In-Home Passive Sensor Data Collection and Its Implications for Social Media Research: Perspectives of Community Women in Rural South Africa

Alastair van Heerden[1,2] iD, Doug Wassenaar[3] iD, Zaynab Essack[1,3] iD, Khanya Vilakazi[1], and Brandon A. Kohrt[4]

Abstract

There has been a recent increase in debates on the ethics of social media research, passive sensor data collection, and big data analytics. However, little evidence exists to describe how people experience and understand these applications of technology. This study aimed to passively collect data from mobile phone sensors, lapel cameras, and Bluetooth beacons to assess people's understanding and acceptance of these technologies. Seven households were purposefully sampled and data collected for 10 days. The study generated 48 hr of audio data and 30,000 images. After participant review, the data were destroyed and in-depth interviews conducted. Participants found the data collected acceptable and reported willingness to participate in similar studies. Key risks included that the camera could capture nudity and sex acts, but family review of footage before sharing helped reduce concerns. The Emanuel et al. ethics framework was found to accommodate the concerns and perspectives of study participants.

Introduction

Rapid changes in technology create new research methods and data sources that do not precisely fit previously defined categories of "text-based" versus "human subjects" research (Samuel et al., 2018). Researchers, research ethics committees (RECs), and Institutional Review Boards (IRBs) are increasingly required to consider the ethical merit of novel approaches which strain current ethics frameworks, often resulting in case-based decisions (Hibbin, Samuel, & Derrick, 2018; Kara, 2018). The lack of clear and specific ethics guidance is compounded by ongoing technological innovation and a proliferation of the type and amount of data produced and stored about our online lives, providing both an opportunity for innovative research and greater potential for abuse. Inexpensive storage solutions mean that the de facto standard in many commercial enterprises is to collect and store as much data as possible from as many sources as possible in what is termed a digital lake (Miloslavskaya & Tolstoy, 2016). Much of these data remain unused for current reporting or analytic purposes, but as new analytic techniques or capacity become available, it is possible to draw the additionally required data from the data lake. This contrasts with research studies that typically require that all collected data

be clearly circumscribed and have a clear purpose, analysis plan, and role in answering the research aim, although the move toward "big data" in recent years has forced RECs/IRBs to increasingly review and approve the research with large preexisting data sets (Salerno, Knoppers, Lee, Hlaing, & Goodman, 2017).

Given that data collection is costlier than data storage, RECs may soon have to grapple with how to apply this new paradigm to research (Anagnostopoulos, Zeadally, & Exposito, 2016; C. Walker & Alrehamy, 2015). This is particularly important as it is plausible that with continued digitization of our everyday lives and advances in unsupervised machine learning data analytics, great value may be found in less targeted data collection. These insights may,

[1]Human Sciences Research Council, Pietermaritzburg, South Africa
[2]University of the Witwatersrand, Johannesburg, South Africa
[3]University of KwaZulu-Natal, Pietermaritzburg, South Africa
[4]George Washington School of Medicine and Health Sciences, Washington, DC, USA

Corresponding Author:
Alastair van Heerden, Human and Social Development, Human Science Research Council, Old Bus Depot, Mbubu Road, Sweetwaters, Pietermaritzburg 3200, South Africa.
Email: avanheerden@hsrc.ac.za

however, come at the expense of increased risk. The most obvious are privacy concerns that could result from a study data breach. Another lesser known risk is that with enough data, almost all de-identified data will become re-identifiable and personal data protection laws, such as those of the General Data Protection Regulation (GDPR), will be triggered for almost everything (Purtova, 2018).

While related research ethics literature currently focuses primarily on either the ethics of social media or digital (big) data collection (Boyd & Crawford, 2012; Hibbin et al., 2018; McKee, 2013; Mittelstadt & Floridi, 2016; Moreno, Goniu, Moreno, & Diekema, 2013), increasing overlap is likely, as progressively more of people's lives play out through online portals, particularly, mobile phones. Social media companies leverage the popularity of their social platforms by designing their mobile applications to collect as much additional user information as possible (Gallagher, 2018; Hartmann, Zaki, Feldmann, & Neely, 2016; Spiekermann, Acquisti, Böhme, & Hui, 2015). One recent example of the increasing overlap between Social media and big data fields is the patent Facebook that recently filed with the U.S. Patent and Trademark Office titled "Offline Trajectories" (Subbian, 2018). Using Global Positioning System (GPS) data about users' last logged location as well as social media data, the Facebook patent claims to be able to build behavioral pattern models that can predict future movements, essentially meaning that Facebook may know where you are going before you do. While this is used to drive revenue for the social media industry, these data also hold potential for researchers interested in understanding behaviors and practices that could inform the development of psychological and behavioral interventions.

The passive collection of data generated by small digital sensors has enabled the unobtrusive measurement and storage of data about patterns of human behavior, cognition, and mood (Martinez-Martin, Insel, Dagum, Greely, & Cho, 2018). Passively collected digital traces are likely to become increasingly important sources of data in intervention research (Estrin, 2014; Insel, 2017). Recent studies exploring the ethics of passive data collection have identified several key concerns, including information privacy, data security, equity in access, informed consent, data ownership, and secondary uses of the data (Maher et al., 2019). Inadequate attention to privacy issues was also identified in a systematic review of the passive data literature (Cornet & Holden, 2018). The Institute for Employment Research SMART (IAB-SMART) study found that when expanding a request for access to limited data to a comprehensive set of digital trace data, including geolocation, accelerometer data, phone and text messaging logs, and app usage, few additional concerns were raised by participants who were willing to share this comprehensive data set. This was partly explained by the fact that participants were found to have minimally differentiated between the different data requests made. Furthermore,

data access was rarely revoked once access had been granted (Kreuter, Haas, Keusch, Bähr, & Trappmann, 2018).

