



Published in final edited form as:

J Biopharm Stat. 2022 May 04; 32(3): 496–510. doi:10.1080/10543406.2022.2080698.

Lessons Learned in the Apple Heart Study and Implications for the Data Management of Future Digital Clinical Trials

Ariadna Garcia^{1,2}, Vidhya Balasubramanian, MS^{1,2}, Justin Lee, MPH^{1,2}, Rebecca M. Gardner, MS^{1,2}, Santosh E. Gummidipundi, MS^{1,2}, Grace Hung, MS³, Todd Ferris, MD³, Lauren Cheung, MD⁴, Christopher B. Granger, MD⁵, Peter Kowey, MD⁶, John S. Rumsfeld, MD PhD⁷, Andrea M. Russo, MD⁸, Mellanie True Hills, BS⁹, Nisha Talati, MBA¹⁰, Divya Nag⁴, Jeffrey W. Stein, PhD⁴, David Tsay, MD PhD⁴, Sumbul Desai, MD⁴, Kenneth W. Mahaffey, MD¹⁰, Mintu P. Turakhia, MD MAS^{2,11}, Marco V. Perez, MD², Haley Hedlin, PhD^{1,2}, Manisha Desai, PhD^{1,2} on behalf of the Apple Heart Study Investigators

¹Quantitative Sciences Unit, Stanford University School of Medicine (Stanford, CA)

²Department of Medicine, Stanford University School of Medicine (Stanford, CA)

³Information Resources and Technology, Stanford University School of Medicine (Stanford, CA)

⁴Apple Inc. (Cupertino, CA)

⁵Duke Clinical Research Institute, Duke University (Durham, NC)

⁶Lankenau Heart Institute and Jefferson Medical College (Philadelphia, PA)

⁷Department of Medicine, University of Colorado School of Medicine (Aurora, CO)

⁸Department of Medicine, Cooper Medical School of Rowan University (Camden, NJ)

⁹[StopAfib.org](https://www.stopafib.org), American Foundation for Women's Health (Decatur, TX)

¹⁰Stanford Center for Clinical Research, Stanford University School of Medicine (Stanford, CA)

¹¹Center for Digital Health, Stanford University School of Medicine (Stanford, CA)

Abstract

The digital clinical trial is fast emerging as a pragmatic trial that can improve a trial's design including recruitment and retention, data collection and analytics. To that end, digital platforms such as electronic health records or wearable technologies that enable passive data collection can be leveraged, alleviating burden from the participant and study coordinator. However, there are challenges. For example, many of these data sources not originally intended for research may be noisier than traditionally-obtained measures. Further, the secure flow of passively collected data and their integration for analysis is non-trivial. The Apple Heart Study was a prospective, single-arm, site-less digital trial designed to evaluate the ability of an app to detect atrial fibrillation.

The study was designed with pragmatic features, such as an app for enrollment, a wearable device

Address of Correspondence: Manisha Desai, Quantitative Sciences Unit, 1701 Page Mill Road, Palo Alto CA 94304, Tel (650) 725-1946, manishad@stanford.edu.

Declaration of interest statement:

None

(the Apple Watch) for data collection, and electronic surveys for participant-reported outcomes that enabled a high volume of patient enrollment and accompanying data. These elements led to challenges including identifying the number of unique participants, maintaining participant-level linkage of multiple complex data streams, and participant adherence and engagement. Novel solutions were derived that inform future designs with an emphasis on data management. We build upon the excellent framework of the Clinical Trials Transformation Initiative to provide a comprehensive set of guidelines for data management of the digital clinical trial that include an increased role of collaborative data scientists in the design and conduct of the modern digital trial.

Keywords

pragmatic trial; digital clinical trial; wearables; atrial fibrillation; digital health

1. Introduction

The incorporation of digital technologies into both medical practice and daily life are increasingly reflected in the current research landscape. Many contemporary clinical trials are designed with various levels of pragmatism including those that rely on digital-based interventions, giving rise to what is referred to as the digital clinical trial (DCT) (Inan et al. 2020). Inan and others (2020) define the DCT as one that leverages digital technology in order to improve critical aspects of a trial including recruitment and retention, data collection and analytics. For example, DCTs bring several advantages related to their ability to passively collect data. First, digital tools may be used to recruit participants with little effort and may further relieve the burden on study participants (e.g. from having to come into the clinic) and on study staff (e.g., from coordinating visits and taking and recording measurements). As such, the DCT may provide an ability to engage many more participants potentially increasing generalizability of findings. For example, in both the Apple Heart Study (AHS) and the fitbit Heart Study, over 400,000 participants were enrolled (Turakhia et al. 2019; Lubitz et al. 2021). Second, once a participant has consented to join the study, passive data collection can be advantageous over a traditional design with on-site data acquisition by collecting data in the participant's normal environment and during activities of daily living. Third, many mobile devices enable a voluminous and rich stream of longitudinal data with measurements sampled much more frequently – opportunistically, or continuously – depending on the device. The frequent sampling of information facilitates monitoring and tracking of activities such as adherence as well as opportunities to incorporate those data into an intervention for example through positive reinforcement or informing participants on their status/performance.

Data capture through digital tools presents challenges, however, and we are not the first to acknowledge this (Clinical Trials Transformation Initiative 2021; Coran et al. 2019; Rosa et al. 2021; Murray et al. 2016). Digital tools often enable research using types of data that were not necessarily intended for research and that therefore differ from the quality of the data collected in the traditional clinical trial setting. The data may be noisier with additional sources of variation above and beyond traditional measures that may be related to participant and device behavior (Pham, Wilier, and Cafazzo 2016). The data generated from

devices will most likely need to be integrated with other types of data to address study goals (Rosa et al. 2015; Cornet and Holden 2018). For example, in the Apple Heart Study (AHS), diverse data (e.g., demographics and electrocardiogram (ECG) data) flowed from multiple types of devices (e.g., the Apple Watch, phones, laptops, ECG monitor) (Figure 1) and data needed to be integrated to address specific research questions posed. Further, assessing and handling incompleteness of data, depending on the device, may not be straightforward. In addition, the timing of measurements may differ from that of a traditional trial where measurements for the latter may be recorded at pre-specified fixed times as opposed to at opportunistic periods. Defining endpoints that leverage and are functions of near continuous data streams – is often novel – and needs validation (Herrington, Goldsack and Landray 2018). Much thought, therefore, needs to go into the study design when incorporating digital tools (Steinhubl, Muse, and Topol 2015; Tomlinson et al. 2013) with particular regard to mitigating any increased burden on the data management team due to challenges with complex data flow, integration, and processing while keeping in mind issues that impact the data to be analyzed downstream.

