Journal of Survey Statistics and Methodology (2021) **00**, 1–23

USING SMARTPHONES TO CAPTURE AND COMBINE SELF-REPORTS AND PASSIVELY MEASURED BEHAVIOR IN SOCIAL RESEARCH

FLORIAN KEUSCH*
FREDERICK G. CONRAD

With the ubiquity of smartphones, it is possible to collect self-reports as well as to passively measure behaviors and states (e.g., locations, movement, activity, and sleep) with native sensors and the smartphone's operating system, both on a single device that usually accompanies participants throughout the day. This research synthesis brings structure to a rapidly expanding body of literature on the combined collection of self-reports and passive measurement using smartphones, pointing out how and why researchers have combined these two types of data and where more work is needed. We distinguish between five reasons why researchers might want to integrate the two data sources and how this has been helpful: (1) *verification*, for example, confirming start and end of passively detected trips, (2) *contextualization*, for example, asking about the purpose of a passively detected trip, (3) *quantifying relationships*, for example, quantifying the association between self-reported stress and passively measured sleep duration, (4) *building composite measures*, for example, measuring components of stress that participants are aware of through self-reports and those they are not through passively measured speech attributes, and (5) *triggering measurement*, for example, asking survey questions contingent on certain passively measured events or participant locations. We discuss challenges of collecting self-reports and passively tracking participants' behavior with smartphones from the perspective of representation (e.g., who owns a smartphone and who is willing to share their data), measurement (e.g.,

Florian Keusch is Professor (interim) in the School of Social Sciences at the University of Mannheim, Mannheim, Germany, and Adjunct Research Assistant Professor in the Joint Program in Survey Methodology at the University of Maryland, College Park, MD, USA. Frederick G. Conrad is Research Professor in the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA and in the Joint Program in Survey Methodology, University of Maryland, College Park, MD, USA.

*Address correspondence to Florian Keusch, School of Social Sciences, University of Mannheim, Mannheim, Germany; E-mail: f.keusch@uni-mannheim.de.

doi: 10.1093/jssam/smab035

different levels of temporal granularity in self-reports and passively collected data), and privacy considerations (e.g., the greater intrusiveness of passive measurement than self-reports). While we see real potential in this approach it is not yet clear if its impact will be incremental or will revolutionize the field.

KEYWORDS: Passive measurement; Smartphones; Data integration; Ecological Momentary Assessment.

Statement of Significance

Smartphones make it possible for researchers to collect self-reports as well as to passively measure behaviors and states with native sensors and the smartphone's operating system, both on a single device that accompanies many people throughout the day. This research synthesis brings structure to the rapidly expanding body of literature on the combined collection of self-reports and passive measurement using smartphones, pointing out how and why researchers have combined these two types of data and where more work is needed.

1. INTRODUCTION

As smartphones have become central to daily life (e.g., Pew Research Center 2015), they have facilitated measurement of social phenomena in new ways (e.g., Link, Murphy, Schober, Buskirk, Hunter Childs, et al. 2014). While not all members of the public own smartphones, their use continues to grow globally and is extremely prevalent in high-income countries. The great advantage of smartphones for conducting research is that users are rarely far from their device (e.g., Patel, Kientz, Hayes, Bhat, and Abowd 2006; Dey, Wac, Ferreira, Tassini, Hong, et al. 2011; Van Laerhoven, Borazio, and Burdinski 2015); for many users, smartphones are in effect "extensions of their bodies" and thus users and their devices are often in the same physical and social context. In addition, smartphones include a large set of native sensors (e.g., GPS and accelerometer) and support multiple modes of communication (e.g., voice and video calls, text messaging, and web browsing). Taken together, smartphones promise to greatly facilitate wide-spread collection of *in situ* data that can be either self-reported or, using the devices' native sensors and its operating system, passively collected. Sensors are built-in hardware that passively detect, that is, without directly interacting with the user, ambient physical states or changes in states (e.g., the accelerometer tracks movements of the phone which can allow researchers to infer steps or mode of transportation taken by a user). Similarly, the operating system automatically logs information about device usage (e.g., incoming and outgoing voice calls and text messages, Internet browsing, and app use). In most cases, neither users nor researchers interact directly with sensors or the operating system but with apps that administer surveys, record sensor output, and log data about device usage (Struminskaya, Lugtig, Keusch, and Höhne 2020).

In situ self-reports and passive measurement are certainly not new research methods. As an alternative to retrospective self-reports, Ecological Momentary Assessments (EMAs)—a few questions about respondents' current state—have been conducted since well before the arrival of smartphones with paper diaries and paging devices (e.g., Shiffman, Stone, and Hufford 2008). Similarly, researchers have provided participants with sensor devices to monitor their behaviors such as activity and mobility (e.g., Lee and Mase 2002; Kapteyn, Banks, Hamer, Smith, Steptoe, et al. 2018) which facilitates more objective, less disruptive, and more continuous measurement than self-reports. However, their joint implementation on smartphones allows researchers to use them in a novel way, that is, to collect both self-reports and passively measure users' behaviors at the same time on a single device that many people already own. For example, Sugie and Lens (2017) studied the job-seeking behavior of recently paroled inmates using smartphones to both track their geolocation and administer EMAs. They found that parolees were able to overcome the mismatch between where they live and where they might find employment by traveling to job-rich areas. The authors indicate that this was a previously understudied topic; their insight was really only possible because they combined these two data sources.

While we see the potential for this approach being widely adopted by empirical social scientists, it is still an emerging paradigm that is not fully defined and articulated. In this article, we synthesize the literature reporting studies in which the investigators have jointly collected self-reports in mobile web surveys through apps or browsers and passively collected data using smartphone sensors and the operating system. Thus, we intentionally exclude studies that use mobile sensors to study human behavior but which do not also collect self-report (e.g., Mohr, Zhang, and Schueller 2017; Chaix 2018). The studies considered here span multiple disciplines using both probability and nonprobability samples. Our primary goal is to help survey researchers navigate this literature and to understand how and why these researchers have jointly collected these two types of data. By doing so, we not only discuss the promise of this new research paradigm but provide a realistic assessment of its limitations and what is yet unknown. In addition, we hope that this synthesis provides a standard terminology and set of concepts that extend beyond the particular sensors used in the studies we discuss.

First, we make explicit the types of phenomena best suited for collection via self-reports or passive measurement on smartphones. Second, we identify several different reasons why researchers might want to combine these two types of data and layout scenarios where each can help promote more accurate and complete investigation than has been previously possible. Third, we discuss challenges and some ways researchers have overcome them concerning representation, measurement, and privacy, arising from the joint collection of self-reports and passive measurement. Finally, we identify promising directions in which this emerging paradigm may evolve.

