frontiers
in Computer Science

Check for updates

Measuring the Effects of Caffeine and L-Theanine on Cognitive Performance: A Protocol for Self-Directed, Mobile N-of-1 Studies

Eddye Golden*, Matthew Johnson, Michael Jones, Ryan Viglizzo, Jason Bobe* and Noah Zimmerman

Institute for Next Generation Healthcare, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Background: The growing consumer digital tools market has made using individual health data to inform lifestyle changes more accessible than ever. The n-of-1 trial–a single participant, multiple crossover, comparative effectiveness trial–offers methodological tools that link interventions directly with personalized outcomes to determine the best treatment for an individual. We have developed a complete digital platform to support self-directed n-of-1 trials, comprised of virtual study on-boarding, visual informed consent, device integrations, in-app assessments, and automated data analysis.

Objective: To evaluate the n-of-1 platform, a pilot study was launched to investigate the effects of commonly consumed substances on cognition. The purpose of the study is to allow an individual to measure the effect of 2 treatments (caffeine alone vs. caffeine + L-theanine) on 3 measures of cognitive performance: creative thinking, processing speed, and visual attention. Upon completion of the study, individuals receive personalized results that compare the impact of the two treatments on each of the cognitive performance measures.

Methods: After the onboarding process, participants are randomized to a study length (5, 15, or 27 days), starting treatment (caffeine or caffeine + L-theanine), and app notification frequency (light, moderate). Each trial begins with a baseline period, during which participants abstain from either treatment, followed by 2 randomized counterbalanced treatment sequences (either ABBA or BAAB). Throughout the trial, daily tests assess participant cognitive performance. These tests are digital versions of the Remote Associates Test, Stroop Test, and Trail Making Test, and are implemented directly in the n-of-1 mobile application ("N1"). Assessments are completed at a fixed time, defined by the individual during study setup. Treatments are taken daily within a fixed time window prior to the user-defined assessment time. Cognitive assessment results are analyzed using a linear model with factors for treatment and block, and each treatment is compared to baseline.

Results: We launched our N1 app on the Apple App Store in mid-October 2019 and recruited over 40 participants within the first month.

Conclusion: This platform provides individuals the opportunity to investigate their response to treatments through n-of-1 methods, empowering them to make data-driven, personalized lifestyle choices.

Trial Registration: www.ClinicalTrials.gov, identifier: NCT04056650.

INTRODUCTION

Technological and medical innovation have brought about a rapid digitalization of research. Digital research has conceptually grown to incorporate a broad set of medical and scientific themes–genomics, AI, wearable technology, mobile apps, longitudinal data capture, among others. With this digitalization of medical research, we see increased interest in digital biomarkers, as well as the growing practice and value of longitudinal data capture, electronic patient reported outcomes, and effective validation strategies for digital health tools (Mathews et al., 2019; Wang et al., 2019).

The growing availability of digital research tools creates a unique opportunity to further develop health research. Individuals are now presented with increased opportunities to participate remotely in research, including going through informed consent processes and interacting with study teams via digital platforms. Significant components of research, such as informed consent and study team communications, are being digitized to enhance user experience, comprehension, and accessibility[1] (U.S. Department of Health Human Services, 2016). Using these methods, some digital studies are able to enroll large numbers of participants in a short time. We see this at play in the app-based mPower study, which enrolled more than 10,000 participants in the first year (Bot et al., 2016), and the Apple Heart Study, in which Stanford researchers enrolled over 400,000 individuals in 1 year for a remote, single-arm study using the Apple Watch to identify cardiac arrhythmias (Turakhia et al., 2019). Apple also recently launched Apple Research, an app designed to better streamline the enrollment and management of mobile health studies[2]. These trends open up the digital research space to the concept of the n-of-1 trial.

An n-of-1 trial monitors the effects of different treatments or interventions on a single participant, where n = 1. It is typically structured as a single-patient, multiple-crossover comparative effectiveness trial. Each participant tests 1 or more interventions multiple times over the course of the trial, and subsequently compares the outcomes of those interventions (Duan et al., 2014; Shamseer et al., 2015). N-of-1 trials have been used in

healthcare when a clinician wants to test different medications, dosages, or treatments on a patient to determine individual response, and thus craft a personalized and effective route to health. The n-of-1 trial is particularly useful where limited evidence exists for a particular treatment or outcome, or where there is variability across individuals in treatment response (Duan et al., 2014). The success of this clinical implementation is dependent on the study design (what is being compared), and the willingness and collaboration of the patient, and the capacity of the clinician to design and implement an n-of-1 trial. With more than 2000 of these trials published to date, examples of previous implementations include an app-based study of the treatment of chronic musculoskeletal pain (Kravitz et al., 2018; Odineal et al., 2019), as well as stimulant effectiveness among people with attention deficit/hyperactivity disorder (Nikles et al., 2006).

We have created a mobile app and study platform that together aims to allow individuals to design, conduct, and analyze methodologically sound, statistically robust n-of-1 trials. We are testing our app and platform by applying the n-of-1 concept to a health outcome (cognitive performance) and to interventions (caffeine vs. caffeine + L-theanine) that are fast-acting, controllable, and easily measurable. Each individual will participate in his/her own study, with treatments applied in sequence to assess whether L-theanine, in addition to caffeine, has a cognitive effect beyond that of caffeine alone for that person. This design choice allows us to use adapted versions of validated cognitive instruments readily available in Apple ResearchKit, and individuals may engage in interventions that are already part of their daily lives (e.g., drinking coffee or tea)[3]. These methods and tools are designed to empower individuals to make more rational, data-driven choices about their own health and wellness. This implementation will also allow us to assess the effectiveness of the n-of-1 trial within the current digital research landscape.