The Clinical Trials Transformation Initiative (CTTI) -- a group of thought leaders in government, academia, industry, and patient advocacy -- developed an excellent set of general guidelines for the design and conduct of DCTs (<https://ctti-clinicaltrials.org/our-work/digital-health-trials/>) with seven areas of focus: (1) supporting decentralized trial approaches, (2) developing novel endpoints, (3) selecting a digital health technology, (4) managing data, (5) delivering an investigational product, (6) preparing a site, and (7) interacting with regulators. While the recommendations provide a suitable framework for the design and conduct of the digital clinical trial, details and examples are needed on how to adhere to the recommendations, particularly for data management. For example, the guidelines under data management state generally that trials should “collect the minimum data set necessary to address the study endpoints”, “proactively address and map data flow, data storage and associated procedures”, and “minimize missing data”. Details on how to do so and in how steps taken would differ from those in a traditional clinical trial would strengthen adoption of the guidelines. Further, the recommendations could benefit from a few additions. For example, we applaud CTTI for prominently featuring feasibility studies under the two categories of Selecting and Testing a Digital Health Technology, and believe these principles should be extended to the data management recommendations, where feasibility studies prior to launch could play a critical role. Coran and others briefly allude to this additional benefit of the feasibility study in their discussion of the CTTI recommendations (Coran et al. 2019).

In this paper, we build upon the CTTI recommendations for the design and conduct of the DCT with focus on key data management issues specific to the DCT. We do this by describing three specific challenges that we faced in the Apple Heart Study (AHS): participant adherence, accounting for the unique number of participants enrolled in the trial; and using timestamp data to establish longitudinal trajectories for participants. These challenges needed solutions to ensure study integrity and high quality data. Our solutions can be considered along with the CTTI guidelines for a more comprehensive set of recommendations for managing data for the DCT.

2. The Apple Heart Study (AHS): Background and Goals

The Apple Heart Study (AHS) was a pragmatic, single arm prospective site-less digital trial designed, conducted, analyzed and reported through an academic-industry partnership between Stanford University and Apple Inc. to evaluate whether an algorithm on the Apple Watch could identify when a participant wearing the watch was experiencing atrial fibrillation (Turakhia et al. 2019; Perez et al. 2019). More specifically, the Apple Watch has an optical sensor that detects pulse waveform to passively measure heart rate when placed on the participant's wrist. The heart rate is measured using the sensor to generate signal within tachograms (periods of one-minute length) during opportunistic periods (i.e., every couple of hours if the participant appears to be resting). If a certain pulse irregularity is detected, heart rate sampling will increase to approximately every 15 minutes. If the irregularity is not observed in the next tachogram, the sampling rate returns to usual. However, if irregularities are detected in 5 out of a series of 6 tachograms, the participant is notified that they may be experiencing an irregular pulse that may be suggestive of atrial fibrillation.

Figure 1 displays the multiple data streams that were generated for each participant throughout the study. Once the study app was downloaded, the participant had the option to enroll and be onboarded into the AHS. If participants were notified of an irregularity, they were to connect with a telehealth doctor (conducted through American Well (Boston, MA)) for their first study visit. Notified participants not deemed in need of urgent care by the telehealth provider were then mailed the gold standard single-lead ECG patch (provided by BioTelemetry, recently acquired by Philips (Malvern, PA)). Once received, participants were to wear the ECG patch along with their Apple Watch for one week and return the ECG patch back to BioTelemetry, who generated reports on arrhythmias. The results were then discussed in the second study visit with the telehealth provider. Notified participants were sent a 90 Day Follow-up Survey to complete to provide insight into what actions, if any, were taken in response to being notified and learning of their ECG patch results. All participants – including the majority who were not notified of an irregular pulse – were asked to complete an End of Study Survey. Thus, different types of data were generated per participant throughout the study duration.

In addition to wanting to understand the proportion of participants who were notified, our goal was to evaluate how well the algorithm identified signals consistent with atrial fibrillation among those notified of an irregularity. To accomplish these goals, we measured the proportion of participants who were notified; the proportion of participants notified who had atrial fibrillation detected in subsequent monitoring via a gold standard heart monitor (ECG patch); the proportion of positive (i.e., irregularly classified) tachograms from the Apple Watch where the ECG patch confirmed signal suggesting atrial fibrillation among notified participants wearing both the Apple Watch and ECG patch simultaneously.

3. Design Challenges Encountered and Solutions

3.1 Participant Adherence, Engagement, and Missing Data

The AHS was designed with an app to facilitate screening and consenting and supported enrollment of 419,297 participants in all 50 states with broad diversity and inclusion over

an 8 month period, a number not likely possible without the use of a digital tool. Of these, only 2,161 participants (0.52%) were notified about a potential episode of atrial fibrillation. These 2,161 participants would form the basis for numerous additional analyses to address other primary goals. The challenge encountered was engaging this group participants throughout the study to collect important outcome data. These participants were prompted by the app to initiate contact with a telehealth doctor. However, only 945 (or 44%) adhered (Figure 2). Of those who did engage a telehealth doctor (945), only 658 (or 70%) were eligible to receive an ECG patch, and of these, only 450 (68%) returned them for analysis of our primary and secondary objectives. Thus, the lack of adherence resulted in a missing data issue that potentially compromised the generalizability of our findings.

Many recognize the potential of DCTs to enroll a more representative sample, especially as gaps in digital abilities across sub-groups continue to decrease, enabling the participation of older individuals and those in more remote areas for example (Rosa et al. 2021). However, adherence and engagement are challenges to the DCT that may differ from that of the traditional setting, particularly if the DCT is conducted in a virtual or site-less manner. Much of the adherence and engagement in a traditional trial setting may come naturally as coordinators directly contact participants to schedule in-person clinic visits. The phone calls and emails to schedule visits as well as the in-person visits themselves often create personal relationships that may increase adherence and engagement. In a DCT where onboarding and measurements may be done virtually one has to develop creative ways to engage and retain participants. Our solution in the AHS was to reach out to those notified via email and phone calls to encourage participants to initiate contact with the telehealth provider and to return the ECG patch. For all participants we additionally relied on reminders through the app to complete the surveys. While this represents a hybrid approach of digital and traditional methods to interface with participants, alternative solutions may be fully digital. Additionally, solutions can involve modest incentives upon data being returned. Other creative approaches may include engagement through the app in the forms of games or informing participants of their results (Steinhubl, Muse and Topol 2015; Geuens et al. 2016; Baumel and Kane 2018; Ludden et al. 2015). In the Metastatic Breast Cancer Project, social media and bidirectional communication to share information was leveraged to achieve 95% adherence in completing the required survey about their cancer, treatment, and demographics (Wagle et al. 2016).