2. SELF-REPORTS AND PASSIVE MEASUREMENT USING SMARTPHONES

There is a small but growing literature in which studies combine both self-reports and passive measurement on smartphones. In these studies, self-reports are often collected via EMAs on a smartphone. Participants are signaled (via an app) at either random or scheduled intervals and asked about their affective or behavioral circumstances in-the-moment, eliminating the need for recall. In some of these studies, the self-reports concern subjective states, for example, stress (Adams, Rabbi, Rahman, Matthews, Voida, et al. 2014), affect (MacKerron and Mourato 2013; Chow, Fua, Huang, Bonelli, Xiong, et al. 2017; Lathia, Sandstrom, Mascolo, and Rentfrow 2017; Scherpenzeel 2017; York Cornwell and Cagney 2017; Boukhechba, Daros, Fua, Chow, Teachman, et al. 2018), emotional well-being (Madan, Cebrian, Lazer, and Pentland 2010; Wang, Chen, Chen, Li, Harari, et al. 2014; Ben-Zeev, Scherer, Wang, Xie, and Campbell 2015; Gao, Li, Zhu, Liu, and Liu 2016; Harari, Gosling, Wang, Chen, Chen, et al. 2016; Wang, Wang, Aung, Ben-Zeev, Brian, et al. 2017; Sugie 2018; English, Zhao, Brown, Catlett, and Cagney 2020), and physical health (Madan et al. 2010; English et al. 2020), because there simply are no smartphone sensors that can directly measure these internal states. In other studies, the self-reports are focused on objective states, particularly location and activity (e.g., MacKerron and Mourato 2013; Lathia et al. 2017; York Cornwell and Cagney 2017; Goodspeed, Yan, Hardy, Vydiswaran, Berrocal, et al. 2018). In some studies in this literature, retrospective self-reports are also collected as daily (e.g., Geurs, Thomas, Bijlsma, and Douhou 2015; Scherpenzeel 2017; Sugie 2018) or monthly summary measures (e.g., Keusch, Leonard, Sajons, and Steiner 2019; Kreuter, Haas, Keusch, Bähr, and Trappmann 2020). While some studies allow respondents to answer on desktop computers (e.g., Stopczynski, Sekara, Sapiezynski, Cuttone, Madsen, et al. 2014; Geurs et al. 2015; Scherpenzeel 2017; Sapiezynski, Stopczynski, Dreyer Lassen, and Lehmann 2019) or in the lab (e.g., Chow et al. 2017; Boukhechba et al. 2018), we are mainly concerned with studies in which all data—whether self-reported or passively measured—are collected from smartphones .

In contrast, we consider the measurement of certain activities to be "passive" in that the data they produce are byproducts of a person's ordinary activities: the user would send a particular text message or walk between the same two locations whether or not they are participating in a research study that is tracking these behaviors. Participants do not intend to passively create data but, instead, to accomplish everyday goals (like communicating or changing location). Of course, the behaviors that are passively measured involve action such as walking or texting but their measurement does not require deliberate action by the participant other than a one-time download of an app and consenting to data collection. This is in contrast to data created in response to a researcher-designed stimulus such as a survey question or a prompt to photograph a receipt (see Wenz, Jäckle, and Couper (2019), for a discussion of active versus passive measurement using smartphones).

The kinds of behaviors that can be passively measured with smartphones fall into two categories. On the one hand, it is possible to measure the participant's use of the smartphone for tasks like texting, making, and receiving phone calls, browsing the Internet, and using other apps. Harari, Müller, Aung, and Rentfrow (2017) call this kind of behavior *mediated* because they are carried out through (i.e., are *mediated* by) the smartphone and are captured by use logs, automatically generated by the device's operating system. This type of measurement does not actually involve any sensors but instead traces the user's interaction with the device (e.g., if the user places a phone call, this action is recorded by existing software). Researchers have used these data to measure smartphone use, including communication and online behavior (e.g., Madan et al. 2010; Gao et al. 2016; Revilla, Ochoa, and Loewe 2017; Scherpenzeel 2017; Wang et al. 2017; Boukhechba et al. 2018; Sugie 2018; Keusch, Leonard, et al. 2019; Kreuter et al. 2020).

On the other hand, native smartphone sensors allow the measurement of attributes of the participant's situation and their behavior without using the phone for communication, rendering these behaviors *non-mediated* (Harari et al. 2017). Many studies collect information about participants' location and mobility using *GPS*, *Wi-Fi*, and *cellular positioning* (e.g., MacKerron and Mourato 2013; Gao et al. 2016; Greene, Flake, Hathaway, and Geilich 2016; Chow et al. 2017; Scherpenzeel 2017; York Cornwell and Cagney 2017; Boukhechba et al. 2018; Goodspeed et al. 2018; Sugie 2018; Wray, Pérez, Celio, Carr, Adia, et al. 2019; Elevelt, Bernasco, Lugtig, Ruiter, and Toepoel 2019; Keusch, Leonard, et al. 2019; English et al. 2020; Kreuter et al. 2020), proximity to others using the smartphone's *Bluetooth* capability (e.g., Eagle and Pentland 2006; Madan et al. 2010; Stopczynski et al. 2014; Wang et al. 2014; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016; Sapiezynski et al. 2019), and physical activity using the *accelerometer* (e.g., Wang et al. 2014, 2017; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016; Lathia et al. 2017; Boukhechba et al. 2018; Höhne, Revilla, and Schlosser 2020; Kern, Höhne, Schlosser, and Revilla 2020; Kreuter et al. 2020). In addition, some studies use the smartphone's native *microphone* to measure frequency and duration of participants' conversation (Wang et al. 2014, 2017; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016) or to analyze psychological stress based on vocal characteristics (Adams et al. 2014). Smartphone audio technology which can be used to infer the specific radio or TV programs that are ambient in a particular setting (i.e., audio-matching or audio-fingerprinting) is used commercially and appears to be rapidly developing. Some studies combine the output from multiple smartphone sensors to infer more complex behaviors or states. The Dartmouth StudentLife study (Wang et al. 2014; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016) and the CrossCheck study (Wang et al. 2017) inferred the duration of participants' sleep using data from the smartphones' light sensor, accelerometer, microphone, and phone use logs.

There are at least two advantages of passive measurement using smartphones. First, this can be conducted much more frequently than researchers can ask survey questions, making it possible to measure and evaluate moment-to-moment change in a behavior (e.g., tracing exact walking routes to determine which neighborhoods participants visit) without increasing participant burden as would be the case when requesting multiple self-reports. This generally requires aggregating the passively collected data in order to promote comparability with self-reports (as discussed below). Second, for some phenomena, passive measurement can lead to more accurate estimates than self-reports. This is certainly the case for smartphone-mediated behaviors such as voice calls, text messages, and app usage, all of which are automatically logged and for which no interpretation is needed, enabling the logs to be treated as a gold standard (Kobayashi and Boase 2012; Boase and Ling 2013; Andrews, Ellis, Shaw, and Piwek 2015; De Reuver and Bouwman 2015; Revilla et al. 2017; Jones-Jang, Heo, McKeever, Kim, Moscowitz, et al. 2020). Even for certain nonmediated behaviors where some interpretation is required to connect the sensor measurement to the phenomenon of interest, passive measurement can be more accurate than self-reports. For example, several mobility/travel studies (e.g., Stopher, FitzGerald, and Xu 2007; Lynch 2017; Scherpenzeel 2017) have found that, compared to passively tracking trips using GPS and other smartphone sensors, respondents in conventional travel surveys tend to overlook or forget certain trips, potentially biasing conclusions.

Table 1 summarizes the distinctions between self-reports and passive measurement in studies that collect both types of data using smartphones. A key distinction concerns the temporal perspective of each type of measure. The temporal perspective of self-reports (two leftmost columns) can be either in-the-moment using EMA or retrospective. The type of phenomenon measured with self-reports could be objective (e.g., behavior), subjective (e.g., internal states such as stress or attitudes), or both. Passive measurement (right column) captures objective phenomena (i.e., behaviors and states) as they are occurring (i.e., in-the-moment) on the smartphone whether mediated by the device's communication capabilities or not.

