

Original Paper

# Understanding the Predictors of Missing Location Data to Inform Smartphone Study Design: Observational Study

Anna L Beukenhorst<sup>1,2</sup>, PhD; Jamie C Sergeant<sup>1,3</sup>, PhD; David M Schultz<sup>4,5</sup>, PhD; John McBeth<sup>1</sup>, PhD; Belay B Yimer<sup>1</sup>, PhD; Will G Dixon<sup>1,6</sup>, PhD

<sup>1</sup>Centre for Epidemiology Versus Arthritis, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

<sup>2</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, United States

<sup>3</sup>Centre for Biostatistics, University of Manchester, Manchester, United Kingdom

<sup>4</sup>Centre for Atmospheric Science, Department of Earth and Environmental Sciences, University of Manchester, Manchester, United Kingdom

<sup>5</sup>Centre for Crisis Studies and Mitigation, University of Manchester, Manchester, United Kingdom

<sup>6</sup>NIHR Greater Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

**Corresponding Author:**

Anna L Beukenhorst, PhD

Centre for Epidemiology Versus Arthritis

Manchester Academic Health Science Centre

University of Manchester

Oxford Road

Manchester, M13 9PL

United Kingdom

Phone: 44 2622757506

Email: [beuk@hsph.harvard.edu](mailto:beuk@hsph.harvard.edu)

## Abstract

**Background:** Smartphone location data can be used for observational health studies (to determine participant exposure or behavior) or to deliver a location-based health intervention. However, missing location data are more common when using smartphones compared to when using research-grade location trackers. Missing location data can affect study validity and intervention safety.

**Objective:** The objective of this study was to investigate the distribution of missing location data and its predictors to inform design, analysis, and interpretation of future smartphone (observational and interventional) studies.

**Methods:** We analyzed hourly smartphone location data collected from 9665 research participants on 488,400 participant days in a national smartphone study investigating the association between weather conditions and chronic pain in the United Kingdom. We used a generalized mixed-effects linear model with logistic regression to identify whether a successfully recorded geolocation was associated with the time of day, participants' time in study, operating system, time since previous survey completion, participant age, sex, and weather sensitivity.

**Results:** For most participants, the app collected a median of 2 out of a maximum of 24 locations (1760/9665, 18.2% of participants), no location data (1664/9665, 17.2%), or complete location data (1575/9665, 16.3%). The median locations per day differed by the operating system: participants with an Android phone most often had complete data (a median of 24/24 locations) whereas iPhone users most often had a median of 2 out of 24 locations. The odds of a successfully recorded location for Android phones were 22.91 times higher than those for iPhones (95% CI 19.53-26.87). The odds of a successfully recorded location were lower during weekends (odds ratio [OR] 0.94, 95% CI 0.94-0.95) and nights (OR 0.37, 95% CI 0.37-0.38), if time in study was longer (OR 0.99 per additional day in study, 95% CI 0.99-1.00), and if a participant had not used the app recently (OR 0.96 per additional day since last survey entry, 95% CI 0.96-0.96). Participant age and sex did not predict missing location data.

**Conclusions:** The predictors of missing location data reported in our study could inform app settings and user instructions for future smartphone (observational and interventional) studies. These predictors have implications for analysis methods to deal with missing location data, such as imputation of missing values or case-only analysis. Health studies using smartphones for data collection should assess context-specific consequences of high missing data, especially among iPhone users, during the night and for disengaged participants.

(*JMIR Mhealth Uhealth* 2021;9(11):e28857) doi: [10.2196/28857](https://doi.org/10.2196/28857)

## KEYWORDS

geolocation; global positioning system; smartphones; mobile phone; mobile health; environmental exposures; data analysis; digital epidemiology; missing data; location data; mobile application

## Introduction

Smartphones offer opportunities to collect sensor data frequently from people's daily lives and to determine their exposures or behaviors. Smartphone location data can be collected frequently (eg, daily, hourly, continuously) over sustained periods of time [1]. Studies have used these data to quantify exposure to weather [2,3], air pollution [4], vicinity to tobacco outlets [5], or to deliver context-aware messages when participants visited health facilities [6,7]. Smartphones can provide complete and accurate location data, especially when participants are provided with study smartphones, studies are short, and data are collected nearly continuously [8,9]. However, in large-scale epidemiological studies, location data are often collected for longer periods, less frequently, and from participants' own smartphones. In these cases, missing data are more common than when using research-grade location trackers [4,10,11]. In observational research studies, missing data can result in the loss of power, selection bias, and misclassification of participants' exposure or behavior [12]. In trials, it could hamper safe and effective delivery of context-aware interventions that rely on location data [13].

To anticipate the potential impact of missing location data on study findings, we need to better understand how often, when, and why location data are missing. Previous smartphone studies have reported the amount of missing location data [4,10,14,15]. However, they typically did not investigate differences in missing data over time [4,10,14,15], between participants [4,10,14,15], or between operating systems [4,14]. In addition, they have limitations of small sample sizes.