In AHS we were concerned about the lack of participant adherence with study procedures that yield missing data and potentially affect interpretation of findings. Missing data can occur for many reasons, some of which may be random (e.g., lost device) or some may be related to underlying health (e.g., concern about cardiovascular health in the wake of an alert). Despite our attempts to keep participants engaged, only 450 (20.8%) of 2,161 notified participants adhered and provided data directly relevant for our primary and secondary goals. Additionally, 1,376 (63.7%) of those notified completed the 90 Day Survey. Interestingly, although only 929 (43%) of those notified completed the End of Study Survey, we saw higher adherence in completing the End of Study Survey among those not notified with 293,015 (70.2%) of 419,297 responding. Our solution to ensuring generalizability was to first consider – during the study design – potential reasons why those notified may not initiate a visit with the telehealth provider or return the ECG patch so that

these reasons could be included in our analysis plans. Second, to provide context for our analysis, we comprehensively described differences in key variables between those notified and analyzed and those notified and excluded for missing data. In the AHS, these two groups of participants were comparable with respect to sociodemographic variables and other key variables including atrial fibrillation burden. This suggested minimal concern around issues of generalizability, assuming we had captured all relevant variables. Third, our analysis plan incorporated sensitivity analyses that considered such reasons for missing data. For example, we had planned to evaluate whether the agreement between signals from the Watch and ECG patch would vary by atrial fibrillation burden or length of ECG patch wear time. Our sensitivity analyses demonstrated our primary results were robust to such variables, bolstering confidence in our findings. Further, although the uncertainty in our estimates was higher than expected due to the lower sample size on which the analysis was based, it was appropriately reflected in wider confidence intervals than originally anticipated. In addition to considering more creative ways to engage participants in a pragmatic site-less study and anticipating reasons for missing data, we encourage explicitly collecting data on reasons for not adhering to study procedures (e.g., prompting participants to explain why they did not initiate contact with a telehealth provider). Such variables may then be incorporated into sensitivity analyses as well as in missing data strategies such as multiple imputation.

While analytic approaches are helpful to addressing missing data issues due to poor adherence or engagement, minimizing missing data through engagement and retention strategies remains a high priority. Treweek and Briel (2020) discuss the need for more formal evaluation of digital tools for retention (Treweek and Briel 2020) in the digital trial. In their comprehensive study of digital tools for recruitment and retention of participants in clinical trials, Blatch-Jones et al. (2020) mentioned that texting and email reminders were the most common digital approaches utilized in retention strategies. The authors urged further research to increase our understanding of effective retention approaches using digital means (Blatch-Jones et al. 2020). Given the challenges, investigators should also pay close attention to which types of trials might be well suited for evaluation of a digital intervention. This may include those trials that require a relatively short follow-up period, or a fast-acting intervention, or a target population comprised of particularly motivated participants (e.g., those with rare or severe disease). Such trials may rely less on sophisticated retention strategies for their success. That said, bridging the digital divide across sub-populations will continue to be crucial for the successful recruitment and retention of participants in the DCT that has strong implications for providing generalizable and useful findings that translate to the clinic.

3.2 Duplicated Participant Identification

An essential task in every study is the ability to count the number of unique participants enrolled. Unique identifiers for each individual in AHS were critical to have the ability to create longitudinal trajectories for each study participant to address primary study goals and to track safety events. A system was established for the study to create two IDs for each enrollment: Device ID (DID) and Participant ID (PID). These two IDs would be helpful in several anticipated scenarios. For example, suppose the app was deleted from the participant's device after a participant had enrolled and the participant subsequently

re-enrolled after re-installing the app. In this case, a new participant ID (PID) would be generated that could be linked to the same DID. However, there were unanticipated pathways that led to the generation of multiple PIDs for an individual. For example, if a participant were to buy a new watch, upload the app again and re-enroll, a new PID and a new DID would be generated. In this case, the system would fail to recognize that the individual was already enrolled in the study. To further complicate matters, participants may share their watch with others, further complicating identification exercises. Deduplication strategies were therefore needed and created during the study to identify unique individuals under a variety of scenarios enabling both an accurate count of the number of participants enrolled and a longitudinal trajectory for each participant.

Our de-duplication algorithm (Figure 3) was developed and evaluated in real time when multiple PIDs for the same individual were discovered early on during standard data quality monitoring procedures. The strategy involved matching two sets of data from two records on multiple pieces of information (two string vectors) through a dissimilarity score based on the Levenshtein string distance, which can be interpreted as the minimum number of edits required to have an exact match between two string vectors (Levenshtein 1966). Data considered included: last name, email address, first name, consent date, state of residence, data of birth, and phone number.

Comparing pairwise vectors among almost half a million records can be computationally challenging and in our case led to close to 96 billion comparisons. To increase computational efficiency, we restricted application of the duplication identification algorithm to four subsets of records as depicted on the left hand side of Figure 3: (1) Different PIDs that correspond to the same DID, (2) Different PIDs associated with the same last name; (3) Different PIDs associated with the same first name, and (4) Different PIDs associated with the same date of birth. Within each subset, we applied the matching algorithm now only resulting in approximately 3 billion comparisons. We flagged multiple records as belonging to the same PID if their string distance was less than a threshold determined through cross validation methods. Our findings demonstrated the positive predictive value of the deduplication algorithm to be 96%.

The first lesson we learned was that it may be helpful to proactively include a verification feature on the app during the onboarding process to verify if whether current enrollee is unique. For example, an algorithm similar to the one we developed could be applied to enrolling participants in an ongoing manner where their information is compared to those among the smaller subset of already enrolled individuals. Individuals flagged by the algorithm could be asked additional questions to verify whether this is their first time enrolling. While we had anticipated that incorporating data from digital tools would necessitate a de-duplication algorithm to be in place, the second lesson we learned from the AHS was that an initial de-duplication algorithm will likely need to be refined and tailored to the specific study and further evaluated to characterize its accuracy. This can be facilitated through a pilot study on a subset of individuals prior to study launch, particularly when consent and onboarding are done electronically. The third lesson is that both PID and DID need to be included on each piece of data whenever possible to facilitate the entire process. Finally, in studies where enrollment is done via an app and where duplication may

be an issue, it may make sense to consider the sample size as a quantity estimated with uncertainty. This is a novel idea for the clinical trial setting, but may be appropriate for large low risk trials such as the AHS where broad outreach is being conducted via an app. A method for estimating the variance of the sample size should account for accuracy of the algorithm and is considered future work.

