BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers’ comments and the authors’ responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open’s open peer review process please email info.bmjopen@bmj.com
# SocialBit: Protocol for a Prospective Study to Validate a Wearable Social Sensor for Persons with Diverse Neurological Abilities

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>BMJ Open</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bmjopen-2023-076297</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Protocol</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>02-Jun-2023</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | White, Kelly; Brigham and Women's Hospital, Neurology; Harvard Medical School  
Tate, Samuel; The University of North Carolina at Charlotte, Department of Computer Science  
Zafonte, Ross; Spaulding Rehabilitation Hospital Boston, Department of Physical Medicine and Rehabilitation  
Narayanan, Shrikanth; University of Southern California, Engineering  
Mehl, Matthias R.; University of Arizona, Department of Psychology  
Shin, Min; The University of North Carolina at Charlotte, Department of Computer Science  
Dhand, Amar; Brigham and Women's Hospital, Neurology; Harvard Medical School |
| Keywords: | Stroke < NEUROLOGY, Social Interaction, Quality of Life |
SocialBit: Protocol for a Prospective Study to Validate a Wearable Social Sensor for Persons with Diverse Neurological Abilities

Kelly White, BS
Research Assistant, Brigham and Women’s Hospital, Harvard Medical School.
65 Landsdowne Street, Cambridge, MA 02139.
kwhite50@bwh.harvard.edu.

Samuel Tate, MS
Software Engineer, Department of Computer Science, University of North Carolina at Charlotte.
9201 University City Boulevard, Charlotte, NC 28223
state18@uncc.edu.

Ross Zafonte, DO
Massachusetts General Hospital & Brigham and Women’s Hospital, Harvard Medical School;
Department of Physical Medicine and Rehabilitation
Spaulding Rehabilitation Hospital
300 First Avenue, Suite 2105, Charlestown, MA 02129
rzafonte@mgh.harvard.edu; 617-952-5227 (t).

Shrikanth Narayanan, PhD
University Professor and Nikias Chair in Engineering, University of Southern California.
3740 McClintock Ave, Los Angeles CA 90089
shri@usc.edu; 213 740 6432 (t).

Matthias R. Mehl, PhD
Professor of Psychology, Department of Psychology, University of Arizona.
1503 East University Boulevard, Building 68, Tucson, AZ 85721.
mehl@arizona.edu; 520-626-2374 (t).

Min Shin, PhD
Professor and Chair of Computer Science, Department of Computer Science, University of North Carolina at Charlotte.
9201 University City Boulevard, Charlotte, NC 28223.
mcsbin@uncc.edu; 704-687-8578 (t); 704-687-3516 (f).

Amar Dhand, MD, DPhil
Associate Professor of Neurology, Brigham and Women’s Hospital, Harvard Medical School;
Affiliate Professor, Network Science Institute, Northeastern University.
65 Landsdowne Street, Cambridge, MA 02139.
adhand@bwh.harvard.edu; 617-732-5330 (t); 617-732-6083 (f).

Corresponding Author: Amar Dhand

Source of Support: This work is supported by the National Institutes of Health grant number R01HD099176. Word Count: 4,337 Figures: 2 Tables: 3
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 characterized 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 hospitalization 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 analyze 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.
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 minimize 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.
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 utilization of preventive health services. [2] Three decades ago, a meta-analysis concluded that social isolation was a major risk factor for health, and rivaled the effect of cigarette smoking, blood pressure, blood lipids, obesity, and physical activity. [3] More recently, a meta-analysis of 148 studies showed that persons with stronger social relationships had a 50% increased likelihood of survival, an odds ratio 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, capitalizing on the ability to monitor patients’ social interactions in real-time for multiple days.

SocialBit is the first wearable social interaction detection sensor customized 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 characterization 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 characterization. [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 was 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 GPS information as well as the individuals' calendar events to analyze 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 analyzes the temporal change of vocal acoustic behavior, 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, analyzed,
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 pre-trained neural network called YAMNet, [21] which is capable of
classifying over 500 audio events from the public AudioSet dataset. [22] YAMNet
produces feature vectors for each 0.96-second segment of audio, which are then
combined and fed through a Transformer network [23] to determine whether the
sequence contains a social interaction. Additionally, SocialBit collects voiceprint
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 the five short sentences from the NIH Stroke Scale. [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 8 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 an additional 3 days of data if a patient transitions to Spaulding Rehabilitation Hospital Boston. We enroll 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 development process.

