

BMJ Open SocialBit: protocol for a prospective observational study to validate a wearable social sensor for stroke survivors with diverse neurological abilities

Kelly White ^{1,2}, Samuel Tate,³ Ross Zafonte,⁴ Shrikanth Narayanan,⁵ Matthias R Mehl,⁶ Min Shin,³ Amar Dhand ^{1,2}

To cite: White K, Tate S, Zafonte R, *et al.* SocialBit: protocol for a prospective observational study to validate a wearable social sensor for stroke survivors with diverse neurological abilities. *BMJ Open* 2023;**13**:e076297. doi:10.1136/bmjopen-2023-076297

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-076297>).

Received 02 June 2023
Accepted 08 August 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Amar Dhand;
adhand@bwh.harvard.edu

ABSTRACT

Introduction Social isolation has been found to be a significant risk factor for health outcomes, on par with traditional risk factors. This isolation is characterised by reduced social interactions, which can be detected acoustically. To accomplish this, we created a machine learning algorithm called SocialBit. SocialBit runs on a smartwatch and detects minutes of social interaction based on vocal features from ambient audio samples without natural language processing.

Methods and analysis In this study, we aim to validate the accuracy of SocialBit in stroke survivors with varying speech, cognitive and physical deficits. Training and testing on persons with diverse neurological abilities allows SocialBit to be a universally accessible social sensor. We are recruiting 200 patients and following them for up to 8 days during hospitalisation and rehabilitation, while they wear a SocialBit-equipped smartwatch and engage in naturalistic daily interactions. Human observers tally the interactions via a video livestream (ground truth) to analyse the performance of SocialBit against it. We also examine the association of social interaction time with stroke characteristics and outcomes. If successful, SocialBit would be the first social sensor available on commercial devices for persons with diverse abilities.

Ethics and dissemination This study has received ethical approval from the Institutional Review Board of Mass General Brigham (Protocol #2020P003739). The results of this study will be published in a peer-reviewed journal.

INTRODUCTION

Social connection has a large role in health outcomes.¹ This effect is independent of socioeconomic status, smoking, alcohol consumption, obesity, physical activity and utilisation of preventive health services.² Three decades ago, a meta-analysis concluded that social isolation was a major risk factor for health, and rivalled the effect of cigarette smoking, blood pressure, blood lipids, obesity and physical activity.³ More recently, a meta-analysis of

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This study introduces a novel machine learning algorithm to detect and quantify social isolation.
- ⇒ The SocialBit wearable sensor accommodates individuals with diverse neurological abilities, promoting inclusivity in social monitoring technologies.
- ⇒ By detecting isolation in real time, this study allows for future interventions to minimise the negative effects of isolation on health outcomes.
- ⇒ Although this study is conducted in a hospital environment, future studies could validate the algorithm in a more naturalistic home setting.
- ⇒ To maintain patient privacy, the SocialBit algorithm does not record raw audio, which poses challenges to the machine learning process.

148 studies showed that persons with stronger social relationships had a 50% increased likelihood of survival, an OR higher than the effect of smoking, alcohol consumption and body mass index on health.⁴

This study is developing and validating a novel social interaction detection framework and an algorithmic implementation, SocialBit. SocialBit is a smartwatch-based mobile sensing application intended to passively and automatically track the amount of daily interactions of the primary person wearing the device. The application tracks social interactions by sampling ambient audio. Importantly, the application never stores the raw audio but rather stores a series of audio features to serve as input for classification by the SocialBit algorithm. Study investigators collect data from inpatient stroke survivors at Brigham and Women's Hospital and Spaulding Rehabilitation Hospital in Boston, Massachusetts, capitalising on the ability to monitor patients' social interactions in real time for multiple days.

SocialBit is the first wearable social interaction detection sensor customised for, and specifically validated with, stroke survivors. After stroke, patients are vulnerable to reduced social interactions and social isolation, which may have negative implications on their physical recovery.^{5–7} This is due to multiple factors, including changing social desires, language dysfunction, loss of shared activities, reduced energy levels, physical disability, depression, anxiety, motor impairment, environmental barriers, embarrassment and social stigma.^{8,9} The period immediately after stroke may be considered a particularly vulnerable time for social isolation because of the inability of family or friends to travel to the hospital, a belief that the patient needs to ‘heal’, and the initial severity of deficits that limit time away from home.¹⁰

It is specifically challenging to detect social interactions in patients with stroke who have cognitive or language deficits (aphasia). Usual methods of social isolation or social network characterisation rely on retrospective surveys or momentary self-report questionnaires.¹¹ Patients with cognitive or language deficits cannot complete these instruments and are typically left out of such studies. Over 35% of patients with stroke have language deficits immediately after stroke,¹² and over 50% of patients have cognitive impairment at 6 months after stroke.¹³ Such deficits make individuals incapable of providing a valid self-report, a limitation to our prior work on social network characterisation.¹⁴ The lack of data on patients with aphasia is especially concerning because of this population’s increased likelihood of social isolation after stroke due to multiple challenges with social integration.⁶ Our goal is to overcome this selection bias by developing a framework to detect interactions regardless of whether the patient is contributing distinguishable words to the conversation, thereby rendering it a viable interaction tracking solution that is universally accessible and inclusive for people with disabilities.

