Journal of Survey Statistics and Methodology (2026) **00**, *1–28*

PARADATA VERSUS SELF-REPORTS FOR THE COLLECTION OF TECHNICAL DATA ABOUT SMARTPHONES: TRADE-OFFS BETWEEN COMPLETENESS AND ACCURACY OF DATA ON SMARTPHONE OPERATING SYSTEM VERSIONS

JIM VINE *
JONATHAN BURTON
MICK P. COUPER
ANNETTE JÄCKLE

Smartphones play a central role in everyday life and offer significant potential for survey data collection. However, their usefulness may be constrained by the smartphone's operating system (OS) and its version, which determine app compatibility. This article investigates methods for accurately determining smartphone OS version, an important factor when evaluating device compatibility for data collection tools. We compare three approaches: (1) passive collection of paradata (user agent strings; UASs) from web respondents, (2) self-reported smartphone make and model matched to OS version information from an online database, and (3) direct self-reporting of OS version by respondents following step-by-step instructions. Using data

JIM VINE is a Senior Research Officer, JONATHAN BURTON is an Associate Director for Surveys, and ANNETTE JÄCKLE is Professor of Survey Methodology, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK. MICK P. COUPER is a Research Professor, Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48106-1248, USA.
We use data from *Understanding Society*: The UK Household Longitudinal Study, including the *Understanding Society* COVID-19 study. *Understanding Society* is an initiative funded by the UK Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research (NatCen) and Verian (formerly Kantar Public). Fieldwork for the *Understanding Society* COVID-19 study was carried out by Ipsos MORI. This study design and analysis were not preregistered. We acknowledge that all research has limitations and that unmeasured error may remain following our analyses. We thank the Associate Editor and two anonymous reviewers for their helpful comments on earlier versions of this article. This research was supported by the Economic and Social Research Council (ESRC) [grant numbers: ES/S007253/1, ES/T002611/1, and ES/Y003071/1]. *The Understanding Society* COVID-19 study was funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). The authors declare no conflicts of interest.
*Address correspondence to Jim Vine, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK; E-mail: jim.vine@essex.ac.uk.

https://doi.org/10.1093/jssam/smag023

from the UK Household Longitudinal Study COVID-19 web survey, a probability sample of UK households, we assess the completeness and accuracy of each method. Our findings suggest the paradata were in some respects inferior to the methods based on self-reports. First, the UAS data only provided valid smartphone OS version data for about 50 percent of respondents; the other 50 percent did not use a smartphone to complete the web survey (compared to 90 percent and 71 percent of valid cases based on methods (2) and (3)). Second, the subset of respondents who completed the survey using their smartphone was not representative of the full set of respondents, while the respondents for whom OS version data were obtained from self-report methods were broadly representative. Third, the UAS data underrepresented older OS versions. Respondents with older smartphones seem unlikely to use them to complete the survey. Our findings also show differences between OSes, suggesting it was easier for iPhone users to provide relevant and valid information than for Android users: they were more likely to report a valid make and model that could be linked to technical data, to report the OS version, and to use their smartphone to complete the survey.

KEY WORDS: Mobile apps; Smartphone compatibility; User agent string.

Statement of Significance

Survey researchers are increasingly using smartphones to collect additional data, whether passively (e.g., GPS, activity trackers) or actively (e.g., diary apps, EMA). To effectively build the tools to maximize the number of participants and minimize bias, researchers need to know what kinds of smartphones survey respondents have and what their capabilities are. The user agent string (UAS) is a reliable method of identifying the smartphone operating system and version, but only works if respondents are actually using their smartphone to complete the survey. We examine two alternative self-report approaches to eliciting this information: one method asked respondents directly for the OS version, the other method asked respondents for the make and model of their smartphone, and we linked this information to an online technical database. We evaluate the three methods using data from a UK probability household panel. Our findings suggest the UAS data are inferior in terms of completeness of the OS version data, representativeness of respondents for whom valid data are obtained, and coverage of older OS versions. Our findings also suggest it is easier for iPhone users to provide relevant and valid information across all methods than it is for Android smartphone users.

1. INTRODUCTION

Smartphones are of substantial importance, both in people's everyday lives and as potential research tools. Survey researchers are increasingly making use of data collected passively or actively from smartphones to supplement or replace survey measurement. However, smartphones use a variety of operating systems (OSes) and come in a wide array of makes and models. Designers of app-based research tools are faced with decisions about which OS(es) and versions they design for. Designers of apps for general public use are less concerned about broad coverage of their app and often design for the latest versions of the major OSes (Android and Apple's iOS), whether for cost reasons or to make maximal use of the latest features of the devices and OSes. To give one example, the COVID-19 test and trace apps developed in many countries in 2020 relied on Bluetooth-based Exposure Notification technology implemented only in iOS version 13.5 (Apple Inc. 2025) and Android version 6.0 or higher (Google 2020). Substantial parts of the population were excluded from using the test and trace apps through not having compatible smartphones.

Researchers are generally more concerned about the potential sample losses arising from such restrictive decisions. Further, *differential* sample loss not only affects precision and limits subgroup analyses but also increases the risk of bias in making inferences to a broader population. Consequently, it is important to understand what mobile devices survey respondents or research participants may have. If participants are using their smartphone at the time, this information could potentially be derived using paradata. However, not all participants may be using their mobile device to complete the survey, and it may not be practical to ask respondents to switch devices to complete the survey, so some form of self-report may be necessary to obtain this information.

In this article, we examine the practical problem of measuring which smartphone OS version people have in surveys where not everyone is using their smartphone to complete the survey. We compare paradata collected passively from web respondents with two methods based on respondents' self-reports. In one method, respondents reported the make and model of their smartphone, and we matched this to information about OS versions obtained from an online database. In the other, respondents were given instructions on how to find their OS version and asked to report it directly in the survey. Knowing what smartphone OSes and versions survey respondents and research subjects have is an important first step in successfully implementing app-based studies. Understanding differences in such access is critical for identifying potential selection biases that may threaten inferences from such studies to broader populations. This article also has practical implications for how best to measure these smartphone features.

2. BACKGROUND AND RESEARCH QUESTIONS

For surveys with a substantive goal of understanding the availability and adoption of technology in a target population, the smartphones in use can be an important question. During the COVID-19 pandemic, a salient aspect of this question arose from many nations' deployment of app-based digital contact tracing. These apps were made available for the two dominant smartphone OSes—Android and iOS (iPhones)—and had OS version requirements on each system (Health and Social Care Northern Ireland 2020; NHS Scotland 2020; NHS Test and Trace n.d.). At a population level, understanding how widely adopted these OS versions were could facilitate understanding of overall potential coverage of the apps; at an individual level, understanding patterns of (non)adoption could facilitate more nuanced understanding of any bias in coverage. Estimates for the UK suggest that in autumn 2020, about 26 percent of the population did not have a smartphone that was compatible with the COVID-19 test and trace app. Of these, most had no smartphone (about 17 percent of the total), some had a smartphone with an OS other than Android or iOS (about 2 percent), and about 7 percent had too old a version of Android or iOS (Vine et al. 2025, table 1).

The question of smartphone adoption is also relevant for survey designers considering new methods of data collection. The use of smartphone apps can open up the potential to collect new forms of data or to collect data in new ways that would be impractical or impossible with traditional survey questions (Struminskaya et al. 2020, 2021). Examples include travel surveys using passive collection of location-based data from smartphones (Abdulazim et al. 2013; Greaves et al. 2015; Assemi et al. 2018; Elevelt et al. 2021; Harding et al. 2021), data collected from health-related app use (Chen et al. 2016; Bach and Wenz 2020), activity tracking (Keusch et al. 2022), expenditure apps (Rodenburg et al. 2025), and ecological momentary assessment, also referred to as experience sampling or ambulatory assessment (Hofmann and Patel 2015; Fortea et al. 2023; Hammoud et al. 2024).

