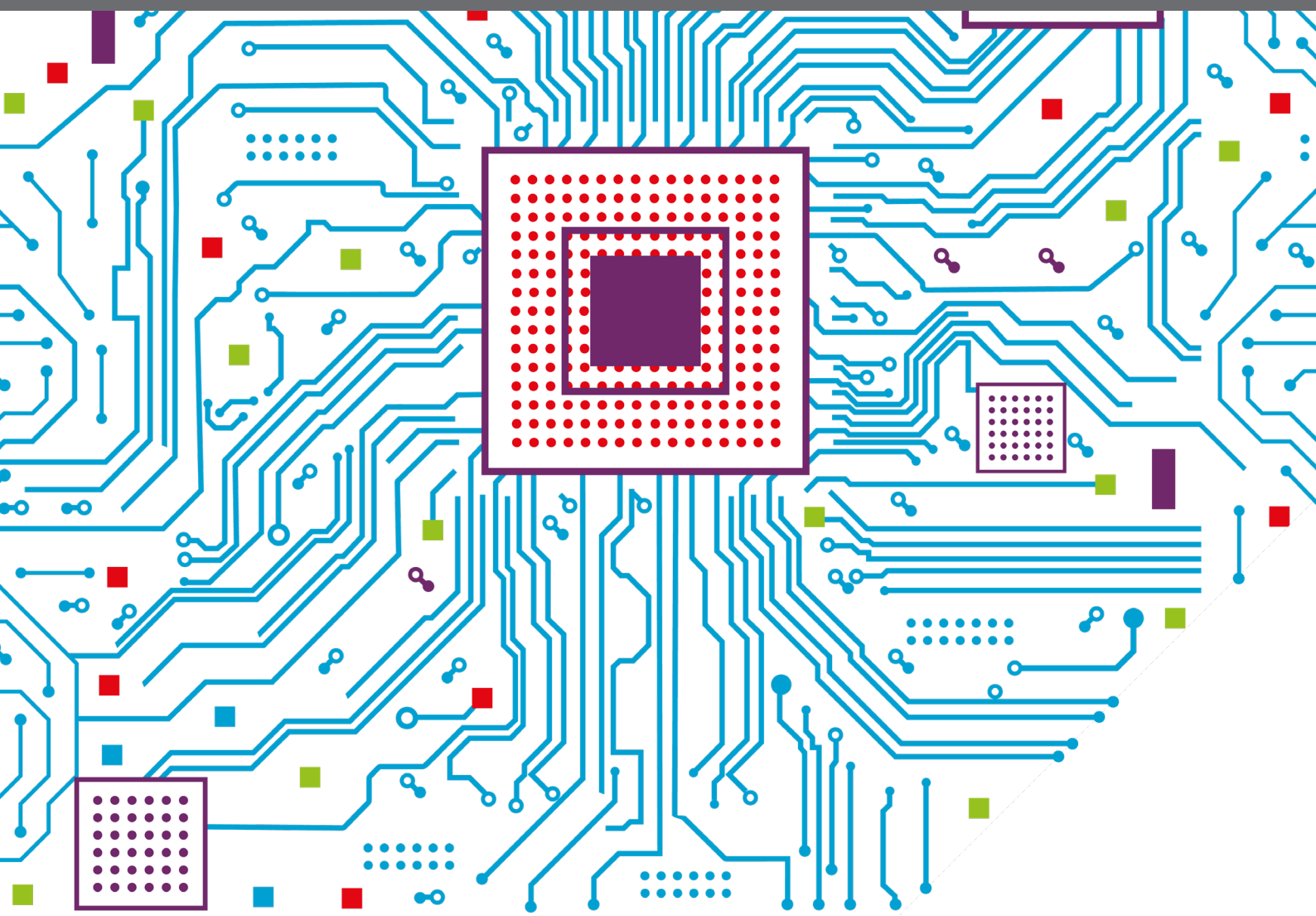
CREATING EVIDENCE FROM REAL WORLD PATIENT DIGITAL DATA

EDITED BY: Jane Nikles, Eric J. Daza, Suzanne McDonald, Eric Hekler and Nicholas Schork

PUBLISHED IN: Frontiers in Computer Science, Frontiers in Psychology, Frontiers Digital Health, Frontiers in Public Health and Frontiers in Psychiatry



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-671-3
DOI 10.3389/978-2-88966-671-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CREATING EVIDENCE FROM REAL WORLD PATIENT DIGITAL DATA

Topic Editors:
Jane Nikles, The University of Queensland, Australia
Eric J. Daza, Evidation Health, Inc., United States
Suzanne McDonald, The University of Queensland, Australia
Eric Hekler, University of California, San Diego, United States
Nicholas Schork, Translational Genomics Research Institute, United States

Topic Editor Dr. Eric Daza is Senior Statistician at Clarify Health Solutions. All other Topic Editors declare no competing interests with regards to the Research Topic subject.

Citation: Nikles, J., Daza, E. J., McDonald, S., Hekler, E., Schork, N., eds. (2021). Creating Evidence from Real World Patient Digital Data. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-671-3

Table of Contents

frontiers in Computer Science

EDITORIAL
published: 24 February 2021
doi: 10.3389/fcomp.2020.636996

Editorial: Creating Evidence From Real World Patient Digital Data

Jane Nikles[1]*, *Eric J. Daza*[2], *Suzanne McDonald*[1], *Eric Hekler*[3] and *Nicholas J. Schork*[4]

[1]The University of Queensland Centre for Clinical Research, The University of Queensland, Brisbane, QLD, Australia, [2]Evidation Health, Inc., San Matteo, CA, United States, [3]Department of Family Medicine & Public Health, University of California, San Diego, CA, United States, [4]Translational Genomics Research Institute, Phoenix, AZ, United States

Keywords: N-of-1 clinical trials, single case experimental designs, digital health, real world data, single case designs

Editorial on the Research Topic

Creating Evidence from Real World Patient Digital Data

Over the last 5 years there has been a tremendous rise in interest in personalized medicine and patient-centred healthcare. This has resulted in an expanding focus on the value of single-case research methods (also known as N-of-1 or Single-Case Designs (SCDs)). SCDs use repeated measures, frequent data collection and patient-reported outcome measures to draw conclusions about an individual.

We use "SCD" as a broad term that includes all SCD sub-designs, encompassing Single-Case Experimental Designs (SCEDs), N-of-1 randomised controlled trials (N-of-1 trials) and Single-Case Observational Designs (SCODs)—all with specific features (see **Box 1**).

SCDs can complement RCTs in a wide range of clinical research and practice contexts (Gabler et al., 2011; Smith, 2012; Punja et al., 2016; McDonald et al., 2017; Shaffer et al., 2018; Hekler et al., 2019). Since 2015, the number of SCD articles published yearly has rapidly increased. High profile articles have recently been featured in Nature (Schork, 2015) and JAMA (Stunneneberg et al., 2018). Guidelines have been published to improve SCD conduct and reporting quality, including CENT (N-of-1 trials) (Vohra et al., 2016), CENT for TCM (N-of-1 trials for traditional Chinese medicine) (Li et al., 2019), SCRIBE (SCDs in behavioral sciences) (Tate et al., 2017), and SPENT (N-of-1 protocol design) (Porcino et al., 2019).

N-of-1 randomised controlled trials (RCTs) provide an opportunity to evaluate individual-person response to interventions, by randomly allocating different time periods within an individual to repeated intervention and control conditions and then comparing responses across these periods.

N-of-1 observational studies involve the repeated measurement of an outcome (e.g., pain) in a person over time, but with no intervention implemented, in order to draw conclusions about naturally occurring patterns and predictors of outcomes over time.

Both N-of-1 RCTs and observational studies can have a 'self-study' design, where an individual conducts the study on themselves, to answer research questions they have generated themselves.