Table 1. Combining Self-Reports and Sensor Measurement within a Study: Temporal Perspective and Type of Phenomenon Investigated

	Self-reports		Passive measurement	
Temporal perspective	In-the-moment (EMA)	Retrospective	In-the-moment	
Type of phenomenon	Subjective, objective, or both	Subjective, objective, or both	Objective	
			Mediated by smartphone (e.g., logs of voice calls or text messages)	Nonmediated (using sensors to capture e.g., location, activity, and sleep)

3. REASONS FOR INTEGRATING PASSIVE MEASUREMENT AND SELF-REPORTS

Going beyond the individual strengths of self-reports and passive measurement on smartphones, combining the two in one study can create synergies allowing substantive questions to be addressed in ways that are potentially novel, more efficient, and more accurate. In this section, we discuss several reasons evident in the literature why researchers have combined these two types of data so far, each for different purposes: (1) verifying, (2) contextualizing, (3) quantifying relationships, (4) building composite measures, and (5) triggering measurement.

3.1 Verifying One Data Source with the Other

First, some studies have used either self-reports or passive measurement to verify the other, treating one as the gold standard. If the passively collected data and the self-reports tell the same story, researchers interpret this as validation. Thus, this joint use of the two data sources really serves as a quality check and will not necessarily enable the researchers to address substantive research questions. This approach has been used to verify self-reported smartphone-mediated behavior (e.g., number of voice calls and text messages, amount of app usage) with usage data that has been automatically logged by the operating system (as discussed above).

However, self-reports have also been used to verify passive measurement. For example, the app designed by Greene et al. (2016) used geocoordinates of participants' travel to group a trip with previously identified trips. The app then asked participants to confirm that these inferences were correct. Similarly, in the Mobile Mobility study (Geurs et al. 2015; Scherpenzeel 2017), participants were provided with an online interface that allowed them to check, and if necessary edit, their trips which were passively registered by a mobility app.

A slightly different approach to verification is evident in Lathia et al. (2017) in which the researchers asked participants what physical activity they were engaged in during the past 15 minutes (self-reports) and measured physical activity using the accelerometer 15 minutes prior to eliciting the self-reports. Goodspeed et al. (2018) used self-reports to study the accuracy of passive location measurement from two devices, a smartphone and a GPS watch. They found that, on average, smartphone GPS measures were significantly more accurate in detecting locations, especially home locations, than the watch data.

3.2 Contextualizing One Data Source through the Other

A second reason why researchers have combined self-reports and passive measurement is that one of the data sources provides context for the other, facilitating its interpretation. For example, in Greene et al. (2016), self-reports are elicited in order to contextualize passively detected information about participants' trips. After each detected trip the app administered a brief questionnaire asking, for example, about the purpose of the trip and who accompanied the participant, aspects of the trip that would not be evident from the passive measurement alone, thus giving them meaning.

The reverse process in which passive measurement provides context for self-reports is also possible. For example, Höhne et al. (2020) and Kern et al. (2020) used accelerometer data to study how respondents' activity level while completing a mobile web survey might have affected measurement error.

3.3 Quantifying Relationships between Measures

A third reason to combine both types of data is to establish the substantive relationship between them. An example of this approach is found in Madan et al. (2010) who found that the severity of students' self-reported health symptoms (sore throat versus fever) as well as self-reported mental health was associated with three passively measured indicators of social contact (number of phone calls and text messages, proximity to other students, and changes in location). Similarly, the StudentLife study (Wang et al. 2014; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016) showed that self-reported stress increased as students' passively measured sleep (using the smartphone's microphone, light sensor, and accelerometer) was reduced.

MacKerron and Mourato (2013) used an app (Mappiness) that measured subjective well-being (via EMA) as well as participants' location (via GPS) at the time of the EMA. The authors associated each EMA response with spatial and environmental indicators such as location, weather, and daylight status derived from external sources. They concluded that participants were substantially happier outdoors, especially in green or natural habitat, and happier when it was sunny than not. A similar approach is used by Lathia et al. (2017), passively measuring activity with the accelerometer and quantifying its association with self-reported affect, for example, whether participants were happier when they had recently engaged in physical exercise. Adams et al. (2014) observed a correlation between ambient sound measured through the smartphone's microphone and self-reported stress.

3.4 Building Composite Measures

Much like a multi-item scale includes slightly different questions to explore different aspects of the same construct, a researcher might collect data on the same phenomenon through self-reports and passive measurement and combine them into one composite measure. This differs from contextualization which is carried out to help researchers interpret data and from quantifying relationships where the goal is to measure conceptually different phenomena with the two measures; in this case, researchers are synthesizing several measures to create a new measure that provides deeper understanding of a single phenomenon. For example, researchers might measure stress through self-report (e.g., EMA) and passively measured speech (e.g., speed, loudness [see, for example, Lu, Frauendorfer, Rabbi, Schmid Mast, Chittaranjan, et al. 2012]). The self-reports are likely to reflect stress that the participant is aware of while the speech data may capture stress that falls below the participant's awareness threshold.

3.5 Triggering Measurement

A fifth reason to combine self-reported and passive measurement is to determine from the latter that the participant is in a particular state—often a location—and defining that state as a triggering condition for self-report. The goal of combining self-report and passive measurement in this way is to collect self-reports in known, and potentially uncommon, contexts (Lathia, Rachuri, Mascolo, and Rentfrow 2013). This approach has the additional virtue of making it unnecessary to repeatedly ask participants whether they are in the intended state, thus reducing their burden and making data collection more efficient.

An example of triggering is found in Kreuter et al. (2020) who used so-called geofencing (Haas, Trappmann, Keusch, Bähr, and Kreuter 2020), that is, they asked a set of questions to job seekers about their experience only when their latitude and longitude indicated they were visiting a job center. Similarly, Wray et al. (2019) deployed location-based questionnaires asking about various characteristics of participants' current drinking location (e.g., type of location, accompanying people, and intoxication level) when study participants (gay and bisexual men who engage in heavy drinking and high-risk sex) were inside the geofence of popular bars. Another example of triggering beyond geofencing is illustrated in Sugie (2018) where any phone calls from or to a new number triggered a survey question about the phone call.

4. CHALLENGES OF COLLECTING DATA USING SMARTPHONES

While we are enthusiastic about the insights that might accrue from combining data for these reasons, as with any data there is error. Here, we turn to errors of representation and measurement in the joint collection and use of self-reports and passive measurement using smartphones from the perspective that one can only reduce these errors if one understands them, building on the Total Survey Error framework (e.g., Groves, Fowler, Couper, Lepkowski, Singer, et al. 2009). We also examine privacy concerns that are potentially introduced by passive data collection and so may be unfamiliar to survey researchers.

4.1 Representation

The ubiquity of smartphones in high income countries is well-established. In the United States, smartphone penetration has reached around 85 percent (Pew Research Center 2021). Similar levels are observed in some European (Eurostat 2019) and East Asian countries (eMarketer Report 2017) while the levels in Africa are substantially lower and there is also considerable variability between countries (Afrobarometer 2018). Even where smartphone penetration is high, there is real risk of coverage error because those who do not own the devices are likely to differ profoundly from owners on many variables (Couper, Gremel, Axinn, Guyer, Wagner, et al. 2018; Antoun, Conrad, Couper, and West 2019). This includes behaviors measured passively such as social engagement (e.g., through number of interactions on smartphones) even after controlling for demographic variables likely associated with smartphone ownership such as age and education (Keusch, Bähr, Haas, Kreuter, and Trappmann 2020). Researchers have dealt with the lack of smartphone ownership by providing smartphones to participants in studies of the general population (Scherpenzeel 2017) and for specific subgroups, including the elderly (York Cornwell and Cagney 2017; English et al. 2020), chronically ill (Goodspeed et al. 2018), men recently released from prison (Sugie 2018), and students who did not own an Android phone (Wang et al. 2014; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016).