3.3 The Role of Time Stamp Data

3.3.1 Time Stamp Data Across Multiple Diverse Data Streams – Assessing Concordance—The coordination of multiple streams of different types of data presents another challenge common in trials with pragmatic elements. In the AHS, data were generated from the following sources (Figure 1): (1) the study app, (2) AmericanWell virtual study visits, (3) the ECG patch (4) a summary report provided by BioTelemetry derived from the gold standard ECG patch data, (5) cardiologist-adjudication of the gold standard ECG patch data, and (6) participant-reported questionnaires. The unique PID was critical for linking all these pieces of data. However, the time of data collection for each data element was also crucial for achieving study goals. For example, one goal was to evaluate whether the rhythms observed in the ECG patch were concordant with those observed using the Apple Watch. To assess concordance, time stamp data from both ECG patch and Apple watch were aligned to match the same time such that the data recorded by the ECG patch intervals matched that recorded by the Apple Watch tachograms. Misalignment of the time stamp data introduces bias, and for example, may increase the probability that we conclude discordance even if there were strong agreement. Our solution in the AHS was to record the date and time that the ECG patch was shipped, received, worn, and returned in order to validate the timing of key measures. For example, noise is introduced into the internal time variable (known as “drift”) that can occur for multiple reasons including if the ECG patch is recording data while the battery is near depletion. For this reason, data were only considered to have high integrity if the ECG patch recording was captured within a certain number of days from shipment. Additionally, BioTelemetry synched the patches right before sending them out to participants to ensure the time and date from the ECG patch was consistent with Coordinated Universal Time (UTC). In addition, to better understand how much variation there was in the time stamp variable from the ECG patch, BioTelemetry calculated the drift using a set of devices corresponding to 26 days of data and found the drift ranged from 7 to 39 seconds with a mean of 20.3 seconds. Drift was considered when we aligned data by adding a window of 60 seconds on either side of a sampled ECG patch interval prior to assessing concordance. More specifically, signal consistent with atrial fibrillation that lasted at least 30 seconds from the Apple Watch (provided in tachograms or periods of one-minute length) were compared to signals from the ECG patch. When assessing concordance from the two devices, this amounted to evaluating ECG patch data that corresponded to the exact time of the Apple Watch tachogram with a period 60-seconds appended on either side of the ECG patch. Thus, if signal suggestive of atrial fibrillation lasting more than 30 seconds occurred on this 3-minute length period from the ECG patch, concordance was noted.

3.3.2 Time Stamp Data Across Multiple Diverse Data Streams – Study Monitoring and Data Integrity—In addition to addressing study aims, time stamp data are critical for both monitoring study conduct and understanding participant engagement

and follow-up length. From a data integrity point of view, many issues can be uncovered by examining the timing of the data being collected. To illustrate the role of time stamp data in both study conduct and data integrity we present real scenarios encountered in the AHS when cleaning the data as part of standard data quality monitoring procedures (Figure 4). Each row in the figure represents a timeline of events throughout the study for a given participant. Various events are marked with different colored symbols. By linking the data over time and indicating when events occur, one can determine important operation metrics for study conduct including how long participants are followed, treated, engaged, and monitored. Data examined from the fourth participant from the top (circled) alerted the data management team to a data integrity issue. In this participant, the order of events indicates that the participant was notified after their study visit and after having received an ECG patch, but this is not possible, leading the study team to troubleshoot potential data integrity issues. Errors such as this may be attributed to a variety of sources including participant and/or device behavior. For example, participants may alter the clock in the device themselves or the clock on the device may be faulty for other reasons including the battery life. This example illustrates the importance of assessing the different types of data jointly and in their longitudinal sequence. Evaluation of time stamp data should not simply be limited to the examination of the range and distribution of single variables, but rather should be done in conjunction with the other data elements once the time stamp data have been incorporated in preparation of analysis. The latter should feature prominently in the data quality monitoring plan.

As illustrated above, time stamp data corresponding to each piece of data are crucial for reaching study goals, monitoring the quality of the data and assessing study conduct such as how long a participant is being followed. In the AHS we included two types of time stamp data: device-generated and server-generated. In addition to having internal clocks on each device to retrieve the time stamp generated by the device, time stamps of when data are recorded (i.e., when data are uploaded) should also be generated by a central server that houses the device data, so that these two types of time stamps are associated with each piece of data. This can be helpful if for example the device's clock is faulty. It may also be helpful in situations when data cannot be uploaded to a central server (e.g., if participants are not in an area connected to a network) in which case one can rely on the device clock for recording when data were measured. Importantly, we recommend relying on standardized timing formats like Coordinated Universal Time (UTC). This will enable integration across different types of data streams for participants worldwide. A Quick Response (QR) scan code may be helpful to estimate the latency between device time stamps. For example, in the AHS, a participant could scan a code or set of codes to record when the ECG patch was received, when the participant started wearing it and when it was returned to BioTelemetry. Other studies have proposed more robust approaches for medical devices that are capable of transmitting data to a server, ensuring appropriate synchronization among smartphones and medical device clocks as well as minimizing data tampering (Siddiqi, Sivaraman, and Jha 2016). These more complex protocols would not only eliminate latency between devices but also ensure data integrity.

4. Study Design and Data Management Planning Recommendations for DCTs

The CTTI guidelines for data management provide an excellent framework for establishing rigor in the planning of the DCT. We borrow heavily from these guidelines, and based on our experience with the AHS, we generalize the principles learned and further build upon this set of recommendations (Table 1). Novel recommendations that we contributed include having a statistical analysis plan where the processing of the data from the digital tool are pre-specified, and where the statistical analysis plan strongly informs the data quality monitoring plan. In particular, the data quality monitoring plan should ensure that -- in addition to univariable distributions -- joint relationships featured in the analysis plan are evaluated on an ongoing basis. While CTTI guidelines wisely suggest that only minimal data necessary for study endpoints should be collected from digital tools, we explicitly mention that this set of data should always include time stamp data. Importantly, it may not be obviously necessary for creation of the endpoint itself, but as we learned from the AHS, these data were critical from a monitoring and data integrity perspective that impacted generation and validity of the endpoint. We additionally mention the inclusion of sensitivity analyses in the analysis plan, as it prompts the study team to understand the assumptions on which findings are based and anticipate what can go wrong with data generation such that key assumptions are violated. Sensitivity analyses consider alternative ways to look at the data when assumptions are challenged and provide context for the reader when interpreting findings. This was particularly true in the AHS. For example, we had to consider the possibility that participants may use the ECG patch outside of requested windows, where drift could be an issue. Our primary analysis involved data from participants who provided ECG patches that were used within desired windows where the time stamp variable was deemed less noisy. This challenges the assumption, however, about generalizability of our findings, particularly if those included in the analysis differ from those excluded. A sensitivity analysis included all individuals who returned their ECG patch regardless of window. As no meaningful differences were observed in the two analyses (primary and sensitivity), this bolstered confidence in our primary findings. If differences were observed, however, this could have suggested the positive predictive value of the Apple Watch signals were not generalizable. On the other hand, if a lower positive predictive value were observed, this may also have been due to noisier time stamp data and less ability to align the data well. While sensitivity analyses may be imperfect, they help provide context for principal interpretations and are particularly critical for the DCT. We believe that use of many digital tools can create issues with ability to uniquely map data to individuals and that all DCTs should therefore consider whether a deduplication algorithm needs to be included in their data quality monitoring plan. We additionally feature the importance of a pilot study prior to launch that includes execution of both the data quality monitoring plan and the statistical analysis plan. As pilot studies have been recognized by others -- including the CTTI -- as crucial for identifying an appropriate digital tool to be selected and tested, we believe such studies are vital to determining the feasibility of obtaining meaningfully analyzable data from the proposed DCT. We also emphasized the importance of including data scientists of diverse type on the leadership of the study team, particularly to help shape and oversee the data management portion of the study design and conduct. Indeed