In addition to the risks participants may face, papers have begun to outline some of the ethical risks social media and passive digital traces hold for researchers undertaking sensitive fieldwork (van Baalen, 2018). Such work has focused on social media accounts and digital traces produced by mobile phones and other research equipment that allow bad actors the opportunity to trace and monitor participant identities, whereabouts, and activities. Depending on country laws, researchers interested in topics such as terrorism, intravenous drug use, sex work, or men who have sex with men (MSM) may be at increased risk of monitoring by security agencies.

Despite the speed at which the field is growing, relatively little is known about people's experience and understanding of passive mobile phone data traces used for research purposes. This is especially true for health research in low- and middle-income countries (LMICs) that is conducted with populations with low literacy and technology exposure. This study was designed, as part of a larger study on mother–child interaction, to address this knowledge gap. The main study was based on the hypothesis that by providing caregivers in low-resource settings with insight into the world as seen through the eyes of their children, caregiver investment can be enhanced in activities predictive of linguistic and cognitive development. Providing this insight for caregivers will be based on leveraging passive sensing technology to obtain behavioral data on caregiver–child interaction. Specifically, for the present study, we address the question of whether mothers in a rural LMIC setting find their participation and participation of their children to be ethical. We operationalize ethics using the Emanuel et al. framework (Emanuel, Wendler, Killen, & Grady, 2004) adapted for social science research by Wassenaar and Mamotte (2012). The eight key ethical benchmarks of the Emanuel et al. (2004) framework (Wassenaar & Mamotte, 2012) are collaborative partnership, social value, scientific validity, favorable risk/benefit ratio, informed consent, independent review, fair selection of participants, and ongoing respect for participants. To explore these domains, this article presents the perspectives of mothers participating in automatic background collection of "social" data about their location (GPS), who they were with (Bluetooth proximity beacons) and what they spoke about (episodic audio recordings).

Method

Context and Setting

The current study on the ethics of passive data collection was initiated in the context of a broader study exploring caregiver–child interactions. Passive data collection presents new opportunities to understand naturally occurring

caregiver–child interactions, which are important to healthy child development. Caregivers' behaviors can lead to differential neurocognitive development early in life with large differences in language and memory development evident before 2 years of age (Grantham-McGregor et al., 2007). For example, differences in patterns of speech between mother and child can result in a cascade of effects that are directly relevant to a child's cortical development during early life (Farah et al., 2006; Ronfani et al., 2015). Early childhood deprivation in these domains is most prevalent in LMICs (Grantham-McGregor et al., 2007; S. P. Walker et al., 2007). In our broader initiative, we hypothesize that by providing caregivers in low-resource settings with insight into the world as seen through the eyes of their children that caregiver investment can be enhanced in activities predictive of linguistic and cognitive development. Providing this insight for caregivers will be based on leveraging passive sensing technology to obtain behavioral data on caregiver–child interaction.

The larger study is being conducted in rural regions of South Africa and Nepal with high risk for developmental, socioemotional, and educational delays of children. The study is divided into three phases. Phase 1 (described in Kohrt et al., 2019) was a qualitative assessment of beliefs and preferences related to passive digital data collection using videos demonstrating different types of technology. Videos were developed in South Africa and Nepal demonstrating the technologies, but participants did not use devices and there was no passive digital data collection. The qualitative data were used to select technologies for piloting in the present study.

Phase 2 (described in this article) was initiated at the suggestion of the South African Human Sciences Research Council (HSRC) REC as an intermediate step before proceeding to data collection with mobile devices, which were ranked as feasible and acceptable in Phase 1. The research question we aimed to answer in Phase 2 was whether the ethical concerns expressed by the REC about embedding passive data recording devices in the home were justified or not. In Phase 2, participants were given devices to collect data. The experience and ethics of collection were then discussed with participants. Then, data were destroyed. No passively collected digital data were retained for analysis by the research team. Phase 2 was conducted only in South Africa due to concerns held by the South African REC. Based on the results of Phase 2, we will decide whether to proceed to Phase 3 and review the methods that will be used in Phase 3. Phase 3 will entail collection of passive data that would then be analyzed by the study team.

In South Africa, the site for all three phases is a peri-urban community located in the Greater Edendale Area (GEA) of KwaZulu-Natal. The sub-district has a population of approximately 500,000 and is characterized by high unemployment as illustrated by the provincial unemployment rate of 39%. Per capita income is low with up to 30% of the households in the sub-district reporting an annual income of less than US$1,200 (Statistics South Africa, 2016).

Participants

Data for this, the second phase of the larger study, were collected in seven households in Pietermaritzburg, KwaZulu-Natal, South Africa. Participants were recruited by capitalizing on existing research infrastructure of the HSRC field office in the GEA. Well-established community engagement processes were followed to introduce the study to the community, its leaders, and other stakeholders. Criteria for household inclusion were the presence of a consistent caregiver and a child between 2 and 5 years old. Seven households were purposefully sampled through consultation with our standing Community Advisory Board who advise on all studies undertaken in the area. They were briefed on the study and the characteristics of the households required. Key characteristics included at least one child aged between 2 and 5 years in the home, a consistent and present caregiver, at least one male in the home, and no known history of violence or abuse in the household.