METHODS

Study Design

We designed and developed a smartphone app and software platform that provides individuals the opportunity to remotely engage in personalized n-of-1 investigations. The platform facilitates enrollment, longitudinal data capture, digital biomarker measurement, administration of validated

---

[1]Home | Usability.gov. Available online at: https://www.usability.gov/ (cited November 15, 2019).

[2]*Apple Launches Three Innovative Studies Today in the New Research App - Apple.* Available online at: https://www.apple.com/newsroom/2019/11/apple-launches-three-innovative-studies-today-in-the-new-research-app/ (accessed November 15, 2019).

[3]*ResearchKit.* Available online at: http://researchkit.org/ (accessed November 15, 2019).

instruments, study task notifications and reminders, statistical analysis, and access to data and results. The platform is designed in a modular way to allow for new studies to be deployed easily, adapting components like e-consent, onboarding, and results reporting (Bobe et al., 2020). To evaluate the n-of-1 platform, we aimed to design a study that is relevant to broad audiences, incorporates ubiquitous and safe treatments, and utilizes validated assessment instruments in a digital format. A cognition study that evaluates the effects of two commonly consumed substances, caffeine and L-theanine, meet these criteria and serves as our first study deployed on the platform.

Treatments

We will measure the effects of two different treatments on daily cognitive function:

- Treatment A: caffeine (50–400 mg, based on choice of beverage/supplement)
- Treatment B: caffeine (50–400 mg, based on choice of beverage/supplement) + L-theanine (∼250 mg, based on choice of beverage/supplement)

These treatments were chosen due to their ubiquity, common daily use, efficacy, and safety. Participants may choose two beverages (e.g., coffee and tea) or they may also use an over-the-counter supplement (e.g., caffeine pills or L-theanine pills). Caffeine is one of the world's most commonly consumed drugs, and is often used to improve alertness and response time (Smith, 2002). L-theanine is an amino acid derived from tea leaves that is believed to have calming physical effects when consumed (Haskell et al., 2008). It is found most commonly in green tea and other teas, or in supplement form. The FDA has given L-theanine Generally Recognized as Safe ("GRAS") status[4]. Studies suggest that L-theanine may improve cognitive performance, and it is claimed that the combination of L-theanine with caffeine allows the consumer to feel the positive cognitive effects of caffeine while counteracting the "jitters," and reducing "mind wandering" (Bryan, 2008). The dosages vary between beverage and supplement choices, and so we provide a reference range to participants based on commonly consumed caffeinated drinks, available supplement dosages, and FDA compound review recommendations[4] (Keenan et al., 2011).

Treatment Blocks

We employ a randomized counterbalanced treatment design for each study length (ABBA or BAAB) (Duan et al., 2014). Participants are randomized into 1 of 3 study lengths —5, 15, and 27 days. Three different study lengths (short, medium, and long) are used in order to assess the effect of trial length on adherence and attrition. The treatment periods are either 1 day, 3 days, or 5 days, depending on study length.

Our 5-day study includes 1 day of baseline, 2 days of treatment A (caffeine) and 2 days of treatment B (caffeine + L-theanine), with treatment periods lasting 1 day. With "N" as baseline, a

participant may be randomized into a 5-day study of either of the following patterns:

N + (A) + (B) + (B) + (A)

N + (B) + (A) + (A) + (B)

Our 15-day study includes 3 days of baseline, 6 days of treatment A, and 6 days of treatment B, with treatment periods lasting 3 days. As such, a participant may be randomized into a 15-day study of either of the following patterns:

NNN + (AAA) + (BBB) + (BBB) + (AAA)

NNN + (BBB) + (AAA) + (AAA) + (BBB)

Our 27-day study includes 7 days of baseline, 10 days of treatment A, and 10 days of treatment B, with treatment periods lasting 5 days. Participants may be randomized into one of the following 27-day studies:

NNNNNNN + (AAAAA) + (BBBBB) + (BBBBB) + (AAAAA)

NNNNNNN + (BBBBB) + (AAAAA) + (AAAAA) + (BBBBB)

Participants must complete 1/3 of treatments and assessments during each treatment period to avoid study failure. The 5-day study is too short for a participant to miss any days of treatment/assessment and still have sufficient data to calculate a result. Individuals in the 15-day study can miss up to 2 days per treatment period before study failure, and those in the 27-day study can miss up to 3 days per treatment period.

Primary Outcome Measures

To assess cognitive function, we use three validated instruments adapted from and implemented using Apple's ResearchKit:

Remote Associates Test (RAT)

A measure of creative thinking (**Figure 1**). The RAT measures an individual's "creative" cognition by presenting them with a word problem consisting of 3 stimulus words and asking them to propose a fourth solution word that ties them together. For example, an individual may be prompted with the following: "sleeping, bean, trash." They would then try to come up with a linking fourth term, which in this case is "bag." It has been shown that problem solvers' success on items from the original RAT reliably correlates with their success on classic insight problems (Mednick, 1968; Dallob and Dominowski, 1993; Schooler and Melcher, 1995; Bowden and Jung-Beeman, 2003).

This test was implemented as a variation of the original RAT, developed as a custom digital ActiveTask within Apple's ResearchKit framework.

Six metrics are collected in our implementation of the RAT: words presented, word difficulty level, participant response/answer, average response time (in seconds), and score percentage (correct answers out of a possible 10)[5]. While multiple metrics are collected in our implementation of the RAT, only score percentage is analyzed and relayed back as an end result to the participant.

---

[4]Food and Drug Administration. *GRAS Notice 000338: L-Theanine.* Available online at: http://wayback.archive-it.org/7993/20171031043741/https://www.fda.gov/downloads/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/UCM269524.pdf (accessed October 3, 2019).