Previous attempts at measuring social interactions have shown promising results under particular assumptions.^{15–17} These assumptions, unfortunately, limit the scope of application. For example, one algorithm required mobile apps with access to Global Positioning System (GPS) information as well as the individuals’ calendar events to analyse social interactions.¹⁶ Another application assumed that each individual wear a sensing device in a bounded environment (e.g., workplace).¹⁵ Another team used cameras to sense interactions, increasing privacy concerns.¹⁷ Another study used a smartphone-based conversation classifier, which assumes close proximity of smartphone to user.¹⁸ Here, we seek a technically parsimonious, privacy-protective and broadly applicable solution to automated social interaction tracking that does not require access to private information (i.e., it does not store raw ambient audio) and only requires the individual in question to be wearing a commercial smartwatch.

The goal of this study is to establish the validity and utility of this tool in detecting social isolation in stroke survivors in the period after stroke when patients may be

observed in the hospital. This allows human observation of social interactions via video-streaming for ground truth determination. There are many implications. Unlike previously tested sensors,^{19,20} SocialBit requires only the patient to wear the device and not conversation partners. The algorithm maintains privacy of conversation content. SocialBit could run on widely accessible commercial devices. The application can be downloaded onto other WearOS devices, and in the future, it may be possible to download onto other platforms such as the Apple Watch, aligning with the vision of such wearables to become health monitoring tools. Lastly, SocialBit can lead to interventions to improve stroke recovery. With accurate social interaction data, SocialBit can provide real-time feedback and coaching to patients, family and clinicians.

METHODS

SocialBit algorithm

The SocialBit algorithm detects social interactions through classifying acoustic features. Specifically, it analyses the temporal change of vocal acoustic behaviour, such as pitch and intensity. It promotes privacy because it does not rely on lexical information or natural language processing (i.e., what words were spoken). Rather, it uses a ‘sound signature’ machine learning method in which features of sound are extracted, analysed and classified as social interaction or not social interaction. The result is quantification of the number of social interaction minutes per day.

In figure 1, we depict the SocialBit algorithm machine learning approach. First, the algorithm converts the audio data into log mel spectrogram representations and passes them through a pretrained neural network called YAMNet,²¹ which is capable of classifying over 500 audio events from the public AudioSet data set.²² YAMNet produces feature vectors for each 0.96 s segment of audio, which are then combined and fed through a transformer network²³ to determine whether the sequence contains a social interaction. Additionally, SocialBit collects voice-print features unique to each patient and passes them to ECAPA,²⁴ a neural network that helps identify when the primary device wearer is speaking. The speech sample used are five short sentences from the National Institutes of Health Stroke Scale (NIHSS).²⁵ In summary, YAMNet discriminates general acoustic features of human voices versus other sounds, and ECAPA identifies when the primary device wearer is talking.

Study design

This is a prospective, observational study of 200 patients with stroke. The data collection includes up to eight inpatient observation days per participant, and a 3-month follow-up assessment for study completion. We observe participants at Brigham and Women’s Hospital until they are discharged or for a maximum of 5 days. We collect up to 3 additional days of data if a patient transitions to Spaulding Rehabilitation Hospital Boston. At Brigham

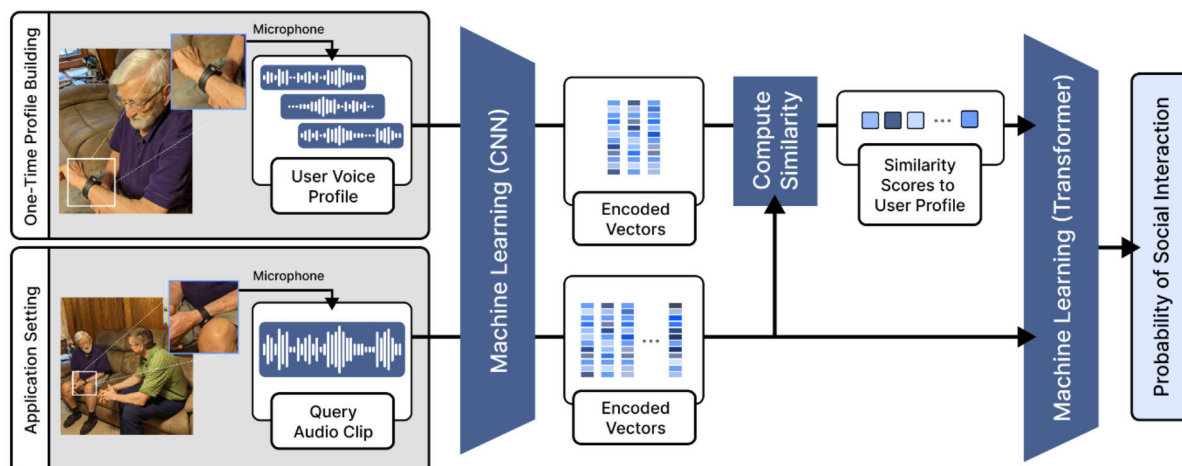


Figure 1 Machine learning algorithm that detects the probability of a social interaction.

and Women's Hospital, participants usually stay in shared rooms with one roommate. The focus is on acute care with modest inpatient rehabilitation services. At Spaulding Rehabilitation Hospital, Boston, participants stay in private rooms and receive at least 3 hours of intensive therapy (physical therapy, occupational therapy and speech therapy). We enrol caregivers of patients for those who cannot engage in surveys or are otherwise available for auxiliary data. We collect data at two time points: (1) In hospital, when research staff collect survey data and ground truth data, and (2) At the 3-month follow-up clinic appointment, when research staff collect additional survey data. [Figure 2](#) shows the timeline of the study. This study was approved by the Institutional Review Board of Mass General Brigham (Protocol #2020P003739).