Beginning on June 15, 2021, and continuing to early 2025, we are recruiting patients with acute ischemic stroke at Brigham and Women’s Hospital in Boston, MA. 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 ischemic 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 requests 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 one hour of surveys with the research staff, including an NIH Stroke Scale, 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 (see Table 2 for the full list of survey measures). We ask patients these same survey questions when they return for their follow-up appointment with the study’s Principal Investigator, Dr. Amar Dhand, 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 am to 5:00 pm 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 3 days. On average, we aim to collect 2-3 days (16-24 hours) of data per patient (i.e., around eight 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 NIH Stroke Scale. We ask the patient to read the five sentences aloud for 30 seconds. 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 HIPAA-certified research staff member is on the other side of the livestream manually coding, with a temporal resolution of one minute, 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, signaling 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 hospitalization 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, including the questions answered each minute. Table 2 is the guidelines for making coding judgements for each question.
Table 1. Ground truth data collection recorded every minute from 9:00 am to 5:00 pm.

<table>
<thead>
<tr>
<th>Ground Truth Table Question (per minute)</th>
<th>Answer (per minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the patient talking to another person?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Who was the patient talking with?</td>
<td>☐ Medical personnel ☐ Caregiver ☐ Other family and friends, adult or child ☐ Patient in other bed ☐ Stranger/Other ☐ Can’t Tell</td>
</tr>
<tr>
<td>How was the patient talking with the person(s)?</td>
<td>○ In-person ○ Phone call ○ Video conference/chat</td>
</tr>
<tr>
<td>How many people was the patient talking with? (i.e. actively in conversation)</td>
<td>○ 1 ○ 2 ○ 3 or more</td>
</tr>
<tr>
<td>Does the conversation contain foreign language?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>What was the tone of the interaction? (-2=Negative, 0=Neutral, 2=Positive)</td>
<td>○ -2 ○ -1 ○ 0 ○ 1 ○ 2 ○ Can’t tell</td>
</tr>
<tr>
<td>What was the depth of the conversation? (1=Superficial to 5=Deep)</td>
<td>○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ Can’t tell</td>
</tr>
<tr>
<td>Was the TV on? (TV includes computer or other devices playing entertainment in a way that is audible in the room.)</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Was there a conversation happening that the patient was not part of?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Was the patient completely silent during this minute? (Did not contribute any words or sounds)</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Anything else that was noteworthy? (i.e. was the patient talking with you, the observer)</td>
<td>________</td>
</tr>
</tbody>
</table>
Table 2. Guidelines for making coding judgements for each question of the ground truth table.