N-of-1 RCTs and observational studies provide individualized findings that can be aggregated to produce results equivalent to those found in traditional group-based RCTs and population-level epidemiological studies, respectively, but may require fewer people for the same statistical power.

Because of their patient-centricity, individualized results, and amenability for use by doctors to tailor therapies to individuals, SCDs are ideally placed to complement, strengthen, and generate advances in precision medicine, patient-centred healthcare, personalized health and digital health (Schork, 2015; Hekler et al., 2020).

Digital health is an exploding field, with over 1,000 relevant studies registered on clinicaltrials.gov in November 2020. Digital health includes digital therapeutics (i.e., actual interventions implemented on a digital device), patient-reported outcomes (e.g., survey responses administered via phone or web app), mobile health tools, such as wearable devices (e.g., worn sensors, implants) that can be used to monitor various aspects of an individual's health in SCDs, along with telehealth, electronic health record

OPEN ACCESS

Edited and reviewed by:
Nilufar Baghaei,
Massey University, New Zealand

***Correspondence:**
Jane Nikles
uqjnikle@uq.edu.au

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *02 December 2020*
Accepted: *23 December 2020*
Published: *24 February 2021*

Citation:
Nikles J, Daza EJ, McDonald S, Hekler E and Schork NJ (2021) Editorial: Creating Evidence From Real World Patient Digital Data. Front. Comput. Sci. 2:636996. doi: 10.3389/fcomp.2020.636996

BOX 1 | Description of some SCD sub-designs
What are Single-Case Designs?
Single-Case Designs (***SCDs***) gather and interpret repeated measures data from a single participant over time.
What are Single-Case Experimental Designs?
Single-Case Experimental Designs (***SCEDs***) are experimental designs that test the effect of an intervention on one participant, using repeated measurements, sequential (±randomised) introduction of an intervention and method-specific data analysis, including visual and statistical techniques. Simultaneous or sequential replications are possible with more individuals.
What are N-of-1 trials?
N-of-1 randomised controlled trials (***RCTs***) provide an opportunity to evaluate individual-person response to interventions, by randomly allocating different time periods within an individual to repeated intervention and control conditions and then comparing responses across these periods.
What are SCODs?
Single-Case Observational Designs (***SCODs***) involve the repeated measurement of an outcome (e.g., pain) in a person over time, but with no intervention implemented, in order to draw conclusions about naturally occurring patterns and predictors of outcomes over time.

systems, and data analytics tools. Digital health data can potentially be monitored continuously during the SCD, and used to help tailor a treatment to the needs and preferences of each patient. Collecting data digitally can be very convenient for participants, especially if they do not have to actively record data. SCDs can be of longer duration than group-based RCTs, so they are well suited to data collection using digital health devices and related technologies.

SCDs can be 'self-study' designs, wherein an individual conducts the study on themselves, to answer research questions important to them. In addition, both use techniques such as meta-analysis to aggregate individual-level findings. This can produce results equivalent to those found in traditional group-based RCTs and population-level epidemiological studies, but may require fewer people to obtain the same statistical power as a larger, group-based study. It is common for a series of SCDs to be conducted, and the results to be pooled to address issues of, e.g., generalisability and population response rate.

To acknowledge the emerging field of digital N-of-1 and SCD research, a team of experts recently completed editing a Frontiers Research Topic entitled 'Creating Evidence from Real World Patient Digital Data'. This Frontiers Research Topic covered digital health applications, delivery, and analysis of SCDs (including self-studies) in any health discipline. It focused on mobile health (mHealth), smart phone applications (apps), wearable devices, sensors and implants, real-time tracking, data analytics, patient experience of digital health and mobile health, patients as collaborators in personalized medicine, and self-tracking efforts in the "citizen science" community.

The articles covered a selection of original research, methodology pieces, opinion pieces, and study protocols, discussing important themes including the significance of technology, emergence of the "self-scientist", and the value of using diverse N-of-1 and SCD designs. The 13 articles written by 60 authors have already generated over 37,000 views to January 2021, reflecting the strong interest in these methods globally. The topic has had viewers from all over the world, particularly from the United States, United Kingdom, Germany, France, and China.

SIGNIFICANCE OF TECHNOLOGY

A key feature of several articles was use of N-of-1 studies enabled by mobile app technology. For example, Bobe et al. discussed the potential for clinicians and patients to collaboratively use an app-based platform for N-of-1 trials, and reported results of a survey exploring perceptions about implementing an app-based N-of-1 trial platform to support data-driven decisions around insomnia treatment. Kravitz et al. reported on feasibility, acceptability, and influence of mHealth-supported N-of-1 trials for enhanced cognitive and emotional well-being in US volunteers. Bauer et al. described a feasibility study protocol for testing individual-level effects of tamsulosin using the PERSONAL app to track daily urinary symptoms and medication side effects. And Golden et al. detailed a protocol for self-directed, mobile-app-based N-of-1 studies to test the effects of caffeine and L-theanine on cognitive performance.

The use of emerging technology in SCDs is not limited to mobile apps. Chrisinger outlined the opportunities to use GPS technology to create geolocated N-of-1 datasets that could be used to explore relationships between individuals, their environment, and their health, or what has been termed "the quantified self-in-place". Chrisinger argues that individual-level information in real-world environmental contexts might lead to a better understanding of how treatments and interventions work, for whom, and under which conditions. A number of logistical, methodological, and ethical challenges were identified.

EMERGENCE OF THE "SELF-SCIENTIST"

Many articles discussed the concept of the "self-scientist" or "personal science", which has been enabled by availability of diverse and accessible digital tools to collect personal real-world data. Wolf and de Groot outlined a 5-stage conceptual framework to guide research and education into practice of "personal science", which they define as using empirical methods to address personal health questions. Important similarities and differences between personal science, citizen science, and single subject (N-of-1/SCD) research were also discussed.

Schwartz et al. discussed the concept of the "digital twin", where individuals have access to self-generated biobehavioural information derived from data collected from various sensors and devices that may reflect their biological and environmental circumstances, and be used to make predictions about their health. Advances in technology have led to more accurate

capture of various biometric, behavioral, emotional, cognitive, and psychological aspects of daily life. Data-driven feedback from their "digital twin" information may inspire users to conduct self-experiments to evaluate their own treatment responses.

Nebeker et al. described the patient perspective of using self-study and peer-to-peer support in conjunction with traditional clinical support guided by external evidence generated from group-based studies. They argue that access to digital health technologies, wearable sensors, affordable lab screenings, etc. may contribute to a paradigm shift wherein "sick" care may become authentic "health" care.

THE VALUE OF DIVERSE N-OF-1 DESIGNS

The articles provided in this Research Topic also covered different SCD types. McMillan and Dixon used a series of digital SCODs to characterize self-regulatory processes, motivation to conserve resources, and activity levels in people with chronic pain. They found that motivational and self-regulatory processes during goal pursuit goal may play a key role in an intervention' success. Similarly, Altman et al. used a series of digital SCODs to characterize processes and mechanisms of change over a course of psychotherapy.

Hendrickson et al. presented findings from statistical simulation studies they conducted to optimize aggregated N-of-1 trial designs for predictive biomarker validation. They described a set of simulation studies comparing the power of four different trial designs to detect relationship between a predictive biomarker (measured at baseline) and subjects' specific responses to the pharmacotherapeutic agent prazosin for Post Traumatic Stress Disorder.

Finally, Munson et al. argued that elicitation of individualized goals and customization of tracking to support those goals are a critical part of conducting N-of-1 studies. Their conclusions serve as an important reminder about the flexibility of these methods and their ability to tailor to preferences and needs of individuals through patient-centred N-of-1 designs.