in the AHS, our leadership team consisted of biostatisticians, information technologists, and software engineers. Finally, while focus may remain on safety of the trial, data integrity, data sharing and trial conduct may face additional challenges in the DCT setting, prompting adjustments to the Data & Safety Monitoring Board composition. We recommend a composition that includes an expert in software engineering who can speak to the sources of variation present in data generated from the device/platform, an expert in informatics who can speak to issues around secure data flow and data integration, an expert in the clinical domain area, as well as an expert in biostatistics familiar with how to analyze such endpoints. It may be additionally advantageous to include a participant stakeholder on the board. In addition to contributions to other design considerations, participant stakeholders can advise on issues around interpretation of data shared during and after the trial, which DCTs may make possible. We believe our novel additions to the CTTI's guidelines on data management provide a comprehensive set of recommendations that other trialists can adopt.

Discussion

The AHS was a pragmatic trial that incorporated the use of digital tools to assess the ability of an app on the Apple Watch to identify signals consistent with atrial fibrillation in the general population of individuals with no known history of atrial fibrillation. The study goals were to calculate proportions to characterize the notification rate, the atrial fibrillation yield among those notified, and the concordance of irregularity detection between the Apple Watch and the ECG patch. Once participants were consented and onboarded, the data management challenges that came with the passive data collection, the high volume of data for each individual from near-continuous monitoring, and the linkage of multiple data streams, were profound. To ensure high quality and compliance we created robust systems during study design as well as refinements during study conduct that support the interpretation of our findings presented previously (Perez et al. 2019). For example, although the targeted statistical precision for estimating the yield of atrial fibrillation on patch monitoring was not met due to low participant adherence, the confidence intervals reflected the uncertainty of our estimands of interest. Generalizability of our findings, however, were a concern that we addressed through planned sensitivity analyses. For example, those who were included and excluded from analyses due to missing data were comparable on key sociodemographic variables. In a retrospective examination, we summarized key lessons that we learned around our challenges, however, that can guide future DCTs. These challenges were related to (1) participant adherence to achieve study goals, (2) how to identify unique individuals enrolled in the trial, and (3) how to link critical pieces of data using timestamps to establish longitudinal trajectories for each participant. Based on our experience with these challenges in the AHS, we further built upon the CTTI's guidelines creating a comprehensive set of recommendations for data management to be considered for future DCTs. Our recommendations arose from data collected from specific devices in the AHS, and thus, as the field of digital health is rapidly expanding, these guidelines may need to further evolve. However, the fundamental principles upon which the recommendations were developed remain.

There were multiple reasons related to the virtual nature of the study for the adherence challenges, particularly among those notified. One was that there was a low effort needed

to enroll in the study, but a higher demand on notified participants. Many participants, therefore, were likely not as invested as they might be in a more traditional trial that may create personal relationships between investigators and participants. In addition, although participants were asked if they had atrial fibrillation (as this was an exclusion criteria), some participants may have already known that they had atrial fibrillation and enrolled anyway. Thus, if these individuals were among those notified, they may have decided not to pursue additional monitoring upon notification. Finally, we excluded those who wore or returned the ECG patch outside a window of 45 days for data integrity reasons (i.e., to avoid drift in the time stamp variable), and thus, engagement plans dealing with drift issues may include incentivizing early participation and return for this reason.

The biggest lesson learned was the need to do a pilot study on the data flow, the data integration and the data integrity. Piloting on a number of participants may have provided us with insight into issues with integrating data across the diverse streams; information on participant duplication; the prevalence of missing data; the need for collecting data on reasons for missingness for certain measures; the need for timestamps; the noise involved in the time stamp variables; and perhaps even the need for a stronger engagement plan. We therefore recommend a pilot study prior to launch where the data flow is tested and the data are analyzed on a sufficient number of participants that may depend on target sample and ability to recruit this number to provide insight into potential areas of concern. Analysis procedures from the pilot should not be limited to ensuring that data can be viewed and evaluated univariably, but rather should include joint investigation of key variables and their interrelationships, as pre-specified in the statistical analysis plan. As part of this data quality monitoring procedure, it is particularly important to establish longitudinal trajectories of events for each participant to enable derivation of key variables that are functions of time. To that end, every piece of data should have a set of time stamps even when they may not be considered necessary. Such exercises will facilitate the discovery of issues with ID duplication, time stamps, and missing data. Key to the pilot study is establishment of a statistical analysis and data quality monitoring plan that can serve as tools for assessment of the pilot data.

Although there were challenges, many of the data-related aspects in the study design went as planned. For example, maintaining the privacy and confidentiality of data from participants was of utmost importance. To that end, Stanford University maintained independence from Apple Inc. and served as the Data Coordinating Center, establishing a firewall between the two entities. Procedures were put into place to ensure that 1) data were encrypted, 2) access to data by Apple Inc. was restricted to limited and minimal data necessary for specific analyses that contained no identifying information, 3) primary and secondary analyses were performed by Stanford University independently. As a result of the systems we put in place, there have been no incidents of privacy or confidentiality breaches.

Incorporating digital devices into clinical trials provides great potential for the research setting. Massive amounts of rich times-series data can be collected pragmatically and can further be used as part of an intervention. Indeed, with the recent COVID-19 pandemic, we see that digital devices may not simply be pragmatic but necessary (Rosa et al. 2021). With their incorporation, however, come challenges that require thoughtful solutions around

the handling of data. Inclusion of data scientists on the leadership of the study team is an important step toward overcoming these challenges. Even in cases where the statistics appear seemingly simple it is important to understand the critical role of collaborative data science in the design and conduct of successful digital health studies.