We therefore investigated the distribution of missing location data over time, predictors of missing location data, and between-participant differences. We used data from a longitudinal smartphone study with 9665 participants using Android phones or iPhones. We anticipate that understanding the predictors of missing location data could inform researchers who want to improve data completeness during study design and data collection.

## Methods

### Ethics Approval and Consent to Participate

The University of Manchester Research Ethics Committee (reference, ethics/15522) and the National Health Service Integrated Research Application System (reference 23/NW/0716) approved this study. Participants were required to provide electronic consent for study inclusion. Further details are available elsewhere [2,3].

### Study Design

We performed a secondary analysis of the data from an observational smartphone study that analyzed the association

between weather conditions and chronic pain in the United Kingdom (study name: *Cloudy with a Chance of Pain*) [3]. In this study, we collected self-reported pain levels from a large group of people with chronic pain such as osteoarthritis, rheumatoid arthritis, or migraine. The exposure of interest was daily average weather conditions (ie, temperature, relative humidity, wind speed, and air pressure). To determine what daily average weather conditions participants were exposed to, the app recorded participants' geolocation, which we could link to weather reports from local weather stations. The analysis of the weather and pain association and the details of data collection are described elsewhere [2,3].

### Data Collection

People with chronic pain downloaded the app onto their Android phones or iPhones, provided informed consent, and reported baseline participant characteristics (eg, sex, year of birth, self-reported weather sensitivity). At local time of 6:24 PM each day, participants received a push notification to complete a survey, rating 10 aspects of symptoms, behavior, and well-being. To obtain weather data from the closest weather station, geolocation was required. The app was programmed to record geolocation each hour on the hour; thus, the app would ideally obtain 24 geolocations each day. The app used GPS (outdoors) and network signals (inside buildings) to determine the latitude and longitude. The app's ability to record geolocations depended on (1) the participant granting the app access to their geolocation and (2) the participant switching on the location services on their phone. Upon downloading the study app, the participants were requested access to their geolocation. Access to geolocation was voluntary; participants who provided the app with access to their geolocation could retract access at any time or switch off location services temporarily or permanently, in which case the app would not be able to record the participant's location. The app recorded the operating system of the smartphone, but this feature was introduced 1 week after the recruitment launch and was not collected for early enrollers.

### Data Preparation and Eligible Participants

We investigated location-data completeness on calendar days that a participant completed the survey. Participants were eligible if they completed the survey at least once, excluding the day of enrollment. This exclusion ensured comparability of participant days, as recording 24 geolocations would be unlikely on the day of download. For each participant, we selected all days with survey data. For each full clock hour, we added indicators for (1) location data (1 if observed, 0 if missing), (2) number of days since the most recent survey completion (0 if less than 24 hours ago, 1 if 24-47 hours ago, etc), (3) time in study (days since first survey submission), and (4) time (weekday or weekend, part of the day where night was considered as midnight to 5:59 AM, morning as 6 AM to 11:59

AM, afternoon as noon to 5:59 PM, evening as 6 PM to 23:59 PM, and hour of the day). In addition, we added indicators for variables that did not change over time: (1) participant characteristics (eg, sex, age, self-reported weather sensitivity) and (2) operating system (eg, iPhone operating system, Android, or unknown).