When considering developing or deploying a smartphone app, the survey designer might face decisions about what OS versions to target: requiring a relatively high minimum version may enable the use of more advanced features of the OS and may be easier than developing or maintaining an app that works on a wider range of OS versions. But that would come at a cost of excluding potential respondents with older OS versions from participation in the data collection. Consequently, it can be important to know who would be covered by a given OS version requirement when surveys make these sorts of app deployment decisions.

In surveys, there are several possible ways of measuring respondents' smartphone OS versions. For respondents completing the survey on their smartphone, the OS version can be derived from paradata (Kreuter 2013;

McClain et al. 2019). For respondents completing the survey online, the web browser provides a "user-agent string" (UAS), which includes a description of the OS and OS version of the device being used (Callegaro 2010, 2013; Ramli et al. 2014; Kline et al. 2017).

Previous scholars have treated UAS data as the "ground truth" when investigating methods of identifying the devices people are using to connect to various services (Hagos et al. 2020; Mulazzani et al. 2013) or treated UAS data preferentially when determining disagreements between methods (Lastovicka et al. 2018). UAS data has also been used in the context of survey studies to identify devices used (Callegaro 2010; Décieux 2024; Décieux and Sischka 2024). Although UAS data are passively collected, so are not affected by reporting errors, and are therefore considered high quality in some respects, these paradata will only provide the relevant information for respondents who complete the survey on their smartphone. Any surveys with in-person, telephone, or paper self-completion questionnaires will inherently not collect any UAS data for respondents completing the survey via those modes. Even for web-only surveys, some respondents complete the survey using non-smartphone devices (PCs, laptops, tablets, etc.), so their UAS paradata might provide information about the device used for the survey, but not their smartphones. The absence of smartphone paradata for a respondent does not distinguish between those having smartphones but choosing to respond on a different device and those without smartphones. Concerns have also been voiced about the accuracy and reliability of parsing browser details from UAS data (Nandakumar et al. 2023).

Rather than collect the UAS, the OS version could be determined by asking respondents. While respondents may be able to report their smartphone OS, it is less likely they will know the particular OS version. In the cognitive response process model (Tourangeau et al. 2000), the latter is unlikely to be encoded in memory and hence is unavailable for recall. Given this, the question could include instructions on how to look up the OS version on the phone. Alternatively, respondents might be asked for relevant information they are more likely to know, such as the make and model of their smartphone, and the researcher could link the relevant technical data from other sources. These approaches might provide more complete data, covering larger parts of the sample, but might be less accurate. In this article, we examine both these approaches and compare trade-offs with the paradata in terms of data completeness and accuracy. We examine accuracy both at the individual level (do the measures derived from paradata and self-reports match?) and at the aggregate level (are the estimated prevalences of different OS versions plausible?). Most existing uses of UAS are uncritical and do not evaluate the quality of the data derived from the UAS. We have found no other studies comparing UAS to other sources of information on device characteristics, including self-reports.

With this background in mind, we address the following research questions:

RQ1: How do paradata and self-reports compare in terms of completeness of the data and reasons for missingness?

The first question addresses the completeness of the information obtained. Inherently, the UAS is uninformative if the respondent is not completing the survey on a smartphone. But self-reports may also be missing. Completeness of information is a necessary but not sufficient condition for determining potential coverage of smartphones used by survey respondents.

RQ2: How do paradata and self-reports compare in terms of representativeness of the sub-samples for whom valid data are obtained?

RQ2 has implications for differential coverage or selection bias and addresses the question of potential bias if a mobile app subsample was restricted to those with valid data on smartphone suitability. Information about smartphones does not guarantee compatibility with the particular app being used or that respondents would in fact install and use such an app. But being able to determine smartphone types is an important first step in guiding app design to mitigate selection biases.

RQ3: How do conclusions about the prevalence of different operating system versions differ between methods?

This research question focuses on aggregate-level estimates of OS versions obtained by the different methods. That is, would using one method over another yield substantially different estimates of the proportion of respondents with a suitable smartphone (however defined)?

RQ4: How do measures from the paradata compare to those from self-reports, for the same respondents?

This is related to RQ3 but focused on individual differences in the information obtained for the subset of respondents with both paradata and self-reported measures of smartphone OS versions. We do not make assumptions about which method represents "ground truth" but rather examine the consistency of information obtained through the different approaches.

From a total survey error perspective (see, e.g., Groves et al. 2009), the primary focus of this article is on coverage error, that is, the potential loss of sample from device restrictions of research apps. More specifically, this article is about how to measure device coverage in a survey context. But the article is also about measurement error: how do different approaches to measuring the potential coverage error differ in their completeness and accuracy?

3. DATA AND METHODS

3.1 Survey and Analysis Sample

We use data from the *Understanding Society* COVID-19 study, a general population study conducted during the COVID-19 pandemic (University of Essex, Institute for Social and Economic Research 2021), supplemented with data from the annual interviews (University of Essex, Institute for Social and Economic Research 2024). Members of the main panel of *Understanding Society*, the UK Household Longitudinal Study, were invited to complete online questionnaires, initially every month for four waves (April to July 2020), and then every two months for four waves (September 2020 to March 2021), with a final ninth wave in September 2021. At waves 2 and 6, some web nonrespondents were invited to complete the survey by telephone.

The study carried questions at waves 6 and 7 (fielded 24 November 2020 to 1 December 2020 and 27 January 2021 to 3 February 2021, respectively) designed to investigate the capabilities of respondents' smartphones. At the time, UK governmental bodies were deploying smartphone apps for COVID tracking, and these apps had certain OS version requirements.

At wave 6, 19,280 sample members were invited to complete the survey online. Of these, 11,802 respondents completed the full survey, for an overall response rate of 61 percent (AAPOR RR1) (American Association for Public Opinion Research 2023). At wave 7, 19,105 sample members were invited to complete the survey online. Of these, 11,797 respondents completed the full survey, for an overall response rate of 62 percent (RR1) (Institute for Social and Economic Research 2021).

Our analysis sample is restricted to respondents who completed waves 6 and 7 online ($n = 10,359$). Of these, we exclude cases where one or more of the covariates used in our analysis was missing, comprising, with some overlap: missing birth year data ($n = 1$); missing race/ethnicity data ($n = 81$); missing household composition (couple status) data ($n = 1$); missing employment status data ($n = 15$); missing income data ($n = 117$); missing qualification data ($n = 371$); and missing housing tenure data ($n = 525$). In total, 973 cases were excluded due to these missing data, leaving an analysis sample of 9,386.

3.2 Methods of Collecting OS Version Data

The two waves adopted different approaches to collect measures of smartphone OS versions. At both waves, the online survey software collected UAS paradata, and respondents were asked whether they had a smartphone, and, if so, whether it was an Android, iPhone, or other type of smartphone. At wave 6, respondents who reported having a smartphone were asked to provide the make and model of the smartphone via a question implemented as a trigram search, whereby respondents would start typing their response, and close matches from a pre-loaded list of options would be available for selection. At wave 7, respondents who reported having Android or iPhone smartphones were routed into questions asking them to provide the OS version, with information on how to find it. The full wording of both sets of questions can be found in Appendix 1 in the supplementary materials online. Both self-report methods enabled us to identify respondents who did not have smartphones, and we included this information in the analyses for RQ1 and RQ2, since identifying respondents who do not have a device can be as important as identifying the OS version of their device. The analyses for RQ3 and RQ4 are based solely on respondents with valid OS version measures, and those without smartphones are excluded.