Acknowledgement:

This work was supported through a partnership sponsored by Apple, Inc; Apple Heart Study [ClinicalTrials.gov](https://clinicaltrials.gov) number, [NCT03335800](https://clinicaltrials.gov/ct2/show/study/NCT03335800).

References:

1. Baumel A and Kane JM. 2018. Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. *Journal of Medical Internet Research*, 20(12), p.e11491. doi: 10.2196/11491. [PubMed: 30552077]
2. Blatch-Jones A, Nuttall J, Bull A, Worswick L, Mullee M, Peveler R, Falk S, Tape N, Hinks J, Lane AJ and Wyatt JC. 2020. Using digital tools in the recruitment and retention in randomised controlled trials: Survey of UK Clinical Trial Units and a qualitative study. *Trials*, 21(1), pp.1–11. doi: 10.1186/s13063-020-04234-0. [PubMed: 31898511]
3. Clinical Trials Transformation Initiative. CTTI Recommendations: Advancing the Use of Mobile Technologies for Data Capture & Improved Clinical Trials. 2021. https://ctti-clinicaltrials.org/wp-content/uploads/2021/06/CTTI_Digital_Health_Technologies_Recs.pdf. Assessed on Dec 10, 2021.
4. Coran P, Goldsack JC, Grandinetti CA, Bakker JP, Bolognese M, Dorsey ER, Vasisht K, Amdur A, Dell C, Helfgott J and Kirchoff M. 2019. Advancing the use of mobile technologies in clinical trials: recommendations from the clinical trials transformation initiative. *Digital Biomarkers*, 3(3), pp.145–154. doi: 10.1159/000503957. [PubMed: 32095773]
5. Cornet VP and Holden RJ. 2018. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, 77, pp.120–132. doi: 10.1016/j.jbi.2017.12.008. Epub 2017 Dec 14. [PubMed: 29248628]
6. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L and Abeele VV. 2016. A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. *JMIR mHealth and uHealth*, 4(4), p.e6286. doi: 10.2196/mhealth.6286.
7. Herrington WG, Goldsack JC and Landray MJ. 2018. Increasing the use of mobile technology-derived endpoints in clinical trials. *Clinical Trials*, 15(3), pp.313–315. doi: 10.1177/1740774518755393. Epub 2018 Feb 5. [PubMed: 29400066]
8. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, Turakhia M, Pletcher MJ, Preston KL, Krumholz HM and Marlin BM. 2020. Digitizing clinical trials. *NPJ Digital Medicine*, 3(1), pp.1–7. [PubMed: 31934645]
9. Levenshtein VI, 1966, February. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics Doklady*, Vol. 10, No. 8, pp. 707–710.
10. Lubitz SA, Faranesh AZ, Atlas SJ, McManus DD, Singer DE, Pagoto S, Pantelopoulos A and Foulkes AS. 2021. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch: The Fitbit heart study. *American Heart Journal*, 238, pp.16–26. doi: 10.1016/j.ahj.2021.04.003. Epub 2021 Apr 15. [PubMed: 33865810]
11. Ludden GD, Van Rompay TJ, Kelders SM and van Gemert-Pijnen JE. 2015. How to increase reach and adherence of web-based interventions: a design research viewpoint. *Journal of Medical Internet Research*, 17(7), p.e4201. doi: 10.2196/jmir.4201.
12. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, Rivera DE, West R and Wyatt JC. 2016. Evaluating digital health interventions: key questions and approaches. *American Journal of Preventive Medicine*. 51(5):843–851. doi: 10.1016/j.amepre.2016.06.008. [PubMed: 27745684]
13. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Balasubramanian V, Russo AM, Rajmane A, Cheung L et al. 2019. Large-scale assessment of a smartwatch to identify

- atrial fibrillation. *New England Journal of Medicine*, 381(20), pp.1909–1917. doi: 10.1056/NEJMoa1901183. [PubMed: 31722151]
14. Pham Q, Wiljer D and Cafazzo JA. 2016. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. *JMIR mHealth and uHealth*, 4(3), p.e5720. doi: 10.2196/mhealth.5720.
 15. Rosa C, Campbell AN, Miele GM, Brunner M and Winstanley EL. 2015. Using e-technologies in clinical trials. *Contemporary Clinical Trials*, 45, pp.41–54. doi: 10.1016/j.cct.2015.07.007. Epub 2015 Jul 12. [PubMed: 26176884]
 16. Rosa C, Marsch LA, Winstanley EL, Brunner M and Campbell AN. 2021. Using digital technologies in clinical trials: Current and future applications. *Contemporary Clinical Trials*, 100, p.106219. doi: 10.1016/j.cct.2020.106219. Epub 2020 Nov 17. [PubMed: 33212293]
 17. Siddiqi M, Sivaraman V and Jha S. 2016, November. Timestamp integrity in wearable healthcare devices. In 2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) (pp. 1–6). IEEE.
 18. Steinhubl SR, Muse ED and Topol EJ. 2015. The emerging field of mobile health. *Science Translational Medicine*, 7(283), pp.283rv3–283rv3. doi: 10.1126/scitranslmed.aaa3487.
 19. Tomlinson M, Rotheram-Borus MJ, Swartz L and Tsai AC. 2013. Scaling up mHealth: where is the evidence?. *PLoS Medicine*, 10(2), p.e1001382. doi: 10.1371/journal.pmed.1001382. Epub 2013 Feb 12. [PubMed: 23424286]
 20. Treweek S and Briel M. 2020. Digital tools for trial recruitment and retention—plenty of tools but rigorous evaluation is in short supply. *Trials*, 21, 476. doi: 10.1186/s13063-020-04361-8. [PubMed: 32498704]
 21. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, Desai S, Nag D, Patel M, Kowey P et al. 2019. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *American Heart Journal*, 207, pp.66–75. doi: 10.1016/j.ahj.2018.09.002. Epub 2018 Sep 8. [PubMed: 30392584]
 22. Wagle N, Painter C, Krevalin M, Oh C, Anderka K, Larkin K, Lennon N, Dillon D, Frank E, Winer EP et al. 2016. The Metastatic Breast Cancer Project: A national direct-to-patient initiative to accelerate genomics research. *Journal of Clinical Oncology*, 34(18). doi:10.1200/JCO.2016.34.18_suppl.LBA1519