However, even if there was universal smartphone penetration, low participation rates can also compromise representation. For example, in the Smartphone Time Use Study (Scherpenzeel 2017), the cumulative participation rate, that is, the percent of members from a Dutch probability online panel who completed an invitation questionnaire, expressed willingness to participate, and actually participated, was 19 percent. Similarly in the Kreuter et al. (2020) study, 16 percent of the invited members of a German national longitudinal survey who owned an Android smartphone downloaded and used a research app that administered survey questions and passively collected data. These rates are likely to be an upper bound on participation given the relatively established relationships between sample members and the panel study from which they were recruited. Trust and familiarity with the data collection organization are likely to be lower when participants are recruited without some familiarity with the researchers, thus reducing participation.

The most frequently reported reasons for nonparticipation in passive mobile measurement seem to be related to the intrusiveness of such data collection and privacy concerns (e.g., Wenz et al. 2019; Keusch, Struminskaya, Antoun, Couper, and Kreuter 2019; Struminskaya, Toepoel, Lugtig, Haan, Luiten, et al. 2020; Struminskaya, Lugtig, Toepoel, Schouten, Giesen, et al. in press). One approach that has been used to address these concerns is to explain to potential participants why these data are required to meet the scientific goals of the study and how the data as well as participants' identities will be protected. In practice, this has been done through in-person one-on-one (Wang et al. 2014; Ben-Zeev et al. 2015) and group (Sugie 2018) information sessions as well as by distributing extensive written information (Kreuter et al. 2020). To further reduce concern about the continuous collection of data, researchers might make passive measurement more discrete, for example, twice daily data collection might feel less intrusive than round-the-clock collection. One approach that has been shown experimentally to increase willingness to participate is to give the participant some control over data collection (Keusch, Struminskaya, et al. 2019; Struminskaya et al. in press). This approach has been used in practice by Sugie (2018), Kreuter et al. (2020), and Keusch, Leonard, et al. (2019), all of whom enabled participants to temporarily turn off the app that monitored their location and phone activities.

Another possible reason for nonparticipation is that many smartphone owners use the device in limited ways. Several studies find that the more activities people use their smartphone for (e.g., downloading apps, browsing the Internet, online banking, and geopositioning) the the more willing they are to participate in passive data collection (Keusch, Struminskaya, et al. 2019; Wenz et al. 2019; Struminskaya, Toepoel, Lugtig, et al. 2020). To demystify the smartphone technology, researchers have provided study-specific in-person training (York Cornwell and Cagney 2017; Sugie 2018).

Irrespective of the reasons for not being willing to participate in studies that passively measure behavior, a natural step to increase participation is to offer a monetary incentive. The benefits of monetary incentives for increasing survey response rates are well known (e.g., Singer and Ye 2013), and to the extent, this has been tested in studies that passively collect data via smartphones, this seems promising (e.g., Pinter 2015; Keusch, Struminskaya, et al. 2019; Haas, Kreuter, Keusch, Trappmann, and Bähr 2021).

Because participation rates are low in studies so far, estimates can be distorted to the extent that nonparticipants differ systematically from participants on the behaviors and momentary attributes (such as location) that are measured in a study. For example, a regular patron of the local red-light district might be unwilling to have their location tracked because doing so will likely expose this contra-normative behavior. If such behavior is the topic of the study, then nonparticipation bias becomes a real possibility. Similarly, older sample members may opt-out because they rarely use their smartphones for anything besides voice calls and might feel unable to download an app. Such people might also be less mobile and less active. Again, if mobility and activity are key measures of the study, nonparticipation increases the risk of bias. This is analogous to the potential bias related to survey nonresponse (e.g., Groves et al. 2009). One piece of relevant evidence comes from the study of Elevelt, Lugtig, and Toepoel (2019). They found in the Dutch Smartphone Time Use Study that people who participated in all parts of the study (i.e., prequestionnaire, diary, pop-up surveys, and GPS tracking) had reported spending significantly more time working (+5 hours) and less time watching television (−4 hours) than people who did not participate.

4.2 Measurement

The consequences of undercoverage and nonparticipation for the accuracy of estimates are essentially the same for both self-reports and passive data collection using smartphones. In contrast, the reasons for measurement error and its consequences differ quite substantially between self-reports and passively collected data. It is well known that self-reports contain measurement error, largely due to limitations in human comprehension, recall, and estimation ability but also self-presentation concerns and disclosure differences depending on survey mode (e.g., Tourangeau, Rips and Rasinski 2000). Questionnaire designers have worked hard for decades to reduce measurement error through improved question wording (e.g., Krosnick and Presser 2010; Schaeffer and Dykema 2011, 2020), pretesting questionnaires (e.g., Willis 2004; Beatty and Willis 2007; Willis 2015) and, more recently, by designing better smartphone interfaces for mobile surveys (Antoun, Katz, Argueta, and Wang 2018).

One might assume that by removing human cognition and social interaction from the data creation process, passive data collection removes measurement error from the individual observations. But this seems unlikely to be the case because (1) sensors are not perfectly accurate, (2) interpretation of sensor data requires inference beyond the actual measures, and (3) smartphone users may not comply with protocol when engaging in a study. Note that while different sensors and logs are by design specific to signals that differ in their modality and corresponding activity, the same type of measurement error is possible whether the sensor detects sound, light, location, movement, etc. For example, auditory data collected by microphones and location data captured by GPS are fundamentally different in substance, but both can suffer from misclassification (e.g., the sound of a barking dog misclassified as human speech and a visit to the liquor store misclassified as a visit to the dentist).

First, with respect to accuracy, measuring the same phenomenon with two different sensors can produce discrepant results. For example, Donaire-Gonzalez, Valentín, de Nazelle, Ambros, Carrasco-Turigas, et al. (2016) and Goodspeed et al. (2018) both found that for most location categories smartphone positioning was more accurate than a GPS watch, but measurement of travel by automobile or walking outdoors was actually more accurately measured by the GPS watch than the smartphone. Similarly, published comparisons of different physical activity trackers such as wristbands, waist-clips, and smartphones, consistently show results that differ by device (e.g., Altamimi, Skinner, and Nesbitt 2015; Ferguson, Rowlands, Olds, and Maher 2015; Toon, Davey, Hollis, Nixon, Horne, et al. 2016; Middelweerd, Van der Ploeg, Van Halteren, Twisk, Brug, et al. 2017; Höchsmann, Knaier, Eymann, Hintermann, Infanger, and Schmidt-Trucksäss 2018) and position of where the device is worn on the body (Sztyler, Stuckenschmidt, and Petrich 2017). Even smartphones from different brands and with different (versions of) operating systems can lead to measurement error for geolocation (Elevelt, Bernasco, et al. 2019; Bähr, Haas, Keusch, Kreuter, and Trappmann 2020) and acceleration (Kuhlmann, Garaizar, and Reips 2021). It is even possible that there is enough noise in passively collected data so that if the same device with the same sensor is used on two different occasions to measure what appears to be the same behavior, the results could differ. However, existing research shows that the reliability of sensor measurement is generally high (e.g., Charlton, Mentiplay, Pua, and Clark 2015; Kuznetsov, Robins, Long, Jakiela, Haran, et al. 2018).