3.2.1 OS versions derived from self-reported smartphone make and model (wave 6).

For the smartphone make and model data collected at wave 6, for each smartphone reported, we sought technical specifications from GSMArena.com, a website that publishes structured information on a large number of smartphone models, including their OS versions. For each smartphone model reported, we conducted an automated search of GSMArena and obtained information on OS versions, supported by manual searches of GSMArena and other online sources. More detail on the process for deriving the measures is provided in Appendix 2 in the supplementary materials online. The analysis set contains 665 unique values of non-missing responses for the smartphone model question at wave 6, from 8,170 respondents. Of these, 630 reported models (from 7341 respondents) were matched to identifiable Android or iPhone smartphones.

For many smartphones, the version information specified a range of versions, describing both the version of the OS supplied on the phone when it was released (minimum version) and the highest version it could be upgraded to (maximum version).

3.2.2 OS versions from respondent self-reports (wave 7).

The question asking respondents to report their OS version in wave 7 varied depending on whether they had an Android or iPhone smartphone and whether they were completing the survey on their smartphone. The questionnaire script used information from browser sniffers to detect whether the respondent was responding on a smartphone. If the sniffer could not detect this, a question was triggered asking the respondent. If they were using their smartphone, they were given instructions on where to find the OS version in their smartphone settings. If they were not using their smartphone to complete the survey, they were given several suggestions: (1) scanning a QR code with their smartphone to open the website whatsmyos.com on their smartphone to display their OS version; (2) typing whatsmyos.com into their smartphone browser with the same result; or (3) checking the settings of their smartphone.

3.2.3 OS versions derived from UASs (waves 6 and 7).
UASs are notoriously difficult to parse consistently (Rudis 2022; Nandakumar et al. 2023). They are inconsistently formatted, and extra processing is required to extract information from them on details like OS versions. In Appendix 2 (please see the supplementary materials online), we provide examples of UASs, details of our parsing process—including details of the parsing software used (Roßmann and Gummer 2020; Roßmann et al. 2020; Rudis et al. 2020; Vine 2021; Rudis 2022)—, and information on spot-checking we conducted to assess the parsing.

For respondents who completed both surveys with Android smartphones or iPhones, we have UAS data from both waves. Appendix 4 (please see the supplementary materials online) documents the extent of changes at the respondent level. For both types of phones, 99 percent of respondents completed wave 7 using smartphones with the same OS as wave 6. Among wave 6 Android smartphone users, 88 percent had the same OS major version in the latter wave; among iPhone users, 91 percent retained the same OS major version.

3.3 Analysis Methods

For RQ1, how paradata and self-reports compare in terms of completeness of data and reasons for missingness, we provide descriptive tables of numbers and proportions of respondents for whom each approach produces valid data about smartphone OS versions.

For RQ2, how paradata and self-reports compare in terms of representativeness of the sub-samples for whom valid data are obtained, we calculate Coefficients of Variation (CVs) of the propensities of producing valid OS version data using the RISQ suite (de Heij et al. 2015) in R (R Core Team 2024). For each method, we estimate logit models using the indicator of whether a valid OS version (including indications of not having a smartphone) was obtained as the dependent variable and a set of characteristics as predictors. The characteristics and their categorizations are: whether they are in a couple (yes, no); housing tenure (own outright, own with mortgage, private rented, social/council rented, other); whether they have a long-term health condition (yes, no); employment status (employed, self-employed, both employed and self-employed, neither employed nor self-employed); highest qualification (no qualifications, secondary school qualifications, degree, other); household income (quintile bands); ethnicity (white, non-white ethnicities); sex (male, female); and age (birth decade pre-1950s, 1950s, 1960s, 1970s, 1980s, post-1980s). Table S1 (please see the supplementary materials online) documents the distributions of these covariates. The data for the characteristics are derived from a combination of respondents' annual interviews for their main Understanding Society surveys and various waves of the COVID study—see Appendix 3 in the supplementary materials online for details. Although our analyses focus on the estimated CVs, we also document results from the underlying logit models. We present both the estimated logit coefficients and average marginal effects, as the latter are more directly interpretable.

Based on these models, we calculate the predicted probability of having a valid OS version (or indication of not having a smartphone) for each respondent and each method. We then calculate the overall CV for these predicted probabilities, by method, as:

$$\widehat{CV}(p_{\boldsymbol{x}}) = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\widehat{p}_i - \widehat{\overline{p}}\right)^2}}{\widehat{\overline{p}}}, \tag{1}$$

where n indicates sample size, $\boldsymbol{x}$ the set of predictors in the model, $\widehat{p}_i$ the predicted probability that respondent i has a valid OS version, and $\widehat{\overline{p}}$ is the average predicted probability. The numerator is the standard deviation of the predicted probabilities. The less the predicted probabilities differ between sample members, the smaller the overall CV. That is, the smaller the CV for a particular method, the more representative the respondents for whom a valid OS version is obtained are, in terms of the predictor variables included in the model. Standard errors for the CVs were calculated using the same software, which implements an analytic approximation for the standard errors (see de Heij et al. 2015, p. 19).

For RQ3, how the prevalences of different OS versions differ between methods, we present descriptive statistics of the proportions of respondents with each version of smartphone OS, separately for Android and iOS.

For RQ4, how measures from the paradata compare to those from the self-reports, we examine those cases where we have values for respondents' OS versions from both the paradata and a survey question. We make these comparisons within a wave, that is, comparing OS versions derived from self-reported smartphone make/model to UAS paradata at wave 6 and comparing self-reported OS version to paradata at wave 7. For the wave 6 questions, in many cases we obtain a range of possible smartphone OS versions, bounded by the minimum OS version for the model (the version with which it was released) and the maximum (the highest version it can be upgraded to). The problem here is that the responses provide no indication of whether respondents have kept their OS versions updated. That is, the responses do not tell us where in the possible range their actual OS version lies. Consequently, we conduct three comparisons in this case:

- Comparing the "min" OS version to the paradata;
- Comparing the "max" (highest) OS version to the paradata;
- Comparing the "min–max range" of possible OS versions to the paradata, declaring agreement if the paradata OS version falls anywhere in that range.

The last of these is of theoretical interest, as it provides a measure of whether the method generates ranges compatible with the UAS data. However, for most use cases, researchers would have to operationalize data gathered via a method like this by collapsing range results down to a single value for each respondent. The most available methods for generating a single value from a range are to treat either all devices as remaining at their release level and receiving no upgrades, or to treat all devices as being fully upgraded to their highest possible OS version. Consequently, we conduct agreement checks on the basis of these options. We also calculate measures of agreement between methods using Cohen's Kappa.

4. RESULTS

RQ1: How Do Paradata and Self-Reports Compare in Terms of Completeness of the Data and Reasons for Missingness?

Table 1 shows the numbers and percentages of respondents for whom valid measures of their OS version were obtained, by method. The table also documents the different reasons for missing information, and how they vary between methods. Respondents identified as not having a smartphone are included as having a "valid" measure for the purposes of this analysis.

The self-reported make and model of the smartphone produced the most complete data. A valid measure was obtained for 90 percent of the sample; missing values are mainly because the reported make and model could not be matched to information about the OS version. This included respondents who had reported only the brand or a series of smartphones but not the specific model (e.g., "Samsung J") or had provided information that could not be coded.