### Data Analysis

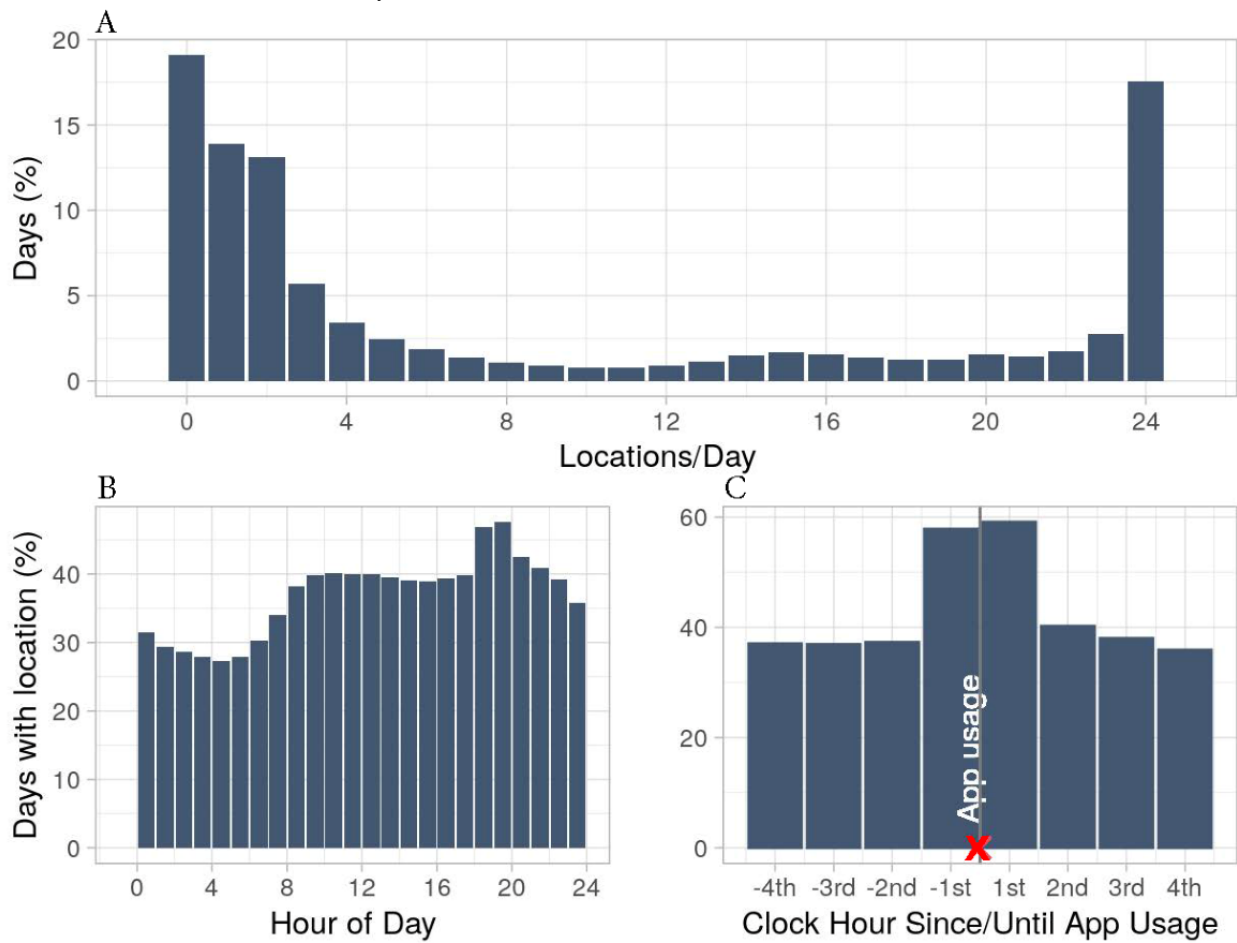
We reported the number of eligible participants and their characteristics. We reported location-data completeness (1) per day (number of recorded locations during a day), (2) per hour for each clock hour (percentage of participant days with a recorded location data at that hour), (3) per hour for the 4 hours before and after survey completion, and (4) averages per participant (median number of recorded locations) for all participants and stratified by operating system. We investigated predictors of the outcome “presence of a location data point” (0 if missing, 1 if observed for a given full clock hour) with a logistic regression model with a participant-specific random intercept for within-participant correlation between repeated measurements [16-18]. A multivariable model identified whether the likelihood of the missing location data were associated with time indicators (ie, weekdays vs weekend days, part of the day), participant characteristics (ie, age, sex, self-reported weather sensitivity dichotomized around the median), operating system on their phone, survey compliance (ie, days since previous survey entry), or time in study (ie, days since first survey entry). Only participants with complete data for all covariates contributed information to the model. We estimated 95% CIs with 1000 simulations as recommended in [19]. Models were fitted in R (R Core Team) version 3.6 with the package lme4

[18]; odds ratios (ORs) and CIs were estimated using the merTools package [20].

## Results

The app was downloaded by 13,207 participants, of which 9665 were eligible for inclusion (mean age 49 [SD 13] years; females, 7211/9665, 74.6%). These participants contributed to 488,400 participant days (median 14 eligible days/participant; IQR 4-60 days/participant). Of 9665 participants, 3109 (32.2%) used an Android phone, 1930 (19.9%) an iPhone, and the operating system was unknown for the remaining 4626 (47.6%) participants. We expected 11.72 million location data points or clock hours: 24 for each hour in the 488,400 participant days. Of 11.72 million hours, the app collected only 4.36 million clock hours (37.2%), resulting in missing data for the remaining 7.36 million clock hours. Data completeness per participant day varied from no location data (0/24) to fully complete data (24/24, Figure 1A, median 3, IQR 1-19). Location data were complete (24/24) for 17.5% (85,606/488,400) of participant days. Participant days with no location (93,255/488,400, 19.1%), 1 location (67,963/488,400, 13.9%), or 2 locations (64,207/488,400, 13.1%) were also common. Location was most often recorded at 7 PM (232,295/488,400, 47.5% of participant days; Figure 1B) just after the default notification of 6:24 PM. Locations were least often recorded between midnight and 6 AM. Location data were often recorded for the hour before survey completion (281,767/487,391, 57.8%) and the hour after survey completion (257,743/436,263, 59.1%; Figure 1C).