[5]Collection of RAT items | What word relates to all three? | *Remote Associates Test of Creativity.* Available online at: https://www.remote-associates-test.com/ (accessed October 3, 2019).

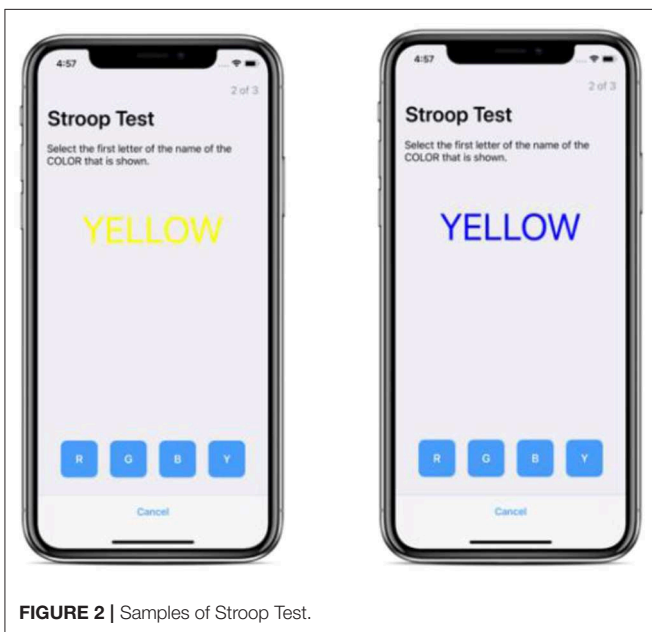**FIGURE 1 |** Sample of Remote Associates Test.

FIGURE 2 | Samples of Stroop Test.

Further details regarding our custom implementation of the RAT can be found in **Appendix A**.

Stroop Test
A measure of selective attention and processing speed (**Figure 2**). The Stroop Test is a measurement of executive function/reaction time. It assesses the ability of the user to distinguish between a printed word that names a color and the color of the actual text (Jensen and Rohwer, 1966). We use an abridged version of the test available from Apple ResearchKit as a predefined ActiveTask. In our abridged mobile version, a single task is presented and

we record a metric that captures both accuracy and reaction speed, known as the rate corrected score[6] (Woltz and Was, 2006). Additional details about the original and Apple-implemented versions of the Stroop Test may be found in **Appendix A**.

Trail Making Test (TMT)
A measure of visual attention and task-switching (**Figure 3**). The TMT is a standard component of many neuropsychological batteries and is one of the most commonly used tests because of its high sensitivity to the presence of cognitive impairment (Reitan, 1958; Spreen and Benton, 1965; Lezak et al., 1995; Kortte et al., 2002).

This test is implemented as a predefined ActiveTask through Apple's ResearchKit. As part of this predefined task, 13 dots are presented by default rather than 25 as in the original test. In this implementation, a line is drawn automatically as participants tap the next labeled dot in ascending order (i.e., when a participant taps "1" and subsequently taps "A," a line is drawn between the two dots.) For this implementation, only Part B of the test is presented to reduce the time commitment required for the participant. Part B is the more difficult of the two parts of the original test and there is evidence that Part B performance is indicative of executive function, where the difficulty of the task may reflect the cognitive flexibility of shifting the course of an ongoing activity (Lamberty et al., 1994; Arbuthnott and Frank, 2000; Kortte et al., 2002).

Two metrics are collected in our implementation of the TMT: number of errors (increased by tapping the incorrect dot), and total time to complete the test (in seconds). Further details regarding our implementation of the TMT can be found in **Appendix A**.

Inclusion and Exclusion Criteria
Eligibility of potential participants is determined through a digital eligibility screener. Individuals may join the study if they are over 18 years old, have an iPhone, and live in the United States.

Individuals may not join the study if they are pregnant or breastfeeding, or if they have a contraindication to caffeine. These exclusion criteria are based on the potential negative health effects of caffeine[7]. Before joining the study, we advise individuals to consult a medical professional if they are unsure of how caffeine may affect them. Due to the ongoing and individual nature of the study, users may enroll at a later date if they are currently ineligible (e.g., due to pregnancy status or age).

We developed our inclusion and exclusion criteria in an attempt to exclude as few potential participants as possible, with primary consideration for health and safety. While certain assessments may pose challenges to certain populations (e.g., RAT to non-native English speakers, and Stroop Test to colorblind individuals), we have decided not to exclude these populations because they may still receive study results based

---

[6]*Lessons Learned Implementing ResearchKit for a Study at Mount Sinai.* Available online at: http://hd2i.org/blog/2019/07/24/researchkit-for-research.html (accessed October 31, 2019).
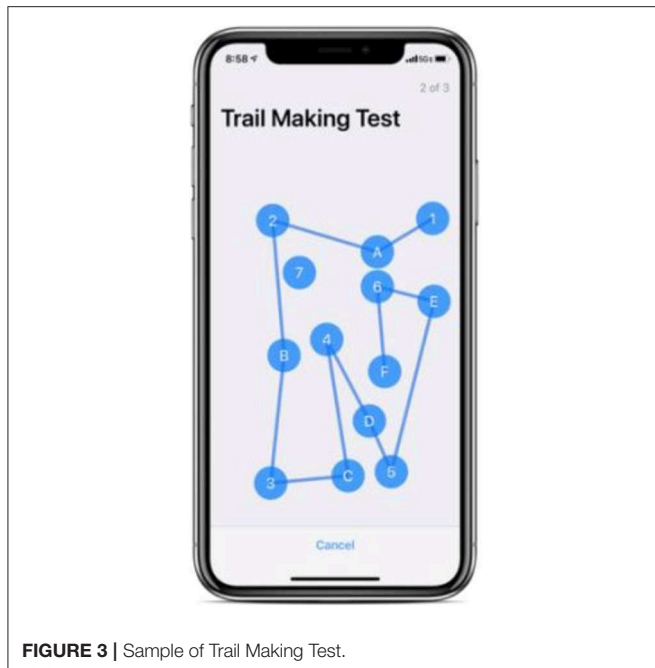[7]*Caffeine.* Available online at: https://medlineplus.gov/caffeine.html (accessed October 3, 2019).