Asking respondents to self-report their smartphone OS version was less successful: valid data were only obtained for 71 percent of respondents. Nearly all the missing information with this method is due to item non-response (28 percent of respondents). Most of this item non-response

Table 1. Completeness of Operating System Version and Reasons for Missingness, By Method

	Self-reported smartphone model (wave 6)		Self-reported OS version (wave 7)		UAS (wave 6)		UAS (wave 7)	
	n	%	n	%	n	%	n	%
Non-missing	8,458	90	6,666	71	4,793	51	4,701	50
Comprising. . .								
Android version	3,613	38	2,574	27	2,313	25	2,211	24
iOS version	3,728	40	2,779	30	2,400	26	2,399	26
No smartphone	1,068	11	1,096	12	0	0	0	0
Other smartphone type	49	1	217	2	80	1	91	1
Missing	928	10	2,720	29	4,593	49	4,685	50
Comprising. . .								
Invalid value	791	8	59	1	0	0	0	0
Item nonresponse	137	1	2,661	28	0	0	0	0
Non-coverage: completing on PC or tablet	0	0	0	0	4,593	49	4,685	50

consisted of “Don’t know” answers (79 percent), with a further 20 percent “Prefer not to say,” and the remainder (1 percent) either missing or inapplicable.

The UAS data produced valid data about smartphone OS versions for all respondents completing the survey on a smartphone (51 percent in wave 6 and 50 percent in wave 7). For those who completed the survey on a PC, laptop, or tablet, this information is missing (49 percent and 50 percent, respectively).

Table 2 shows the completeness of OS version data by method, separately for those reporting having Android versus iPhone smartphones, and excludes respondents without smartphones or having other smartphone types. For each data collection method, respondents are grouped by the OS they reported in the survey when asked whether they had a smartphone and what type it was. In wave 6, 4,163 respondents reported having an Android smartphone and 4,051 an iPhone. In wave 7, 4,125 respondents reported having Android smartphones, and 3,933 iPhones.

All methods seemed to work better, in the sense of producing higher rates of valid OS version data, for iPhones than Android smartphones. When asked to report their smartphone make and model, 92 percent of iPhone users provided information that was successfully matched with

Table 2. Completeness of OS Version Data by Method, By Reported Smartphone Type (Android/iPhone)

	Self-reported smartphone model		Self-reported OS version		UAS wave 6		UAS wave 7	
	Android ($n = 4163$)	iPhone ($n = 4051$)	Android ($n = 4125$)	iPhone ($n = 3933$)	Android ($n = 4163$)	iPhone ($n = 4051$)	Android ($n = 4125$)	iPhone ($n = 3933$)
	%	%	%	%	%	%	%	%
Valid Android version	84	1	62	0	51	1	49	1
Valid iOS version	0	92	0	71	1	58	1	59
Other smartphone type[a]	0	0	0	0	0	1	0	2
Invalid value	14	5	0	1	0	0	0	0
Item nonresponse	2	2	37	28	0	0	0	0
Non-coverage: completing on PC or tablet	0	0	0	0	48	40	50	39

[a]The 1 percent and 2 percent of self-reported iPhone users identified as having other smartphones are higher than expected. Manual inspection, described in Appendix 2 in the supplementary materials online, suggests some of these should be classified as non-coverage due to completing on a PC.

the technical data, compared to 84 percent of Android smartphone users ($\chi^2(1) = 112$, $p < .001$). Android smartphone users were more likely to provide non-valid answers to the make and model question (14 percent) than iPhone users (5 percent). When asked to report their smartphone's OS version, 71 percent of iPhone users provided valid information, compared to 62 percent of Android smartphone users ($\chi^2(1) = 61$, $p < .001$). Even the UAS data are more complete for users of iPhones than of Android smartphones. They produced OS version data for 58/59 percent of iPhone users in waves 6/7, compared to 51/49 percent of Android smartphone users ($\chi^2(1) = 33$, $p < .001$; $\chi^2(1) = 74$, $p < .001$). This suggests iPhone users were more likely to complete the survey on their smartphone than Android smartphone users.

RQ2: How Do Paradata and Self-Reports Compare in Terms of Representativeness of the Sub-Samples for Whom Valid Data Are Obtained?

The differences in CVs estimated for each method are striking. The first row in table 3 is based on all respondents for whom a valid measure is obtained. For the self-report methods, this includes respondents for whom we identify that they do not have a smartphone, as identifying that a respondent does not have a smartphone at all is just as important for most practical purposes as identifying the OS version if they do have one. For both methods based on self-reports, the CVs are comparatively close to zero: 0.048 for the method based on self-reported smartphone make and model and 0.083 for the self-reported OS version. That is, for these two methods, the characteristics of those for whom valid OS version measures are obtained are relatively similar to the characteristics of the full set of respondents. The probabilities of having a valid measure do not vary much between respondents with different characteristics. In contrast, the CVs for the UAS-based measures are 0.437 and 0.439 for waves 6 and 7, respectively. This shows that respondents who used their smartphone to complete the survey were not representative of the overall sample of respondents in terms of the characteristics included in the propensity models.

The second row in table 3 shows CVs calculated using only the acquisition of an Android or iOS smartphone OS version as a valid outcome. That is, the 12–14 percent of cases where respondents are identified by the self-report measures as having neither an Android smartphone nor an iPhone are classified as not having valid data; for the 1 percent of UAS methods, only those cases identified as having another smartphone type change classification as the UAS data does not identify respondents who do not have a smartphone. In this comparison, the self-reports perform less well,

Table 3. Overall Coefficients of Variation in Probabilities of Producing a Valid OS Version Measure, By Method

	Self-reported smartphone model—wave 6	Self-reported OS version—wave 7	UAS—wave 6	UAS—wave 7
Any valid measure obtained	0.048 (0.004)	0.083 (0.007)	0.437 (0.009)	0.439 (0.009)
Valid OS version number obtained	0.205 (0.006)	0.248 (0.008)	0.451 (0.009)	0.456 (0.009)

Note.—Coefficients of variation with associated standard errors in brackets. Tables S6–S9 in Appendix 5 in the supplementary materials online document the associated logistic regression models' coefficients and associated statistics, along with average marginal effects estimates from those logistic regression models.

but still better than the UAS measures. As expected, it has a much smaller impact on the CVs of the UAS methods.

The logistic regression models from which the first set of CVs are calculated (including respondents identified as not having a smartphone as a valid outcome) are documented in tables S6 to S9 in Appendix 5 in the supplementary materials online. Across methods and controlling for other characteristics, age is most consistently associated with whether a valid measure of OS version was obtained. For all methods, the probability of obtaining a valid measure decreases with age, though the age gradient is stronger for the UAS measures than the self-reports. Other associations are discussed in Appendix 5 in the supplementary materials online.

RQ3: How Do Conclusions about the Prevalence of Different OS Versions Differ between Methods?

Figures 1 and 2 show the distributions of OS versions obtained with the different methods, separately for Android smartphones and iPhones. For the make and model matched to technical data, the maximum OS version is shown. The counts and percentages underlying figures 1 and 2 are available in tables S10 and S11 in Appendix 6 in the supplementary materials online.