Second, even if passively measured data are highly accurate, researchers need to bring meaning to the data. For example, electrical signals from an accelerometer need to be converted into units of movement such as steps which, in turn, need to be converted into distance walked, becoming more meaningful with each transformation (Elhoushi, Georgy, Noureldin, and Korenberg 2017; Duncan, Wunderlich, Zhao, and Faulkner 2018). The development of algorithms that classify raw data into meaningful units requires human judgment, for example, a decision on what qualifies as a step, potentially introducing measurement error. Typically, raw data are processed (aggregated and classified) as they are collected; while this conserves storage space and thus costs it also results in the loss of information if only the outcome of these processes is available to researchers (e.g., Harari, Lane, Wang, Crosier, Campbell, et al. 2016). Thus what seems to be objective raw output is often actually the result of considerable processing involving layers of interpretation and aggregation. Consumer-grade devices such as smartphones typically run proprietary algorithms and are subject to the development priorities of the manufacturer; thus, these algorithms can change without warning, potentially reducing the reliability of sensor measurement over time.

A third source of error in passive measurement involves noncompliance. If the participant forgets to turn the device on after powering it down or forgets to bring their smartphone with them during the day or it runs out of charge no data will be collected for potentially long intervals (Bähr et al. 2020). Similarly, if the user forgets to turn on a specific app or sensor (required in some designs) no data will be collected. This noncompliance is unintentional. However, some participants may intentionally keep their phone turned off, for example, if traveling internationally without a data plan or because they simply do not want their actions to be tracked. In general, noncompliance will lead to a downward bias in tracking activities because measures that should be taken are not. And the measurement error is more likely to be confined to specific types of activities if the noncompliance is intentional, for example, if the phone is turned off to preempt measurement of visits to a casino. The evidence so far suggests that intentional noncompliance is rare (e.g., Sugie, 2018). Related to noncompliance is the phenomenon of sharing a device with another individual (e.g., household member). This would imply that part of what is observed by the researcher is actually not coming from the unit of interest (see, e.g., Revilla et al. 2017). So far, the evidence that participants regularly share their smartphones with others suggests this practice is rare (Bähr et al. 2020).

Over and above the error inherent in each of the two types of measures, *combining* them may introduce additional challenges. For example, if researchers want to establish a relationship between a variable based on self-reports and a variable based on passive measurement it will be necessary to harmonize the two variables so that they are expressed at compatible levels of granularity. Because passive measurement is usually collected at higher frequencies than self-reports, harmonizing the measures most likely requires converting a large number of passive measurements into a single summary measure that aligns with the temporal granularity of the self-reports. For example, Lathia et al. (2017) summarized participants' passively measured activity over 15 minutes and correlated this aggregated measure with self-reported happiness at the end of the corresponding 15-minute period. Elevelt, Bernasco, et al. (2019) used the 10-minute slots for which self-reports were collected in a time use survey as the interval in which geolocation was passively measured. In other cases, it makes sense to aggregate both measures to a common unit of time that is meaningful for both. For example, Wang et al. (2014) asked participants to self-report stress eight times a day and measured sleep by continuously monitoring three smartphone sensors. They then correlated daily means of self-reported stress with the total amount of sleep on the corresponding days. Depending on the goals of the project, a daily mean might be a sufficient summary, but in other cases, one would want to preserve the rate and direction of change over the time periods being summarized, suggesting the use of slopes or derivatives (see, e.g., Wang et al. 2017). This process requires researchers to consider a potential tradeoff between losing information through aggregation on the one hand and harmonizing data to allow their analysis on the other. But standardizing the units of measurement is almost certainly necessary to reap the benefits of combining these two types of data.

4.3 Implications for Privacy

Just as surveys create privacy concerns for respondents and researchers (Singer 1978; Plutzer 2019) this is at least as much of a concern with respect to passive data collection (Keusch, Struminskaya, Kreuter, and Weichbold 2021). Consider the case of location detection. The more frequent the measurement the more likely researchers will be able to infer the function of certain locations in participants' lives, like their home or where their children go to school. Many participants might refuse to answer survey questions that directly ask about these locations because they consider the answers too personal to share with others, including the researchers. Thus in general, researchers using passively collected data are obliged to take the same steps they would when collecting self-report data but also to address participants' privacy concerns that are specific to passive measurement. See section 4.1 above for approaches taken to alleviate concerns about privacy in studies involving passive measurement (e.g., providing information about study goals and how data are used, reducing frequency of measurement, allowing participants to temporarily turn off data collection).

Once a participant has begun a study, they may forget that their behavior is being passively tracked in the background. It therefore may be appropriate for researchers to occasionally remind participants about the continuous measurement along the lines of Almuhimedi, Schaub, Sadeh, Adjerid, Acquisti, et al.'s (2015) "mobile privacy nudging" in which users are repeatedly informed about how their data are being shared and given the opportunity to adjust their privacy settings. This can help address the fact that the blanket proactive consent typically obtained in surveys may not fit the intensive longitudinal character of passive data collection.

Researchers might also proactively restrict the information they passively measure in order to avoid unnecessarily collecting personal information about participants and members of their social networks. For example, several studies have measured the presence of conversation using the smartphone's microphone without capturing the content of the conversations or the identity of the speakers (Adams et al. 2014; Wang et al. 2014, 2017; Ben-Zeev et al. 2015; Harari, Gosling, et al. 2016). The approach in these studies (developed by Wyatt, Chandhury, and Bilmes 2007) was to use "privacy-sensitive audio and conversation classifiers" developed to determine that the participant was "around conversation" but not to reconstruct the content of the speech or to identify individual speakers. Similarly, some studies collect information about participants' social networks (e.g., size and frequency of contact) based on logs of phone and text communication (e.g., Sugie 2018; Kreuter et al. 2020). To conceal the identities of members of the participants' social networks while still providing researchers with enough information to satisfy the goals of the project, Sugie (2018) recorded only the first name of each conversational partner while Kreuter et al. (2020) encrypted all phone numbers and only stored information automatically derived from names (likely gender and ethnicity).

5. DISCUSSION AND OUTLOOK

It is clear from the literature that we have discussed here that researchers have successfully conducted studies collecting and combining self-reports and passive measurements using smartphones. It is our impression that this approach is both viable and promising, but on closer inspection, it is actually not a single approach. We have identified five reasons why researchers have done this, each of which adds value by simultaneously exploiting the relative strengths of the two types of data but the implementation is slightly different for the different reasons. For example, for *verification* and *building composite measures* researchers collect information about the same phenomenon with the two types of data; for *contextualizing*, *quantifying relationships*, and *triggering measurement* different phenomena are measured.

Given the state of the literature, a next step could be to generalize passively measured behaviors by collecting them in a representative sample to a larger population (whatever that population might be), leveraging the infrastructure of probability-based surveys. This also makes it possible to carry out nonparticipation adjustments using participants' demographics and to conduct subgroup analyses in order to produce population estimates. In the literature we review here, this has rarely been done, mostly because the studies have been based on convenience samples. Two notable exceptions are the studies by Scherpenzeel (2017) who passively collected location in a time use study from members of the LISS panel (a probability panel representing the Dutch national population) and Kreuter et al. (2020) who passively measured location, activity, and smartphone use for a subsample of respondents from a yearly survey on labor market activity and poverty in Germany based on a national, probability sample.