Local political and traditional leadership were then consulted in the community from which the households were to be drawn. Initial participation interest was sought from the household head and the child's caregiver. If the head of household expressed interest in the study, a date and time were set for the research assistant to return to speak with other household members, the caregiver, and child to establish whether the entire family was in agreement about participation. During this time, the household head was asked to discuss the study with all potentially affected household members so that on returning, a final informed decision could be made as to whether the family wished to participate. Reimbursement equivalent to US$8 for participation was approved by the REC that reviewed and approved this study.

If interest was expressed on the return visit, a short screening survey was administered to ensure the eligibility of child and caregiver for inclusion in the study. Inclusion criteria were that the child must have been born full-term, healthy birth weight, no known disabilities or impairments, and no expressed caregiver concerns about the child's development. The technical elements and associated passive data collection devices (see Figure 1) were introduced to participants through the use of pre-recorded videos. The three videos each showed the devices in action and the data that each generated to increase participants' understanding of the technology.

We selected a sample size of seven caregiver–child pairs. Hagaman and Wutich (2017) suggest that six qualitative interviews can generate meta-themes in cross-cultural research. We planned to conduct seven interviews, and
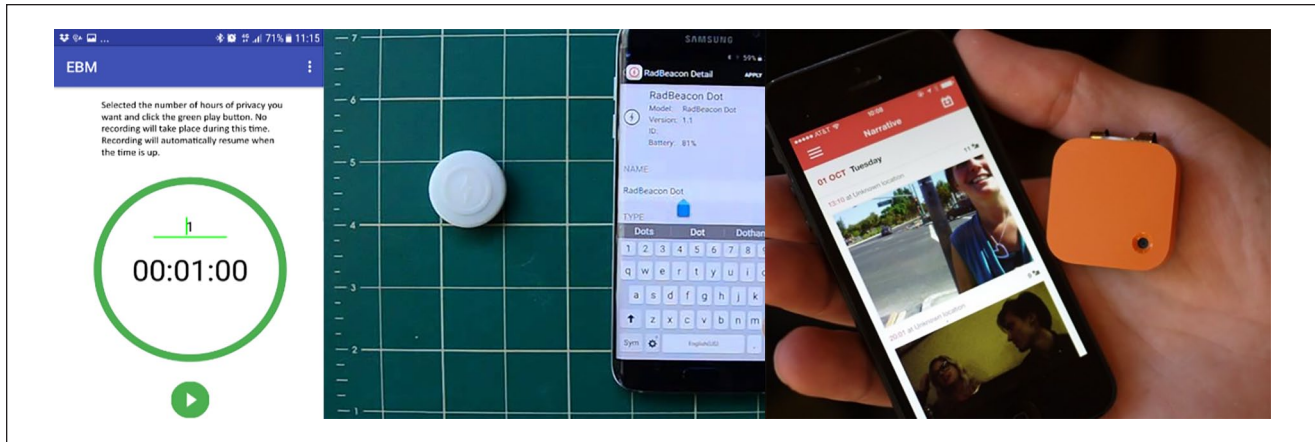
Figure 1. From left to right: The time-lapse camera worn by the child, Electronic Behavior Monitor app used to collect episodic audio, Bluetooth proximity and GPS data, and the Bluetooth beacon worn by the child.
Note. GPS = Global Positioning System.

explore whether consistent ethical perspectives were reported among participants and whether any major concerns were raised. If views were inconsistent, it would have been possible to increase the number of interviews.

Passive Sensing Technology

Video 1: Periodic photographic image capture. A wearable time-lapse camera was demonstrated as a technology that could be used to capture images from the child's point of view (see Figure 1). The video demonstrated various ways the device could be attached to the child's clothing, how to turn it on and off, and examples of the resulting images (https://bit.ly/2O7ObWB).

Video 2: Periodic audio recordings. Participants were introduced to the Electronic Behavior Monitor (EBM v1.2; Figure 1) app, which was designed to collect episodic audio samples in the home environment. Recordings were automatically captured by the EBM app every 15 min for 30 s at a time between 6:00 and 19:00. Audio samples were compressed and saved in a user accessible folder on the device with each 30-s clip being roughly 200 KB in size. A video demonstrated audio recording in the home (https://bit.ly/2pz9RAe).

Video 3: Bluetooth proximity beacon recordings. A Bluetooth beacon (Figure 1) was introduced in the video as a way to evaluate when the target caregiver and target child were in close proximity. This was illustrated through a coin-sized plastic toy that could be attached to the child's clothing and that sent out a signal that could be received by a mobile phone carried by the caregiver. A mock output was shown in the video to simulate data that could be collected while the caregiver and child spent time together (https://bit.ly/2pvdrv3).

Procedures

After the demonstration videos were reviewed, informed consent was obtained and a baseline survey undertaken. Appointments were made for a researcher to return to discuss the research with other household members to establish their willingness to participate. After consent was obtained from all adult family members, the devices were installed. The devices remained in the home for 10 days. During the first 5 days, three visits were made to the home by the research team (Days 1, 2, and 5) to assess whether the devices had caused discomfort or tension. At each visit, caregivers were reminded of their right to withdraw without penalty, how to turn off devices should they wish to do so, and how to review and delete images and sound clips that had been recorded. Households were then visited on the 11th day for the exit interview, prior to which the caregiver reviewed the audio, image, and proximity data with household members. All digital data were then deleted from the devices and the only data retained for analysis were the participants' exit ($n = 7$) and in-depth ($n = 5$) interviews. In-depth interviews focused on understanding caregivers' experiences of having the devices in the home and provided an opportunity to provide feedback on the audio, image, and proximity data that they had collected and reviewed. Thematic areas covered were feasibility of using the devices in the home, acceptability of using the devices in the home, and thoughts and feelings (after review) toward the data collected (see the appendix). Qualitative data were coded according to seven of the eight domains of the Emanuel et al. (2004) ethical framework (Wassenaar & Mamotte, 2012): social value, scientific validity, favorable risk/benefit ratio, informed consent, independent review, fair selection of participants, and ongoing respect for participants. One of Emanuel et al.'s domains "collaborative partnership" was not

addressed as it was already a requirement through the engagement with a Community Advisory Board. Again, it must be emphasized that in the phase of the study reported here, the research team did *not* view the recorded data.