For Android smartphones, the self-reported OS version produces a much higher rate of older versions: 18 percent report having OS versions 1–5. With the self-reported make and model, only 4 percent have OS versions 1–5, with the UAS data only 1 percent. This pattern is also visible for iPhones, though less strong: 4 percent self-reported OS versions 5–11,

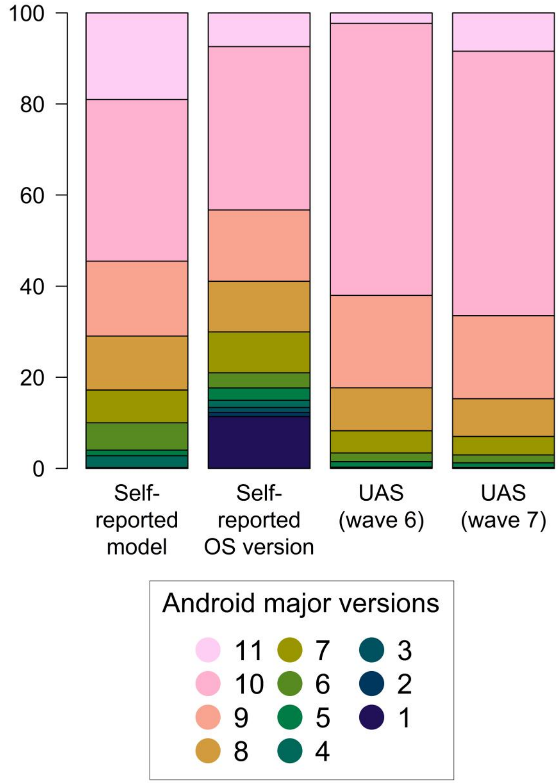

Figure 1. Distribution of OS Versions Obtained by Method (Android Smartphones).

compared to 3 percent based on the make and model, and 1 percent in the UAS data. The differences between the self-reported OS versions and the UAS data might, in part, be due to respondents with older smartphones being less likely to use their phone to complete the survey.

RQ4: How Do Measures from the Paradata Compare to Those from Self-Reports, for the Same Respondents?

Tables 4 and 5 report the individual-level agreement between the OS version measured by the UAS and each of the two methods based on self-reports, separately for Android smartphones and iPhones. For each table, the analyses are restricted to respondents for whom both types of OS version data are available. Each comparison only uses data from the same wave: the OS version derived from self-reported smartphone make and model is compared to the UAS measure from wave 6; the self-

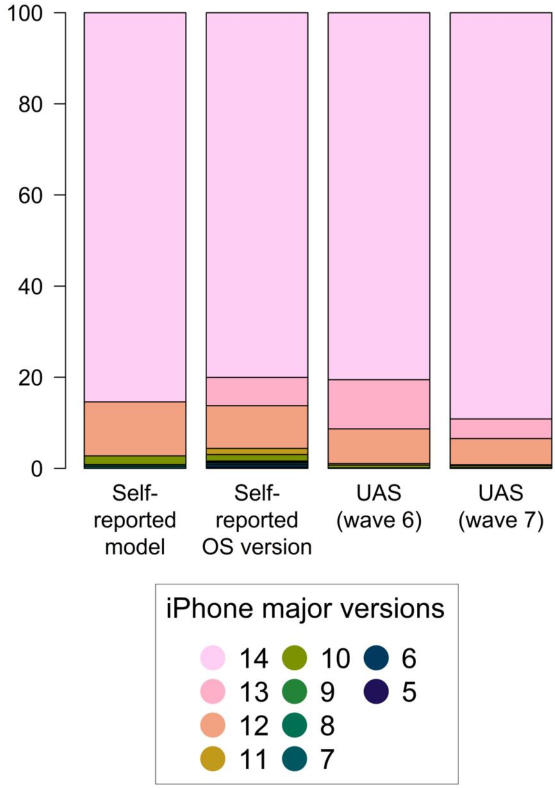

Figure 2. Distribution of OS Versions Obtained by Method (iPhones). Because Apple's upgrade approach tends to upgrade generations of iPhones to the maximum iOS version, certain OS versions were not possible values for the self-reported model method based on maximum OS version. iPhone 5 and iPhone 5c had maximum OS version 10. The next newest phones, iPhone 5s and iPhone 6/6 Plus, had maximum OS version 12. All iPhones newer than those could go up to iOS14 at the time. No phones had iOS 11 or 13 as their maximum OS version. Consequently, the first bar in this figure does not include any iPhone OS version 11 or 13 results.

reported OS version is compared to the UAS measure from wave 7. Looking at the self-reported make and model for Android phones (table 4), when each model's minimum OS version is used, only 18 percent of cases match the OS version derived from UAS data. When the maximum version is used, 50 percent match, and when the range of possible OS versions is used, 71 percent match. Comparing the self-reported OS version to the UAS data, 73 percent of cases match. For all comparisons, the most likely discrepancy is the version detected by the UAS being higher than the

Table 4. Agreement of OS Version Between UAS and Self-Report Methods (Android)

Agreement	Self-reported smartphone model: Min OS version *n*	%	Max OS version *n*	%	Min–max range *n*	%	Self-reported OS version *n*	%
Agree with UAS to at least major level	349	18	972	50	1,379	71	890	73
OS different	26	1	26	1	26	1	18	1
OS major version > UAS	36	2	443	23	36	2	54	4
OS major version < UAS	1,539	79	509	26	509	26	251	21

Note.—"major" refers to the first number in the OS version string.

Table 5. Agreement of OS Version Between UAS and Self-Report Methods (iPhone)

Agreement	Self-reported smartphone model: Min OS version *n*	%	Max OS version *n*	%	Min–max range *n*	%	Self-reported OS version *n*	%
Agree with UAS to at least major level	128	6	1,811	81	2,069	93	1,527	91
OS different	16	1	16	1	16	1	14	1
OS major version > UAS	16	1	274	12	16	1	37	2
OS major version < UAS	2,074	93	133	6	133	6	78	5

Note.—"major" refers to the first number in the OS version string.

version derived from the self-report method. There are also a few cases (1 percent for each comparison) where the OS detected by the UAS does not match the self-reported OS.

For iPhones (table 5), the agreement rates are higher than for Android smartphones, except when using the models' minimum OS version, which only matches in 6 percent of iPhone cases. When maximum OS versions matched to makes and models are used, 81 percent of cases match the UAS data; when the full ranges of possible OS versions are used, 93 percent match. When self-reported OS versions are compared to UAS data, 91 percent of cases match. The UAS data only imply higher versions than the self-report for 5 percent of cases, and lower versions for 2 percent.

Tables 6 and 7 contain estimated agreement rates between methods based on Cohen's Kappa, separately for Android and iOS. Agreement rates between the two self-report measures are higher for iOS than Android, as are agreement rates between each self-report measure and the UAS measure collected in the same wave. However, the agreement rate between the two UAS measures is lower for iOS than Android. This suggests the OS version was more likely to be updated for iOS users than Android users during the two-month interval between waves. Comparing the different measures at the individual level suggests the self-report measures obtained

Table 6. Cohen's Kappa for Each Pair of Methods (Android)

Measure	Self-reported OS version (wave 7)	UAS (wave 6)	UAS (wave 7)
Self-reported smartphone model (wave 6)[a]	0.455	0.564	0.534
Self-reported OS version (wave 7)	–	0.488	0.621
UAS (wave 6)	–	–	0.806

[a]For each comparison involving the self-reported smartphone model at wave 6, we compute the agreement based on the min–max range (declaring agreement for a value anywhere in the range) for the associated OS version number.