This project has three primary aims: (1) Establish the accuracy of the SocialBit application for use in research with stroke survivors. (2) Determine the association between social interaction times and social isolation and stroke outcomes at 3 months (e.g., physical function, mood, disability). (3) Examine the influence of medical factors (e.g., depression, delirium and stroke severity) on social interaction time. The central hypothesis is that the SocialBit algorithm accurately detects social interactions of stroke survivors in hospital.

Recruitment

One goal of this project is to create social sensing technology universally accessible and inclusive for people with disabilities. Therefore, we focus on people with a range of neurological deficits after stroke as an integral part of the

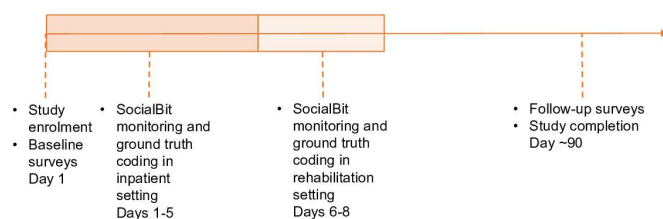


Figure 2 Timeline for the SocialBit study spanning 3 months.

development process. Beginning on 15 June 2021, and continuing to early 2025, we are recruiting patients with acute ischaemic stroke at Brigham and Women's Hospital in Boston, Massachusetts, USA. This urban setting provides a racially and economically diverse sample population for the study. We aim to recruit patients with a variety of neurological deficits including aphasia, dysarthria, cognitive changes and paralysis.

The following inclusion criteria apply: (1) Diagnosed with an acute ischaemic stroke defined clinically with support from imaging and (2) 18 years old or older. Exclusion criteria include the following: (1) On Comfort Measures Only (a patient end-of-life care plan that focuses on pain relief and quality of life), (2) Diagnosed with dementia prior to stroke in the medical record, (3) Unable to obtain informed consent from the patient or patient decision maker, and (4) Patient or patient decision maker is unable to understand or speak English well enough to complete surveys.

We screen patients for eligibility daily via the Brigham and Women's Hospital inpatient neurology lists on Epic Systems. The research staff then request permission from the nursing staff to approach patients who meet formal inclusion and exclusion criteria for the study. The nurse presents the study to the patient and asks if the patient is willing to discuss the study further with the research team. After acquiring appropriate permission through the nursing staff, research staff approach qualifying patients and/or their family members to further explain the study.

We obtain signed informed consent from all patients who are consentable, meaning they do not have aphasia or confusion. For the patients who do have aphasia, we collect signed informed consent from caregivers who are willing to participate and answer survey questions on behalf of the patient. Patients who agree to participate complete about 1 hour of surveys with the research staff, including an NIHSS, a Montreal Cognitive Assessment (MoCA), and additional surveys about their social network, their perceived loneliness and depression, their subjective physical function, their life satisfaction, and

their personality. We ask patients these same survey questions when they return for their follow-up appointment with the study's principal investigator, AD, at ~90 days. At this visit, the patient completes his/her participation in the study and is compensated for his/her time.

After the patient completes the initial surveys, we ask the patient to wear a SocialBit-equipped Samsung Galaxy Watch5 Pro from 9:00 to 17:00 until he/she is discharged or for a maximum of 5 days during their inpatient stay at Brigham and Women's Hospital. Additionally, if the patient is discharged to Spaulding Rehabilitation Hospital in Boston, we collect data in that setting for up to 3 days. On average, we aim to collect 2–3 days (16–24 hours) of data per patient (i.e., around 8 hours of expected monitoring per day, a constraint imposed by both the ground-truth data collection and watch battery considerations).

To identify the primary speaker, the patient creates a voice profile on the SocialBit application by reading five short sentences from the NIHSS. We ask the patient to read the five sentences aloud for 30 s. For patients who cannot read or speak, no voice profile is created. Following the voice profile, SocialBit runs in the background for passive detection of interactions without any interface with users. If the patient were to take a shower, leave for imaging or perform another task that may interfere with the functioning of the smartwatch, we instruct the patient and/or nursing staff to take the watch off before leaving and put the watch back on upon return.

Study staff set up an iPad with a Zoom livestream in the patient's room and put signage in the room to alert other people of the livestream video and provide people with a contact number in case they have any questions or concerns. A Health Insurance Portability and Accountability Act (HIPAA) certified research staff member is on the other side of the livestream manually coding, with a temporal resolution of 1 min, all social interactions that the patient has throughout the day. If there are any moments throughout the day that the patient does not want their conversation to be overheard, we encourage the patient and/or nursing staff to cover up the iPad, signalling to the research team to pause data collection and stop listening.

Exposure and outcome measurements

To address the three aims of this study, we collect a variety of measures about patients' social interactions, social connectedness, stroke severity, physical function and cognitive function. This section outlines all the measures we collect with patients and their caregivers during both the hospitalisation period and the 3-month follow-up appointment.