The great variety of articles illustrates the versatility of this design and the opportunity to use digital methods to collect real world health data. We look forward to seeing the impact of this research on digital health and personalized medicine worldwide.

AUTHOR CONTRIBUTIONS

JN drafted the editorial and revised it after feedback from the other authors. SM, EH, EJD and NJS provided important input and critical review of the manuscript.

ACKNOWLEDGMENTS

A similar article has been published as a blog on the website of The International Collaborative Network (ICN) for N-of-1 Clinical Trials and Single-Case Designs (SCDs). JN and SM are co-chairs of this network.

REFERENCES

Gabler, N. B., Duan, N., Vohra, S., and Kravitz, R. L. (2011). N-of-1 trials in the medical literature: a systematic review. *Med. Care* 49 (8), 761–768. doi:10.1097/MLR.0b013e318215d90d

Hekler, E., Tiro, J. A., Hunter, C. M., and Nebeker, C. (2020). Precision health: the role of the social and behavioral sciences in advancing the vision. *Ann. Behav. Med.* 54 (Issue 11), 805–826. doi:10.1093/abm/kaaa018

Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., and Sim, I. (2019). Why we need a small data paradigm. *BMC Med.* 17, 133. doi:10.1186/s12916-019-1366-x

Li, J., Hu, J.-y., Zhai, J.-b., Niu, J.-q., Kwong, J. S. W., Ge, L., et al.CENT for TCM Working Group. (2019). CONSORT extension for reporting N-of-1 trials for traditional Chinese medicine (CENT for TCM): recommendations, explanation and elaboration. *Compl. Ther. Med.* 46, 180–188. doi:10.1016/j.ctim.2019.08.014

McDonald, S., Quinn, F., Vieira, R., O'Brien, N., White, M., Johnston, D. W., et al. (2017). The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol. Rev.* 11 (4), 307–323. doi:10.1080/17437199.2017.1316672

Porcino, A. J., Shamseer, L., Chan, A. W., Kravitz, R. L., Orkin, A., Punja, S., et al.SPENT group (2020). SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. *BMJ* 368, m122. doi:10.1136/bmj.m122

Punja, S., Bukutu, C., Shamseer, L., Sampson, M., Hartling, L., Urichuk, L., et al. (2016). The design, analysis and meta-analysis of N-of-1 trials: a tapestry of heterogeneity. *J. Clin. Epidemiol.* 76, 47. doi:10.1016/j.jclinepi.2016.03.023

Schork, N. J. (2015). Personalized medicine: time for one-person trials. *Nature* 520, 609. doi:10.1038/520609a

Shaffer, J. A., Kronish, I. M., Falzon, L., Cheung, Y. K., and Davidson, K. W. (2018). N-of-1 randomized intervention trials in health psychology: a systematic review and methodology critique. *Ann. Behav. Med.* 52 (9), 731–742. doi:10.1093/abm/kax026

Smith, J. D. (2012). Single-case experimental designs: a systematic review of published research and current standards. *Psychol. Methods* 17 (4), 510–550. doi:10.1037/a0029312

Stunnenberg, B. C., Raaphorst, J., Groenewoud, H. M., Statland, J. M., Griggs, R. C., Woertman, W., et al. (2018). Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated N-of-1 trials. *J. Am. Med. Assoc.* 320 (22), 2344–2353. doi:10.1001/jama.2018.18020

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., et al. (2017). The single-case reporting guideline in BEhavioural interventions (SCRIBE) 2016 statement. *Neuropsychol. Rehabil.* 27 (1), 1–15. doi:10.1080/09602011.2016.1190533

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., et al. CENT Group (2016). CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *J. Clin. Epidemiol.* 76, 9–17. doi:10.1016/j.jclinepi.2015.05.004

Conflict of Interest: JN and SM have a commercial interest in N-of-1 Hub Pty Ltd. consultancy company. Author EJD is employed by the company Evidation Health.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

METHODS
published: 13 February 2020
doi: 10.3389/fcomp.2020.00004

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.

OPEN ACCESS

Edited by:
Jane Nikles,
University of Queensland, Australia

Reviewed by:
Joyce Samuel,
University of Texas Health Science Center at Houston, United States
Ian Kronish,
Columbia University, United States

****Correspondence:***
Eddye Golden
eddye.golden@mssm.edu
Jason Bobe
jason.bobe@mssm.edu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *18 November 2019*
Accepted: *27 January 2020*
Published: *13 February 2020*

Citation:
Golden E, Johnson M, Jones M, Viglizzo R, Bobe J and Zimmerman N (2020) Measuring the Effects of Caffeine and L-Theanine on Cognitive Performance: A Protocol for Self-Directed, Mobile N-of-1 Studies. Front. Comput. Sci. 2:4. doi: 10.3389/fcomp.2020.00004

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.

Keywords: n-of-1 trials, cognition, digital health, caffeine, nootropics, cognitive, mobile app, mhealth

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).

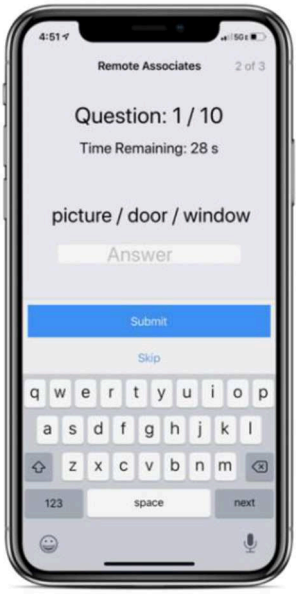


FIGURE 1 | Sample of Remote Associates Test.

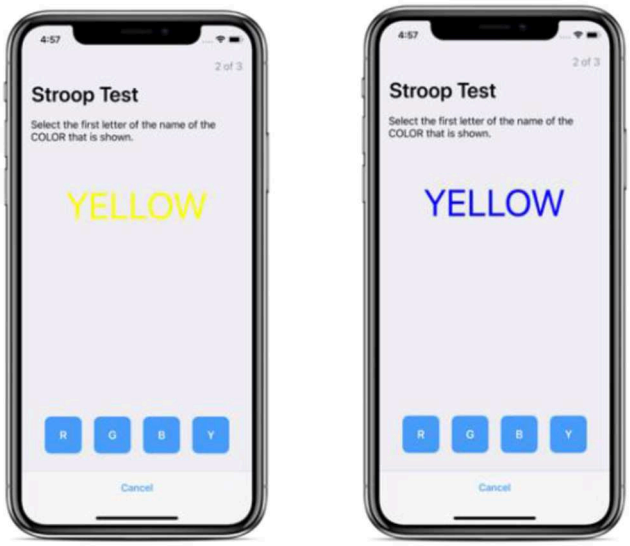


FIGURE 2 | Samples of Stroop Test.

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).

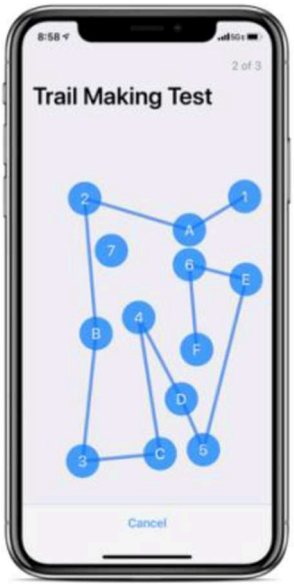


FIGURE 3 | Sample of Trail Making Test.