Results

Willingness to Participate and Demographics

Eleven households were approached to participate in the study. During recruitment, four households declined participation due to concerns about privacy, confidentiality, and the lack of study incentives. In one instance, the caregiver and the head of the household disagreed about participation a few days into the study. After discussion between household members, the caregiver withdrew their consent and data collection was ended. Two participants declined to participate in the in-depth interviews, one due to time constraints and the other due to loss of interest. The median age of children who wore the proximity beacon was 4 years (range = 3-5 years, four females), and the median age of the caregivers was 29 years (range = 22-50 years, seven females).

Domains of Emanuel Et al.'s Ethical Framework

Social value. To be socially valuable, the research should address questions relevant and of importance to society or to particular communities (Emanuel et al., 2004; Wassenaar & Mamotte, 2012). Participants observed that data generated could provide insight into the world as seen through their children's eyes, potentially improve the amount of quality time they spend with their children, and facilitate more constructive interactions with children, for example:

> I thank you for the project like this to arrive here [sound of clapping hands in the background] it will help a lot of mothers, because it is important to have time with the child and see it, there is that tendency of coming back from work then you go to your neighbor because you saw that the child is right but this project it will help you to know that at 9 or 10 I am with him, ok that means for 4 hours I am absent it will help us as parents to see, to create more time with the child. (Respondent 1)

> I attend a lot of things that have to do with children and mothers in this place that I moved to since I am a caregiver; I can see that this place still needs a lot of help, this place is like umm . . . when you hear of places like those that are rented out (*imqasho*), they lack adequate ways of raising children, even the language they use when talking to their children is not at their children's level, it's like at home for instance there are certain things that are not meant for children or there are things that aren't allowed in it, or since we are Christians there are things God does not want us talking about. (Respondent 3)

Respondent 3 emphasized the important gap that could be addressed by the data generated by the study. She identified a need for skills and capacity building in terms of parenting and caregiving ("they lack adequate ways of raising children"). This points to the value that such technologies and data present for developing tailored behavioral interventions for communities. Furthermore, a key indication of their perception of the value of the research was that participants found the technology and surveillance acceptable and reported willingness to participate in similar studies in the future.

Scientific validity. Scientific validity is also an important ethics indicator (Emanuel et al., 2004). While participants did not reflect on the scientific validity of the study design and analysis, a few discussed whether the presence of the devices led to any significant behavior changes or socially desirable behaviors. While some noted that they were more aware of the whereabouts of their child, none believed that the presence of the devices caused them to moderate their behaviors toward their children. One participant described the process as "normal" and another noted,

> Because I am sure if you were to tell her (the child) what is this, I am sure he would have changed her behavior because my mother wanted to say "the police are watching you" then I said don't say that because she would then change the way she does things, because they want to see what kind of a person is she, but she frequently asked "mama what is this" I would say it is a clip that makes you beautiful that's all. (Respondent 1)

Favorable risk/benefit ratio. A favorable risk/benefit ratio requires a fair distribution of risks and benefits of research (Emanuel et al., 2004; Wassenaar & Mamotte, 2012). Respondents in our study identified the potential for improved understanding of parent–child dynamics and increased awareness of the amount of quality time spent with children as important potential benefits (see "Social Value" section). Key risks were identified by all respondents and included that the camera could capture nudity and sex acts, as well as recording and capturing private conversations. However, in efforts to minimize such risks, all families were encouraged to review all footage before being interviewed by the researchers and also encouraged to pause the recordings for conversations or footage that they did not want to be recorded. For example,

> What I noticed is that [name of child] father, who hardly visited the house asked me when I explained the study to him was if this records everything and if all of us in the house would be heard, I then said yes. There are some conversations that we want to have alone, so we would remove the phone or step outside because we couldn't have our discussions which weren't part of the research present. (Respondent 3)

One participant spoke about the potential for physical harms as a result of having a mobile phone:

> R5: . . . I was careful when it comes to the phone.
> I: How?
> R5: So that the para's [slang for local thugs] won't take it from me when I have it at school, for it not to fall and crack the screen.

Informed consent. Informed consent requires that all participants are provided with sufficient information about the study, including risks and benefits. Participants must have capacity to consent, understand what they are consenting to, provide consent voluntarily, and document the consent in writing (Wassenaar & Mamotte, 2012). Careful and thorough consent processes were highlighted by participants as critical, particularly when introducing such novel technologies. One participant articulated the need for clearer more detailed information in similar future studies:

> From the start when they introduced it I was very hesitant because I had not received an adequate explanation about it as well as the videos that would be deleted, but when I started I told myself that what I did with my child here at home, I don't think it's disgraceful in anyway, so I think it would be a learning curve for us parents. There are things that we say to children but if there is something which could expose the type of person you are to other people, it would make us better people around children and lead to us speaking in an appropriate manner towards them. (Respondent 3)

Clear explanations for third-party consent to be captured on camera or audio-recorded were also emphasized by our participants. In the present study, participants received training on how to deal with third-party consent. It was explained that in the presence of non-family members, it was important to explain the study and make sure people were comfortable that they may be both audio-recorded and appear in photographs. Where third parties did not agree to have themselves recorded by the various data collection tools, then study participants would turn off the devices for that period.