Table 7. Cohen's Kappa for Each Pair of Methods (iPhone)

Measure	Self-reported OS version (wave 7)	UAS (wave 6)	UAS (wave 7)
Self-reported smartphone model (wave 6)[a]	0.735	0.807	0.689
Self-reported OS version (wave 7)	–	0.422	0.637
UAS (wave 6)	–	–	0.693

[a]For each comparison involving the self-reported smartphone model at wave 6, we compute the agreement based on the min–max range (declaring agreement for a value anywhere in the range) for the associated OS version number.

from iPhone users are more consistent with the measures derived from the UAS data than those from Android smartphone users.

5. DISCUSSION

In this article, we examine the challenge of how to measure people's smartphone OS versions. Knowledge of OS versions is of interest, both as an indicator of technology adoption, and because the OS and version constrain the apps with which a smartphone is compatible. This affects not only personal uses but also potential uses of mobile applications for research purposes and data collection.

We use data from the UK Household Longitudinal Study COVID-19 survey, a probability sample surveyed online during 2020 and 2021, to compare OS version data derived from paradata with data derived from two self-report methods.

We first compare the completeness of data on OS versions collected with different methods (RQ1). Asking respondents to report the make and model of their smartphone, and linking those to data obtained from online databases, produced the most complete data: a valid OS version was obtained for 90 percent of respondents. Missing observations are mainly because of non-conformities in the answers respondents gave, for example, reporting only the make but not the model of their smartphone. These errors could possibly be reduced further through validation checks and follow-up questions within the survey. Asking respondents to self-report the OS version of their smartphone was the second most successful method in terms of completeness: a valid measure was obtained for 71 percent of respondents. Missing observations are all due to item nonresponse. This could possibly be reduced through further investigation of the reasons why respondents did not answer the question. For example, whether they had technical difficulties finding the relevant information, which could be remedied by improved instructions, or whether they thought this information was too intrusive. Although the UASs provided valid information for all respondents who completed the web survey on their smartphone, about 50 percent of respondents completed the survey on a PC, laptop, or tablet, and so have missing information for the smartphone OS version. The UAS was therefore the least successful method, though this could be improved if respondents were encouraged to complete the survey on their smartphone rather than on other devices.

For RQ2, we examined the selectiveness of respondents for whom valid measures of OS versions were obtained with the different methods. Both methods based on self-reports produced subsets of respondents that were broadly representative of all respondents, based on a set of sociodemographic and health characteristics. The subset of respondents for whom valid OS version data were obtained from the UAS data was much more selective.

Comparing the distributions of OS versions obtained with the different methods (RQ3) suggests the self-reported OS version resulted in higher prevalences of old OS versions. Some caution is suggested, especially for Android respondents, by the finding of a substantial group of respondents reporting having an OS version over a decade old at the time of the survey.

Our findings from RQ4 suggest the self-reported OS versions resulted in more accurate (when provided) measures of smartphone OS version, compared to the measures derived from the passively collected UAS data, than our attempts to derive an OS version from self-reported smartphone make/model. However, this finding should be interpreted in light of the higher proportion of item non-response for the self-reported smartphone OS version.

Taken together, our findings suggest the paradata were in some respects inferior to the methods based on self-reports for the purposes of identifying the smartphone OS version of respondents. First, UAS data only provided valid smartphone OS version data for about 50 percent of respondents. Although all respondents completed the survey online, only half completed it on their smartphone. In contrast, the self-reported make and model (wave 6) yielded valid data for 90 percent of cases, while the self-reported smartphone OS version yielded valid data for 71 percent of cases. Second, the subset of respondents who completed the survey using their smartphone was not representative of the full set of respondents, meaning valid OS version data were only obtained from the UAS data for a selected set of respondents. In comparison, the respondents for whom OS version data were obtained from the self-report methods were broadly representative of all respondents. Third, the UAS data under-represented older OS versions. It seems respondents with older smartphones were unlikely to use them to complete the survey.

Considering these features of the performance of the methods from a total survey error perspective, we can see there are tradeoffs in the likely scale and nature of error sources between the methods. In this study, the likely high accuracy of UAS-based approaches is (partially) negated by substantial coverage and selection limitations. Direct OS/OS version self-reports are more complete but more vulnerable to measurement error and retain some gaps due to non-response. Self-report of smartphone make/model exhibits fewer data gaps but instead increases the potential for error at the post-data-collection processing stage.

Although the method based on self-reported smartphone make and model produced the most complete data, it only provided the range of OS versions the device was compatible with. The actual OS version, however, also depends on users' behaviors and how regularly their devices are updated. This, in turn, is influenced by the platform ecosystem (Muchmore and Zamora 2025). Apple has strong platform control and pushes out frequent default overnight OS updates. Android updates are fragmented between smartphone manufacturers. Only a minority of Android devices run the latest version at any given time, and several earlier versions coexist with non-negligible market shares. Using the maximum assumes all devices are up to date. Some respondents will therefore be misclassified as having newer OS features than they actually do. This could bias conclusions about access to particular features and is more of an issue for Android than iOS. Using the minimum misclassifies those who do update as running older OS versions than they actually have. This underestimates OS capabilities (e.g., app compatibility, security, permissions) and is more problematic for iOS than Android users. In both cases, misclassifications are not random: they correlate with user characteristics and device ecosystems (iOS versus Android). Using both the minimum and maximum supported OS versions provides a lower and upper bound for the true value. Researchers can use these bounds for sensitivity analyses, estimating key models using the lower and upper bounds separately to check the robustness of results in “low-update” versus “high-update” scenarios. Our results for RQ4 show that for both Android and iOS devices, the maximum is more likely to match the OS version derived from the UAS than the minimum. This suggests the maximum is a better approximation of the current OS version than the minimum.

Our findings also reflect differences between OSes, suggesting it was easier for iPhone users to provide relevant and valid information than for Android smartphone users. iPhone users were more likely than Android smartphone users to report a valid make and model that could be linked to technical data; they were more likely to self-report the OS version; they were more likely to use their smartphone to complete the survey; and the self-reports were more likely to match the UAS-derived OS versions. The quality of data about smartphone OS versions therefore differs systematically between platforms. Because iPhone and Android smartphone users differ in characteristics relevant to many studies, these measurement errors are unlikely to be random. These differences between OSes are at least partly due to differences in user interface design. iOS settings and system information screens follow a tightly controlled, relatively stable design language across devices, which can make instructions on where to find information in the device’s settings easier to follow for most users. Android devices vary more in how settings are organized, including manufacturer-specific menus and icons. Generic survey instructions may therefore fit some Android users poorly, making it more difficult for them to find and report the relevant information.

The results from our analyses invite further investigations. If survey designers decide to use self-report measures rather than paradata to collect OS version data, how are those approaches best implemented? Are there groups of respondents who have more difficulty providing these data than others? What are the most prevalent types of problems? How can the methods and instructions be tailored to differentially support different groups of respondents in providing these data?

One limitation is the self-report measures were collected two months apart. While there are some changes between the two waves, these do not invalidate the overall conclusions from this study. Another limitation of the approaches used in this study is that the technical information on smartphones was accessed online after the data collection period, potentially reducing the proportion of successful matches to smartphone specification data.

At the time of data collection for this study, we found essentially no instances of smartphone UAS data that did not report an identifiable OS version. However, in the subsequent period, privacy concerns have led to browser developers increasingly constraining the identifiable information contained in UAS data. This may limit future use of UAS paradata for identifying participants' smartphone OS/OS versions. Methods based on self-reports might therefore become more important for gauging the prevalence of compatible devices in the population if mobile applications used for research have stringent OS version requirements.