<table>
<thead>
<tr>
<th>Question</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the patient talking with another person?</td>
<td>If any social interaction occurs with the patient, even for a couple of seconds, during this one-minute timeframe, the research staff select “Yes.” All utterances spoken by or to a patient implies 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.”</td>
</tr>
<tr>
<td>Who was the patient talking with?</td>
<td>The research staff categorize 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.</td>
</tr>
<tr>
<td>How was the patient talking with the person(s)?</td>
<td>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.</td>
</tr>
<tr>
<td>How many people was the patient talking with?</td>
<td>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”.</td>
</tr>
<tr>
<td>Does the conversation contain foreign language?</td>
<td>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.</td>
</tr>
<tr>
<td>What was the tone of the interaction?</td>
<td>The research staff select one answer from a five-point scale by making a judgment 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.</td>
</tr>
<tr>
<td>What was the depth of the conversation?</td>
<td>The research staff select one answer from a five-point scale by making a judgment 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 (e.g., 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.</td>
</tr>
<tr>
<td>Was the TV on?</td>
<td>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.</td>
</tr>
<tr>
<td>Was there a conversation happening that the patient was not part of?</td>
<td>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 amongst nurses. Staff select “Yes” to this question even if the outside conversation is barely audible.</td>
</tr>
<tr>
<td>Was the patient completely silent during this minute?</td>
<td>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.”</td>
</tr>
<tr>
<td>Anything else that was noteworthy?</td>
<td>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 (e.g., medical alarms). Additionally, staff note whether a patient was on speakerphone during a phone call.</td>
</tr>
</tbody>
</table>
To ensure study staff are taking a standardized 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 standardized answer key. The 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 study lead coordinator re-trains auxiliary study staff every six 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 Scale [27] and 20-item Center for Epidemiological Studies Depression Scale [28] to assess patients’ subjective feelings of loneliness and depression. These surveys are conducted at patient enrollment 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 Scale [29] and NIH 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 questions [29] 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 enrollment as well as at their 3-month follow-up visit. Data from enrollment is compared to 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 NIH Stroke Scale (NIHSS) [34] and cognitive abilities through the reliable and validated Montreal Cognitive Assessment (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 enrollment and at the 3-month follow-up to determine a patient’s physical and cognitive
improvement over the first three 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] This data is 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.
**Table 3.** List of all measures collected in the SocialBit study, including measures for exposures, outcomes, and covariates.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measure</th>
<th>Participant</th>
<th>T1 at Intake</th>
<th>T2 at 3-Months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction Time</td>
<td>Interaction detected by SocialBit application</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Social interaction Time</td>
<td>Ground Truth Data collected by Study Staff</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Isolation (objective)</td>
<td>Personal Netowork Survey for Clinical Research (PERSNET)</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Social isolation (subjective)</td>
<td>UCLA Loneliness Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Disability</td>
<td>Modified Rankin Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Physical Function (subjective)</td>
<td>NIH Patient-Reported Outcome Measurement Information System (PROMIS), Physical Function</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Depression</td>
<td>Center for Epidemiological Studies Depression (CES-D)</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Life Satisfaction</td>
<td>Satisfaction with Life Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cognition</td>
<td>Montreal Cognitive Assessment</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stroke Severity</td>
<td>National Institutes of Health Stroke Scale (NIHSS)</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Delirium</td>
<td>Confusion Assessment Method</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caregiver Burden</td>
<td>Caregiver Burden Scale</td>
<td>Caregiver</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>Demographics</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Charlson Comorbidity Index</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lesion Characteristics</td>
<td>Medical record, neuroimaging</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Big Five Personality</td>
<td>Big Five Inventory-2 XS</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Analysis plan

The analysis focuses on determining the accuracy of the SocialBit algorithm in detecting social interactions between the primary device-wearer and others. Secondly, 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, five-fold cross validation is used with 40 patients per fold. The unit of analysis is social interaction or no social interaction in every 1-minute interval for 8 hours a day. If social interaction occurs partially within an interval, then the entire 1-minute interval is marked as social interaction. To be consistent with the ground truth, the automated algorithm processes at the 1-minute interval as well. However, only 1 minute every 5 or 6 minutes depending on the battery capacity is analyzed by the algorithm. We minimize the overfitting to training data through optimizing hyperparameters (e.g., regularization 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 datasets. 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 odds ratio. 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 x 24 hours x 60 minutes/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 4,800 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) [22] and ImageNet (~1,000 average samples
per class, ~3,000 for the mode). [41] Furthermore, deep learning algorithms have
achieved good accuracy for each of these datasets. For example, a deep learning
algorithm like ours achieved an 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 analyze the dataset 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 the Health Insurance Portability and
Accountability Act (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 authorized
study staff and our collaborators at 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 authorized study staff access the REDCap database.

Patient and Public Involvement

Study participants were not involved in the design of this study. The Principal Investigator and Co-Investigators are responsible for the conceptualization and design of the study, and this study was approved by the Institutional Review Board at Mass General Brigham. Study participants can request information on the results upon 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 multi-fold. 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 upon 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 behavior 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. Firstly, 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 check for their time. We offer a $50 check 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 upon reasonable request.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to patient confidentiality and ongoing data collection but are available from the corresponding author on reasonable request.

Acknowledgements

We would like to 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.
Author Contributions

AD, MM, and MS are responsible for the conceptualization and design of the study. KW wrote the manuscript. AD made substantial contributions to the manuscript, and all authors reviewed and edited the manuscript.

Funding

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

Competing Interests

All authors declare no competing interests for this paper.
REFERENCES


   https://doi.org/10.1371/journal.pmed.1000316

   doi: 10.1161/JAHA.117.008029


   https://doi.org/10.1002/9781118445112.stat06823.