Independent review. The sixth principle specifies that research should be reviewed by an independent and competent REC/IRB prior to implementation (Emanuel et al., 2004; Wassenaar & Mamotte, 2012). This study was approved by the HSRC REC (REC 6/18/05/16).

Fair selection of participants. Fair participant selection requires that the population selected for the study should be those to whom the research question applies (Emanuel et al., 2004). Participants in this study were purposively selected in line with the scientific goals of the study, anticipating future

child development and child/caregiver interaction studies with the technology involved.

Ongoing respect for participants. Finally, ongoing respect for participants and study communities requires that participants be treated with respect during and after the study by ensuring confidentiality and informing them of study results (Emanuel et al., 2004). Despite concerns about privacy and confidentiality, most participants (*n* = 4) were comfortable, in principle, with allowing the researchers to review the collected audio, proximity, and camera data:

> You can listen to voice clips and you can even look at the pictures there is no problem. (Respondent 5)

Discussion

The Emanuel framework (Emanuel et al., 2004; Wassenaar & Mamotte, 2012) was found to accommodate the concerns and perspectives of study participants. Participants did not directly raise certain issues such as fair subject selection, independent review, and scientific validity. Of considerable importance was the consent procedure as participants had limited familiarity with the processes and implications of sensor data captured by embedding technologies into their daily lives. Participants were also not always clear about how the underlying technologies worked and what information could be recorded. Rather than voicing their concerns and asking for clarity, most participants chose to simply accept the information they were given. Digital devices are not culturally neutral and beyond providing an understanding of the underlying technology, a meaningful consent process must take into account cultural norms and family preferences.

In an effort to address some of these issues and to improve the quality of the consent process for researchers wanting to use passive digital sensing devices, we developed the Qualitative Cultural Assessment for Passive sensing Data collection Technology (QualCAPDT) procedure (Kohrt et al., 2019). This approach builds on common cultural anthropology methods that include a suite of structured elicitation tasks and the development of videos describing in detail how the technology works, as described in the "Method" section of this article. People's views of the research use of their social media data may be similarly overly trusting with many not appreciating the extent to which rapid advances in data science and machine learning can be linked to create full and rich profiles that can be mined for profit or bona fide research. To address this problem, we propose that in addition to careful and detailed consent written in clear non-technical jargon and the use of methods such as QualCAPDT, a shift may be required in the thinking of researchers and RECs. A test of understanding might similarly be considered (e.g., Lindegger et al., 2006).

To be ethical, research must meaningfully involve and engage communities in decisions about the research throughout the research cycle—from conceptualization to results dissemination (Emanuel et al., 2004). To highlight the ethical implications, for researchers and RECs/IRBs on the growing wave of passive digital data streams that will soon be collectable about people's behavior, we provide two hypothetical examples of possible ways forward. First, rather than trying to control and regulate the data collection phase, the post-collection analysis and use phases of the research cycle need to be more carefully targeted and scrutinized. Giving access to and control of the data to participants, even after it has been collected, might empower participants to regulate ongoing use of their data. This control could extend to consent being required for each new analysis. Initial consent would be required for data collection to begin but once enrolled, passive data collected through social media and digital sensors could become available to researchers who might then consent participants for each use of the data they contribute to an analysis. Hand (2018) argues that if consent is to be meaningful, it requires (a) an understanding of what the data might be used for in the future and (b) an understanding of how the data are to be used. Although these principles may, on initial review, appear to contradict the proposed approach, they could rather be seen as complementary. Initial consent could lay out possible future uses of the data with a clear statement that any future use would be preceded by a reengagement that specifically addresses for what purpose and exactly how the data will be used.

Alternately, "Broad consent" is another option being discussed in the field of data sharing (Bull et al., 2015). The literature is currently mixed on the universal applicability of broad consent (Moodley, Sibanda, February, & Rossouw, 2014). An extreme contrasting stance might be that passive digital data collected for research purposes should either not be shared or that they should be destroyed soon after the initial analysis. The social credit or trust assessment systems introduced by the Chinese government, if mismanaged, give a particularly Orwellian feel to how these data could be abused by governments to control and punish citizens. The national system will use mass surveillance to access citizens' economic and social reputation. This score will then be used to reward or punish citizens for specific behaviors. When fully operational, punishments generated by thes system might include travel bans, slow Internet service and exclusion from certain hotels (Chen & Cheung, 2017). In a similar vein, Google recently had an internal thought piece, titled the "Selfish Ledger," leaked to the public. In the video, they speculate that "digital DNA," the digital model of a person that is built up from online presence and passive data collected from mobile phones and other sensors embedded in the environment, might be used to predict, influence, and possibly shape human behavior in a direction toward a desired result (Sedkaoui, 2019). If the risk existed that researchers might be forced or have data taken by governments or corporations, it might increase the risk/benefit ratio such that the research is unethical and ineligible for ethics approval.