FIGURE 3 | Sample of Trail Making Test.

on significant outcomes from one of the cognitive tests. Each test outcome is analyzed separately, so a treatment outcome is possible where focus is given to only one test in the cognitive assessment. Additionally, since a person serves as their own control in an n-of-1 trial, other possibly disadvantageous factors, such as mild cognitive impairment, should not matter in achieving a significant result.

Informed Consent

After a user downloads the app, they register for an account and go through a study onboarding process. The onboarding includes a study-specific eligibility screener, a brief introduction to n-of-1 studies, and a digital informed consent process.

Our informed consent process is modeled on Sage Bionetworks' multimedia eConsent framework (Doerr et al., 2016). It includes a short, self-guided digital consent module that clearly presents screens outlining the following parts of the consent form: study procedures, data privacy and security, data sharing, benefits, risks, withdrawal process, and consent review (**Figure 4**).

Procedures

After the consent module, the participant is presented with a PDF version of the full consent document. Participants type their name and provide an electronic signature after reviewing the full consent form. Their signature and timestamp are digitally placed on the consent and the signed version is subsequently available to the user for viewing within the study app at any time. Currently, informed consent is available in English only. The participant is provided the option of contacting the research staff during regular business hours if they have questions about the consent form or the study.

After a digital eligibility screening, onboarding to the n-of-1 concept, and informed consent via the study app, participants engage in a study that is randomized by study length (5, 15, or 27 days), treatment sequence (starting with treatment A or B), and app notification frequency (light or moderate). Participants log their choice of beverages or supplements for the study and choose a daily fixed time at which to measure their cognitive performance via the app. Participants are reminded to choose a treatment that they will be able to repeat in the same dose for the duration of the study. They are instructed to consume their treatment 1 h prior to the cognitive assessment. Participants are given a 2-h window in which to complete the cognitive assessment (**Figure 5**). This window is based on the quick wash-in/wash-out period of caffeine, and the time it takes for individuals to consume a beverage (coffee, tea) if chosen as a treatment (Institute of Medicine, 2001). Participants receive daily reminders via notifications in the app to take their treatments and complete the cognitive assessment. The daily cognitive assessment takes ∼5 min total to complete.

The trial begins with a baseline period, where participants take the daily cognitive assessment but do not take either treatment. Participants then alternate between treatment A and treatment B for the duration of the study, according to the instructions in the app. We employ a randomized counterbalanced design for each participant's study (either ABBA or BAAB) (Duan et al., 2014). Participants may log any daily occurrences related to the study, such as missed treatments, changed beverages, or interruptions that may have impacted their performance on the cognitive assessments. Numerous additional data categories and variables are collected during the study (see **Appendix B**). To mitigate the risk of influencing the study outcomes, the scores of daily cognitive assessments are not returned at the time of completion. Instead, individual study results are processed and provided to the participant at the conclusion of the study.

To improve the likelihood that individuals are able to obtain "actionable results," we also invite participants to enroll in a longer study after completing their initial study. While this feature is aimed at users initially enrolled in the 5-day study because they are the most likely group to obtain a statistically inconclusive result, it will be made available to all users who have completed a study. If the participant wants to continue, they will be able to choose their subsequent study length from the 3 available study lengths (5, 15, and 27 days). Currently, data from multiple studies will not be grouped and will be analyzed as independent studies. We anticipate adding a "grouped analysis" feature in the future.

Recruitment

Our N1 app is available for download in the Apple App Store. To ensure a diverse sample, the app and cognition study will be promoted using Mount Sinai Health System recruitment channels, related conferences (e.g., Quantified Self conference), and through promotional messaging on online message boards and social media (e.g., Reddit, Twitter, Facebook).