Ground truth coding system

To accomplish the primary aim of this study, establishing the accuracy of the SocialBit algorithm in detecting social interaction, the SocialBit algorithm and trained research staff independently collect social interaction data. [Table 1](#) shows the ground truth data coded by research staff,

Table 1 Ground truth data collection recorded every minute from 9:00 to 17:00

Ground truth table question (per minute)	Answer (per minute)
Was the patient talking to another person?	<ul style="list-style-type: none"> ▶ Yes ▶ No
Who was the patient talking with?	<ul style="list-style-type: none"> ▶ Medical personnel ▶ Caregiver ▶ Other family and friends, adult or child ▶ Patient in other bed ▶ Stranger/other ▶ Can't tell
How was the patient talking with the person(s)?	<ul style="list-style-type: none"> ▶ In person ▶ Phone call ▶ Videoconference/chat
How many people was the patient talking with? (i.e., actively in conversation)	<ul style="list-style-type: none"> ▶ 1 ▶ 2 ▶ 3 or more
Does the conversation contain foreign language?	<ul style="list-style-type: none"> ▶ Yes ▶ No
What was the tone of the interaction? (−2=negative, 0=neutral, 2=positive)	<ul style="list-style-type: none"> ▶ −2 ▶ −1 ▶ 0 ▶ 1 ▶ 2 ▶ Can't tell
What was the depth of the conversation? (1=superficial to 5=deep)	<ul style="list-style-type: none"> ▶ 1 ▶ 2 ▶ 3 ▶ 4 ▶ 5 ▶ Can't tell
Was the TV on? (TV includes computer or other devices playing entertainment in a way that is audible in the room)	<ul style="list-style-type: none"> ▶ Yes ▶ No
Was there a conversation happening that the patient was not part of?	<ul style="list-style-type: none"> ▶ Yes ▶ No
Was the patient completely silent during this minute? (did not contribute any words or sounds)	<ul style="list-style-type: none"> ▶ Yes ▶ No
Anything else that was noteworthy? (i.e., was the patient talking with you, the observer)	

including the questions answered each minute. [Table 2](#) lists the guidelines for making coding judgements for each question.

To ensure study staff are taking a standardised approach to coding the ground truth table, we provide training to all new study staff. First, current staff teach new study staff how to code the ground truth table and new staff observe the coding process for at least two full days. Then, new study staff watch two 15-minute-long sample videos and complete sample ground truth tables for each scenario. The principal investigator and team created these two sample videos along with a standardised answer key. The

Table 2 Guidelines for making coding judgements for each question of the ground truth table

Question	Approach
Was the patient talking with another person?	If any social interaction occurs with the patient, even for a couple of seconds, during this 1 min timeframe, the research staff select 'Yes'. All utterances spoken by or to a patient imply that a social interaction is occurring with the patient, meaning that even patients with aphasia who are limited to fragmented speech or grunts are included as having talked with another person for this question. If no words or sounds are exchanged between the patient and another person, research staff select 'No'.
Who was the patient talking with?	The research staff categorise the person(s) that the patient is interacting with during each minute. 'Medical Personnel' includes any person involved in delivery of medical care, such as doctors, nurses, phlebotomists, medical assistants and certified nursing assistants. The 'caregiver' is a single person, often a spouse, who is usually listed as the patient's healthcare proxy in his/her chart. 'Stranger/Other' includes auxiliary staff, such as research staff, food delivery and housekeeping.
How was the patient talking with the person(s)?	The research staff choose the mode of communication that occurs for the majority of time during the minute. If multiple modes of communication are being used throughout the minute, research staff note that in the 'Anything else that was noteworthy?' section.
How many people was the patient talking with?	The research staff select the total number of people that are talking with the patient throughout the minute. If the patient is speaking to only one person, staff select '1'. If the patient is on the phone with someone and a medical professional speaks to them outside of their phone conversation in the same minute, staff select '2'.
Does the conversation contain foreign language?	The research staff select 'Yes' if even one word spoken is foreign. It does not matter if the foreign language is spoken by the patient or someone speaking with the patient. Incomprehensible speech does not count as a foreign language.
What was the tone of the interaction?	The research staff select one answer from a 5-point scale by making a judgement about the tone of a conversation. Study staff consider the tone, actual speech, and 'feeling' of the conversation. The rating is made on a scale from -2 (very negative) to +2 (very positive), with 0 being neutral (-1: somewhat negative; +1: somewhat positive). This is coded even when the participant is minimally involved in the conversation. Most conversations are neutral or mildly positive in tone. 'Can't tell' is coded when the coder is not able to make a tone determination, which may happen when the conversation is too short to be able to tell the tone. The tone coding should reflect the predominant tone during the captured segment. Positive tones include cheerful, upbeat, excited, loving, caring, supportive. Negative tones include tense, irritated, disapproving, sad/depressed, unsupportive, mean, sarcastic, angry and frustrated.
What was the depth of the conversation?	The research staff select one answer from a 5-point scale by making a judgement about the depth of a conversation. Research staff consider the extent of the thoughts, information, values or ideas being exchanged. The rating is made on a scale from 1 (very superficial) to 5 (very deep), with 3 being a conversation at mid-level depth (2: somewhat superficial; 4: somewhat deep). This is coded even when the participant is minimally involved in the conversation. Many routine interactions are mildly or very superficial (eg, standard greetings). In these interactions, no real information is exchanged; if the conversation had not happened, the involved parties would know the same about each other. Some conversations neither stand out as superficial nor deep (3; 'midway'). Other conversations are deep with substantial/meaningful information exchange, sharing and disclosure. Few conversations are very deep with personally meaningful information being exchanged. 'Can't tell' is coded when the coder is not able to make a depth determination due to inability to understand conversation. The depth coding should reflect the predominant depth during the captured segment.
Was the TV on?	Regardless of if an interaction occurs within the minute, research staff answer 'Yes' if a television is audible. Staff also answer 'Yes' to this question if a computer or other device is playing entertainment that is audible in the room.
Was there a conversation happening that the patient was not part of?	The research staff select 'Yes' if they can hear a conversation that the patient is not a part of, such as a conversation between the patient's roommate and their provider or a hallway conversation among nurses. Staff select 'Yes' to this question even if the outside conversation is barely audible.
Was the patient completely silent during this minute?	If the patient makes no noise and the other person in the interaction is the only one speaking, research staff select 'Yes'. If the patient says anything or makes any noise during the minute, research staff select 'No'.
Anything else that was noteworthy?	The research staff include anything here that is important for collaborators. For example, staff note if the observation was halted, if there were any unusual noise (eg, medical alarms). Additionally, staff note whether a patient was on speakerphone during a phone call.