Figure 1. Machine learning algorithm that detects the probability of a social interaction.
Figure 2. Timeline for the SocialBit study spanning 3 months.
SocialBit: Protocol for a Prospective Observational Study to Validate a Wearable Social Sensor for Stroke Survivors with Diverse Neurological Abilities

<table>
<thead>
<tr>
<th>Journal:</th>
<th>BMJ Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bmjopen-2023-076297.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Protocol</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>01-Aug-2023</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>White, Kelly; Brigham and Women's Hospital, Neurology; Harvard Medical School Tate, Samuel; The University of North Carolina at Charlotte, Department of Computer Science Zafonte, Ross; Spaulding Rehabilitation Hospital Boston, Department of Physical Medicine and Rehabilitation Narayanan, Shrikanth; University of Southern California, Engineering Mehl, Matthias R.; University of Arizona, Department of Psychology Shin, Min; The University of North Carolina at Charlotte, Department of Computer Science Dhand, Amar; Brigham and Women's Hospital, Neurology; Harvard Medical School</td>
</tr>
<tr>
<td>Primary Subject Heading:</td>
<td>Research methods</td>
</tr>
<tr>
<td>Secondary Subject Heading:</td>
<td>Neurology, Communication, Diagnostics, Sociology</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Stroke &lt; NEUROLOGY, Social Interaction, Quality of Life</td>
</tr>
</tbody>
</table>
SocialBit: Protocol for a Prospective Observational Study to Validate a Wearable Social Sensor for Stroke Survivors with Diverse Neurological Abilities

Kelly White, BS
Research Assistant, Brigham and Women’s Hospital, Harvard Medical School.
65 Landsdowne Street, Cambridge, MA 02139.
kwhite50@bwh.harvard.edu.

Samuel Tate, MS
Software Engineer, Department of Computer Science, University of North Carolina at Charlotte.
9201 University City Boulevard, Charlotte, NC 28223
state18@uncc.edu.

Ross Zafonte, DO
Massachusetts General Hospital & Brigham and Women’s Hospital, Harvard Medical School;
Department of Physical Medicine and Rehabilitation
Spaulding Rehabilitation Hospital
300 First Avenue, Suite 2105, Charlestown, MA 02129
rzafonte@mgh.harvard.edu; 617-952-5227 (t).

Shrikanth Narayanan, PhD
University Professor and Nikias Chair in Engineering, University of Southern California.
3740 McClintock Ave, Los Angeles CA 90089
shri@usc.edu; 213 740 6432 (t).

Matthias R. Mehl, PhD
Professor of Psychology, Department of Psychology, University of Arizona.
1503 East University Boulevard, Building 68, Tucson, AZ 85721.
mehl@arizona.edu; 520-626-2374 (t).

Min Shin, PhD
Professor and Chair of Computer Science, Department of Computer Science, University of North Carolina at Charlotte.
9201 University City Boulevard, Charlotte, NC 28223.
mcshin@uncc.edu; 704-687-8578 (t); 704-687-3516 (f).

Amar Dhand, MD, DPhil
Associate Professor of Neurology, Brigham and Women’s Hospital, Harvard Medical School;
Affiliate Professor, Network Science Institute, Northeastern University.
65 Landsdowne Street, Cambridge, MA 02139.
adhand@bwh.harvard.edu; 617-732-5330 (t); 617-732-6083 (f).

Corresponding Author: Amar Dhand

Source of Support: This work is supported by the National Institutes of Health grant number R01HD099176. Word Count: 4,368 Figures: 2 Tables: 3
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 characterized 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 hospitalization 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 analyze 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.
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 minimize 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.
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 utilization of preventive health services. [2] Three decades ago, a meta-analysis concluded that social isolation was a major risk factor for health, and rivaled the effect of cigarette smoking, blood pressure, blood lipids, obesity, and physical activity. [3] More recently, a meta-analysis of 148 studies showed that persons with stronger social relationships had a 50% increased likelihood of survival, an odds ratio 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, capitalizing on the ability to monitor patients’ social interactions in real-time for multiple days.

SocialBit is the first wearable social interaction detection sensor customized 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 characterization 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 characterization. [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 GPS information as well as the individuals' calendar events to analyze 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 analyzes the temporal change of vocal acoustic behavior, 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, analyzed,
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 pre-trained neural network called YAMNet, [21] which is capable of
classifying over 500 audio events from the public AudioSet dataset. [22] YAMNet
produces feature vectors for each 0.96-second segment of audio, which are then
combined and fed through a Transformer network [23] to determine whether the
sequence contains a social interaction. Additionally, SocialBit collects voiceprint
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 the five short sentences from the NIH Stroke Scale. [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 8 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 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 enroll 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 development process.