While there is certainly promise in this emerging paradigm, there is also much that is still unknown about errors of representation and measurement. For example, the analysis of nonparticipation biases, that is, differences between survey respondents who are willing to have their behavior passively tracked and those in the full sample, has so far primarily concerned demographic characteristics and not the behaviors one would want to measure passively (e.g., activity level and mobility). With respect to measurement, an open question is whether participation produces reactivity. For example, it is possible that participants increase their steps or avoid certain locations because their activities or geolocation are being tracked. Similarly, misclassification is an inherent risk when using passively measured data (e.g., data from a smartphone left at home in a bag may be mistaken as evidence that the participant is asleep). Because this enterprise is new, the nature and extent of such biases are still under investigation.

Going forward, a broader range of sensors and smartphone logs are almost certain to be used—and used more frequently—either independently or in combination with one another. In fact, new technology may make it possible to measure phenomena that previously could only have been self-reported. For example, using sensors that are capable of detecting when people prepare meals or are eating by measuring "clatter" of dishes (Bi, Xing, Hao, Huh, Peng, et al. 2017) or by sensing the smell of certain foods (Gonzalez-Jimenez, Monroy, and Blanco 2011), it may be possible to trigger self-reports of dietary intake. Another direction in which passive measurement in the context of self-reports seems poised to advance is assessing affect, mood, stress, or other subjective phenomena. For example, by first establishing a relationship between self-reported depression and passively measured conversation frequency, amount of social contact, and sleep patterns, it may be possible to infer negative affect or psychological stress from the sensor data alone.

In summary, we believe that there is real potential in combining passive measurement with self-reports on smartphones but it is not clear if this practice will benefit all aspects of social and behavioral research or just some. For example, studies of public opinion might not benefit as much as studies of mobility. But as new and better sensing technologies emerge and researchers develop creative ways to combine sensor data with self-reports, we anticipate progress addressing research questions that were previously out of reach.

REFERENCES

Adams, P., M. Rabbi, T. Rahman, M. Matthews, A. Voida, G. Gay, T. Choudhury, and S. Voida (2014), "Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild," Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, pp. 72–79.

Afrobarometer (2018), "Afrobarometer," A Pan-African Series of National Public Attitude Surveys on Democracy, Governance, and Society [online], available at https://afrobarometer.org/online-data-analysis/analyse-online. Accessed August 29, 2021.

Almuhimedi, H., F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal (2015), "Your Location has been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging," Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 787–796.

Altamimi, R. I., G. D. Skinner, and K. V. Nesbitt (2015), "A Position Paper on Managing Youth Screen Time versus Physical Activity," *GSTF Journal on Computing (JoC)*, 4, 10.

Andrews, S., D. A. Ellis, H. Shaw, and L. Piwek (2015), "Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use," *PLoS One*, 10, e0139004.

Antoun, C., F. G. Conrad, M. P. Couper, and B. T. West (2019), "Simultaneous Estimation of Multiple Sources of Error in a Smartphone-Based Survey," *Journal of Survey Statistics and Methodology*, 7, 93–117.

Antoun, C., J. Katz, J. Argueta, and L. Wang (2018), "Design Heuristics for Effective Smartphone Questionnaires," *Social Science Computer Review*, 36, 557–574.

Bähr, S., G.-C. Haas, F. Keusch, F. Kreuter, and M. Trappmann (2020), "Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data," *Social Science Computer Review*. doi:10.1177/0894439320944118.

Beatty, P. C., and G. B. Willis (2007), "Research Synthesis: The Practice of Cognitive Interviewing," *Public Opinion Quarterly*, 71, 287–311.

Ben-Zeev, D., E. A. Scherer, R. Wang, H. Xie, and A. T. Campbell (2015), "Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health," *Psychiatric Rehabilitation Journal*, 38, 218–226.

Bi, C., G. Xing, T. Hao, J. Huh, W. Peng, and M. Ma (2017), "FamilyLog: A Mobile System for Monitoring Family Mealtime Activities," *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 21–30.

Boase, J., and R. Ling (2013), "Measuring Mobile Phone Use: Self-Report versus Log Data," *Journal of Computer-Mediated Communication*, 18, 508–519.

Boukhechba, M., A. R. Daros, K. Fua, P. I. Chow, B. A. Teachman, and L. E. Barnes (2018), "DemonicSalmon: Monitoring Mental Health and Social Interactions of College Students Using Smartphones," *Smart Health*, 9–10, 192–203.

Chaix, B. (2018), "Mobile Sensing in Environmental Health and Neighborhood Research," *Annual Review of Public Health*, 39, 367–384.

Charlton, P. C., B. F. Mentiplay, Y.-H. Pua, and R. A. Clark (2015), "Reliability and Concurrent Validity of a Smartphone, Bubble Inclinometer and Motion Analysis System for Measurement of Hip Joint Range of Motion," *Journal of Science and Medicine in Sport*, 18, 262–267.

Chow, P. I., K. Fua, Y. Huang, W. Bonelli, H. Xiong, L. E. Barnes, and B. A. Teachman (2017), "Using Mobile Sensing to Test Clinical Models of Depression, Social Anxiety, State Affect, and Social Isolation among College Students," *Journal of Medical Internet Research*, 19, e62.

Couper, M. P., G. Gremel, W. Axinn, H. Guyer, J. Wagner, and B. T. West (2018), "New Options for National Population Surveys: The Implications of Internet and Smartphone Coverage," *Social Science Research*, 73, 221–235.

De Reuver, M., and H. Bouwman (2015), "Dealing with Self-Report Bias in Mobile Internet Acceptance and Usage Studies," *Information & Management*, 52, 287–294.

Dey, A. K., K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos (2011), "Getting Closer: An Empirical Investigation of the Proximity of User to Their Smartphones," Proceedings of the 13th International Conference on Ubiquitous Computing, pp. 163–172.

Donaire-Gonzalez, D.,. A. Valentín, A. de Nazelle, A. Ambros, G. Carrasco-Turigas, E. Seto, M. Jerrett, and M. J. Nieuwenhuijsen (2016), "Benefits of Mobile Phone Technology for Personal Environmental Monitoring," *JMIR mHealth and uHealth*, 4, e126.

Duncan, M. J., K. Wunderlich, Y. Zhao, and G. Faulkner (2018), "Walk This Way: Validity Evidence of iPhone Health Application Step Count in Laboratory and Free-Living Conditions," *Journal of Sports Sciences*, 36, 1695–1704.

Eagle, N., and A. S. Pentland (2006), "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, 10, 255–268.

eMarketer Report (2017), "Internet and Mobile Users in Asia-Pacific: eMarketer's Country-by-Country Forecast for 2017-2021," available at https://www.emarketer.com/Report/Internet-Mobile-Users-Asia-Pacific-eMarketers-Country-by-Country-Forecast-20172021/2002155. Accessed August 29, 2021.

Elevelt, A., W. Bernasco, P. Lugtig, S. Ruiter, and V. Toepoel (2019), "Where You at? Using GPS Locations in an Electronic Time Use Diary Study to Derive Functional Locations," *Social Science Computer Review*, 39, 509–526.

Elevelt, A., P. Lugtig, and V. Toepoel (2019), "Doing a Time Use Survey on Smartphones Only: What Factors Predict Nonresponse at Different Stages of the Survey Process?," *Survey Research Methods*, 13, 195–213.