new study staff's ground truth coding should match the key, except for coding ± 1 for tone and depth due to the subjectivity of these variables. If new study staff members do not accurately fill out the practice ground truth tables,

they must complete further practice and training before coding for an actual patient. Study staff also continually work together to ensure that they are consistently answering the ground truth questions similarly. The

Table 3 List of all measures collected in the SocialBit Study, including measures for exposures, outcomes and covariates

Construct	Measure	Participant	T1 at intake	T2 at 3 months
Exposures				
Social Interaction time—SocialBit	Interaction detected by SocialBit application	Patient	X	
Social interaction time—ground truth coded	Ground truth data collected by study staff	Patient	X	
Outcomes				
Social isolation (objective)	Personal Network Survey for Clinical Research (PERSNET)	Patient and caregiver	X	X
Social isolation (subjective)	UCLA Loneliness Scale	Patient and caregiver	X	X
Disability	Modified Rankin Scale	Patient and caregiver	X	X
Physical function (subjective)	NIH Patient-Reported Outcome Measurement Information System (PROMIS), Physical Function	Patient and caregiver	X	X
Depression	Centre for Epidemiological Studies Depression (CES-D)	Patient and caregiver	X	X
Life satisfaction	Satisfaction with Life Scale	Patient and caregiver	X	X
Cognition	Montreal Cognitive Assessment	Patient	X	X
Stroke severity	National Institutes of Health Stroke Scale (NIHSS)	Patient	X	X
Delirium	Confusion assessment method	Patient	X	X
Caregiver burden	Caregiver Burden Scale	Caregiver		X
Covariates				
Demographics	Demographics	Patient and caregiver	X	
Comorbidities	Charlson Comorbidity Index	Patient	X	
Lesion characteristics	Medical record, neuroimaging	Patient	X	
Big Five Personality	Big Five Inventory-2 XS	Patient and caregiver	X	X

NIH, National Institutes of Health.

study lead coordinator re-trains auxiliary study staff every 6 months.

Subjective social connectedness

To address the second aim of determining the association between social isolation and stroke outcomes, we perform a series of self-report surveys (table 3) with the participants (or caregivers) to get their subjective interpretation of their social networks and social connectedness. We use the Personal Network Survey for Clinical Research, developed by Dhand *et al*,²⁶ to assess the social connectedness of patients with stroke. We also administer the 20-item UCLA Loneliness Scale²⁷ and 20-item Center for Epidemiological Studies Depression Scale²⁸ to assess patients' subjective feelings of loneliness and depression. These surveys are conducted at patient enrolment as well as at their 3-month follow-up visit, to evaluate trends in stroke survivors' social connectedness during recovery.

Subjective physical function

To understand patients' stroke outcomes, we use the simplified Modified Rankin Scale²⁹ and National Institutes of Health Patient-Reported Outcome Measurement

Information System (PROMIS) Physical Function Surveys through self-report with patients and/or their caregivers (table 3). These measures are widely used and consensually accepted measures of functional ability in patients with stroke.^{30–32} The Modified Rankin Scale is a 6-point disability scale using dichotomous questions²⁹ and the PROMIS Physical Function is a computer adaptive test that selects items from a large PROMIS bank.³³ We conduct these surveys at patient enrolment as well as at their 3-month follow-up visit. Data from enrolment is compared with data at the 3-month follow-up to assess the patient's physical improvements over the 3-month period.

Stroke severity and cognitive function

To address the third aim of determining the impact of medical factors on social interaction time, we assess stroke severity through the broadly used NIHSS³⁴ and cognitive abilities through the reliable and validated MoCA (table 3).^{35 36} The NIHSS is a 15-item scale used to assess the physical and cognitive effects of an acute stroke.²⁵ The MoCA is a short screening tool for cognitive impairment.³⁵ We compare NIHSS and MoCA scores

at enrolment and at the 3-month follow-up to determine a patient's physical and cognitive improvement over the first 3 months of recovery. We also record patient's delirium from the Confusion Assessment Method,³⁷ an instrument used in a clinical setting to detect delirium, noted in patients' charts by nursing staff during their inpatient stay.