Recruitment is largely targeted to individuals who have shown interest in and/or previous experience using L-theanine and
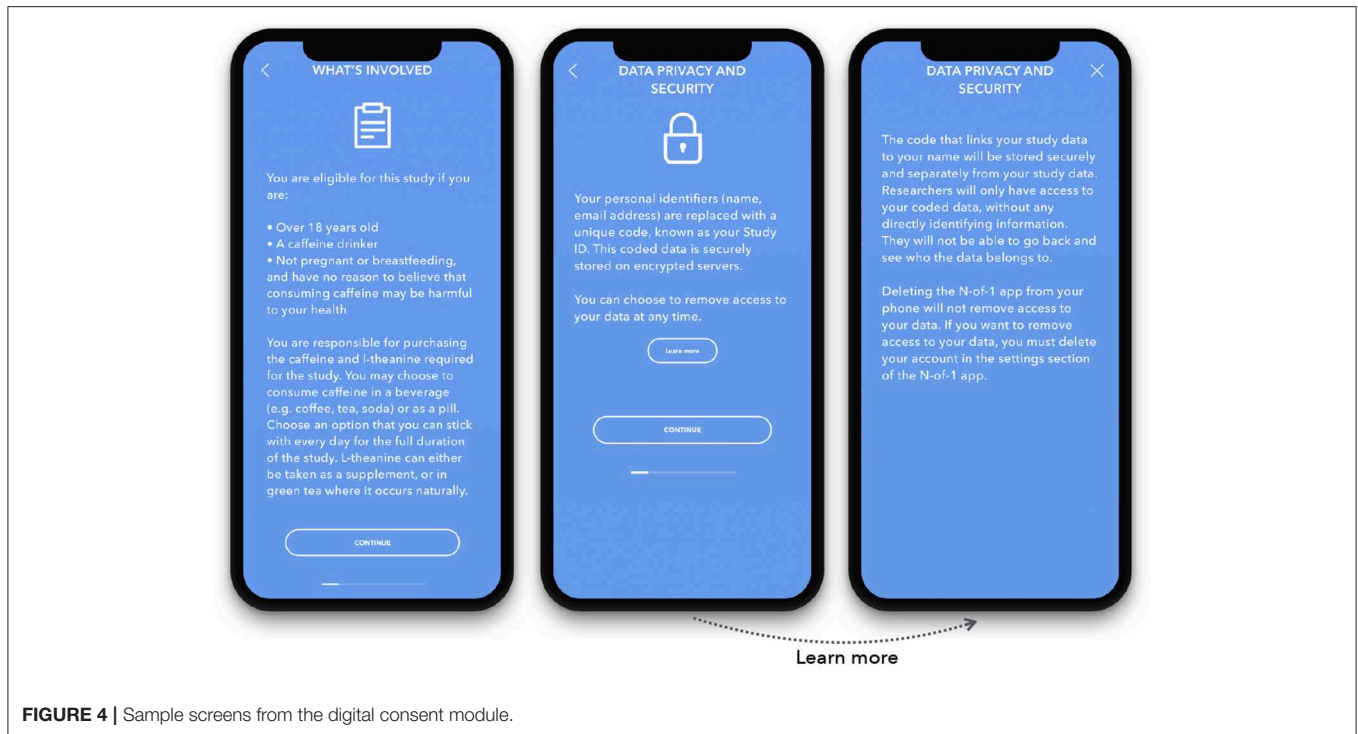
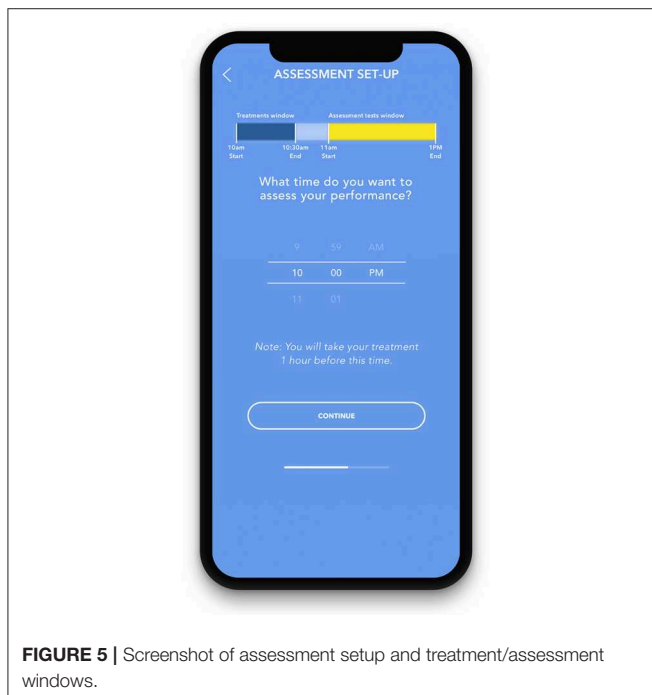**FIGURE 4 |** Sample screens from the digital consent module.

FIGURE 5 | Screenshot of assessment setup and treatment/assessment windows.

similar compounds through posting on specific L-theanine-related web pages and social media groups. However, people may join our study without prior experience or knowledge of L-theanine, as recruitment will also include more general websites. Our recruitment strategy is broad, with no population-based restrictions.

Safety Monitoring

Considering this is an online, remote, and individualized study, monitoring participants for safety and risk must differ from conventional studies where participants and researchers interact face-to-face. Thus, we provide participants with an electronic method in which they can email and report any questions, concerns, or abnormal events they believe to be related to the study, at any time. We also provide participants with a specific avenue within our app to send health-related questions, as opposed to general questions. This allows us to filter and expedite safety and risk-related concerns. If a health concern is submitted via email, a healthcare professional on the research team will contact the participant as soon as possible for additional information and will subsequently inform Mount Sinai of the event.

Analysis
Individual Study Performance

For individual cognition study results, we will use the outcomes of each cognitive test: (a) RAT, (b) Stroop Test, and (c) Trail Making Test. Each test outcome will be analyzed separately. Because both caffeine and L-theanine are short-acting, we do not anticipate carryover effects between treatment periods. For each test in the cognitive assessment, results are analyzed using a linear model with factors for treatment and block, and each treatment is compared to baseline. At present, we are not comparing the treatments to each other.

For each study duration, we will measure the proportion of completed n-of-1 trials that yield statistically meaningful results, for the comparisons (a) caffeine vs. baseline, (b) caffeine + L-theanine vs. baseline. A study is considered complete if a participant reaches the end of a trial without a study failure

or voluntary withdrawal. A study failure occurs when there is insufficient data generated during baseline or any treatment period due to missed treatments (self-reported) or incomplete assessments. Participants must complete 1/3 of treatments and assessments during each treatment period to avoid study failure and involuntary withdrawal. A study will be considered to have yielded a statistically-meaningful result if the coefficient on the treatment effect is significantly different from zero at the 80% confidence level in at least one of the three models; that is, if taking caffeine (relative to baseline, with or without L-theanine) produces an effect on cognitive performance measured by at least one of the three cognitive tests. The current confidence level was chosen arbitrarily, with a plan to develop a feature in the future that allows individuals to set their own preferred level of statistical significance.