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

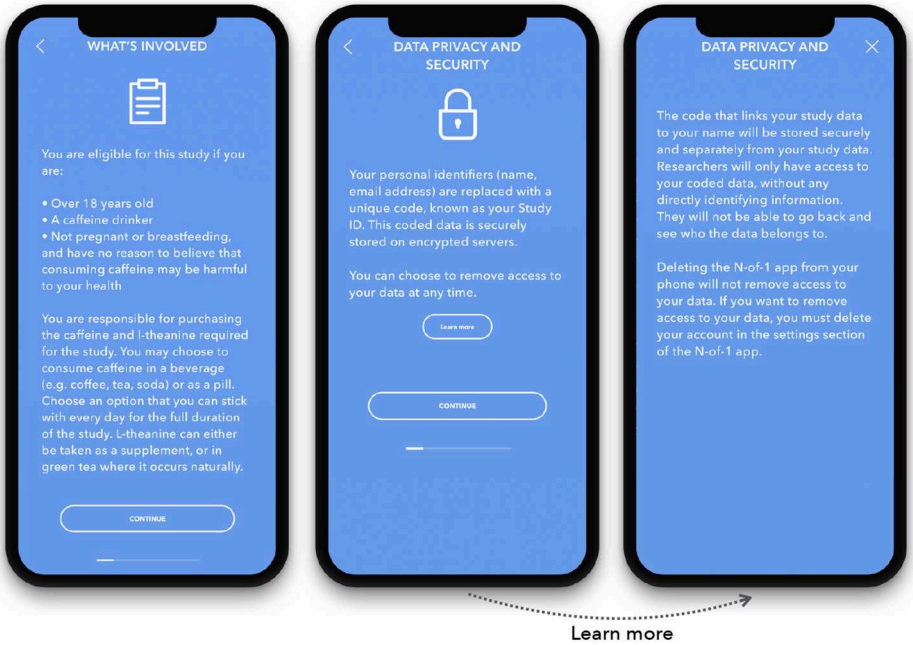


FIGURE 4 | Sample screens from the digital consent module.

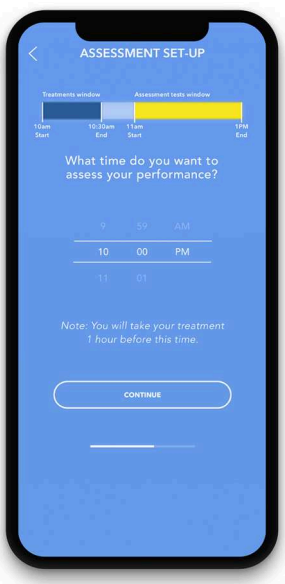


FIGURE 5 | Screenshot of assessment setup and treatment/assessment windows.

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

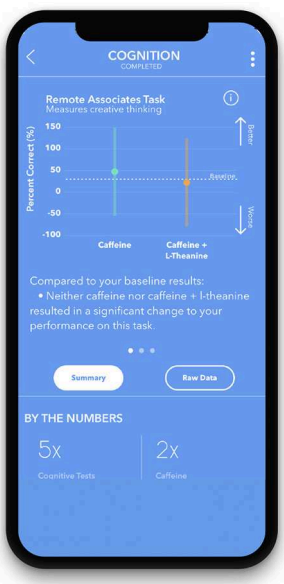


FIGURE 6 | Sample of summary results screen.

[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

EG wrote the manuscript with support from JB, MJoh, MJon, and NZ. NZ conceived the original idea. NZ and JB designed the project with support from all authors. EG developed the regulatory framework and is program manager for the project. RV created the design framework for the project and MJoh and MJon carried out the development. MJoh created the front-end of the platform and MJon built the back-end. NZ developed the analytical methods. NZ and JB directed the project.

FUNDING

This project was fully supported by internal funding from the Icahn School of Medicine at Mount Sinai.

ACKNOWLEDGMENTS

We thank Joshua and Marjorie Harris, the Harris Family Charitable Foundation, and Julian Salisbury, for their generous donations to support the development and success of this project. Additionally, we thank all of our volunteer beta testers and lab members for providing continuous feedback and support.

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 April 2020
doi: 10.3389/fpsyg.2020.00782

Why Does Therapy Work? An Idiographic Approach to Explore Mechanisms of Change Over the Course of Psychotherapy Using Digital Assessments

Allison Diamond Altman[1], Lauren A. Shapiro[2] and Aaron J. Fisher[1]*

[1] Idiographic Dynamic Lab, Psychology Department, University of California, Berkeley, Berkeley, CA, United States,
[2] Psychology Department, The Wright Institute, Berkeley, CA, United States

Background and Objective(s): While psychotherapy treatments are largely effective, the processes and mechanisms underlying such positive changes remain somewhat unknown. Focusing on a single participant from a treatment outcome study that used a modular-based cognitive behavior therapy protocol, this article aims to answer this question by identifying changes in specific symptomatology over the course of the treatment. Using quantitative data derived from digital health methodology, we analyzed whether a given therapeutic intervention was related to downstream effects in predicted symptom domains, to assess the accuracy of our interventions.

Methods: This case study employed an observational N-of-1 study design. The participant (n = 1) was a female in the age range of 25–35 years. Using digital health data from ambulatory assessment surveys completed prior to and during therapy, separate linear regression analyses were conducted to assess if hypothesized treatment targets reduced after a given module, or intervention.

Results: Support was found for some of the hypothesized quantitative changes (e.g., decreases in avoidance after exposures module), yet not for others (e.g., decreases in rumination following the mindfulness module).

Conclusion: We present data and results from our analyses to offer an example of a novel design that may allow for a greater understanding of the nature of symptom changes with increased granularity throughout the course of a psychological treatment from the use of digital health tools.

Keywords: case study, digital assessments, mechanisms, cognitive behavioral therapy, ambulatory assessment

OPEN ACCESS

Edited by:
Eric Hekler, University of California, San Diego, United States

Reviewed by:
David Pincus, Chapman University, United States
Javier Fernández-Álvarez, Catholic University of the Sacred Heart, Italy

****Correspondence:***
Allison Diamond Altman
adiamond@berkeley.edu

Specialty section:
This article was submitted to Psychology for Clinical Settings, a section of the journal Frontiers in Psychology

Received: *30 October 2019*
Accepted: *30 March 2020*
Published: *24 April 2020*

Citation:
Altman AD, Shapiro LA and Fisher AJ (2020) Why Does Therapy Work? An Idiographic Approach to Explore Mechanisms of Change Over the Course of Psychotherapy Using Digital Assessments. Front. Psychol. 11:782. doi: 10.3389/fpsyg.2020.00782

INTRODUCTION

The field of clinical psychology has undergone many changes in the past decade. After the introduction of the Research Domain Criteria (RDoC; Insel et al., 2010) in 2010, researchers seeking external funding have been incentivized to move away from investigating psychopathological constructs at a disorder-level, in order to explore a dimensional system that encompasses multiple

levels units and of analysis, from genes at the most basic, granular level to behavior at the most macroscopic. While the development of the RDoC was not intended to replace the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) and serve as a diagnostic guide, its introduction nonetheless serves as a reminder that our current diagnostic system, and treatment development efforts, may be structurally flawed by the sheer heterogeneity underlying diagnostic labels as they currently stand. Taking the diagnosis of Posttraumatic Stress Disorder (PTSD) as an example, current DSM-5 criteria allow 636,120 combinations of presenting symptoms to exist in order to meet criteria for the diagnosis, meaning that it is possible for 636,120 individuals to meet criteria for PTSD, with no repeats in the exact constellation of symptoms from person to person (Galatzer-Levy and Bryant, 2013). Treatments for such conditions have been historically developed by researchers based on diagnostic categories and group-level (i.e., nomothetic) information. These often fail to produce significant change in a large subset of individuals, and outcomes for the treatment of many mental health conditions are lower than desired.

