Case Example: Using Remote, Smartphone-Based Data Collection to Share Health Insights

Study Overview

In a research study evaluating the feasibility of performing remote, objective assessments of Parkinson's Disease (PD), sensor data was collected from smartphones during self-administered activities designed to assess PD symptoms, including dexterity, balance, gait, and phonation. The goal of the study was to evaluate the utility of smartphone-based remote assessments to monitor disease symptoms. If successful, such tools could be used by clinicians, researchers, and patients to monitor PD.

The study was also designed to provide broad access to study data so that qualified researchers with an interest in digital health assessments could gain a first-hand understanding of the nature and nuance of such data. The goal of the data-sharing program was to provide open data resources and transparent analytical tools to establish early parameters to guide researchers across the digital health field in terms of what, how, and why to use digital data capture.

Collecting and Sharing Data

Throughout the study, data sharing was managed directly by participants through the study application. Upon completion of the self-guided visual consent process, participants were explicitly required to state their preferences for data use. Participants were given the option to share data either with the primary research team alone or more broadly with qualified researchers. Of the 9,000 participants that enrolled in the first six months of the study, more than 75% opted to broadly share their study data.

The lead center curated the first six months of study data for use, within months of completion of data collection. This data was released through a qualified researcher program in keeping with the consent participants provided. Data governance structures were designed and put in place to:

1) Balance the expected protection of participants' privacy with their desire for optimal data use and reuse;
2) Emphasize transparency, so that anyone can know how data are being used and by whom;
3) Describe and cultivate a clear set of behavioral norms for working with participant-donated data sets;
4) Assess data requester's knowledge of research ethics and appropriate conduct for accessing, using, and managing participant-donated data; and
5) Emphasize return of information, data, and results to participants and the research community.

To maximize the usability of these data, a companion data descriptor paper was published that provided detailed descriptions of study design and data collection to support external use. This system has been used to distribute six digital health data sets, to date.

Data Governance

The data governance model focuses on evaluation and monitoring of qualified researchers to maximize transparency in data use. Individual researchers are "qualified" based on how they

plan to use the data, with a specific focus on transparency, rather than uses based on traditional scholarly metrics. As such, any researcher interested in accessing the data must complete the following steps:

1) Demonstrate their awareness and understanding of the data-sharing framework and applied ethics through a short, 18-question examination;
2) Validate their identity through a variety of approved methods, such as an academic letter from a signing official, a notarized letter attesting to identity, or a copy of a professional license;
3) Make a public statement of intended data use, which we can then use to provide feedback to participants in the spirit of engagement and transparency; and
4) Explicitly agree to a "contract" of data sharing, including the following:
    (i) Downloading, initialing, signing, scanning, and uploading a researcher oath to adhere to a code of behavior;
    (ii) Complying with any data-specific conditions of use. More than 150 researchers have used this system to access the data set over the past two years for use in independent research programs.

Crowdsourcing to Conduct Better Research

To further derive utility from these data contributions, use and analysis across the research community were actively catalyzed through a crowd-sourced analytical challenge. In this project, 440 researchers joined together to evaluate existing analytical approaches to derive biologically-meaningful features from smartphone-based sensor data collected during gait and balance tests. Research teams included multiple organizations that are working to develop their own digital PD data collection technologies—all of whom will need to understand how best to effectively process these data. Using the qualified researcher program to access the sensor data, researchers identified a signal processing mechanism for feature extraction that could detect PD status with an AUROC[*] of 0.87—significantly better than any researcher was doing before this challenge.

Positive Results

In conclusion, broad sharing of these data enabled a community of researchers to unite around common questions instead of a single institution controlling what insights are developed from this resource.

Reference: Relevant CTTI Recommendations

Sponsors should ensure that they are aware of and comfortable with the ways in which data generated by mobile technologies used in their trials may be accessed and used by the technology manufacturer and any additional third parties. This information should be clearly stipulated in the outsourcing agreement(s) and a clear accounting of which parties will have access to each level of data should be included in the informed consent and HIPAA research authorization form. CTTI recommends that sponsors engage potential participants in these discussions regarding access to and use of data by external entities to reach a decision that ultimately meets the patients' level of comfort and expectation of privacy. (Click here for more)

---

[*] The Area Under the curve of the Receiver Operating Characteristic (AUROC) of a test can be used as a criterion to measure the test's discriminative ability, i.e. how good is the test in a given clinical situation.