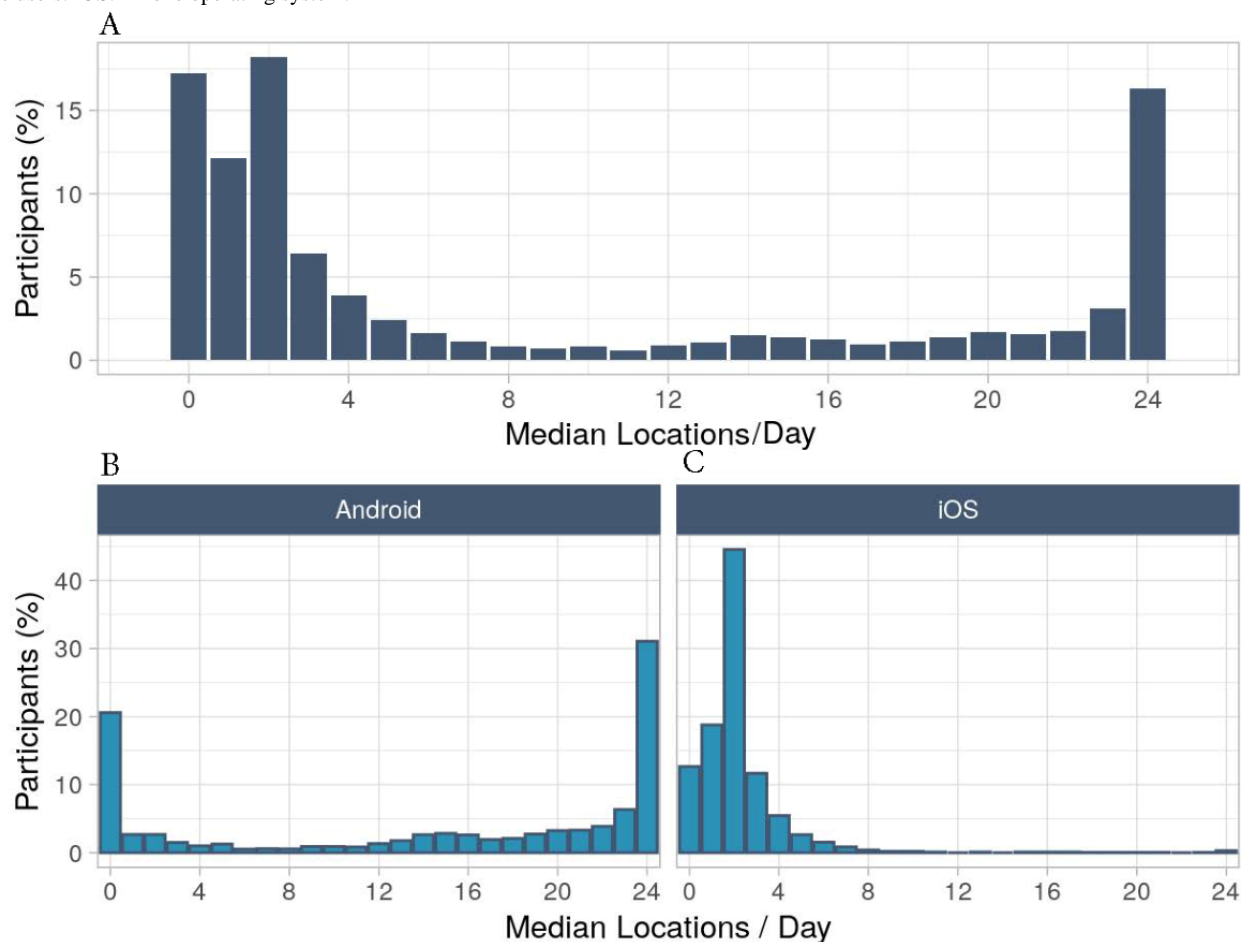
**Figure 1.** Data completeness. A: Distribution of participant days with a recorded location, stratified per hour of the day (N=488,400). B: Data completeness per hour of the day (N=24 x 488,400). C: Data completeness around the moment of survey completion (N=24 x 488,400). The red X marks app usage, and 1st is the first full clock hour after data entry.



For most participants, the app collected a median of 2 out of a maximum of 24 locations (1760/9665, 18.2% of participants), no location data (1664/9665, 17.2%), or complete location data (1575/9665, 16.3%; Figure 2A). Stratification by phone operating system and participant characteristics showed that 31.0% (965/3109) of Android users had 24 recorded locations

on average versus less than 1% (6/1930) of iPhone users (Figure 2B). Android users usually had averages of 24 (out of 24) locations (965/3109, 31.0%) or 0 (out of 24) locations (640/3109, 20.6%). iPhone users usually had averages of 2 (out of 24) (859/1930, 44.5%) or 1 (out of 24) (362/1930, 18.8%) location, while only 0.03% (6/1930) had averages of 24 (out of 24) locations.

