

# Using the Socialise app to collect smartphone sensor data for mental health research: A feasibility study

Tjeerd W. Boonstra<sup>1, \*</sup>, Jennifer Nicholas<sup>1</sup>, Quincy J. J. Wong<sup>1</sup>, Frances Shaw<sup>1</sup>, Samuel Townsend<sup>1</sup>,  
Helen Christensen<sup>1</sup>

<sup>1</sup>Black Dog Institute, UNSW Sydney, Sydney, Australia

## Abstract

To investigate the feasibility of collecting smartphone sensor data for mental health research, we tested the Socialise app that was developed at the Black Dog Institute in a group of people with a lived experience of mental health challenges (n=32). Bluetooth, GPS and battery status data were collected at regular intervals (3, 4, 5 or 8 minutes) for 4 weeks. In addition, survey data was collected using the app to investigate the views of participants on user experience and the acceptability of passive data collection for mental health research. No mental health data was collected as part of the feasibility study.

## General

This study was approved by the University of New South Wales Human Research Ethics Committee (HC17203). Participants were recruited through advertisements disseminated through the online Black Dog Institute volunteer research register. To be able to participate, individuals had to be 18 years or older, reside in Australia and speak English, and have a smartphone running Android version 4.4 or newer or running iOS8 or newer. Interested individuals received a link to the study website where they could read the detailed participant information and provide consent.

Participants installed the Socialise app on their own smartphone and were instructed to use the Socialise app for four weeks. Bluetooth and GPS data were collected during scans that occurred at intervals of 8, 5, 4, or 3 minutes (equivalent to 7.5, 12, 15, 20 scans per hour, respectively). Each scanning rate was tested for one week and participants were instructed to use their phones as normal for the duration of the study. Of 32 individuals consented to participated in the study, 30 completed the entry survey and we obtained sensor data from 28 devices participants. The participant demographics are provided in Table 1.

One participant did not consent to have their data made available on a public repository. We do not share GPS data as this would allow re-identification of participants. Likewise, the MAC address of devices detected using Bluetooth were cryptographically hashed to ensure that detected people cannot be identified.

**Table 1.** Participant demographics

	n	percentage
Sex (n=30)		
Male	7	23%
Female	23	77%
Age (n=30)		
18-24	5	17%
25-34	6	20%
35-44	5	17%
45-54	4	13%
55-64	7	23%
65+	3	10%
Mental health diagnosis (n=32)	23	72%
Depression	22	69%
Bipolar disorder	9	28%
Anxiety disorder	17	53%
Schizophrenia	0	0%
Personality disorder	2	6%
Drug and alcohol problems	5	16%
Eating disorder	7	22%
Autism Spectrum Disorder	1	3%
Post-traumatic stress disorder	2	6%
Attention deficit hyperactivity disorder	1	3%
Daily phone usage (n=30)		
Less than 30 min	2	7%
30min – 1 hour	7	23%
1–2 hours	4	13%
2–3 hours	6	20%
More than 3 hours	11	37%

## Files

### *participant\_info.csv*

This file contains a table with information about the participants that completed the study. The first column (*pid*) contains the unique participant ID, which is consistent across data files. The second column (*os*) contains the operating system of their smartphone. The third column (*model*) contains the model of their smartphone. The fourth column (*time*) contains the time and date at which they enrolled in the study. Columns 5-8 contain the answers to the following questions which they had to answer at the start of the study:

1. age: How old are you?
2. gender: What is your gender?
3. phone\_use: How much time do you spend on your mobile phone on average in a day?
4. phone\_age: How old is your phone?

### *bluetooth.csv*

This file contains a table with all the Bluetooth data that was obtained during the study. The first column (*pid*) contains the unique participant ID. The second column (*detected*) contains the hashed MAC address of the device that was detected. To only record other smartphones, detected devices were filtered according to the Bluetooth Core Specification. This involved removing any devices not matching the Class of Device (CoD) 0x200 during the Bluetooth scan. If no smartphones were detected, this is indicated by 'None Detected'. The third column (*time*) contains the local time and date at which the scan was made.

### *battery.csv*

This file contains a table with all the battery data that was obtained during the study. The first column (*pid*) contains the unique participant ID. The second column (*time*) contains the local time and date at which the scan was made. The third column (*level*) contains the battery level as a percentage. The fourth column (*status*) describes whether the phone was charging or discharging. The text that is provided differs between Android and iOS.

#### *app\_survey.csv*

This file contains the responses of participants to survey questions that were delivered during the study using the Socialise app. The first column (*pid*) contains the unique participant ID. The second column (*time*) contains the local time and date at which the survey was completed. Columns 3 to 7 contain the answers to the following questions:

1. *bluetooth\_use*: In the last week, did you use Bluetooth on your phone?
2. *running\_problem*: In the last week, did you experience any problems when running the app?
3. *issue*: What issues did you experience? (open question)
4. *settings*: In the last week, did you change any settings on your phone?
5. *battery*: In the last week, did the app impact the battery life of your phone? (1: not much to 7: very much)
6. *data\_use*: In the last week, did the app impact the data usage of your phone? (1: not much to 7: very much)

#### *ethics\_multiple\_choice.csv*

This file contains the responses of participants to survey questions that were delivered at the end of the study. The first column (*Participant*) contains the participant number. Columns 2 to 5 contain the answers to the following multiple-choice questions “How comfortable were you with the following aspects of the study?”:

1. GPS: Having your GPS data collected
2. Bluetooth: Having your Bluetooth connections collected
3. Background: Having the app installed on your phone and running in the background
4. Questionnaire: Filling out self-report questionnaires

Columns 6 to 9 contain the answers to the following multiple-choice questions “How comfortable are you, in general terms, with the following practices?”:

5. Research: Social media data collection for health / medical research
6. Advertising: Social media data collection for advertising and market research
7. Medical: Social media monitoring in order to deliver evidence-based targeted medical interventions
8. Legal: Social media monitoring for targeted legal interventions

#### *ethics\_open\_questions.csv*

This file contains the responses of participants to the corresponding open questions that were delivered at the end of the study. The first column (*Participant*) contains the participant number. The second column (*Question*) indicates the question that was answered. The third column (*Response*) contains the answer that was provided. The question numbers (column 2) relate to the following questions:

- 1: Can you tell us more about how you felt about having your GPS data collected? Why were you comfortable or not comfortable with this aspect of the study?
- 2: Can you tell us more about how you felt about having your Bluetooth data collected? Why were you comfortable or not comfortable with this aspect of the study?
- 3: Can you tell us more about how you felt about having the app installed on your phone and running in the background? Why were you comfortable or not comfortable with this aspect of the study?

4: Can you tell us more about how you felt about filling out self-report questionnaires? Why were you comfortable or not comfortable with this aspect of the study?

5: Would you continue to use this app if it was available to you after the completion of this study? Why or why not?

10: If you said that you were uncomfortable with any of the above practices, what concerns do you have about social media data collection or social media monitoring?