Beginning on June 15, 2021, and continuing to early 2025, we are recruiting patients with acute ischemic stroke at Brigham and Women’s Hospital in Boston, MA. 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 ischemic 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 requests 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 one hour of surveys with the research staff, including an NIH Stroke
Scale, 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, Dr. Amar Dhand, 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 am to 5:00 pm 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 eight 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 NIH Stroke Scale. We ask the patient to read the five sentences aloud for 30 seconds. 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 HIPAA-certified research staff member is on the other side of the livestream manually coding, with a
temporal resolution of one minute, 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, signaling 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 hospitalization 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, including the questions answered each minute. **Table 2** is the
guidelines for making coding judgements for each question.
Table 1. Ground truth data collection recorded every minute from 9:00 am to 5:00 pm.

<table>
<thead>
<tr>
<th>Ground Truth Table Question (per minute)</th>
<th>Answer (per minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the patient talking to another person?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Who was the patient talking with?</td>
<td>○ Medical personnel ○ Caregiver ○ Other family and friends, adult or child ○ Patient in other bed ○ Stranger/Other ○ Can’t Tell</td>
</tr>
<tr>
<td>How was the patient talking with the person(s)?</td>
<td>○ In-person ○ Phone call ○ Video conference/chat</td>
</tr>
<tr>
<td>How many people was the patient talking with? (i.e. actively in conversation)</td>
<td>○ 1 ○ 2 ○ 3 or more</td>
</tr>
<tr>
<td>Does the conversation contain foreign language?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>What was the tone of the interaction? (-2=Negative, 0=Neutral, 2=Positive)</td>
<td>○ -2 ○ -1 ○ 0 ○ 1 ○ 2 ○ Can’t tell</td>
</tr>
<tr>
<td>What was the depth of the conversation? (1=Superficial to 5=Deep)</td>
<td>○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ Can’t tell</td>
</tr>
<tr>
<td>Was the TV on? (TV includes computer or other devices playing entertainment in a way that is audible in the room.)</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Was there a conversation happening that the patient was not part of?</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Was the patient completely silent during this minute? (Did not contribute any words or sounds)</td>
<td>○ Yes ○ No</td>
</tr>
<tr>
<td>Anything else that was noteworthy? (i.e. was the patient talking with you, the observer)</td>
<td>__________</td>
</tr>
</tbody>
</table>
Table 2. Guidelines for making coding judgements for each question of the ground truth table.