Caregiver burden

At the 3-month follow-up, we ask caregivers who agree to participate in the study a series of questions regarding the burden they have felt since taking care of their loved one after their stroke. The Caregiver Burden Scale (table 3) used in this study is a short-form 6-item scale adopted from the Zarit Burden Interview. Previous studies have validated the reliability and usefulness of the shortened scale in detecting caregiver burden.^{38 39} These data are useful information to put into perspective how illness and disease have an impact on not only the patient but also their social network.

Covariates

We also collect data on explanatory variables that may impact health outcomes (table 3). In this category, we collect sociodemographic information, including age, sex, ethnicity, race, education level, household income, employment status and marital status. We also collect a patient's personality index by using the validated 15-item Big Five Inventory2 XS Survey.⁴⁰ Caregivers who agree to participate answer this personality survey twice, once regarding the patient's personality and once regarding their own personality. This provides the research team with a better understanding of the personality traits of both the patient and their caregiver. We collect information on comorbidities and stroke characteristics, which can impact stroke severity and recovery time, from the patient's chart.

Analysis plan

The analysis focuses on determining the accuracy of the SocialBit algorithm in detecting social interactions between the primary device wearer and others. Second, we determine the correlation between the amount of social interaction and stroke outcomes, as well as the association of medical factors with the amount of social interaction.

Quantitative data analysis

The overall analysis plan is to evaluate accuracy (% of correct classification), specificity, sensitivity, positive predicted value (PPV) and negative predictive value (NPV) of the social interaction times measured by the SocialBit algorithm versus the ground truth. For the machine learning, fivefold cross validation is used with 40 patients per fold. The unit of analysis is social interaction or no social interaction in every 1 min interval for 8 hours a day. If social interaction occurs partially within an interval, then the entire 1 min interval is marked as social interaction. To be consistent with the ground

truth, the automated algorithm processes at the 1-min interval as well. However, only 1 min every 5 min or 6 min depending on the battery capacity is analysed by the algorithm. We minimise the overfitting to training data through optimising hyperparameters (eg, regularisation and stop training when the validation set is maximum). To increase the diversity of the training corpus, we add audio extracts of voices with different pitch, tone and volume from public data sets. We also train using examples of TV shows and ambient healthcare setting noise from recordings in empty hospital rooms.

We are conducting all regression and longitudinal statistical analyses in R with biostatistical consultation at Harvard Medical School. For the primary outcome, we are assessing baseline characteristics and any differences between patients. We are also performing checks for outlying values. For diagnostic accuracy determination, we are constructing a 2-by-2 table, and then determining overall accuracy, sensitivity, specificity, PPV, NPV and diagnostic OR. For the secondary outcomes, we are using multivariate linear regression to determine the association between social interactions and stroke outcomes and the influence of medical factors on social interaction times. All analyses are accounting for gender and socioeconomic status variables at baseline including education, income and occupation. We aim to include a diverse set of participants including both men and women and a mix of race and ethnicity. We are completing stratified and interaction analyses by these variables to assess whether patterns are seen within and across these categories.

Power analysis

In machine learning projects, high-quality and diverse data with positive and negative examples are needed for successful model building. Therefore, contrary to traditional power analysis, sample size determination is based on iterative performance metrics during training. In our project, we plan for 200 patients \times 24 hours \times 60 min/hour divided by 6 (to conserve battery life) which is equal to 48 000 samples. The ratio of social interaction to non-social interaction samples is unknown, but even at extreme levels (9 to 1), this results in 4800 samples in a minority class. Based on the literature, this sample size is comparable to gold standard studies such as Audio Set (most classes contain less than 10 000 samples)⁴² and ImageNet (~1000 average samples per class, ~3000 for the mode).⁴¹ Furthermore, deep learning algorithms have achieved good accuracy for each of these data sets. For example, a deep learning algorithm like ours achieved an Area Under the Curve (AUC) level of 0.96 for Audio Set.⁴² Therefore, our strong preliminary data and literature-based estimates justify our sample size determination. For future use of these data, we also plan to analyse the data set size needed to reach reliable classification. We will compute a learning curve that measures the trade-off between the size of training set and the classification accuracy.⁴³ This would allow evaluation of the feasibility and scalability of the algorithm.

Data management

One of the audio features collected through the SocialBit application includes the patient reading a passage. This is considered a voice print, or a visual record of speech, which is Protected Health Information (PHI) as defined by HIPAA.⁴⁴ We treat such data in the same way as other PHI in terms of privacy and security when collecting and transferring their PHI. In general, all audio features are encrypted and only temporarily stored on the watch. These encrypted audio features are uploaded to Amazon Web Services (AWS), which is a HIPAA-compliant platform. To respect two-party consent laws and to avoid storing PHI from anyone other than the patient, we store only the similarity score (0–1) between the primary device wearer's voice print and subsequent audio recordings. Only authorised study staff and our collaborators at the University of North Carolina at Charlotte have access to the data on AWS.

We collect and store ground truth data and survey data in a secure online database, Research Electronic Data Capture (REDCap).^{45 46} REDCap is a commonly used platform in clinical research with HIPAA-compliant data security. Only authorised study staff access the REDCap database.

Patient and public involvement

Patients and the public were not involved in the design of this study. Study participants can request information on the results on study completion.