These findings also raise important considerations about privacy, culture, and gender. In the context of caregiving roles, especially motherhood, there may be different trade-offs for privacy and benefit. If information is being collected to improve health and developmental outcomes of one's children, then there may be less concern about privacy. This would predict that persons in a household with fewer caregiving duties might be less supportive of the same technology and data collection. For example, in Phase 1 of this study, we found that although mothers and community health workers in Nepal were enthusiastic about the passive data collection, some husbands and mothers-in-law were not supportive of the data collection (Kohrt et al., 2019). Therefore, future studies should examine the influence of culture, gender, and other factors on perception of risk–benefit trade-offs and willingness to participate in this type of research.

While these partially formed examples introduce their own issues and ethical quandaries, they were intended to invoke thought and spark debate, both which will undoubtedly be required to reimagine the research ethics process in a world where data about people and their behavior are continually being produced by their devices, online usage patterns, and interaction with the environment. This study had several limitations. First, although large amounts of sensor data were collected and then discarded, the study had a small number of participants. This was reduced further by two of the seven caregivers not being available for in-depth interviews. That said, no major concerns were raised by participants, which allowed proceeding to Phase 3. In Phase 3, qualitative data will be collected from a larger sample to further explore these themes. Moreover, because data were destroyed after the qualitative interviews, participants did not have the opportunity to reflect on ethics after their data were used by the study team. In Phase 3, we will be able to obtain more real-world perspectives because the data will be collected and analyzed by the team, thus allowing participants more time to reflect on sharing their data. Another limitation was that the qualitative data collected expressed only the views of those caregivers in the study. In Phase 1, we previously collected qualitative data from family members (Kohrt et al., 2019). Because these were all female, it is possible that male caregivers would report different experiences of the data collection technique.

Best Practices

Our data suggest that prior engagement ("Collaborative partnership"; Emanuel et al., 2004) with research participants regarding intrusive digital research technologies can assist in their understanding of the technologies used and the data produced. Our data also suggest that an existing research ethics framework (e.g., Emanuel et al., 2004) can

be applied to elucidate the particular ethical issues inherent in such digital research. This may not of course be the case for digital studies involving technologies and data sets other than those described in this article.

Although not arising directly from our data, researchers and RECs will need to consider how privacy will be protected in the reporting of studies using passive sensor data, particularly where the researcher wants to publish captured images rather than just use images as raw data. Images pose a much greater threat to privacy and anonymity than do narrative description of images. This also has implications for the informed consent process.

Educational Implications

The new wave of researchers that will increasingly be using digital research technologies of various kinds will need to be exposed in advance to careful discussion of the particular ethical issues associated with each particular technology, applying principles such as the Emanuel et al. (2004) framework if no more specific guidance is available.

Research Implications

It is clear that further empirical research with researchers, REC members (Hibbin et al., 2018), and target communities is required to uncover specific ethical concerns associated with specific emerging digital research technologies—for example, using captured images in reports as opposed to treating images only as raw data. Such data, in turn, should be carefully interpreted using known ethics principles and frameworks to inform best research practices and guide ethics reviewers (Ives et al., 2018).

Conclusion

In conclusion, this study suggests that even relatively disadvantaged communities, if sensitively approached and carefully and systematically engaged, can enthusiastically understand and allow the use of an invasive digital research technology. This is more likely to be the case if researchers embrace and apply the required ethical sensitivity.

Appendix

Interview Guide

Through a child's eyes: Technological solutions to improve early child development in low-resource settings—Interview Guide

Introduction (2-5 min). (*Following consent from participants*) Thank you for agreeing to participate in this discussion. My name is (facilitator name) and this is my colleague (community caregiver name). We are interested in learning more about your opinions and experiences with the audio and camera devices you recently had in your home. We also want to hear your thoughts about how to improve the experience of having these devices in your home. We are going to ask you some questions about these things.

This discussion should take about an hour to an hour and a half. Keep in mind that there are no right or wrong answers. If you feel uncomfortable at any time, you can choose not to answer a question or end the interview. Everything that you say in here will be kept confidential. The information may be shared with other members of the research staff, but the information that you share with me today will only be used for this research project.

We will be taping this conversation so that we can obtain complete and accurate answers. No personal identifiers, like you name, will be attached to either this tape or any other documents related to this interview during the transcription process.

Do you have any questions? May we begin the discussion?

A. Feasibility (20-30 min)
   1. To start, can you tell us about the 2 weeks in which you had the devices in your home. We are interested in your overall experience, including how long it took for you to get use to them being there, how others felt about the devices, and how your child responded to wearing the camera.
   2. Did you have any technical challenges and if so what where they?
   3. How did you resolve these challenges?
   4. Did you find one device better and easier to use than the other? If so which one and why?
   5. Do you think it is feasible for other people in this community to have these devices in their home for a short period of time? Please elaborate on either why or why you do not think so.
   6. How often did you have to charge the devices and did they ever stop working?
   7. How often and under what circumstances did you turn the devices off?
   8. Did the device make you feel unsafe or uncomfortable during the study? Probes: devices safety, personal safety, interpersonal safety.
   9. Were there any days that were atypical or unusual during this week period?

B. Acceptability (20-30 min)
   1. Do you think you or your child change the way you do things because the device was in your home? How do you think it changes the child's behavior, your behavior, other family member's behavior?
   2. Can you tell us about your general experience of having the devices in the home?

3. Can you share with us an interesting story that happened in your home over the past 2 weeks that was in some way related to having the devices in the home?

4. Could you describe people's reaction to having snippets of their conversation recorded?

5. Could you describe people's reaction to having your child wearing the camera and taking pictures?

6. Do you think these devices would be accepted for use in the homes of your friends and family? If so why?

7. How do you think the information you have gained from the images and audio have impacted the way you interact with your child?