Analysis of individual cognition studies may change in the future, as we will retain the raw data for each trial. For instance, we may choose to add a comparison of treatments to each other.

Of note, since the study is looking at individual outcomes, the consistency of caffeine and L-theanine dosage across individuals is not an outcome-related concern. Furthermore, the precise quantity of each treatment is not critical to the study design. Individual participant consistency in caffeine and L-theanine intake is the most important treatment factor related to the study outcome.

The code for calculating an individual numerical result will be made publicly available.

Platform Performance

As previously described, we will also evaluate the performance of the platform across several additional outcome measures: (a) proportion of studies completed, (b) proportion of studies yielding a statistically significant result, and (c) adherence, defined as the proportion of total actions (treatments + assessments) completed by a participant during a study[8] (Bobe et al., 2020). We will use a multiple logistic regression model to assess whether the randomized elements of the study (length, notifications) affect study completion, adjusting for any variance in age and sex. Study adherence will be assessed using a Bayesian survival-style model with semi-competing risks over the course of the study (Bobe et al., 2020).

Sharing of Individual Data and Results

Individuals will receive their personalized study results at the end of their study, on the app. They will be presented with graphical, numerical, and textual representations of the results, comparing both treatments to baseline measurements (**Figure 6**). Upon completion of the study, participants will also have the option to download their raw data.

There is a precedent for research participants to dynamically set and adjust their data sharing preferences. Providing the option for global sharing allows participants to contribute to

open science and creates increased inferential reproducibility (Plesser, 2018). Such dynamic preference-setting is a feature of the mPower Mobile Parkinson Disease Study, led by researchers at Sage Bionetworks (Bot et al., 2016). While pooled analysis of study data related to treatment response does not make sense for this study due to the variability of treatments and doses across individuals, other elements of the study may prove useful for researchers. Adherence data, treatment choices, and baseline cognitive assessment scores may be utilized by researchers for additional investigation. Data sharing language in our consent form is similar to that in existing approved protocols, in order to follow this precedent.

Participants may choose to share their cognitive assessment scores with friends and others by exporting the data from the app or by saving results images displayed in the app to their phone. Additionally, participants may choose to share their study data with external researchers. This goes into effect once a participant opts in to global sharing in the app settings. Name, contact information, and other directly identifiable information will never be included in externally shared study data.

Aggregated study results will be shared with app users via email once published.

RESULTS

Progress to Date

As of June 2019, we completed a "soft" beta test of our cognition study with 13 diverse participants recruited slowly over a few months. This helped us assess usability of the platform and study. Testers continuously shared feedback with the study team during their participation, in person and through online messaging. This process allowed us to fix some general platform issues and
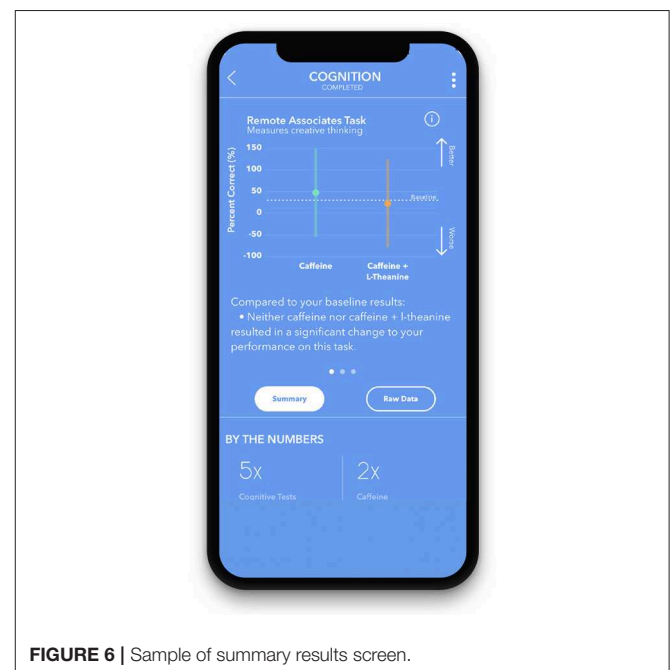

FIGURE 6 | Sample of summary results screen.

---

[8]*Assessing the Effectiveness of an N-of-1 Platform Using Study of Cognitive Enhancers..* Available online at: https://clinicaltrials.gov/ct2/show/NCT04056650?term=NCT04056650&rank=1 (accessed October 22, 2019).

address frequent study-related questions, which we subsequently clarified within the app through consent adjustments and the addition of a study-specific FAQ page. We used the issue-tracking feature on GitHub to record all noted problems and necessary fixes over time. We have iterated and improved the N1 app with more than 75 builds over the past 2 years.

In mid-October 2019, we publicly launched our platform and study on the Apple App Store, with ∼40 enrolled participants within the first month. Individual study results will be provided on an ongoing basis, and initial platform performance results will be expected upon completion of ∼100 studies (Bobe et al., 2020). While we do not have a recruitment goal for the study specifically due to the individual nature of the results, we do aim to recruit 640 participants in order to evaluate platform performance and the relationships between study completion, study duration and notification level (Bobe et al., 2020).

Ethics Approval
Ethics approval has been requested and granted by the Program for the Protection of Human Subjects at Icahn School of Medicine at Mount Sinai in New York City (IRB-18-00343; IRB-18-00789). This study is conducted in accordance with HIPAA regulations.