**Figure 2.** Median locations per day per participant. A: All participants (N=9665). B: Stratified by operating system for 3109 Android users and 1930 iPhone users. iOS: iPhone operating system.



The generalized linear mixed-effects model estimated whether time indicators, operating system, time since previous survey completion, or participant characteristics predicted the presence of a location data point (N=4435). The presence of a location data point was strongly predicted by the operating system and the part of the day (Table 1). The odds of a recorded location were the highest for Android phones (OR 21.91, 95% CI 19.53-26.87, referent: iPhone operating system) and during the

afternoon (OR 1.18, 95% CI 1.18-1.20, referent: morning). The odds of a recorded location were lower in the weekends (OR 0.94, 95% CI 0.94-0.95, referent: weekdays) and if previous survey completion was longer ago (OR 0.95 per additional day, 95% CI 0.95-0.95) and marginally lower if a participant's time in study was longer (OR 0.998, 95% CI 0.9984-0.9985). Participant characteristics (eg, age, sex, self-reported weather sensitivity) did not predict the probability of location data.

**Table 1.** Results of the generalized linear mixed-effects model estimating the odds of having a recorded location (N=4435).

Variable	category	Odds ratio (95% CI)
<b>Day of the week</b>		
	Weekdays	Referent
	Weekend	0.94 (0.94-0.95)
<b>Part of the day</b>		
	Morning	Referent
	Afternoon	1.19 (1.18-1.20)
	Evening	1.11 (1.10-1.11)
	Night	0.37 (0.37-0.38)
Time in study	Per day	0.99 (0.99-1.00)
<b>Operating system</b>		
	iPhone operating system	Referent
	Android	22.91 (19.53-26.87)
Time since previous survey completion	Per day	0.96 (0.96-0.96)
Age	Per 10 years	1.00 (1.00-1.01)
<b>Sex</b>		
	Female	Referent
	Male	0.99 (0.80-1.22)
<b>Weather sensitivity</b>		
	Weak	Referent
	Strong	0.98 (0.84-1.15)

## Discussion

### Principal Findings

In our study, location data collected from participants' smartphones were missing for 63% of the intended hours (7.36 million/11.72 million). This percentage is higher than that reported in 5 other studies, reporting 26% [4], 28% [14], and 50% [10,15,21] of missing data. This difference may be due to the choices during the analysis: 3 studies excluded participants with the highest amounts of missing data and only investigated Android users, possibly resulting in an underestimation of the overall percentage of missing location data [4,14,21]. The other 2 studies sampled location continuously multiple times per hour for a few minutes, suggesting that our findings may not generalize to higher frequencies of location data collection [10,15].

### Why Do Time Indicators and Operating System Predict Location-Data Completeness?

Missing data were predicted by part of the day, time since previous survey completion, and participants' operating system. Missing data at night might be caused by people being indoors where GPS signals are unavailable [11] or by their phones being switched off in airplane mode or out of battery [11,22]. Location data were most complete in the hour before and after survey completion, showing that apps are more likely to record the last known location upon restarting the app and the location on the clock hour after. In addition, we found a small but significant

reduction in odds of a recorded location over time. Lower location-data completeness when people stay longer in a study is in line with the findings reported previously [22]. Less than 1% of iPhone users had complete location data. Other studies of smartphone data corroborate our finding of higher missing sensor data in iPhone users compared to Android users. iPhone's operating system refuses geolocation requests by apps more often compared to Android. Reasons for refusing geolocation requests are, for example, to reduce the phone's power consumption or to prioritize sensor data collection by other apps [10,15,23,24]. Of note, some studies have succeeded in obtaining higher coverage location data from iPhones compared to Android phones in spite of these operating system-specific differences [22,25]. This finding suggests that the research app used to collect data and the way this app interacts with the operating system may influence the amount of missing data. Experimental studies could further investigate this, as we cannot exclude the role of other differences between this study and our own study, such as the investigated population (eg, mean age 48 years in our study, but mean age 25 years in [22]) and sampling frequency (once an hour in our study; continuously for 1 minute every 10 minutes in [15,22]).

### Implications: Consequences of Missing Data Are Context-Specific

Although missing location data reduce precision, they do not necessarily reduce a study's validity. For example, missing data during the night may not be a problem for a study interested in

identifying daytime behaviors from location data. In our study, we calculated daily average exposure to the weather based on the 24-hourly weather reports from participants' location [3]. For days with missing data, we imputed participant location. As UK weather stations are approximately 40 km apart, missing information on small relocations would not result in assigning participants to the wrong weather station. Furthermore, misclassification would only occur if the weather conditions at the "wrong" weather station were sufficiently different to change a participant's daily average exposure. Most previous studies investigating weather and pain measured participants' location only once and used daily weather reports, rather than hourly [26]. Compared to those studies, weather exposure in our study is less likely to be misclassified, even for participants with only 1 or 2 observed locations per day.