5.1 Practical Implications

The use of UAS paradata for identifying smartphone OS/OS versions is only available for those completing the survey on their smartphone. We need other ways to learn about the capabilities of users' smartphones (and indeed, whether they even have a smartphone). We tried two self-report approaches, but better ways need to be found that do not tax the willingness or ability of respondents to find this information. Smartphone make/model self-reporting could be improved by ensuring in advance that all models available for respondents to select are matched to known, unique database entries, for example, using year identifiers where manufacturers repeat the same model name. While paradata alone is insufficient, some combination of paradata and self-report may be optimal.

We have also identified how mapping self-reported smartphone make/model responses to OS versions might imply mapping to a range of potential versions. There is not a single best way to map each smartphone to an OS version within its supported range, and the implications of different approaches vary between iPhone and Android devices. Our results suggest mapping a given iPhone model to its maximum OS version will often be reasonable, while using the minimum iOS version would typically be unduly conservative. For Android phones, assuming Android smartphones are all running their highest supported OS versions would often be optimistic; assuming the minimum supported OS version is in use will almost always be safe. However, even for Android phones, this will produce downward-biased estimates of the OS versions respondents' phones are running. Conservative coverage estimates may be appropriate in some circumstances while in others an optimistic assessment may be preferable. For example, when making an eligibility assessment for a survey, it might be preferable to err on the side of offering participation to some who might not be able to participate rather than risking excluding some eligible cases. In this situation, assuming everyone's smartphone is running its maximum supported OS version may be suitable. Conversely, at the point of designing an app study and specifying minimum smartphone OS requirements the app should support, taking a conservative estimate might be necessary if there is a goal of ensuring that certain people (or a certain proportion of panel members, for example) will be covered given their OS version.

The differences between our results for iPhone and Android phone users, which recur throughout our analyses, have implications for equity and differential exclusion in mobile app and mobile-first studies. Further analyses could contribute to the literature comparing the characteristics of iOS versus Android users and of users with up-to-date versus out-of-date OS versions.

There are also implications for designers of app studies. These issues are particularly important for researchers aiming at representative samples of the general population. Aiming at exploiting the capabilities of the latest smartphone OS versions may result in sample loss and (of potentially greater concern) *differential* sample loss, potentially biasing inferences from such studies. There are ways to mitigate some of these effects, for example, by providing smartphones to those without their own smartphone or one that is sufficiently updated to run the app. But knowing how many would need a provided device—and who they are—is important both for operational efficiency and to maximize participation across all key subgroups of interest to the study.

For survey researchers, while app-based studies have the potential to increase the amount of data collected without over-burdening participants, they run the risk of additional selection biases (over and above survey nonresponse or attrition). Having a compatible smartphone is a necessary but not sufficient condition for getting value from these new methods. Understanding sources of selection bias potentially arising with these enhancements would help to find ways to minimize differential sample loss, potentially affecting conclusions derived from these kinds of studies. Measurement of the issue is an important first step, but more work remains.

SUPPLEMENTARY MATERIALS

Supplementary materials are available online at academic.oup.com/jssam.

DATA AVAILABILITY

This article uses data from the Understanding Society COVID-19 study dataset and the main Understanding Society annual interviews. The data and documentation for the COVID-19 study and the main study are available to researchers from the UK Data Service (SN8644 and SN6614, respectively) at https://ukdataservice.ac.uk/. The analysis code for this article is available from https://doi.org/10.17605/OSF.IO/UV52C.

REFERENCES

Abdulazim, T., Abdelgawad, H., Habib, K. M. N., and Abdulhai, B. (2013), "Using Smartphones and Sensor Technologies to Automate Collection of Travel Data," *Transportation Research Record*, 2383, 44–52. https://doi.org/10.3141/2383-06.

American Association for Public Opinion Research (2023), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, Alexandria, VA: AAPOR.

Apple Inc. (2025), "Exposure Notification: Implement a COVID-19 exposure notification system that protects user privacy," [online], Available at https://developer.apple.com/documentation/exposurenotification.

Assemi, B., Jafarzadeh, H., Mesbah, M., and Hickman, M. (2018), "Participants' Perceptions of Smartphone Travel Surveys," *Transportation Research Part F: Traffic Psychology and Behaviour*, 54, 338–348. https://doi.org/10.1016/j.trf.2018.02.005.

Bach, R. L., and Wenz, A. (2020), "Studying Health-Related Internet and Mobile Device Use Using Web Logs and Smartphone Records," *PLOS ONE*, 15, e0234663. https://doi.org/10.1371/journal.pone.0234663.

Callegaro, M. (2010), "Do You Know Which Device Your Respondent Has Used to Take Your Online Survey?," *Survey Practice*, 3, 1. https://doi.org/10.29115/SP-2010-0028.

———. (2013), "Paradata in Web Surveys," in *Improving Surveys with Paradata: Analytic Uses of Process Information, Wiley Series in Survey Methodology*, ed. F. Kreuter, Hoboken, NJ: Wiley & Sons, pp. 261–279. https://doi.org/10.1002/9781118596869.ch11.

Chen, J., Bauman, A., and Allman-Farinelli, M. (2016), "A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research," *Telemedicine and e-Health*, 22, 655–665. https://doi.org/10.1089/tmj.2015.0159.

de Heij, V., Schouten, B., and Shlomo, N. (2015), "RISQ manual v2.1: Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation." Available at https://hummedia.manchester.ac.uk/institutes/cmist/risq/RISQ-manual-v21.pdf.

Décieux, J. P. (2024), "Sequential On-Device Multitasking within Online Surveys: A Data Quality and Response Behavior Perspective," *Sociological Methods & Research*, 53, 1384–1411. https://doi.org/10.1177/00491241221082593.

Décieux, J. P., and Sischka, P. E. (2024), "Comparing Data Quality and Response Behavior Between Smartphone, Tablet, and Computer Devices in Responsive Design Online Surveys," *SAGE Open*, 14, 21582440241252116. https://doi.org/10.1177/21582440241252116.

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., and Toepoel, V. (2021), "Where You at? Using GPS Locations in an Electronic Time Use Diary Study to Derive Functional Locations," *Social Science Computer Review*, 39, 509–526. https://doi.org/10.1177/0894439319877872.

Fortea, L., Tortella-Feliu, M., Juaneda-Seguí, A., De la Peña-Arteaga, V., Chavarría-Elizondo, P., Prat-Torres, L., Soriano-Mas, C., Lane, S. P., Radua, J., and Fullana, M. A. (2023), "Development and Validation of a Smartphone-Based App for the Longitudinal Assessment of Anxiety in Daily Life," *Assessment*, 30, 959–968. https://doi.org/10.1177/10731911211065166.

Google (2020), "Exposure Notifications Android API documentation v1.3.2." [online], Available at https://web.archive.org/web/20200603200341/https://static.googleusercontent.com/media/www.google.com/en//covid19/exposurenotifications/pdfs/Android-Exposure-Notification-API-documentation-v1.3.2.pdf.

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., and Crane, M. (2015), "A Web-Based Diary and Companion Smartphone App for Travel/Activity Surveys," *Transportation Research Procedia*, Transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods, 16–21 November 2014, Leura, Australia (Vol. 11), pp. 297–310. https://doi.org/10.1016/j.trpro.2015.12.026.

Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2009), *Survey Methodology, Wiley Series in Survey Methodology*, Hoboken, NJ: John Wiley & Sons, Inc.

Hagos, D. H., Løland, M., Yazidi, A., Kure, Ø., and Engelstad, P. E. (2020), "Advanced Passive Operating System Fingerprinting Using Machine Learning and Deep Learning," in 2020 29th International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, pp. 1–11. https://doi.org/10.1109/ICCCN49398.2020.9209694.