DISCUSSION AND EXPECTED IMPACT

SocialBit is a machine learning algorithmic framework designed to detect and quantify social interactions, and as inverse measure, social isolation. The goal of SocialBit is to be an objective, valid and easy-to-interpret metric of social interactions for individuals with diverse abilities. The current study aims to validate the accuracy of SocialBit in detecting social interactions in a sample of stroke survivors with a variety of speech, cognitive and physical impairments. This study takes advantage of the unique opportunity to directly observe patients for multiple days in a hospital setting.

The implications of this technology are multifold. SocialBit could be a useful tool for social sensing in individuals with diverse abilities, as well as detecting social isolation in vulnerable individuals with high accuracy. The algorithm could be used as a basis for social therapeutics in which social isolation is detected and acted on quickly, leading to improved health outcomes. These ideas connect with trends in the social sensing literature including the importance of social sensing for disease surveillance, health behaviour monitoring and intervention design.⁴⁷ These measurement possibilities also answer the call for greater understanding of the effects of social isolation and loneliness on public health.⁴⁸

There are some limitations to the current study. First, the hospital environment in which the study is conducted

is not entirely natural, and interaction frequency, types of persons and types of interactions may differ from what an individual experiences in their day-to-day life. Additionally, the hospital environment is often noisy with monitors and televisions, which may interfere with the algorithm's ability to detect social interactions. Lastly, due to HIPAA concerns, the algorithm does not store raw audio, missing the ability to understand pitfalls in specific cases.

In conclusion, the validation of SocialBit in stroke survivors represents an important step forward in the development of an objective, valid and easy-to-interpret metric of social interactions. The ability to detect social isolation with high accuracy and sensitivity could lead to improved health outcomes for vulnerable individuals. Further studies are needed to determine the utility of SocialBit in other populations and settings, and to determine how best to use SocialBit in social therapeutics to improve health outcomes.

Ethics and dissemination

We received ethical approval from the Institutional Review Board at Mass General Brigham (Protocol #2020P003739). We obtain written informed consent from patients and their caregivers, and we reimburse patients who complete the study with a \$100 cheque for their time. We offer a \$50 cheque to patients who do not complete the entire study.

Once this study is complete, results will be submitted for publication in a peer-reviewed journal and data will be available from the corresponding author on reasonable request.

Author affiliations

¹Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, USA

²Harvard Medical School, Boston, Massachusetts, USA

³Department of Computer Science, The University of North Carolina at Charlotte, Charlotte, North Carolina, USA

⁴Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital Boston, Boston, Massachusetts, USA

⁵Department of Engineering, University of Southern California, Los Angeles, California, USA

⁶Department of Psychology, University of Arizona, Tucson, Arizona, USA

Acknowledgements The authors thank all past, current and future participants in the SocialBit Study for their contributions towards creating a metric of social interactions to help detect social isolation.

Contributors AD, MRM and MS are responsible for the conceptualisation and design of the study. KW wrote the manuscript. AD, ST, RZ, SN, MRM and MS made substantial contributions to the manuscript. KW, ST, RZ, SN, MRM, MS and AD reviewed and edited the manuscript.

Funding This work is supported by the National Institutes of Health (grant number R01HD099176).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s)

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Kelly White <http://orcid.org/0000-0003-0761-549X>