Over the past 30 years there has been a long history of researchers seeking to understand underlying mechanisms of psychotherapy success, and over time numerous models of change have emerged, including but not limited to psychotherapy integration (Stricker and Gold, 1996), the common factors approach (CF; Frank and Frank, 1991; Wampold, 2007), theoretical integration (Stricker and Gold, 1996), phase models (Howard et al., 1993), and the transtheoretical states of change model (Prochaska and DiClemente, 1983). From the psychotherapy integration approach, which aims to look beyond single approaches and instead hopes to integrate multiple perspectives, to the common factors approach, which proposes that different approaches in psychotherapy share common factors that account for the majority of the effectiveness of a psychological treatment, each model has developed their own ways of assessing and understanding change in psychotherapy. While these models indicate the efforts of researchers to understand why treatments may work, the majority of work in treatment development still focuses on groups, rather than individuals.

Conversely, clinicians typically focus on single patients, often using case conceptualization methods such as the case formulation approach to cognitive-behavior therapy (Persons, 2012). These often involve series of linked interventions typically following comprehensive assessments of the patient, and allow clinicians to choose techniques based on the best match to the presenting needs and problems of the patient.

To bridge the gap between research and practice, it is imperative for researchers to gather information that will be immediately useful to clinicians. Such information may come from idiographic treatment models, wherein researchers investigate change in individual patients, rather than diagnostic groups, to explore mechanisms of change over the course of a given treatment. Recently, and following along the footsteps of other medical domains such as oncology, there has been a push toward an idiographic, personalized approach to psychotherapy research, focusing on the precise symptomatology of an individual patient instead of broad diagnostic categories for generating treatment decisions. One such approach, outlined by Fisher (2015), calls for an idiographic, dynamic methodology whereby clinicians conduct person-specific dynamic assessments that yield information about syndrome structures and states to provide actionable information for personalized interventions. This work requires intensive, repeated digital assessment measures for individual patients, with the hope that this intensive measurement will yield prescriptive information for improved results. In a recent open trial of personalized modular CBT using the Unified Protocol (UP; Barlow et al., 2011), Fisher et al. (2019) demonstrated a Hedge's g of 2.33 over an average of 10 sessions, outperforming an historical average effect size for from a recent-meta-analysis (Johnsen and Friborg, 2015). The UP is typically delivered over 16 sessions, and has not been shown to demonstrate equivalent effects in trials to date (c.f. Farchione et al., 2012). While additional work is required to substantiate such an intensive approach to personalization, this trial illustrates that treatment based off of idiosyncratic structures of psychopathology may be an important part of improving overall treatment efficacy in the mental health domain. This information then can be immediately useful to practicing clinicians hoping to understand how best to approach singular cases.

The following case study aims to use the same person-specific ambulatory assessment data to investigate changes occurring in an individual over the course of a treatment, focusing on a single participant from the Fisher et al.'s (2019) open trial—participant 007 (P007). The idiographic approach outlined by Fisher et al. (2019) involves intensive, repeated digital assessment measures, captured four-times-daily for approximately 30 days. This provides sufficient data to facilitate person-specific factor analyses and dynamic factor modeling. In the open trial, pre-therapy analyses were used to determine predominance among latent symptom clusters in order to generate targeted therapies (using existing modules of the UP) person by person. Participants were also given the chance to extend these surveys and continue to complete them throughout the course of therapy, as did P007. The current article aims to use P007's data to identify changes in specific symptomatology over the course of treatment in order to identify if a given therapeutic intervention, or module (e.g., mindfulness) was related to downstream effects in predicted symptom domains (e.g., reduced restlessness). It should be acknowledged that some researchers have proposed that efficacy of psychotherapy is not due to specific interventions or techniques, but rather from factors of psychotherapy common to all treatments, referred to as common factors (Luborsky et al., 1975; Wampold, 2001, 2007). Yet, the current research aims to define specific effects that can be attributed to certain interventions, rather than common factors. This novel design may allow for a greater understanding of the nature of symptom changes with increased granularity throughout the course of treatment, and may serve as a model for clinicians and researchers to incorporate such work in their own research and practice.

As noted above, in the treatment trial from which P007's data was draw, delivery of the UP was personalized based on

TABLE 1 | Sessions and module orders for P007.

Session(s)	Therapy module
1	Motivation and enhancement
2	Emotional awareness and tracking
3	Emotional awareness and tracking
4	Mindfulness
5	Mindfulness and non-judgmental awareness
6	Mindfulness and non-judgmental awareness
7	Exposures (imaginal and *in vivo*)
8	Exposures (imaginal and *in vivo*)
9	Cognitive appraisals and reappraisals
10	Cognitive appraisals and reappraisals
11	Emotion-driven behaviors and emotional avoidance
12	Emotion-driven behaviors and emotional avoidance
13	Relapse prevention
14	Relapse prevention

TABLE 2 | Module-specific hypothesized relationships in digital assessment survey data over the course of therapy.

Therapy module	Quantitative survey hypothesis
Pre-therapy period	N/A
Motivation enhancement	N/A
Emotional awareness and tracking	N/A
Mindfulness and non-judgmental awareness	Reduced restlessness, dwelling on the past, worry; reduced worthlessness and guilt
Exposures (imaginal and *in vivo*)	Reduced difficulty concentrating, avoiding activities, avoiding people, procrastination, reassurance seeking
Cognitive appraisals and reappraisals	Reduced worry, depression; increased positivity, contentedness
Emotion driven behaviors and emotional avoidance	Reduced worry, depression
Relapse prevention	N/A

data gathered prior to therapy, which was then subjected to an analysis for the identification of latent symptom dimensions and dynamic factor modeling to determine the dynamics and module delivery order (see Fernandez et al., 2017; Fisher et al., 2019). Each individual patient received a specific delivery order of the modules based on their presenting symptoms and relationships among symptom dimensions. The module order for P007 is outlined in **Table 1**. For the present study, hypotheses were developed based on specific modules in order to investigate symptom changes throughout therapy, with expected changes to appear after a given module was delivered. Each hypothesis is outlined in **Table 2** and briefly reviewed below.

Prior to delivery of the first module, the participant underwent a 30-day pre-therapy assessment. Although this data collection was intended to reflect stationary processes, engagement with treatment providers can have distress-reducing effects for dysphoric individuals. Therefore, stationarity, the assumption that the mean, variance, and auto-correlation remain relatively stable over time, may be violated because of a process known as remoralization (Howard et al., 1993). Howard's remoralization theory (1993), proposes that psychotherapy entails sequential changes and the first change is an enhancement in the patient's sense of subjective well-being, which typically occurs before the process of formal psychotherapy beings (Howard et al., 1993). In this study, remoralization may have occurred during engagement with the phone surveys prior to the start of treatment delivery. While no formal hypotheses were made during this initial pre-therapy period, data from it is included in this study.

Therapy began with an emotional awareness and tracking module, and no significant changes were expected after this module, as the intervention was primarily targeting overall emotional awareness across both positive and negative affect domains. Because this relates to processes already in place from the pre-therapy assessment, we believed that – to the degree that self-monitoring may elicit symptomatic change – these changes would have already occurred.

Hypothesis 1: The second module delivered was a mindfulness module, and it was hypothesized that after the delivery of this module the participant would report reductions in feelings of restlessness and dwelling on the past on their survey responses. Extant work across a variety of domains has illustrated the success of a mindfulness-based approach in reducing physiological restlessness, including using a mindfulness-based stress reduction (MBSR) paradigm to reduce symptoms of restless leg syndrome (Bablas et al., 2016) and using a MBSR approach to reduce levels of pre-sleep arousal (Cincotta et al., 2010). Extant work in the literature has similarly demonstrated a negative correlation between mindfulness and rumination (Jain et al., 2007; Svendsen et al., 2017), hence we expected reductions in the survey item "dwelled on the past" following delivery of this module.