Hammoud, R., Tognin, S., Smythe, M., Gibbons, J., Davidson, N., Bakolis, I., and Mechelli, A. (2024), "Smartphone-Based Ecological Momentary Assessment Reveals an Incremental Association between Natural Diversity and Mental Wellbeing," *Scientific Reports*, 14, 7051. https://doi.org/10.1038/s41598-024-55940-7.

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J., and Nurul Habib, K. (2021), "Are We There yet? Assessing Smartphone Apps as Full-Fledged Tools for Activity-Travel Surveys," *Transportation*, 48, 2433–2460. https://doi.org/10.1007/s11116-020-10135-7.

Health and Social Care Northern Ireland (2020), "StopCOVID NI – FAQs," COVID-19 (Coronavirus) Northern Ireland [online], Available at https://web.archive.org/web/20201202214841/https://covid-19.hscni.net/stopcovid-ni-faqs/.

Hofmann, W., and Patel, P. V. (2015), "SurveySignal: A Convenient Solution for Experience Sampling Research Using Participants' Own Smartphones," *Social Science Computer Review*, 33, 235–253. https://doi.org/10.1177/0894439314525117

Institute for Social and Economic Research (2021), Understanding Society COVID-19, user guide, Version 10.0. Available at https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/user-guides/8644_covid_19_user_guide.pdf.

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., Trappmann, M., and Eckman, S. (2022), "Non-Participation in Smartphone Data Collection Using Research Apps," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185, S225–S245. https://doi.org/10.1111/rssa.12827.

Kline, J., Barford, P., Cahn, A., and Sommers, J. (2017), "On the Structure and Characteristics of User Agent String," in Proceedings of the 2017 Internet Measurement Conference, London, UK: ACM, pp. 184–190. https://doi.org/10.1145/3131365.3131406.

Kreuter, F. (ed.) (2013), *Improving Surveys with Paradata: Analytic Uses of Process Information, Wiley Series in Survey Methodology*, Hoboken, NJ: Wiley & Sons. https://doi.org/10.1002/9781118596869.

Lastovicka, M., Jirsik, T., Celeda, P., Spacek, S., and Filakovsky, D. (2018), "Passive OS Fingerprinting Methods in the Jungle of Wireless Networks," in NOMS 2018—2018 IEEE/

IFIP Network Operations and Management Symposium, Taipei, Taiwan, pp. 1–9. https://doi.org/10.1109/NOMS.2018.8406262.
McClain, C. A., Couper, M. P., Hupp, A. L., Keusch, F., Peterson, G., Piskorowski, A. D., and West, B. T. (2019), "A Typology of Web Survey Paradata for Assessing Total Survey Error," *Social Science Computer Review*, 37, 196–213. https://doi.org/10.1177/0894439318759670.
Muchmore, M., and Zamora, G. (2025), "Android vs. iOS: Which Phone OS is Actually the Better Pick?," PCMag. [online], Available at https://web.archive.org/web/20251216223159/https://uk.pcmag.com/comparison/146901/android-vs-ios-which-mobile-os-is-best.
Mulazzani, M., Reschl, P., Huber, M., Leithner, M., Schrittwieser, S., and Weippl, E. (2013), "Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting." in Web 2.0 Security & Privacy 2013 (W2SP 2013), San Francisco, Available at https://www.ieee-security.org/TC/W2SP/2013/papers/s2p1.pdf.
Nandakumar, D., Murli, S., Khosla, A., Choi, K., Rahman, A., Walsh, D., Riede, S., Dull, E., and Bowen, E. (2023), "A Novel Approach to User Agent String Parsing for Vulnerability Analysis Using Multi-Headed Attention," in 2023 International Conference on Machine Learning and Cybernetics (ICMLC), Adelaide, Australia: IEEE, pp. 105–113. https://doi.org/10.1109/ICMLC58545.2023.10327975.
NHS Scotland (2020), "How it Works | Protect Scotland," NHS Scotland Test & Protect [online], Available at https://web.archive.org/web/20201202213950/https://protect.scot/how-it-works.
NHS Test and Trace (n.d), "Which phones will the NHS COVID-19 app not work on?," NHS COVID-19 app support [online], Available at https://web.archive.org/web/20210519105529/https://faq.covid19.nhs.uk/article/KA-01116/.
R Core Team (2024), "R: A language and environment for statistical computing," Vienna, Austria: R Foundation for Statistical Computing.
Ramli, N. I., Ibrahim, M. F., Yahya, S., Taib, M. N., and Binti Buja, A. G. (2014), "A Dynamic Approach for Device Characteristics Based on String Matching Technique," in 2014 IEEE 2nd International Symposium on Telecommunication Technologies (ISTT), pp. 368–372, Langkawi, Malaysia. https://doi.org/10.1109/ISTT.2014.7238237.
Rodenburg, E., Schouten, B., and Struminskaya, B. (2025), "Nonresponse and Dropout in an App-Based Household Budget Survey: Representativeness, Interventions to Increase Response, and Data Quality," *Survey Research Methods*, 19, 25–41. https://doi.org/10.18148/srm/2025.v19i1.8263.
Roßmann, J., and Gummer, T. (2020), "*PARSEUAS: Stata Module to Extract Detailed Information from User Agent Strings*," Chestnut Hill, MA: Boston College.
Roßmann, J., Gummer, T., and Kaczmirek, L. (2020), "Working with User Agent Strings in Stata: The Parseuas Command," *Journal of Statistical Software*, 92, 1–16. https://doi.org/10.18637/jss.v092.c01.
Rudis, B. (2022), Package 'uaparserjs' (reference manual), Available at https://web.archive.org/web/20250209201451/https://cran.r-project.org/web/packages/uaparserjs/uaparserjs.pdf.
Rudis, B., Simon, L., and Langel, T. (2020), "uaparserjs: Parse 'User-Agent' Strings." Available at https://cran.r-project.org/package=uaparserjs.
Struminskaya, B., Lugtig, P., Keusch, F., and Höhne, J. K. (2020), "Augmenting Surveys with Data from Sensors and Apps: Opportunities and Challenges," *Social Science Computer Review*. Advance online publication December 20, 2020. https://doi.org/10.1177/0894439320979951.
Struminskaya, B., Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., and Dolmans, R. (2021), "Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Nonparticipation Bias," *Public Opinion Quarterly*, 85, 423–462. https://doi.org/10.1093/poq/nfab025.
Tourangeau, R., Rips, L. J., and Rasinski, K. (2000), *The Psychology of Survey Response*, New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511819322.

University of Essex, Institute for Social and Economic Research (2021), "Understanding Society: COVID-19 Study, 2020-2021," UK Data Service [dataset], Available at https://doi.org/10.5255/UKDA-SN-8644-11.

University of Essex, Institute for Social and Economic Research (2024), "Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009," UK Data Service [dataset], Available at https://doi.org/10.5255/UKDA-SN-6614-20.

Vine, J. (2021), "uasparseplus — An R package for deriving additional variables based on the parsing of User-Agent strings" [online], Available at https://github.com/jimvine/uasparseplus.

Vine, J., Burton, J., Couper, M. P., Jäckle A., and Horn, D. (2025), "Selectivity of Digital Trace Data: The Case of the UK COVID-19 Contact Tracing Apps," Understanding Society Working Paper Series, Working Paper, University of Essex, Colchester, Essex. Available at https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2025-07/.

Journal of Survey Statistics and Methodology, 2026, 00, 1–28
https://doi.org/10.1093/jssam/smag023
Article