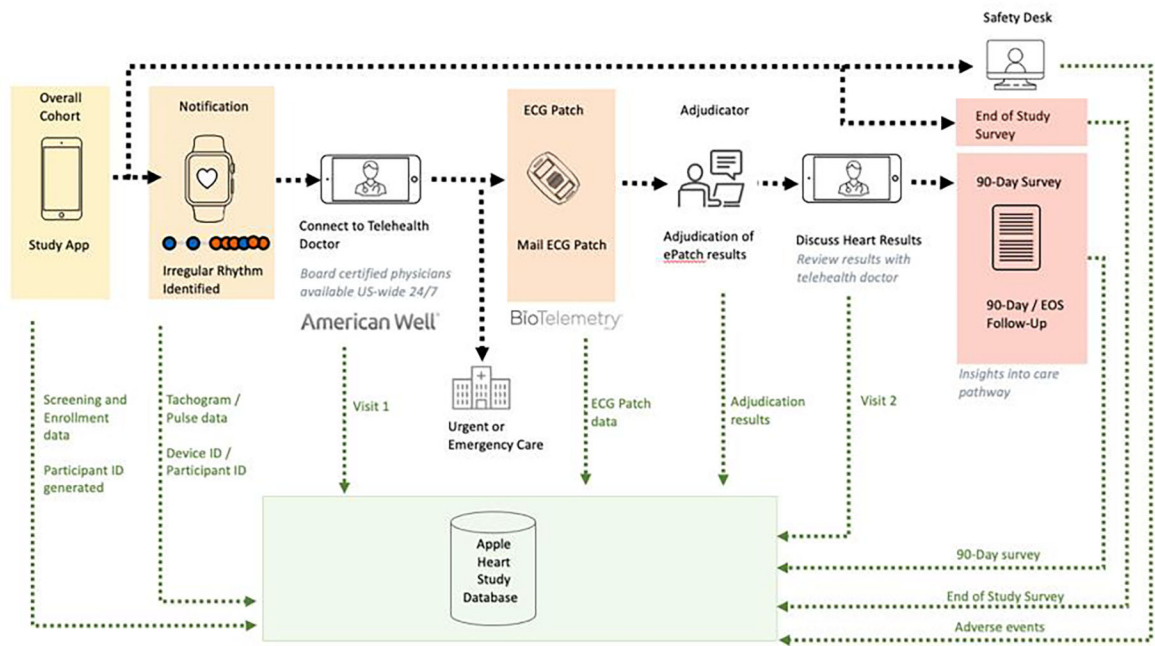


Figure 1. Data Flow of Multiple Data Streams for Participants in the Apple Heart Study