Hypothesis 2: The third module was an *in vivo* exposure intervention, aimed at facilitating habituation and inhibition of fear-conditioning. We hypothesized that the participant would report reductions in avoidance-related items (avoiding people, avoiding activities, procrastinating, and seeking reassurance) after delivery of this module, based on an abundance of previous work illustrating the effect of exposures on reducing avoidance (e.g., Foa and Kozak, 1986).

Hypothesis 3: The fourth module was a cognitive appraisal and reappraisals module, and it was hypothesized that after delivery the participant would report reductions in worry, and depression, and increases in positivity and contentedness, as reappraisals have previously been shown to reduced symptoms of stress and stress-related symptoms (Moore et al., 2008).

Hypothesis 4: Lastly, the fifth and final module was an emotion driven behaviors and emotional avoidance module. It was hypothesized that after this

module, the participant would report further reductions in feelings of worry and depression, as a large body of work has indicated reductions in anxiety and depressive symptoms following emotional exposures (Foa and Kozak, 1986; Hayes et al., 2005).

This quantitative, survey-based approach allows for an in-depth investigation and quantification of therapeutic changes across the course of a modularized individualized therapy in a single participant. We propose that this novel design will lend greater insight into individual symptom perturbations throughout the course of therapy.

MATERIALS AND METHODS

Setting and Participant

Data was obtained as part of an ongoing research study at the University of California, Berkeley. The participant, a female in the age range of 25–35 years old, was initially recruited to participate in a multi-phase personalized treatment study, for which she completed an initial diagnostic assessment, 14 weeks of individualized therapy treatment, and phone surveys from after the diagnostic assessment through the conclusion of therapy. Inclusion criteria included the following: principal diagnosis of either GAD or MDD; no concurrent psychosocial treatment; the participant had not previously received CBT; no medical conditions were identified as contributors to anxiety problems (e.g., hypoglycemia, thyroid problems); and mania and/or psychosis were absent. All procedures of the study were conducted under the approval of the University of California, Berkeley Institutional Review Board. Additional demographic information can be found in **Table 3**.

Procedure

The participant was initially recruited via paper and electronic advertisements placed in the community. After obtaining verbal consent, she completed a brief telephone screening, and based on this preliminary information, the participant was invited for an in-person structured clinical interview. They presented to the University of California, Berkeley's Department of Psychology building for clinical assessment, during which the anxiety and related disorders interview schedule for DSM-5 (Brown and Barlow, 2014), Hamilton Anxiety Rating Scale (Hamilton, 1959), and Hamilton Rating Scale for Depression (Hamilton, 1960) were administered by an advanced graduate student in clinical psychology. Results were reviewed with a supervising psychologist before the participant was invited to enroll in the study. After consent paperwork was reviewed, the participant took part in a two-phase study: Phase 1 required the completion of daily surveys to assess mood and anxiety disorder symptoms, and Phase 2 involved a 14-week cognitive-behavioral therapy treatment based on The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (Barlow et al., 2011; details described in Fisher, 2015). During Phase 2, the participant was instructed to continue the daily surveys in order to track progress in treatment. In both phases of the study, the individual received four text messages per day, each one containing a hyperlink to a web-based survey, resulting in four surveys per day. P007 completed surveys for a total of 42 days during Phase 1 (pre-therapy) and 140 days during Phase 2 (during therapy), with 158 and 437 total viable, non-missing observations for Phase 1 and Phase 2, respectively. The participant completed 96% of their surveys throughout Phase 1, and 78% throughout Phase 2, with an overall compliance rate of 82%.

TABLE 3 | Demographic information for participant 007.

Age range	25–35
Gender	Female
Marital status	Single
Race/ethnicity	African American
Education level	Some college

Measures

- *Anxiety and Related Disorders Interview Schedule for DSM-5 (ADIS-5; Brown and Barlow, 2014).* The ADIS-5 is a semi-structured clinical interview designed to diagnose current anxiety, mood, and related disorders according to new DSM-5 criteria. This updated version of the ADIS-5 builds upon previous versions, which all exhibited well-established reliability. The ADIS-5 demonstrates good-to-excellent interrater reliability for DSM-5 disorders (kappa ranging from 0.67 to 0.86, with the exception of dysthymia, kappa = 0.31) (DiNardo et al., 1994).

- *Hamilton Anxiety Rating Scale (HAM-A; Hamilton, 1959).* The HAM-A is a 14-item clinician administered scale that assesses severity of anxious symptoms. This scale provides a severity rating of each overarching symptom cluster on a scale from 0 (not present) to 4 (very severe). Research has shown that retest reliability for the HAM-A was good (intraclass correlation coefficient 0.86) and interrater reliability ranged from an intraclass correlation coefficient of 0.74–0.96 (Bruss et al., 1994). Construct validity has also been demonstrated in clinical samples (Beck and Steer, 1991).

- *Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960).* The HAM-D is a 13-item clinician administered scale developed to assess the severity of depressive symptoms. This scale provides severity ratings of each overarching symptom cluster on a scale from 0 (not present) to 4 (very severe/incapacitating). Internal consistency of the HAM-D ranges from adequate to good (0.73–0.81; Steer et al., 1987; Moras et al., 1992). HAM-D have also been shown to correlate significantly with self-report measures of depression in clinical samples (Steer et al., 1983).

- *Digital Assessment Daily Survey Items.* In addition to the extant DSM-5 GAD and MDD symptom criterion, daily surveys included four behavioral symptoms: (a) avoiding activities with possible negative outcomes, (b) preparing for possible negative outcomes, (c) procrastinating about taking action or

TABLE 4 | Daily digital assessment survey items.

item
1. To what degree have you felt energetic
2. To what degree have you felt enthusiastic
3. To what degree have you felt content
4. To what degree have you felt irritable
5. To what degree have you felt restless
6. To what degree have you felt worried
7. To what degree have you felt worthless or guilty
8. To what degree have you felt frightened or afraid
9. To what degree have you experienced loss of interest or pleasure
10. To what degree have you felt angry
11. To what degree have you felt hopeless
12. To what degree have you felt down or depressed
13. To what degree have you felt positive
14. To what degree have you felt fatigued
15. To what degree have you experienced muscle tension
16. To what degree have you had difficulty concentrating
17. To what degree have you felt accepted or supported
18. To what degree have you felt threatened; judged; or intimidated
19. To what degree have you dwelled on the past
20. To what degree have you avoided activities
21. To what degree have you sought reassurance
22. To what degree have you procrastinated
23. To what degree have you avoided people

decision making, and (d) seeking reassurance, as recent data have illustrated these behavioral symptoms to be significant features of GAD and MDD phenomenology (Beesdo-Baum and Knappe, 2012). The participant rated their experience of each symptom domain over the preceding 4 h (the surveys were randomized to roughly a 4-h interval schedule) on a 0–100 visual analog slider, with anchors of *not at all* and *as much as possible* anchored at the 0 and 100 positions, respectively. Survey items are presented in **Table 4**.

Statistical Approach to Quantitative Survey Items

Data for each specific module-based hypothesis was subset for individual hypotheses. For analyses done on each specified module, data from that module and throughout the rest of therapy was used in order to assess the degree to which each *specific* module, or intervention, was associated with changes in the specific hypothesized downstream variables. For example, the Exposures module was the 4th module for this participant, so data was subset from the date that module started through the last day of therapy for the exposure module-based hypotheses, and data prior to delivery of that module was excluded. To then assess changes in the participant's survey responses over the course of therapy, ordinary least squares (OLS) linear regression was employed to test response trajectories of each item. That is, separate linear regressions were applied to test the relationship between Time (coded in days) and changes in the dependent variable in question (e.g., worry, rumination, procrastination). The decision to use OLS regressions for trends over time instead of a multilevel approach was chosen due to much prior work in our lab indicating that one can handle intraindividual temporal dependence equally well with trend or AR components.