Participant age and sex did not predict missing location data, suggesting that data completeness is not associated with those 2 demographic factors. However, the difference in location-data completeness between iPhone and Android users could be a source of bias. Just-in-time interventions that depend on location data could be less safe and effective for iPhone users compared to Android users. On average, Android users have a lower socioeconomic status than iPhone users—a factor that is related to many health outcomes and may be associated with health disparities in underprivileged groups [27-29]. In observational studies, this difference could introduce selection bias. For example, exclusion of participants with incomplete data (complete case analysis) could lead to results that do not generalize to wealthier iPhone users.

Observational studies could impute missing location data based on participants' past behavior [30,31]. In that case, it is important to assess whether the imputation algorithm is also valid for iPhone users who may have fewer past data points available. If imputation is not feasible, researchers may want to consider using different devices to collect location data, such

as a GPS tracker, which may be more suitable to answer certain research questions requiring complete location data for short periods of time [4,9]. Of note, although the imputation would mitigate some threats to internal validity due to selection bias, they do not address external validity. Study results may still not be generalizable to the wider population, especially not to underserved communities that tend to use health technologies less and may have fewer financial resources to purchase smartphones and pay for connection maintenance [29].

### Improving Location-Data Completeness

At study design, researchers should optimize app settings and user instructions to improve location-data completeness. Our study showed that location was more often recorded around survey completion and around push notifications. Thus, encouraging participants to complete surveys and sending push notifications may improve location-data completeness as well as survey responses. As Android phone users have higher location-data completeness than iPhone users, restricting participation to Android users could improve location-data completeness. However, it could introduce important limitations to generalizability, given that many people have iPhones (market share 27% worldwide [32] and 54% in the United States [33]).

### Conclusion

Missing hourly smartphone location data is common: in our study, 63% of hourly data points were missing. Missing data were more likely for iPhone users, during the night, on weekend days, and if participants had not recently used the app to complete a survey. Participant age and sex did not predict missing location data. Differences in location-data completeness between iPhone and Android users may impact the validity of observational or interventional studies. The predictors of missing data can help researchers at study design to optimize app settings and user instructions for higher location-data completeness. In addition, it may inform their assessment of context-specific consequences of missing location data.

### Acknowledgments

We thank Sabine van der Veer for providing feedback on this manuscript. ALB is supported by a Medical Research Council Doctoral Training Partnership grant (MR/N013751/1). *Cloudy with a Chance of Pain* was supported by Versus Arthritis (grant reference 21225). This research was further supported by the Centre for Epidemiology Versus Arthritis (grant 21755) and the National Institute for Health Research Manchester Biomedical Research Centre.

### Conflicts of Interest

WGD has received consultancy fees from AbbVie and Google, which is unrelated to this work.

### References

1. Amor J, James C. Setting the scene: Mobile and wearable technology for managing healthcare and wellbeing. 2015 Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 25-29 Aug; Milan, Italy p. 7752-7755. [doi: [10.1109/EMBC.2015.7320189](https://doi.org/10.1109/EMBC.2015.7320189)]
2. Schultz DM, Beukenhorst AL, Yimer BB, Cook L, Pisaniello H, House T, et al. Weather Patterns Associated With Pain In Chronic-Pain Sufferers. *Bull Am Meteorol Soc* 2020 May 01;101(5):E555-E566. [doi: [10.1175/BAMS-D-19-0265.1](https://doi.org/10.1175/BAMS-D-19-0265.1)]
3. Dixon WG, Beukenhorst AL, Yimer BB, Cook L, Gasparrini A, El-Hay T, et al. How the weather affects the pain of citizen scientists using a smartphone app. *NPJ Digit Med* 2019;2:105 [FREE Full text] [doi: [10.1038/s41746-019-0180-3](https://doi.org/10.1038/s41746-019-0180-3)] [Medline: [31667359](https://pubmed.ncbi.nlm.nih.gov/31667359/)]