C. Image and data review and feedback (20-30 min)

1. Did you delete any pictures or audio clips or find any of them disturbing? If so, please explain.

2. What emotions did you experience when reviewing the audio clips and images? PROBE, if necessary, for additional information: Was anything embarrassing recorded? Was anything unexpected recorded?

3. How would you feel about having to hand the images and audio over to the research team for review?

4. In addition to being able to delete any images or audio clips and knowing how to turn the devices on and off, could anything else be done to make you feel more at ease about sharing the data with the research team in the future?

5. Were there lots of images or sounds recorded of visitors to the home?

D. Conclusion (5-10 min)

We are now near the end of the interview.

1. Are there any potential areas or topics that you feel are important to your community that we have not asked about?

2. Are there any other questions or comments that you have for us?

That is the end of the interview. We really appreciate your willingness to speak to us about these issues.

Authors' Note

All authors approved the final version of the manuscript.

Acknowledgments

Declaration of Conflicting Interests

Funding

ORCID iDs

Alastair van Heerden https://orcid.org/0000-0003-2530-6885
Doug Wassenaar https://orcid.org/0000-0003-0839-9231
Zaynab Essack https://orcid.org/0000-0002-8867-3415

References

Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: Research challenges and future directions. *The Journal of Supercomputing*, *72*, 1494-1516. doi:10.1007/s11227-016-1677-z

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*, 662-679. doi:10.1080/1369118X.2012.678878

Bull, S., Cheah, P. Y., Denny, S., Jao, I., Marsh, V., Merson, L., . . . Parker, M. (2015). Best practices for ethical sharing of individual-level health research data from low- and middle-income settings. *Journal of Empirical Research on Human Research Ethics*, *10*, 302-313. doi:10.1177/1556264615594606

Chen, Y., & Cheung, A. S. Y. (2017). The transparent self under big data profiling: Privacy and Chinese legislation on the social credit system. *The Journal of Comparative Law*, *12*, 356-378. doi:10.2139/ssrn.2992537

Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, *77*, 120-132. doi:10.1016/j.jbi.2017.12.008

Emanuel, E. J., Wendler, D., Killen, J., & Grady, C. (2004). What makes clinical research in developing countries ethical? The benchmarks of ethical research. *The Journal of Infectious Diseases*, *189*, 930-937. doi:10.1086/381709

Estrin, D. (2014). Small data, where n = me. *Communications of the ACM*, *57*(4), 32-34. doi:10.1145/2580944

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., . . . Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, *1110*, 166-174. doi:10.1016/j.brainres.2006.06.072

Gallagher, S. (2018, March 25). Facebook scraped call, text message data for years from Android phones. *Ars Technica*. Retrieved from https://arstechnica.com/information-technology/2018/03/facebook-scraped-call-text-message-data-for-years-from-android-phones/

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet, 369*(9555), 60-70. doi:10.1016/S0140-6736(07)60032-4

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big Data*, *6*, 176-190. doi:10.1089/big.2018.0083

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data—A taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, *36*, 1382-1406. doi:10.1108/IJOPM-02-2014-0098

Hagaman, A. K., & Wutich, A. (2017). How Many Interviews Are Enough to Identify Metathemes in Multisited and Cross-cultural Research? Another Perspective on Guest, Bunce, and Johnson's (2006) Landmark Study. *Field Methods*, *29*(1), 23-41. doi:10.1177/1525822X16640447

Hibbin, R. A., Samuel, G., & Derrick, G. E. (2018). From "a fair game" to "a form of covert research": Research ethics committee members' differing notions of consent and potential risk to participants within social media research. *Journal of Empirical Research on Human Research Ethics*, *13*, 149-159. doi:10.1177/1556264617751510

Insel, T. R. (2017). Digital phenotyping. *Journal of the American Medical Association*, *318*, 1215-1216. doi:10.1001/jama.2017.11295

Ives, J., Dunn, M., Molewijk, B., Schildmann, J., Bærøe, K., Frith, L., . . .Widdershoven, G. (2018). Standards of practice in empirical bioethics research: Towards a consensus. *BMC Medical Ethics*, *19*(1), Article 68. doi:10.1186/s12910-018-0304-3

Kara, H. (2018). *Research ethics in the real world: Euro-Western and Indigenous perspectives*. Bristol, UK: Policy Press.

Kohrt, B., Rai, S., Vilakazi, K., Thapa, K., Bhardwaj, A., & van Heerden, A. (2019). Procedures to select digital sensing technologies for passive data collection with children and their caregivers: Findings from a qualitative cultural assessment in South Africa and Nepal. *Journal of Medical Internet Research: Pediatrics and Parenting*, *2*, e12366. doi:10.2196/12366

Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2018). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439318816389

Lindegger, G., Milford, C., Slack, C., Quayle, M., Xaba, X., & Vardas, E. (2006). Beyond the checklist: Assessing understanding for HIV vaccine trial participation in South Africa. *Journal of Acquired Immune Deficiency Syndromes*, *43*, 560-566. doi:10.1097/01.qai.0000247225.37752.f5

Maher, N. A., Senders, J. T., Hulsbergen, A. F. C., Lamba, N., Parker, M., Onnela, J.-P., . . . Broekman, M. L. D. (2019). Passive data collection and use in healthcare: A systematic review of ethical issues. *International Journal of Medical Informatics*, *129*, 242-247. doi:10.1016/j.ijmedinf.2019.06.015

Martinez-Martin, N., Insel, T. R., Dagum, P., Greely, H. T., & Cho, M. K. (2018). Data mining for health: Staking out the ethical territory of digital phenotyping. *npj Digital Medicine*, *1*(1), Article 68. doi:10.1038/s41746-018-0075-8