Elhoushi, M., J. Georgy, A. Noureldin, and M. Korenberg (2017), "A Survey on Approaches of Motion Mode Recognition Using Sensors," *IEEE Transactions on Intelligent Transportation Systems*, 18, 1662–1686.

English, N., C. Zhao, K. L. Brown, C. Catlett, and K. Cagney (2020), "Making Sense of Sensor Data: How Local Environmental Conditions Add Value to Social Science Research," *Social Science Computer Review*. doi:10.1177/0894439320920601.

Eurostat (2019), "Digital Economy and Society: ICT Usage in Households and by Individuals," http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database. Accessed August 29, 2021.

Ferguson, T., A. V. Rowlands, T. Olds, and C. Maher (2015), "The Validity of Consumer-Level, Activity Monitors in Healthy Adults Worn in Free-Living Conditions: A Cross-Sectional Study," *International Journal of Behavioral Nutrition and Physical Activity*, 12, 42.

Gao, Y., A. Li, T. Zhu, X. Liu, and X. Liu (2016), "How Smartphone Usage Correlates with Social Anxiety and Loneliness," *PeerJ*, 4, e2197.

Geurs, K. T., T. Thomas, M. Bijlsma, and S. Douhou (2015), "Automatic Trip and Mode Detection with Move Smarter: First Results from the Dutch Mobile Mobility Panel," *Transportation Research Procedia*, 11, 247–262.

Gonzalez-Jimenez, J., J. G. Monroy, and J. L. Blanco (2011), "The Multi-Chamber Electronic Nose. An Improved Olfaction Sensor for Mobile Robotics," *Sensors*, 11, 6145–6164.

Goodspeed, R., X. Yan, J. Hardy, V. G. V. Vydiswaran, V. G. Berrocal, P. Clarke, D. M. Romero, I. N. Gomez-Lopez, and T. Veinot (2018), "Comparing the Data Quality of Global Positioning System Devices and Mobile Phones for Assessing Relationships between Place, Mobility, and Health. Field Study," *JMIR Mhealth and Uhealth*, 6, e168.

Greene, E., L. Flake, K. Hathaway, and M. Geilich (2016), "A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana," paper presented at the Transportation Research Board 95th Annual Meeting, January 10–14, 2016, Washington, DC.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009), *Survey Methodology* (2nd ed.), Hoboken, NJ: Wiley & Sons.

Haas, G.-C., F. Kreuter, F. Keusch, M. Trappmann, and S. Bähr (2021), "Effects of Incentives in Smartphone Data Collection," in *Big Data Meets Survey Science*, eds. C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, and L. E. Lyberg, pp. 387–414, Hoboken, NJ: Wiley.

Haas, G.-C., M. Trappmann, F. Keusch, S. Bähr, and F. Kreuter (2020), "Using Geofences to Collect Survey Data: Lessons Learned from the IAB-SMART Study," *Survey Methods: Insights from the Field*, available at https://surveyinsights.org/?p=13405.

Harari, G. M., S. D. Gosling, R. Wang, F. Chen, Z. Chen, and A. T. Campbell (2016), "Patterns of Behavior Change in Students over an Academic Term: A Preliminary Study of Activity and Sociability Behaviors Using Smartphone Sensing Methods," *Computers in Human Behavior*, 67, 129–138.

Harari, G. M., N. D. Lane, R. Wang, B. S. Crosier, A. T. Campbell, and S. D. Gosling (2016), "Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges," *Perspectives on Psychological Science*, 11, 838–854.

Harari, G. M., S. R. Müller, M. S. H. Aung, and P. J. Rentfrow (2017), "Smartphone Sensing Methods for Studying Behavior in Everyday Life," *Current Opinion in Behavioral Sciences*, 18, 83–90.

Höchsmann, C., R., Knaier, J. Eymann, J. Hintermann, D. Infanger, and A. Schmidt-Trucksäss (2018), "Validity of Activity Trackers, Smartphones, and Phone Applications to Measure Steps in Various Walking Conditions," *Scandinavian Journal of Medicine & Science in Sports*, 28, 1818–1827.

Höhne, J. K., M. Revilla, and S. Schlosser (2020), "Motion Instructions in Surveys: Compliance, Acceleration, and Response Quality," *International Journal of Market Research*, 62, 43–57.

Jones-Jang, S. M., Y.-J. Heo, R. McKeever, J. H. Kim, L. Moscowitz, and D. Moscowitz (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, 346–363.

Kapteyn, A., J. Banks, M. Hamer, J. P. Smith, A. Steptoe, A. van Soest, A. Koster, and S. H. Wah (2018), "What They Say and What They Do: Comparing Physical Activity across the USA, England and The Netherlands," *Journal of Epidemiology & Community Health*, 72, 471–476.

Kern, C., J. K. Höhne, S. Schlosser, and M. Revilla (2020), "Completion Conditions and Response Behavior in Smartphone Surveys: A Prediction Approach Using Acceleration Data," *Social Science Computer Review*. doi:10.1177/0894439320971233.

Keusch, F., S. Bähr, G.-C. Haas, F. Kreuter, and M. Trappmann (2020), "Coverage Error in Data Collection Combining Mobile Surveys with Passive Measurement Using Apps: Data from a German National Survey," *Sociological Methods & Research*. doi:10.1177/0049124120914924.

Keusch, F., M. L. Leonard, C. Sajons, and S. Steiner (2019), "Using Smartphone Technology for Research on Refugees—Evidence from Germany," *Sociological Methods & Research*. doi:10.1177/0049124119852377.

Keusch, F., B. Struminskaya, C. Antoun, M. P. Couper, and F. Kreuter (2019), "Willingness to Participate in Passive Mobile Data Collection," *Public Opinion Quarterly*, 83, 210–235.

Keusch, F., B. Struminskaya, F. Kreuter, and M. Weichbold (2021), "Combining Active and Passive Mobile Data Collection: A Survey of Concerns," in *Big Data Meets Survey Science*, eds. C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, and L. E. Lyberg, pp. 657–682, Hoboken, NJ: Wiley.

Kobayashi, T., and J. Boase (2012), "No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use," *Communication Methods and Measures*, 6, 126–143.

Kreuter, F., G.-C. Haas, F. Keusch, S. Bähr, and M. Trappmann (2020), "Collecting Survey and Smartphone Sensor Data with an App: Opportunities and Challenges around Privacy and Informed Consent," *Social Science Computer Review*, 38, 533–549.

Krosnick, J. A., and S. Presser (2010), "Question and Questionnaire Design," in *Handbook of Survey Research*, eds. P. V. Marsden and J. D. Wright, pp. 263–314, Bingley, UK: Emerald Group Publishing.

Kuhlmann, T., P. Garaizar, and U.-D. Reips (2021), "Smartphone Sensor Accuracy Varies from Device to Device in Mobile Research: The Case of Spatial Orientation," *Behavior Research Methods*, 53, 22–32.

Kuznetsov, N. A., R. K. Robins, B. Long, J. T. Jakiela, F. J. Haran, S. E. Ross, G. W. Wright, and C. K. Rhea (2018), "Validity and Reliability of Smartphone Orientation Measurement to Quantify Dynamic Balance Function," *Physiological Measurement*, 39, 02NT01.

Lathia, N., K. K. Rachuri, C. Mascolo, and P. J. Rentfrow (2013), "Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods," Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 183–192.