4. Glasgow ML, Rudra CB, Yoo E, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. *J Expo Sci Environ Epidemiol* 2016 Jun;26(4):356-364. [doi: [10.1038/jes.2014.78](https://doi.org/10.1038/jes.2014.78)] [Medline: [25425137](https://pubmed.ncbi.nlm.nih.gov/25425137/)]
5. Lee EW, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, et al. The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study. *J Med Internet Res* 2020 Jul 07;22(7):e17451 [FREE Full text] [doi: [10.2196/17451](https://doi.org/10.2196/17451)] [Medline: [32673252](https://pubmed.ncbi.nlm.nih.gov/32673252/)]
6. Marsh A, Hirve S, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Validating a GPS-based approach to detect health facility visits against maternal response to prompted recall survey. *J Glob Health* 2020 Jun;10(1):010602 [FREE Full text] [doi: [10.7189/jogh.10.010602](https://doi.org/10.7189/jogh.10.010602)] [Medline: [32426124](https://pubmed.ncbi.nlm.nih.gov/32426124/)]
7. Dale LP, White L, Mitchell M, Faulkner G. Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake. *Vaccine* 2019 Jul 26;37(32):4594-4600. [doi: [10.1016/j.vaccine.2018.04.018](https://doi.org/10.1016/j.vaccine.2018.04.018)] [Medline: [29699784](https://pubmed.ncbi.nlm.nih.gov/29699784/)]
8. Donaire-Gonzalez D, Valentín A, de Nazelle A, Ambros A, Carrasco-Turigas G, Seto E, et al. Benefits of Mobile Phone Technology for Personal Environmental Monitoring. *JMIR Mhealth Uhealth* 2016 Nov 10;4(4):e126 [FREE Full text] [doi: [10.2196/mhealth.5771](https://doi.org/10.2196/mhealth.5771)] [Medline: [27833069](https://pubmed.ncbi.nlm.nih.gov/27833069/)]
9. Goodspeed R, Yan X, Hardy J, Vydiswaran VV, Berrocal VJ, Clarke P, et al. Comparing the Data Quality of Global Positioning System Devices and Mobile Phones for Assessing Relationships Between Place, Mobility, and Health: Field Study. *JMIR Mhealth Uhealth* 2018 Aug 13;6(8):e168 [FREE Full text] [doi: [10.2196/mhealth.9771](https://doi.org/10.2196/mhealth.9771)] [Medline: [30104185](https://pubmed.ncbi.nlm.nih.gov/30104185/)]
10. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using Mobile Phone Sensor Technology for Mental Health Research: Integrated Analysis to Identify Hidden Challenges and Potential Solutions. *J Med Internet Res* 2018 Jul 30;20(7):e10131 [FREE Full text] [doi: [10.2196/10131](https://doi.org/10.2196/10131)] [Medline: [30061092](https://pubmed.ncbi.nlm.nih.gov/30061092/)]
11. Vhaduri S, Poellabauer C, Striegel A, Lizardo O, Hachen D. Discovering places of interest using sensor data from smartphones and wearables. In: *IEEE Xplore.*: IEEE; 2017 Presented at: SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI; June 28; San Francisco. [doi: [10.1109/UIC-ATC.2017.8397495](https://doi.org/10.1109/UIC-ATC.2017.8397495)]
12. Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. *Am J Prev Med* 2011 Nov;41(5):508-515 [FREE Full text] [doi: [10.1016/j.amepre.2011.06.046](https://doi.org/10.1016/j.amepre.2011.06.046)] [Medline: [22011423](https://pubmed.ncbi.nlm.nih.gov/22011423/)]
13. Marsh A, Hirve S, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Validating a GPS-based approach to detect health facility visits against maternal response to prompted recall survey. *J Glob Health* 2020 Jun;10(1):010602 [FREE Full text] [doi: [10.7189/jogh.10.010602](https://doi.org/10.7189/jogh.10.010602)] [Medline: [32426124](https://pubmed.ncbi.nlm.nih.gov/32426124/)]
14. Do TMT, Dousse O, Miettinen M, Gatica-Perez D. A probabilistic kernel method for human mobility prediction with smartphones. *Pervasive and Mobile Computing* 2015 Jul;20:13-28. [doi: [10.1016/j.pmcj.2014.09.001](https://doi.org/10.1016/j.pmcj.2014.09.001)]
15. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ Digit Med* 2018;1:15 [FREE Full text] [doi: [10.1038/s41746-018-0022-8](https://doi.org/10.1038/s41746-018-0022-8)] [Medline: [31304300](https://pubmed.ncbi.nlm.nih.gov/31304300/)]
16. Gelman A. Analysis of variance—why it is more important than ever. *Ann. Statist* 2005 Feb 1;33(1):1-53. [doi: [10.1214/009053604000001048](https://doi.org/10.1214/009053604000001048)]
17. Vittinghoff E, Glidden D, Shiboski SM. Repeated measures and longitudinal data analysis. In: *Regression Methods in Biostatistics*. Boston: Springer; 2012:261-308.
18. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Package ‘lme4’. Linear mixed-effects models using S4 classes. R Package Version 2021. 2021. URL: <https://cran.r-project.org/web/packages/lme4/index.html> [accessed 2021-12-10]
19. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press; 2006.
20. Knowles J, Frederick C. merTools: tools for analyzing mixed effect regression models. R Package version 050 2019. URL: <https://cran.r-project.org/web/packages/merTools/index.html> [accessed 2021-10-12]
21. Bähr S, Haas G, Keusch F, Kreuter F, Trappmann M. Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data. *Social Science Computer Review* 2020 Aug 06;089443932094411. [doi: [10.1177/0894439320944118](https://doi.org/10.1177/0894439320944118)]
22. Kiang MV, Chen J, Krieger N, Buckee C, Alexander M, Baker J, et al. Sociodemographic Characteristics of Missing Data in Digital Phenotyping. *Scientific Reports* 2021 Jul 29;11(1):15408 (forthcoming). [doi: [10.1101/2020.12.29.20249002](https://doi.org/10.1101/2020.12.29.20249002)]
23. Beukenhorst AL, Schultz DM, McBeth J, Lakshminarayana R, Sergeant JC, Dixon WG. Using Smartphones for Research Outside Clinical Settings: How Operating Systems, App Developers, and Users Determine Geolocation Data Quality in mHealth Studies. *Stud Health Technol Inform* 2017;245:10-14. [Medline: [29295042](https://pubmed.ncbi.nlm.nih.gov/29295042/)]
24. Torous J, Kiang MV, Lorme J, Onnela JP. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health* 2016 May 05;3(2):e16 [FREE Full text] [doi: [10.2196/mental.5165](https://doi.org/10.2196/mental.5165)] [Medline: [27150677](https://pubmed.ncbi.nlm.nih.gov/27150677/)]
25. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ Digit Med* 2018;1:15 [FREE Full text] [doi: [10.1038/s41746-018-0022-8](https://doi.org/10.1038/s41746-018-0022-8)] [Medline: [31304300](https://pubmed.ncbi.nlm.nih.gov/31304300/)]