Amar Dhand <http://orcid.org/0000-0001-6470-7548>

REFERENCES

- Pantell M, Rehkopf D, Jutte D, *et al*. Social isolation: a predictor of mortality comparable to traditional clinical risk factors. *Am J Public Health* 2013;103:2056–62.
- Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol* 1979;109:186–204.
- House JS, Landis KR, Umberson D. Social relationships and health. *Science* 1988;241:540–5.
- Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. *PLOS Med* 2010;7:e1000316. 10.1371/journal.pmed.1000316 Available: <https://doi.org/10.1371/journal.pmed.1000316>
- Dhand A, Longstreth WT Jr, Chaves PHM, *et al*. Social network trajectories in myocardial infarction versus ischemic stroke. *J Am Heart Assoc* 2018;7:e008029.
- Northcott S, Marshall J, Hilari K. What factors predict who will have a strong social network following a stroke *J Speech Lang Hear Res* 2016;59:772–83.
- Hilari K, Northcott S. Struggling to stay connected”: comparing the social relationships of healthy older people and people with stroke and aphasia. *Aphasiology* 2017;31:674–87.
- Northcott S, Hilari K. Why do people lose their friends after a stroke *Int J Lang Commun Disord* 2011;46:524–34.
- Northcott S, Hirani SP, Hilari K. A typology to explain changing social networks post stroke. *Gerontologist* 2018;58:500–11.
- Longman J, Passey M, Singer J, *et al*. The role of social isolation in frequent and/or avoidable hospitalisation: rural community-based service providers' perspectives. *Aust Health Rev* 2013;37:223–31.
- Reis HT, Wheeler L. Studying social interaction with the Rochester interaction record. *Adv Exp Soc Psychol* 1991;24:269–318.
- Berthier ML. Poststroke Aphasia epidemiology, pathophysiology and treatment. *Drugs Aging* 2005;22:163–82.
- Mellon L, Brewer L, Hall P, *et al*. Cognitive impairment six months after ischaemic stroke: a profile from the ASPIRE-S study. *BMC Neurol* 2015;15:31.
- Dhand A, Luke D, Lang C, *et al*. Social networks and risk of delayed hospital arrival after acute stroke. *Nat Commun* 2019;10:1206.
- Choudhury T, Pentland A. Sensing and modeling human networks using the sociometer. Seventh IEEE International Symposium on Wearable Computers, 2003; :216–22White Plains, NY, USA.
- Li Q, Chen S, Stankovic JA. Multi-modal in-person interaction monitoring using smartphone and on-body sensors
- Chen D, Yang J, Malkin R, *et al*. Detecting social interactions of the elderly in a nursing home environment. *ACM Trans Multimedia Comput Commun Appl* 2007;3:6.
- Wang R, Chen F, Chen Z, *et al*. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. in: proceedings of the 2014 ACM international joint conference on pervasive and Ubiquitous computing (Ubicomp '14). New York, NY: Association for Computing Machinery, September 13, 2014:3–14
- Dhand A, Dalton AE, Luke DA, *et al*. Accuracy of wearable cameras to track social interactions in stroke survivors. *J Stroke Cerebrovasc Dis* 2016;25:2907–10.
- Onnela J-P, Waber BN, Pentland A, *et al*. Using sociometers to quantify social interaction patterns. *Sci Rep* 2014;4:6278.
- Plakal M, Ellis D. Yamnet. 2020. Available: <https://github.com/tensorflow/models/tree/master/research/audioset/yamnet>
- Gemmeke JF, Ellis DPW, Freedman D, *et al*. Audio set: an ontology and human-labeled dataset for Audio events. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); New Orleans, LA, New Orleans, LA.
- Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. *NIPS* 2017.
- Dawalatabad N, Ravanelli M, Grondin F, *et al*. ECAPA-TDNN embeddings for speaker diarization. Interspeech 2021; ISCA, :3560–4
- Ortiz GA, Sacco RL. National Institutes of health stroke scale (NIHSS). *Wiley StatsRef: Statistics Reference* 2014. 10.1002/9781118445112 Available: <https://doi.org/10.1002/9781118445112.stat06823>
- Dhand A, White CC, Johnson C, *et al*. A Scalable online tool for quantitative social network assessment reveals potentially modifiable social environmental risks. *Nat Commun* 2018;9:3930.
- Russell DW. UCLA loneliness scale (version 3): reliability, validity, and factor structure. *J Pers Assess* 1996;66:20–40.
- Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385–401.
- Bruno A, Akinwuntan AE, Lin C, *et al*. Simplified modified Rankin scale questionnaire. *Stroke* 2011;42:2276–9.
- Banks JL, Marotta CA. Outcomes validity and reliability of the modified rankin scale: implications for stroke clinical trials. *Stroke* 2007;38:1091–6.
- Katzan IL, Fan Y, Uchino K, *et al*. The PROMIS physical function scale: a promising scale for use in patients with ischemic stroke. *Neurology* 2016;86:1801–7.
- van Swieten JC, Koudstaal PJ, Visser MC, *et al*. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988;19:604–7.
- Cella D, Riley W, Stone A, *et al*. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63:1179–94.
- Lyden P. Using the National Institutes of health stroke scale. *Stroke* 2017;48:513–9.
- Nasreddine ZS, Phillips NA, Bédirian V, *et al*. The Montreal cognitive assessment, Moca: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53:695–9.
- Julayanont P, Phillips N, Chertkow H, *et al*. Montreal cognitive assessment (Moca): concept and clinical review. *Cognitive Screening Instruments: A Practical Approach* 2013:111–51.
- Inouye SK, van Dyck CH, Alessi CA, *et al*. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med* 1990;113:941–8.
- Higginson IJ, Gao W, Jackson D, *et al*. Short-form Zarit Caregiver burden interviews were valid in advanced conditions. *J Clin Epidemiol* 2010;63:535–42.
- Yu J, Yap P, Liew TM. The optimal short version of the Zarit burden interview for dementia caregivers: diagnostic utility and externally validated cutoffs. *Aging Ment Health* 2019;23:706–10.
- Soto CJ, John OP. Short and extra-short forms of the big five Inventory-2: the BFI-2-S and BFI- 2-XS. *J Res Personal* 2017;68:69–81.
- Deng J, Dong W, Socher R, *et al*. Imagenet: a large-scale Hierarchical image database. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops); Miami, FL.Miami, Florida, June 20, 2009
- Hershey S, Chaudhuri S, Ellis DPW, *et al*. CNN architectures for large-scale audio classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); New Orleans, LA.
- Figueroa RL, Zeng-Treitler Q, Kandula S, *et al*. Predicting sample size required for classification performance. *BMC Med Inform Decis Mak* 2012;12:8.
- United States Department of Health and Human Services. Standards for privacy of individually identifiable health information. office of the federal register, national archives and records administration. 2002;67:53181–273.
- Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (Redcap) – A Metadata-driven methodology and Workflow process for providing Translational research Informatics support. *J Biomed Inform* 2009;42:377–81.
- Harris PA, Taylor R, Minor BL, *et al*. The Redcap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:S1532-0464(19)30126-1.
- Dhand A, Podury A, Choudhry N, *et al*. Leveraging social networks for the assessment and management of neurological patients. *Semin Neurol* 2022;42:136–48.
- Holt-Lunstad J, Perissinotto C. Social isolation and loneliness as medical issues. *N Engl J Med* 2023;388:193–5.