RESULTS

For OLS regression analyses, rows of data with missing surveys were excluded as a function of listwise deletion. In order to explore the presence of remoralization as predicted by Howard's theory, we tested the degree to which the client experienced reductions in negative-affect items and increases in positive-affect items. Thus, the data was subset into the portion of time prior to the start of therapy, and then separate linear regression analyses were employed to predict specific negative affect items as a function of time. Results are presented in **Table 5**. Models for negative affect items of "dwelled on the past," "felt worthless or guilty," "felt worried," "felt irritable," "experienced a loss of interest or pleasure," "felt threatened, judged, or intimidated," "felt down or depressed," and "felt angry" were all significant at the $p < 0.05$ level, indicating significant decreases in these negative affect items during the pre-therapy period.

Positive affect items were tested in the same manner, with separate simple linear regressions to test trajectories over time during the pre-therapy period. All positive affect items emerged as significant; during the pre-therapy period, feeling positive significantly increased over time [$\beta = 0.30$, $F(1,156) = 176.4$, $p < 0.01$, $R^2 = 0.53$], feeling energetic significantly increased over time [$\beta = 0.30$, $F(1,156) = 184$, $p < 0.01$, $R^2 = 0.54$], feeling enthusiastic significantly increased over time [$\beta = 0.32$, $F(1,156) = 173.9$, $p < 0.01$, $R^2 = 0.52$], feeling content significantly increased over time [$\beta = 0.33$, $F(1,156) = 154.6$, $p < 0.01$, $R^2 = 0.49$], and feeling accepted or supported significantly increased over time [$\beta = 0.17$, $F(1,156) = 29.69$, $p < 0.01$, $R^2 = 0.15$].

In order to test our first hypothesis, that participant would report reductions in restlessness and rumination following the mindfulness interventions, the data was subset to the time period after the module was delivered, and separate linear

TABLE 5 | Separate linear regression models for the trajectories of negative affect items as a function of time during the pre-therapy period.

	Negative Affect items over Time			
	p	SE	*t*	*p*
Dwelling on the past	−0.304	0.0275	−11.03	<0.00*
Worthless or guilty	−0.10	0.04	−2.72	0.01*
Hopelessness	−0.07	0.04	−1.083	0.07
Worry	−0.21	0.03	−6.69	<0.00*
Irritability	−0.23	0.03	−7.93	<0.00*
Loss of interest or pleasure	−0.19	0.04	−5.18	<0.00*
Threatened, judged, intimidated	−0.19	0.03	−6.18	<0.00*
Down or depressed	−0.18	0.03	−6.66	<0.00*
Anger	−0.10	0.04	2.79	0.01*

** indicates significance at $p < 0.05$.*

regression analyses were employed to test the trajectory of feelings of restlessness and dwelling on the past over time. Results were significant for restlessness, yet in the opposite direction then hypothesized; after delivery of the module, the participant reported significantly increased feelings of restlessness [$\beta = -0.08$, $F(1,167) = 8.43$, $p < 0.01$, $R^2 = 0.04$].

In order to test the hypothesis that P007 would report reductions in avoidance-related items (difficulty concentrating, avoiding people, avoiding activities, procrastinating, and seeking reassurance) following the exposure module, data was again subset for after the module was delivered, and separate linear regression analyses were employed to test the trajectory of each avoidance-related item over time. Significant reductions were observed for difficulty concentrating [$\beta = -0.16$, $F(1,122) = 7.82$, $p = 0.01$, $R^2 = 0.05$], avoiding people [$\beta = -0.39$, $F(1,122) = 34.93$, $p \leq 0.01$, $R^2 = 0.22$], and procrastinating [$\beta = -0.25$, $F(1,122) = 19.18$, $p \leq 0.01$, $R^2 = 0.13$]. Seeking reassurance significantly increased over time following the exposures module [$\beta = 0.18$, $F(1,122) = 7.88$, $p = 0.01$, $R^2 = 0.05$], which we hypothesize may be due to the participant conceptualizing reassurance seeking as a pro-social quality rather than a safety behavior (more in discussion).

In order to test the hypothesis that the participant would report reductions in worry and depression and increases in positivity and contentedness following the reappraisal module, the data was again subset for after the module was delivered, and separate linear regression analyses were employed to test the trajectory of worry, depression, positivity, and contentedness over time. No significant findings emerged from these models. However, it should be emphasized that this may be due to the embedded nature of these constructs in all modules and symptomatic experiences. Thus, as a secondary analysis, we examined the entire therapy section of the time series to assess the degree of change in depression, worry, positivity, and contentedness across the complete treatment period. Results indicate significant reductions in depression [$\beta = -0.04$, $F(1,434) = 28.88$, $p \leq 0.00$, $R^2 = 0.06$] and worry [$\beta = -0.07$, $F(1,434) = 95.23$, $p \leq 0.00$, $R^2 = 0.18$], and a significant increase in contentedness [$\beta = 0.02$, $F(1,434) = 4.69$, $p = 0.03$, $R^2 = 0.01$] over the complete treatment period.

In order to test the final hypothesis, that the participant would report further reductions in feelings of worry and depression, the data was again subset for after the final module was delivered, and separate linear regression analyses were employed to test the trajectory of worry and depression over time. Feelings of worry significantly decreased over time following delivery of this module [$\beta = -2.05$, $F(1,17) = 5.71$, $p = 0.03$, $R^2 = 0.21$]. No significant findings emerged for feelings of depression.

In addition to the module-specific hypotheses, we also employed separate linear regressions for each survey item as a function of time over the entire therapy period to the data. Results are depicted in **Table 6**. Significant reductions were observed for the following items: dwelling on the past, loss of interest or pleasure, procrastinated, and feeling worthless or guilty, hopeless, worried, irritable, threatened or judged, down or depressed, restless, fatigued, and energetic. Significant increases were observed for the following items: sought reassurance, feeling content, and feeling accepted or supported.

TABLE 6 | Separate linear regression models for the trajectories of all survey items as a function of time during the entire therapy period.

	Change in all survey items over entire therapy *y* Period			
	p	**SE**	***t***	***p***
Dwelling on the past	−0.20	0.01	−4.22	<0.00*
Worthless or guilty	−0.20	0.01	−4.22	<0.00*
Hopelessness	−0.40	0.01	−8.98	<0.00*
Worry	−0.42	0.01	−9.76	<0.00*
Irritability	−0.28	0.01	−6.04	<0.00*
Loss of interest or pleasure	0.06	0.01	1.27	0.21
Threatened or judged	−0.12	0.01	−2.51	0.01*
Down or depressed	−0.25	0.01	−5.37	<0.00*
Anger	−0.02	0.01	−0.32	0.75
Frightened or afraid	−0.01	0.02	−0.15	0.88
Restless	−0.16	0.01	−3.44	0.00*
Fatigued	−0.18	0.01	−3.91	0.00*
Muscle tension	−0.08	0.01	−1.69	0.09
Difficulty concentrating	−0.41	0.01	−9.34	<0.00*
Avoided activities	0.01	0.01	0.29	0.78
Sought reassurance	0.23	0.01	4.94	<0.00*
Procrastinated	−0.42	0.01	−9.62	<0.00*
Avoided people	0.29	0.01	6.25	<0.00*
Energetic	−0.11	0.01	−2.40	0.02*
Enthusiastic	0.08	0.00	1.62	0.11
Accepted or supported	0.10	0.01	2.20	0.03*
Positive	0.01	0.01	0.15	0.88
Content	0.10	0.01	2.17	0.03*