26. Beukenhorst AL, Schultz DM, McBeth J, Sergeant J, Dixon W. Are weather conditions associated with chronic musculoskeletal pain? Review of results and methodologies. *Pain* 2020 Apr;161(4):668-683. [doi: [10.1097/j.pain.0000000000001776](https://doi.org/10.1097/j.pain.0000000000001776)] [Medline: [32195783](https://pubmed.ncbi.nlm.nih.gov/32195783/)]
27. Dorsey ER, Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. *Acad Med* 2017 Feb;92(2):157-160. [doi: [10.1097/ACM.0000000000001205](https://doi.org/10.1097/ACM.0000000000001205)] [Medline: [27119325](https://pubmed.ncbi.nlm.nih.gov/27119325/)]
28. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology* 2021 Jan;46(1):45-54. [doi: [10.1038/s41386-020-0771-3](https://doi.org/10.1038/s41386-020-0771-3)] [Medline: [32679583](https://pubmed.ncbi.nlm.nih.gov/32679583/)]
29. Lee E, Viswanath K. Big Data in Context: Addressing the Twin Perils of Data Absenteeism and Chauvinism in the Context of Health Disparities Research. *J Med Internet Res* 2020 Jan 07;22(1):e16377 [FREE Full text] [doi: [10.2196/16377](https://doi.org/10.2196/16377)] [Medline: [31909724](https://pubmed.ncbi.nlm.nih.gov/31909724/)]
30. Barnett I, Onnela JP. Inferring mobility measures from GPS traces with missing data. *Biostatistics* 2020 Apr 01;21(2):e98-e112 [FREE Full text] [doi: [10.1093/biostatistics/kxy059](https://doi.org/10.1093/biostatistics/kxy059)] [Medline: [30371736](https://pubmed.ncbi.nlm.nih.gov/30371736/)]
31. Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. *Science* 2010 Feb 19;327(5968):1018-1021. [doi: [10.1126/science.1177170](https://doi.org/10.1126/science.1177170)] [Medline: [20167789](https://pubmed.ncbi.nlm.nih.gov/20167789/)]
32. O'Dea S. Mobile operating systems: market share worldwide from January 2012 to October 2021. Statista. 2012. URL: <https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/> [accessed 2021-05-05]
33. O'Dea S. Market share of mobile operating systems in North America from January 2018 to June 2021. Statista. URL: <https://www.statista.com/statistics/1045192/share-of-mobile-operating-systems-in-north-america-by-month/> [accessed 2021-05-05]

## Abbreviations

**OR:** odds ratio

*Edited by L Buis; submitted 16.03.21; peer-reviewed by I Barnett, E Lee, H Mehdizadeh; comments to author 22.05.21; revised version received 11.06.21; accepted 27.08.21; published 16.11.21*

*Please cite as:*

*Beukenhorst AL, Sergeant JC, Schultz DM, McBeth J, Yimer BB, Dixon WG*

*Understanding the Predictors of Missing Location Data to Inform Smartphone Study Design: Observational Study*

*JMIR Mhealth Uhealth* 2021;9(11):e28857

URL: <https://mhealth.jmir.org/2021/11/e28857>

doi: [10.2196/28857](https://doi.org/10.2196/28857)

PMID:

©Anna L Beukenhorst, Jamie C Sergeant, David M Schultz, John McBeth, Belay B Yimer, Will G Dixon. Originally published in JMIR mHealth and uHealth (<https://mhealth.jmir.org>), 16.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on <https://mhealth.jmir.org/>, as well as this copyright and license information must be included.