Lathia, N., G. M. Sandstrom, C. Mascolo, and P. J. Rentfrow (2017), "Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity," *PLoS One*, 12, e0160589.

Lee, S.-W., and K. Mase (2002), "Activity and Location Recognition Using Wearable Sensors," *IEEE Pervasive Computing*, 1, 24–32.

Link, M. W., J. Murphy, M. F. Schober, T. D. Buskirk, J. Hunter Childs, and C. Langer Tesfaye (2014), "Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research," *Public Opinion Quarterly*, 78, 779–787.

Lu, H., D. Frauendorfer, M. Rabbi, M. Schmid Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury (2012), "Stresssense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones," Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 351–360.

Lynch, J. (2017), "Smartphone GPS Apps as a Mode of Travel Survey Data Collection," paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, May 18–21, 2017, New Orleans, LA.

MacKerron, G., and S. Mourato (2013), "Happiness Is Greater in Natural Environments," *Global Environmental Change*, 23, 992–1000.

Madan, A., M. Cebrian, D. Lazer, and A. Pentland (2010), "Social Sensing for Epidemiological Behavior Change," Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 291–300.

Middelweerd, A., H. P. Van der Ploeg, A. Van Halteren, J. W. R. Twisk, J. Brug, and S. J. Te Velde (2017), "A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals," *Medicine & Science in Sports & Exercise*, 49, 1270–1279.

Mohr, D. C., M. Zhang, and S. M. Schueller (2017), "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning," *Annual Review of Clinical Psychology*, 13, 23–47.

Patel, S. N., J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd (2006), "Farther Than You May Think: An Empirical Investigation of the Proximity of Users to Their Mobile Phones," Proceedings of the 8th International Conference on Ubiquitous Computing, pp. 123–140.

Pew Research Center (2015), "The Smartphone Difference," available at http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015. Accessed August 29, 2021.

Pew Research Center (2021), "Mobile Fact Sheet," available at http://www.pewinternet.org/fact-sheet/mobile/. Accessed August 29, 2021.

Pinter, R. (2015), "Willingness of Online Access Panel Members to Participate in Smartphone Application-Based Research," in *Mobile Research Methods*, eds. P. de Pedraza, R. Pinter, and D. Toninelli, pp. 141–156, London, UK: Ubiquity Press.

Plutzer, E. (2019), "Privacy, Sensitive Questions, and Informed Consent: Their Impacts on Total Survey Error, and the Future of Survey Research," *Public Opinion Quarterly*, 83, 169–184.

Revilla, M., C. Ochoa, and G. Loewe (2017), "Using Passive Data from a Meter to Complement Survey Data in Order to Study Online Behavior," *Social Science Computer Review*, 35, 521–536.

Sapiezynski, P., A. Stopczynski, D. Dreyer Lassen, and S. Lehmann (2019), "Interaction Data from the Copenhagen Networks Study," *Scientific Data*, 6, 315.

Schaeffer, N. C., and J. Dykema (2011), "Questions for Surveys: Current Trends and Future Directions," *Public Opinion Quarterly*, 75, 909–961.

———. (2020), "Advances in the Science of Asking Questions," *Annual Review of Sociology*, 46, 37–60.

Scherpenzeel, A. (2017), "Mixing Online Panel Data Collection with Innovative Methods," in *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*, eds. S. Eifler and F. Faulbaum, pp. 27–49, Wiesbaden: Springer.

Shiffman, S., A. A. Stone, and M. R. Hufford (2008), "Ecological Momentary Assessment," *Annual Review of Clinical Psychology*, 4, 1–32.

Singer, E. (1978), "Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys," *American Sociological Review*, 43, 144–162.

Singer, E., and C. Ye (2013), "The Use and Effects of Incentives in Surveys," *The Annals of the American Academy of Political and Social Science*, 645, 112–141.

Stopczynski, A., V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann (2014), "Measuring Large Scale Social Networks with High Resolution," *PLoS One*, 9, e95978.

Stopher, P., C. FitzGerald, and M. Xu (2007), "Assessing the Accuracy of the Sydney Household Travel Survey with GPS," *Transportation*, 34, 723–741.

Struminskaya, B., P. Lugtig, F. Keusch, and J. K. Höhne (2020), "Augmenting Surveys with Data from Sensors and Apps: Opportunities and Challenges," *Social Science Computer Review*. doi:10.1177/0894439320979951.

Struminskaya, B., P. Lugtig, V. Toepoel, B. Schouten, D. Giesen, and R. Dolmans (in press), "Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Non-Participation Bias," *Public Opinion Quarterly*.

Struminskaya, B., V. Toepoel, P. Lugtig, M. Haan, A. Luiten, and B. Schouten (2020), "Understanding Willingness to Share Smartphone Sensor Data," *Public Opinion Quarterly*, 84, 725–759.

Sugie, N. F. (2018), "Utilizing Smartphones to Study Disadvantaged and Hard-to-Reach Groups," *Sociological Methods & Research*, 47, 458–491.

Sugie, N. F., and M. C. Lens (2017), "Daytime Locations in Spatial Mismatch: Job Accessibility and Employment at Reentry from Prison," *Demography*, 54, 775–800.

Sztyler, T., H. Stuckenschmidt, and W. Petrich (2017), "Position-Aware Activity Recognition with Wearable Devices," *Pervasive and Mobile Computing*, 38, 281–295.

Toon, E., M. J. Davey, S. L. Hollis, G. M. Nixon, R. S. C. Horne, and S. N. Biggs (2016), "Comparison of Commercial Wrist-Based and Smartphone Accelerometers, Actigraphy, and PSG in a Clinical Cohort of Children and Adolescents," *Journal of Clinical Sleep Medicine*, 12, 343–350.

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

Van Laerhoven, K., M. Borazio, and J. H. Burdinski (2015), "Wear Is Your Mobile? Investigating Phone Carrying and Use Habits with a Wearable Device," *Frontiers in ICT*, 2, 10.

Wang, R., F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell (2014), "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones," Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3–14.

Wang, R., W. Wang, M. S. H. Aung, D. Ben-Zeev, R. Brian, A. T. Campbell, T. Choudhury, et al. (2017), "Predicting Symptom Trajectories of Schizophrenia Using Mobile Sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 1(3):110, pp. 1–24.

Wenz, A., A. Jäckle, and M. P. Couper (2019), "Willingness to Use Mobile Technologies for Data Collection in a Probability Household Panel," *Survey Research Methods*, 13, 1–22.

Willis, G. B. (2004), *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, Thousand Oaks: Sage.

———. (2015), *Analysis of the Cognitive Interview in Questionnaire Design*, New York: Oxford University Press.

Wray, T. B., A. E. Pérez, M. A. Celio, D. J. Carr, A. C. Adia, and P. M. Monti (2019), "Exploring the Use of Smartphone Geofencing to Study Characteristics of Alcohol Drinking Locations in High-Risk Gay and Bisexual Men," *Alcoholism: Clinical and Experimental Research*, 43, 900–906.

Wyatt, D., T. Choudhury, and J. Bilmes (2007), "Conversation Detection and Speaker Segmentation in Privacy Sensitive Situated Speech Data," Proceedings of the Eighth Annual Conference of the International Speech Communication Association, pp. 586–589.

York Cornwell, E., and K. A. Cagney (2017), "Aging in Activity Space: Results from Smartphone-Based GPS-Tracking of Urban Seniors," *Journals of Gerontology: Series B*, 72, 864–875.