DISCUSSION
Relevance
To our knowledge, this is one of the first attempts to take app-based n-of-1 investigations outside of the clinic and into the participant's hands with ongoing enrollment on a publicly available app. Bringing n-of-1 studies outside of the clinical realm allows individuals to engage in regulated experiments about everyday health conditions and outcomes of interest on their own terms in a manner that aims to also ensure methodological rigor and safety. It remains a challenge for individuals or small groups to marshal the resources necessary to study themselves through rigorous n-of-1 investigations. Conversely, experimental rigor sometimes introduces complexities or burdens (e.g., daily actions, lengthy trials, etc.) that may be uninviting to a study's target population. With consideration for these challenges, n-of-1 experiments show promise and require further exploration. We hope that the development of this tool, and the introduction of more studies on the platform, will provide individuals with increased agency over their health and allow them to make conscious health-positive decisions that align with their lifestyle.

Our platform is designed in a modular fashion that allows new studies to be deployed by adapting existing components (e.g., e-consent, onboarding, notifications, reporting) (Bobe et al., 2020). This allows for easy implementation of future studies, which may open the door to clinician-driven protocols and collaborative research across institutes. Success of the platform in wellness-related treatments may further set the stage for its implementation in clinical medicine as an alternative approach to "therapy by trial" for some treatments (Kravitz et al., 2009).

Strengths and Limitations
One of the strengths of this study is that it introduces a novel model of robust self-investigation that can be built upon in future research and practice. Additionally, we are using versions of validated cognitive assessments (modified for digital use instead of on paper), which allows us to be more confident in our potentially actionable results.

Our choice of study and treatments was informed by what would be implemented most smoothly and successfully as a pilot study on the platform. It was also influenced by availability of assessments and ease of access to treatments. The ubiquity of coffee, tea, and over-the-counter supplements makes the study accessible to many individuals. However, this cognition study introduces a potential limitation due to our choice of L-theanine as part of a treatment. While commonly available in tea, L-theanine is most widely known within a population of individuals interested in "nootropics," substances believed to enhance cognition. These individuals, due to their existing interest in utilizing cognition-enhancing supplements, may also be inclined toward self-experimentation. We risk losing a generalizable and diverse participant population through treatment choice. We aim to mitigate this by also recruiting through broad-audience websites and social media. While this potential lack of diversity will not affect individual study results, it may impact the generalizability of our platform performance results if the primary users are not representative of the general population.

While longer study durations are desirable for generating statistically meaningful results, they may also be more likely to suffer from drop-out (Eysenbach, 2005), especially for a study that intrudes upon the caffeine ritual that some may find difficult to abstain from. The 15-day study is a good balance between these tradeoffs, which is why we allocate 60% of participants into this study duration, as described elsewhere (Bobe et al., 2020). We acknowledge that a 5-day study may not provide statistically meaningful results, which is why we unlock all study durations for participants after completion of their first study.

Another important note is that our individual analyses that are reported back to participants currently only include comparison of the two treatments to baseline (e.g., caffeine vs. baseline). While we are collecting data that will allow us to analyze comparative treatment responses (caffeine vs. caffeine + L-theanine) and plan to report them in our results paper, these individual results are not currently available on our app for participants to see. We plan to add this feature to our results visualization in the future, but it is still a work-in-progress.

Additionally, qualitative research with users to assess comprehension and preferences may benefit future iterations on the visualizations and reporting of results. Future user research, along with the additions of new app functionalities (e.g., choosing among numerous treatments to compare against each other) will strengthen this program moving forward.

Implications
In this study protocol, we described the methods for developing and launching a digital, remote n-of-1 study. To our knowledge,

this study is the first of its kind. It employs a statistical model that accounts for many variables, bias, and learning effects. The results of the study are expected to be relevant to individual participants who want to make positive lifestyle changes, as well as clinicians and researchers interested in exploring n-of-1 methodology. N1 platform and study implementation can be enhanced by learning about what draws people toward self-investigation and behavior change, as well as what causes digital study drop-out. We anticipate that these insights will become clearer as our cognition study progresses, and we will use participant feedback along with these insights about adherence to inform the next iteration of study on this platform (Bobe et al., 2020).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

This study was carried out in accordance with HIPAA regulations. The protocol was approved by the Program for the Protection of Human Subjects at Icahn School of Medicine at Mount Sinai in New York City (IRB-18- 00343; IRB-18-00789). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00004/full#supplementary-material

REFERENCES

Arbuthnott, K., and Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. *J. Clin. Exp. Neuropsychol.* 22, 518–528. doi: 10.1076/1380-3395(200008)22:4;1-0;FT518

Bobe, J. R., Buros, J., Golden, E., Johnson, M., Jones, M., Percha, B., et al. (2020). Factors associated with trial completion and adherence in app-based n-of-1 trials: protocol for a randomized trial evaluating study duration, notification level, and meaningful engagement in the brain boost study. *JMIR Res. Protoc.* 9:e16362. doi: 10.2196/16362

Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., et al. (2016). The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* 3:160011. doi: 10.1038/sdata.2016.11

Bowden, E. M., and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instr. Comput.* 35, 634–639. doi: 10.3758/bf03195543

Bryan, J. (2008). Psychological effects of dietary components of tea: Caffeine and L-theanine. *Nutr. Rev.* 66, 82–90. doi: 10.1111/j.1753-4887.2007.00011.x

Dallob, P. I., and Dominowski, R. L. (1993). "Erroneous solutions to verbal insight problems: effects of highlighting critical material," *Paper Presented at the 73rd Annual Meeting of the Western Psychological Association* (Portland, OR).