McKee, R. (2013). Ethical issues in using social media for health and health care research. *Health Policy*, *110*, 298-301. doi:10.1016/j.healthpol.2013.02.006

Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, *88*, 300-305. doi:10.1016/j.procs.2016.07.439

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, *22*, 303-341. doi:10.1007/s11948-015-9652-2

Moodley, K., Sibanda, N., February, K., & Rossouw, T. (2014). "It's my blood": Ethical complexities in the use, storage and export of biological samples: Perspectives from South African research participants. *BMC Medical Ethics*, *15*, Article 4. doi:10.1186/1472-6939-15-4

Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking*, *16*, 708-713. doi:10.1089/cyber.2012.0334

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, *10*, 40-81. doi:10.1080/17579961.2018.1452176

Ronfani, L., Vecchi Brumatti, L., Mariuz, M., Tognin, V., Bin, M., Ferluga, V., . . . Barbone, F. (2015). The complex interaction between home environment, socioeconomic status, maternal IQ and early child neurocognitive development: A multivariate analysis of data collected in a newborn cohort study. *PLoS ONE*, *10*(5), e0127052. doi:10.1371/journal.pone.0127052

Salerno, J., Knoppers, B. M., Lee, L. M., Hlaing, W. M., & Goodman, K. W. (2017). Ethics, big data and computing in epidemiology and public health. *Annals of Epidemiology*, *27*, 297-301. doi:10.1016/j.annepidem.2017.05.002

Samuel, G., Ahmed, W., Kara, H., Jessop, C., Quinton, S., & Sanger, S. (2018). Is it time to re-evaluate the ethics governance of social media research? *Journal of Empirical Research on Human Research Ethics*, *13*, 452-454. doi:10.1177/1556264618793773

Sedkaoui, S. (2019). *Big data analytics for entrepreneurial success*. Hershey, PA: IGI Global. doi:10.4018/978-1-5225-7609-9

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*, 161-167. doi:10.1007/s12525-015-0191-0

Statistics South Africa. (2016). *Quarterly Labour Force Survey: Quarter 1, 2016*. Pretoria: Author.

Subbian, K. (2018). Patent No. *20180352383*. Retrieved from http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220180352383%22.PGNR.&OS=DN/20180352383&RS=DN/20180352383

van Baalen, S. (2018). "Google wants to know your location": The ethical challenges of fieldwork in the digital age. *Research Ethics*, *14*(4), 1-17. doi:10.1177/1747016117750312

Walker, C., & Alrehamy, H. (2015, August 26-28). *Personal data lake with data gravity pull*. 2015 IEEE Fifth International

Conference on Big Data and Cloud Computing, Dalian, China. doi:10.1109/BDCloud.2015.62

Walker, S. P., Wachs, T. D., Meeks Gardner, J., Lozoff, B., Wasserman, G. A., Pollitt, E., & Carter, J. A. (2007). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet, 369*(9556), 145-157. doi:10.1016/S0140-6736(07)60076-2

Wassenaar, D. R., & Mamotte, N. (2012). Ethical issues and ethics reviews in social science research. In A. Ferrero, Y. Korkut, M. M. Leach, G. Lindsay, & M. J. Stevens (Eds.), *The Oxford handbook of international psychological ethics* (pp. 268-282). New York: Oxford University Press. doi:10.1093/oxfordhb/9780199739165.013.0019

Author Biographies

Alastair van Heerden is an associate professor in the School of Clinical Medicine, Faculty of Health Sciences at the University of the Witwatersrand, and a research director at the Human Sciences Research Council, South Africa. His area of expertise is in the application of pervasive computing to public health in low- and middle-income countries. He is the study co-PI (co-principal investigator) and was responsible for the initial draft of the manuscript, and has contributed to the collection of the empirical data and to the analysis, drafting, and revision of the manuscript.

Doug Wassenaar is professor in the School of Applied Human Sciences at the University of KwaZulu-Natal and is the principal investigator (PI) of SARETI—the South African Research Ethics Training Initiative. He has chaired two research ethics committees and has published widely on research ethics. He contributed in various ways to the work reported in this article, including data interpretation, writing, and approval of final version.

Zaynab Essack is a senior research specialist in the Human and Social Development (HSD) research program at the Human Sciences Research Council of South Africa. She holds an MSocSci in research psychology at the University of KwaZulu-Natal, and obtained a PhD in psychology from the University of KwaZulu-Natal (2015) on the topic titled "An Empirical Study of Standards of Prevention in South African HIV Vaccine Trials: Norms, Perspectives and Practices." She performed the data analysis and assisted in drafting and revision of the manuscript.

Khanya Vilakazi is a master's intern in the HSD research program at the Human Sciences Research Council of South Africa. He was responsible for study management, data collection, and quality assurance. He assisted in reviewing the draft manuscript, provided comment on data analysis, and approved the final version.

Brandon A. Kohrt is associate professor of psychiatry and behavioral sciences. As an internationally recognized global mental health expert, he works with populations affected by war-related trauma, torture, environmental disasters, and chronic stressor of poverty, discrimination, and lack of access to health care. He has worked in Nepal for 16 years using a biocultural developmental perspective integrating epidemiology, cultural anthropology, ethnopsychology, and neuroendocrinology. He is the study co-PI and was responsible for reviewing the initial draft of the manuscript. He contributed to the collection of the empirical data, analysis, drafting, and revision of the manuscript, and gave approval to the final submitted version.