oliphant, T. E. (2006) *A guide to NumPy* (Vol. 1). Trelgol Publishing USA.
- OpenCV. (2015). *Open source computer vision library*.
- Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 5, 1535–1547. <https://doi.org/10.1038/s41562-021-01117-5>.
- Pew Research Center (2019). *Mobile fact sheet*. <https://www.pewresearch.org/internet/fact-sheet/mobile/>
- Rafaeli, A., Ashtar, S., & Altman, D. (2019). Digital traces: New data, resources, and tools for psychological-science research. *Current Directions in Psychological Science*, 28(6), 560–566. <https://doi.org/10.1177/0963721419861410>
- Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., ..., & Gagneja, A. (2019). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human-Computer Interaction*, 36(2), 150–201. <https://doi.org/10.1080/07370024.2019.1578652>
- Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, 31(2), 226–235. <https://doi.org/10.1037/pas0000662>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). All our data will be health data one day: The need for universal data protection and comprehensive consent. *Journal of Medical Internet Research*, 22(5), Article e16879. <https://doi.org/10.2196/16879>
- Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., & Siegers, P. (2021). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behavior. SocArXiv <https://doi.org/10.31235/osf.io/dz93u>
- Silver, L. (2019) *Smartphone ownership is growing rapidly around the world, but not always equally*. Pew Research Center. <https://www.pewresearch.org/global>
- Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *Plos One*, 14(11), Article e0224240. <https://doi.org/10.1371/journal.pone.0224240>
- Smith, R. (2007). An overview of the tesseract OCR engine. Ninth international conference on document analysis and recognition (ICDAR 2007) (Vol. 2, pp. 629–633). IEEE.
- Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, 117(30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- Statcounter (2020). *Mobile operating system market share worldwide*. <https://gs.statcounter.com/os-market-share/mobile/worldwide>
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319828011>
- Trifan, A., Oliveira, M., & Oliveira, J. L. (2019). Passive sensing of health outcomes through smartphones: Systematic review of current solutions and possible limitations. *JMIR mHealth and uHealth*, 7(8), e12649. <https://doi.org/10.2196/12649>
- Van der Zouwen, J., & Van Tilburg, T. (2001). Reactivity in panel studies and its consequences for testing causal hypotheses. *Sociological Methods & Research*, 30(1), 35–56. <https://doi.org/10.1177/0049124101030001003>

Author Biographies

Susanne E. Baumgartner, Amsterdam School of Communication Research, University of Amsterdam, the Netherlands

Sindy R. Sumter, Amsterdam School of Communication Research, University of Amsterdam, the Netherlands

Vladislav Petkevich, Amsterdam School of Communication Research, University of Amsterdam, the Netherlands

Wisnu Wiradhany, Psychology Department, Faculty of Humanities, Bina Nusantara University.