Doerr, M., Suver, C., and Wilbanks, J. (2016). *Developing a Transparent, Participant-Navigated Electronic Informed Consent for Mobile-Mediated Research*. Available online at: https://ssrn.com/abstract=2769129

Duan, N., Eslick, I., Gabler, N. B., Kaplan, H. C., Kravitz, R. L., Larson, E. B., et al. (2014). *Design and Implementation of N-of-1 Trials: A User's Guide*. Kravitz, R., Duan, N., Eds, and the DEcIDE Methods Center N-of-1 Guidance Panel. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality. Available online at: https://

effectivehealthcare.ahrq.gov/products/n-1-trials/research-2014-4/ (accessed October 3, 2019).

Eysenbach, G. (2005). The law of attrition. *J. Med. Internet Res.* 7:e11. doi: 10.2196/jmir.7.1.e11

Haskell, C. F., Kennedy, D. O., Milne, A. L., Wesnes, K. A., and Scholey, A. B. (2008). The effects of l-theanine, caffeine and their combination on cognition and mood. *Biol. Psychol.* 77, 113–122. doi: 10.1016/j.biopsycho.2007.09.008

Institute of Medicine (2001). *Caffeine for the Sustainment of Mental Task Performance: Formulations for Military Operations*. Washington, DC: The National Academies Press. doi: 10.17226/10219

Jensen, A. R., and Rohwer, W. D. (1966). The stroop color-word test: a review. *Acta Psychol.* 25, 36–93. doi: 10.1016/0001-6918(66)90004-7

Keenan, E. K., Finnie, M. D. A., Jones, P. S., Rogers, P. J., and Priestley, C. M. (2011). How much theanine in a cup of tea? Effects of tea type and method of preparation. *Food Chem.* 125, 588–594. doi: 10.1016/j.foodchem.2010.08.071

Kortte, K. B., Horner, M. D., and Windham, W. K. (2002). The trail making test, part B: cognitive flexibility or ability to maintain set? *Appl. Neuropsychol.* 9, 106–109. doi: 10.1207/S15324826AN0902_5

Kravitz, R. L., Paterniti, D. A., Hay, M. C., Subramanian, S., Dean, D. E., Weisner, T., et al. (2009). Marketing therapeutic precision: Potential facilitators and barriers to adoption of n-of-1 trials. *Contemp. Clin. Trials.* 30, 436–445. doi: 10.1016/j.cct.2009.04.001

Kravitz, R. L., Schmid, C. H., Marois, M., Wilsey, B., Ward, D., Hays, R. D., et al. (2018). Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern. Med.* 178, 1368–1377. doi: 10.1001/jamainternmed.2018.3981

Lamberty, G. J., Putnam, S., Chatel, D., Bieliauskas, L., and Adams, K. M. (1994). Derived trail making test indices: a preliminary report. *Neuropsychiatry Neuropsychol. Behav. Neurol.* 7, 230–234.

Lezak, M. D., Howieson, D. B., and Loring, D. W. (1995). *Neuropsychological Assessment,* 3rd Edn. New York, NY: Oxford University Press.

Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., and Cohen, A. B. (2019). Digital health: a path to validation. *npj Digit Med.* 2:38. doi: 10.1038/s41746-019-0111-3

Mednick, S. A. (1968). The remote associates test. *J. Creat. Behav.* 2, 213–214. doi: 10.1002/j.2162-6057.1968.tb00104.x

Nikles, C. J., Mitchell, G. K., Del Mar, C. B., Clavarino, A., and McNairn, N. (2006). An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. *Pediatrics* 117, 2040–2046. doi: 10.1542/peds.2005-1328

Odineal, D. D., Marois, M. T., Ward, D., Schmid, C. H., Cabrera, R., Sim, I., et al. (2019). Effect of mobile device-assisted n-of-1 trial participation on analgesic prescribing for chronic pain: randomized controlled trial. *J. Gen. Intern. Med.* 35, 102–111. doi: 10.1007/s11606-019-05303-0

Plesser, H. E. (2018). Reproducibility vs. replicability: a brief history of a confused terminology. *Front. Neuroinform.* 11:76. doi: 10.3389/fninf.2017.00076

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Percept. Mot. Skills* 8, 271–276. doi: 10.2466/pms.1958.8.3.271

Schooler, J., and Melcher, J. (1995). "The ineffability of insight," in *The Creative Cognition Approach,* eds S. M. Smith, T. B. Ward, and R. A. Finke (Cambridge, MA: MIT Press), 97–133.

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., et al. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015: explanation and elaboration. *BMJ* 350:h1793. doi: 10.1136/bmj.h1793

Smith, A. (2002). Effects of caffeine on human behavior. *Food Chem. Toxicol.* 40, 1243–1255. doi: 10.1016/s0278-6915(02)00096-0

Spreen, O., and Benton, A. L. (1965). Comparative studies of some psychological tests for cerebral damage. *J. Nerv. Ment. Dis.* 140, 323–333. doi: 10.1097/00005053-196505000-00002

Turakhia, M. P., Desai, M., Hedlin, H., Rajmane, A., Talati, N., Ferris, T., et al. (2019). Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *Am. Heart J.* 207, 66–75. doi: 10.1016/j.ahj.2018.09.002

U.S. Department of Health and Human Services (2016). *Use of Electronic Informed Consent: Questions and Answers*. Tel. Available online at: https://www.fda.gov/media/116850/download (accessed October 3, 2019).

Wang, T., Azad, T., and Rajan, R. (2019). *The Emerging Influence of Digital Biomarkers on Healthcare*. Available online at: https://rockhealth.com/reports/the-emerging-influence-of-digital-biomarkers-on-healthcare/ (accessed October 11, 2019).

Woltz, D. J., and Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Mem Cogn.* 34, 668–684. doi: 10.3758/bf03193587