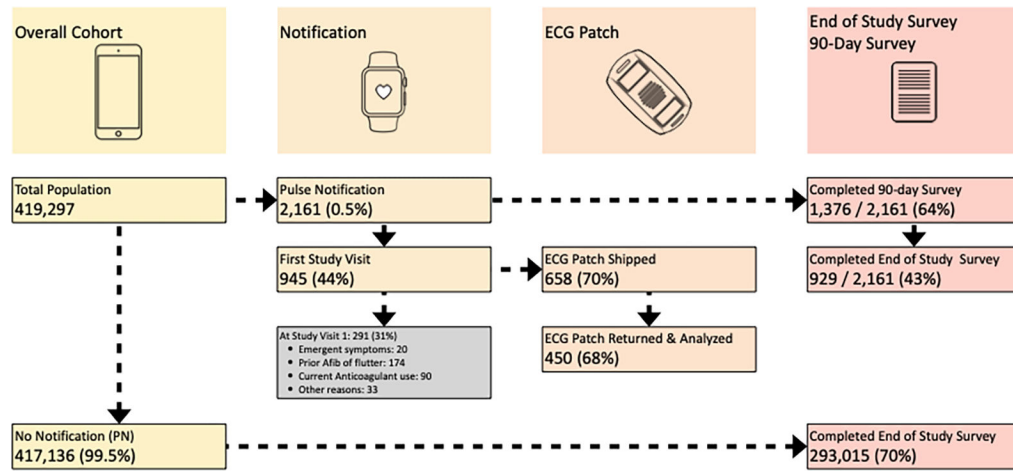


Figure 2.
Participant Engagement in the Apple Heart Study

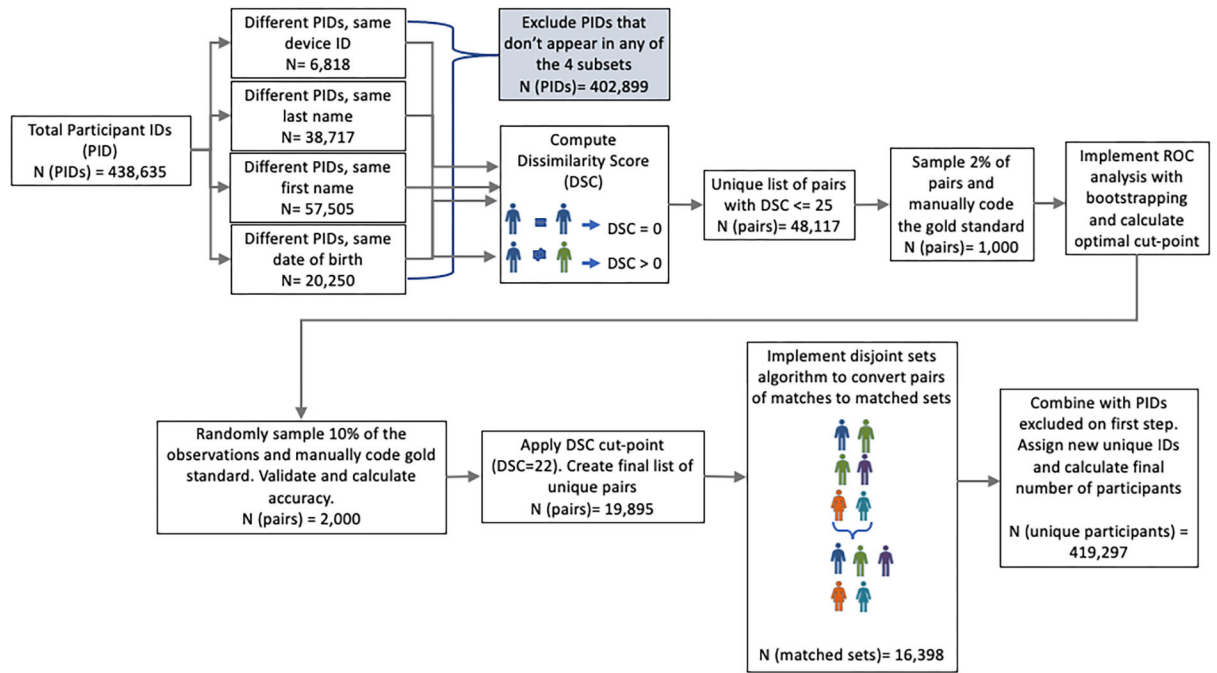


Figure 3.
Schematic of Deduplication Algorithm

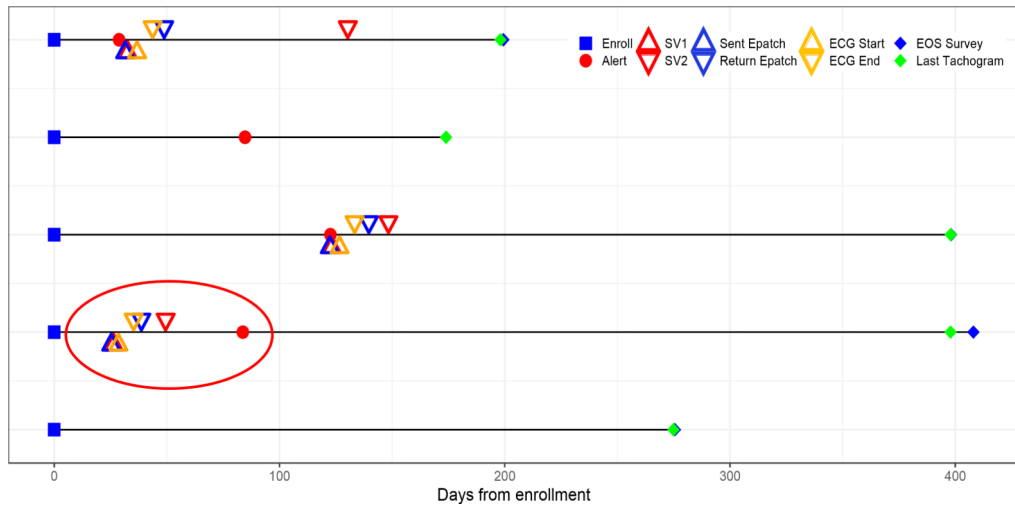


Figure 4.
Hypothetical Illustration of the Use of Timestamping in Data Monitoring

Table 1.

Comprehensive recommendations for data management of the digital clinical trial building on the CTTI guidelines*.

Category	CTTI Recommendation	Enhanced Recommendation
1.General	N/A	Consider whether data storage needs to be FDA 21 CFR Part 11 Compliant as this will inform the structure of the electronic database platform
	Proactively address and map data flow, data storage, and associated procedures	Proactively address and map data flow, data storage, and associated procedures
	Plan appropriately for the statistical analysis of data captured using digital technologies (Moved from Section 4)	Plan appropriately for the processing and analysis of all data in the digital clinical trial including but not limited to data captured using digital technologies where all primary, secondary, and sensitivity analyses are prespecified including how to handle missing data
	N/A	Plan to obtain time stamps using Coordinated Universal Time on each piece of data deemed necessary to address study goals while understanding sources of variation in the time stamp variable
	Collect the minimum data set necessary to address the study endpoints	Collect the minimum data set necessary to address the study goals, recognizing that time stamp data should always be considered necessary even if not part of the endpoint definition
	Identify acceptable ranges and mitigate variability in endpoint values collected via mobile technologies	Identify acceptable ranges and mitigate variability in endpoint values collected via mobile technologies including the timing of when the data are obtained, as battery life of the device may impact the variation in data obtained for example
2.Access to Data	Optimize data accessibility while preventing data access from unauthorized users	Optimize data accessibility while preventing data access from unauthorized users
	Ensure that access to data meets your needs prior to contracting an electronic service vendor	Ensure that access to data meets your needs prior to contracting an electronic service vendor
	Address data attribution proactively with patient input	Address data attribution proactively with patient input
	Ensure that site investigators have access to data generated by their participants	Ensure that site investigators have access to data generated by their participants
	Return value to participants throughout the trial, including return of outcomes data collected by digital technologies	Return value to participants throughout the trial, including return of outcomes data collected by digital technologies, when appropriate interpretation is possible and can be provided
	Let data sharing decisions be driven by safety and trial integrity	Let data sharing decisions be driven by safety and trial integrity
3.Security & Confidentiality	Apply an end-to-end, risk-based approach to data security	Apply an end-to-end, risk-based approach to data security
	Ensure the authenticity, integrity, and confidentiality of data over its entire lifecycle	Ensure the authenticity, integrity, and confidentiality of data over its entire lifecycle
	Be prepared to collaboratively identify and evaluate privacy risks	Be prepared to collaboratively identify and evaluate privacy risks
	Ensure that participants understand the privacy and confidentiality implications of using digital technologies	Ensure that participants understand the privacy and confidentiality implications of the particular digital tool(s) involved in the study of interest
4.Monitoring for Safety & Quality	N/A	Consider a novel composition of Data & Safety Monitoring Board members that includes members of the community (e.g., participant stakeholders critical for discussions on sharing and interpretation of data) and multiple quantitative experts who can speak to the data integrity of the trial including software engineers familiar with the device/platform and sources of variation in data generated, information technology experts familiar with data flow, clinicians with expertise in the relevant disease area, and biostatisticians familiar with trial monitoring and analysis of the key endpoints

Category	CTTI Recommendation	Enhanced Recommendation
	N/A	Include data scientists with diverse background in the study leadership who will shape and oversee the monitoring plans including biostatisticians analyzing the data, information technology experts facilitating data flow and secure data capture, and software engineers well versed in data generation from the digital tools
	Include appropriate strategies for monitoring and optimizing data quality	Pre-specify appropriate strategies for early and frequent monitoring of data quality that involve consideration of relevant joint relationships among variables as described in Statistical Analysis Plan.
	N/A	Consider a deduplication process to ensure mapping and linkage of data to unique individuals (e.g., if relying on an app or an electronic health care system in enrollment plan) employed in real-time (e.g., as participants are enrolled) that includes a plan to refine and evaluate updated algorithm's ability to identify duplications
	Set clear expectations with participants about safety monitoring during the trial	Set clear expectations with participants about safety monitoring during the trial and about sharing clinical findings during and after the trial
	Monitor data quality centrally through automated processes	Monitor data quality centrally through automated processes
	Minimize missing data	Minimize missing data
	N/A	Design a pilot study prior to launch that collects data from participants within the target population if possible to test the data flow beyond end-to-end testing, and that executes the prespecified analysis plan and data quality monitoring plan so that data are processed, endpoints derived, missing data are observed and joint relationships evaluated in order to identify potential issues with the database and data integrity with an emphasis on ability to identify non-unique participants

* Items were borrowed and adapted from CTTI guidelines (https://ctti-clinicaltrials.org/wp-content/uploads/2021/07/CTTI_Managing_Data_Recommendations.pdf)