<table>
<thead>
<tr>
<th>Question</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the patient talking with another person?</td>
<td>If any social interaction occurs with the patient, even for a couple of seconds, during this one-minute timeframe, the research staff select “Yes.” All utterances spoken by or to a patient implies 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.”</td>
</tr>
<tr>
<td>Who was the patient talking with?</td>
<td>The research staff categorize 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.</td>
</tr>
<tr>
<td>How was the patient talking with the person(s)?</td>
<td>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.</td>
</tr>
<tr>
<td>How many people was the patient talking with?</td>
<td>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”.</td>
</tr>
<tr>
<td>Does the conversation contain foreign language?</td>
<td>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.</td>
</tr>
<tr>
<td>What was the tone of the interaction?</td>
<td>The research staff select one answer from a five-point scale by making a judgment 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.</td>
</tr>
<tr>
<td>What was the depth of the conversation?</td>
<td>The research staff select one answer from a five-point scale by making a judgment 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 (e.g., 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.</td>
</tr>
<tr>
<td>Was the TV on?</td>
<td>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.</td>
</tr>
<tr>
<td>Was there a conversation happening that the patient was not part of?</td>
<td>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 amongst nurses. Staff select “Yes” to this question even if the outside conversation is barely audible.</td>
</tr>
<tr>
<td>Was the patient completely silent during this minute?</td>
<td>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.”</td>
</tr>
<tr>
<td>Anything else that was noteworthy?</td>
<td>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 (e.g., medical alarms). Additionally, staff note whether a patient was on speakerphone during a phone call.</td>
</tr>
</tbody>
</table>
To ensure study staff are taking a standardized 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 standardized answer key. The 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 study lead coordinator re-trains auxiliary study staff every six 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 Scale [27] and 20-item Center for Epidemiological Studies Depression Scale [28] to assess patients’ subjective feelings of loneliness and depression. These surveys are conducted at patient enrollment 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 Scale
[29] and NIH 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 questions [29] 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 enrollment as well as at their 3-month follow-up visit.
Data from enrollment is compared to 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 NIH Stroke Scale
(NIHSS) [34] and cognitive abilities through the reliable and validated Montreal
Cognitive Assessment (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
enrollment and at the 3-month follow-up to determine a patient’s physical and cognitive
improvement over the first three 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] This data is 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.
### Table 3. List of all measures collected in the SocialBit study, including measures for exposures, outcomes, and covariates.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measure</th>
<th>Participant</th>
<th>T1 at Intake</th>
<th>T2 at 3-Months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction Time –</td>
<td>Interaction detected by SocialBit application</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SocialBit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social interaction Time –</td>
<td>Ground Truth Data collected by Study Staff</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ground Truth Coded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Isolation (objective)</td>
<td>Personal Network Survey for Clinical Research (PERSNET)</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Social isolation (subjective)</td>
<td>UCLA Loneliness Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Disability</td>
<td>Modified Rankin Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Physical Function (subjective)</td>
<td>NIH Patient-Reported Outcome Measurement Information System (PROMIS), Physical Function</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Depression</td>
<td>Center for Epidemiological Studies Depression (CES-D)</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Life Satisfaction</td>
<td>Satisfaction with Life Scale</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cognition</td>
<td>Montreal Cognitive Assessment</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stroke Severity</td>
<td>National Institutes of Health Stroke Scale (NIHSS)</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Delirium</td>
<td>Confusion Assessment Method</td>
<td>Patient</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caregiver Burden</td>
<td>Caregiver Burden Scale</td>
<td>Caregiver</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>Demographics</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Charlson Comorbidity Index</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lesion Characteristics</td>
<td>Medical record, neuroimaging</td>
<td>Patient</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Big Five Personality</td>
<td>Big Five Inventory-2 XS</td>
<td>Patient and Caregiver</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Analysis plan

The analysis focuses on determining the accuracy of the SocialBit algorithm in detecting social interactions between the primary device-wearer and others. Secondly, 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, five-fold cross validation is used with 40 patients per fold. The unit of analysis is social interaction or no social interaction in every 1-minute interval for 8 hours a day. If social interaction occurs partially within an interval, then the entire 1-minute interval is marked as social interaction. To be consistent with the ground truth, the automated algorithm processes at the 1-minute interval as well. However, only 1 minute every 5 or 6 minutes depending on the battery capacity is analyzed by the algorithm. We minimize the overfitting to training data through optimizing hyperparameters (e.g., regularization 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 datasets. 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 odds ratio. 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 x 24 hours x 60 minutes/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 4,800 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) [22] and ImageNet (~1,000 average samples per class, ~3,000 for the mode). [41] Furthermore, deep learning algorithms have achieved good accuracy for each of these datasets. For example, a deep learning algorithm like ours achieved an 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 analyze the dataset 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 the Health Insurance Portability and Accountability Act (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 authorized
study staff and our collaborators at 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 authorized 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 upon 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 multi-fold. 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 upon 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 behavior 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. Firstly, 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 check for their time. We offer a $50 check 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 upon reasonable request.

**Data Availability**
The datasets generated and/or analyzed during the current study are not publicly available due to patient confidentiality and ongoing data collection but are available from the corresponding author on reasonable request.

**Acknowledgements**
We would like to 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.

**Author Contributions**
AD, MM, and MS are responsible for the conceptualization and design of the study. KW wrote the manuscript. AD, ST, RZ, SN, MM, and MS made substantial contributions to the manuscript. KW, ST, RZ, SN, MM, MS, and AD reviewed and edited the manuscript.
Funding

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

Competing Interests

All authors declare no competing interests for this paper.
REFERENCES


https://doi.org/10.1002/9781118445112.stat06823


Figure 1. Machine learning algorithm that detects the probability of a social interaction.
Figure 2. Timeline for the SocialBit study spanning 3 months.