** indicates significance at $p < 0.05$.*

DISCUSSION

The current study uses an observational N-of-1 case study design on intensive repeated digital measures data to investigate the nature of change throughout the course of a modularized therapy. Analysis of the intensive repeated measures data revealed improvements in the pre-therapy period, providing additional evidence for Howard's remoralization theory (1993). This theory states that the first of three sequential changes throughout the course of psychotherapy is an improvement of the patient's sense of well-being (*remoralization*), and typically occurs quickly in response to setting up an appointment, getting advice, and other occurrences that tend to happen prior to and at the onset of psychotherapy, including the work done within the first three sessions. Previous work in a variety of treatment settings has found support of this theory, including early gains in optimism very early in a depression treatment study (Schwartz, 1997), and early increases of subjective well-being, followed by reduction of symptom distress, in a study applying the phase model to short-term psychodynamic psychotherapy (Hilsenroth et al., 2001). In the present study, during the first 30-day monitoring period, the participant exhibited significant decreases in negative affect and accompanying significant increases in positive affect. This suggests that the survey paradigm employed in the present study—which might be considered a form of self-monitoring—may serve as a first-step intervention, capable

of improving well-being. Extant work has also found support for self-monitoring as a first step in behavior change (Spates and Kanfer, 1977). This understanding that improvement in symptoms can result solely from self-monitoring is important to the first therapy sessions with a patient, and may provide evidence that ecological momentary assessment techniques used prior to therapy may have an intrinsic therapeutic effect in and of themselves.

Pertaining to the module hypotheses, support was found for some, but not all, of our initial hypotheses. After the mindfulness module, it was predicted that restlessness would decrease, yet findings supported change in the opposite direction than hypothesized; after delivery of the module, the participant reported significantly *increased* feelings of restlessness. However, we should note that this increase accounted for only 4% of the variance in restlessness, leaving 96% unexplained. Thus, this increase may be secondary to other, predominant phenomena. Nevertheless, one explanation for this finding is that we did not assess whether or not the participant continued to use the mindfulness exercises following this module, and perhaps they did not incorporate the mindfulness work into any more of their treatment. A second explanation may be that restlessness can occur as one tries to quiet the mind in the beginning stages of meditation; a study investigating the effects of an MBSR on nurse stress and burnout similarly found that, while the program was overall effective at reducing stress, when the participants were asked about the challenges of the program the most common response was restlessness, which was mentioned by 52% of the nurses, with comments such as, "my mind is everywhere," "my body feels restless," and "it's so hard to concentrate!" (Cohen-Katz et al., 2004). Perhaps instituting a more frequent mindfulness practice following this module could mitigate these issues in future work.

Our hypotheses following the exposure module were supported. Significant reductions were observed in difficulty concentrating, avoiding activities, and procrastinating following delivery of the module, illustrating actual changes in the hypothesized downstream targets of the intervention. Of note, endorsement of reassurance seeking significantly *increased* over time following the exposures module, however, we hypothesize that this may be due to the participant conceptualizing reassurance seeking as a pro-social quality rather than a safety behavior. Support was not found for our hypothesis following delivery of the cognitive appraisal and reappraisals module, which we believe may be due to the fact that the predicted targets of worry, depression, positivity, and contentedness were too broad, and that future work should include more specific questions aimed at assessing the success of this intervention (e.g., questions aimed at assessing catastrophizing, overconfidence, and flexible thinking). As noted in the results, an exploratory analysis that examined changes in these constructs over the complete treatment period revealed significant change—in the expected direction—for each, demonstrating that these variables were affected by the treatment generally. Lastly, our hypothesis that worry would decrease after the emotion driven behaviors and emotional avoidance module was supported, illustrating that targeting avoidance in this one individual subsequently improved their worry over the course of this treatment.

In addition to the individual hypothesized changes and subsequent findings, the overall treatment prescription for this individual participant was successful at reducing her symptoms of major depressive disorder and generalized anxiety disorder. The participant began therapy with HAM-A and HAM-D scores of 17 and 11, respectively, indicating that the participant was in the mild severity range of both depressive and anxious symptomatology. One week after the participant completed therapy, her scores for the HAM-A and HAM-D were 4 and 3, respectively, indicating that she fell in the normal ranges for both assessments. Furthermore, separate linear regressions for each survey items as a function of time were conducted over the entire therapy period. The results (**Table 6**) show significant reductions for the majority of negative affect items, and significant increases for many positive affect items. The participant therefore, upon self-report of the surveys, felt a reduction in negative affect and increase in positive affect throughout therapy.

Limitations

Limitations of the present study include the use of a single participant, as well as utilizing a method of assessing change per specific module. The changes may have been due to overall changes across the entire therapy period and not due to the specific intervention taking place. Our method of subsetting the data attempted to minimize this from occurring. Other time series designs, including multiple baseline and interrupted time series, may be useful for addressing these questions in the future; since we were performing secondary analyses to a primary study that did not employ these types of designs, we were not able to utilize one here.

Future Directions

Idiographic approaches to treatment research are growing in popularity. In order to develop more efficient and targeted interventions, researchers and clinicians alike have called for idiographic hypothesis testing to investigate mechanisms of change within individuals over the course of a treatment period. This approach is not new to clinicians, as evidenced by existing conceptualization methods that integrate different modalities to meet the needs of the presenting problems of the individual client, including case formulation driven approaches for cognitive behavior therapy (Persons, 2012), and psychotherapy integration approaches (Stricker and Gold, 1996). Many recent research groups have demonstrated the utility of such approaches, primarily investigating quantitative changes to investigate whether alterations on certain treatment parameters or symptoms predict subsequent changes over time (Brown et al., 2019). This work provides yet another important avenue by which to investigate treatment changes idiographically, serving as a model for a quantitative approach.

It is important to note that while conducted on a single individual, this work was analyzed ideographically, and thus findings are not meant to generalize to other individuals, but rather are meant to illustrate how idiographic work such as

this can be utilized perhaps for prediction models (i.e., used to improve prediction of response in the future for that specific N-of-1 unit). Previous inferences made from psychological and medical research (e.g., treatment development, personality research) are typically drawn from statistical tests conducted on aggregated, group-level data, with the implicit assumption that group-level inferences, or findings, will generalize to the individuals who comprise those groups. Often overlooked in this assumption is the problem of ergodicity. Broadly speaking, ergodicity refers to a process by which individual variation can be inferred from group-level data. Historically, the field of psychology has assumed that most processes are ergodic in nature. But this assumption is not always upheld, and recent work by Fisher et al. (2018) found that in self-reported emotion data (and other types of data) the processes were not ergodic. In fact, they found that the variance at the individual level of analysis was up to four times larger than at the group level. Assuming ergodicity for non-ergodic processes leads to misinterpretations of findings that stall the pace of progress in the field. Idiographic work such as this can help to mitigate this gap and provide a groundwork for personalized prediction models.

Furthermore, as noted in the introduction, some researchers believe that therapeutic elements of therapy are due to common factors across all techniques, not specific interventions themselves. Our approach, and supporting evidence, however illustrates the ability to discern specific effects attributed to certain interventions.

The methods employed and findings indicated here also have the potential to aid psychotherapists in routine care by helping them assess whether their interventions are working. By employing routine progress monitoring, whether through daily phone surveys or other methods such as one-time daily diaries, psychotherapists can visualize reductions in symptoms over time and assess whether they are targeting the symptoms they hope, and thus conclude that their prescribed treatment course is effective, or if they need to change course. Methods to collect intensive repeated measures prior to therapy delivery have already been employed in naturalistic settings such as a University health center (e.g., the UC Berkeley Psychology Clinic). These therapists could continue to collect similar data and employ the methods outlined here in order to assess the efficacy and accuracy of their interventions.

CONCLUSION

In conclusion, future work should continue to utilize digital health tools to administer quantitative surveys, such as this, as well as other methodologies (e.g., multiple baselines and interrupted time series designs) in order to better understand the nature of change in psychotherapeutic treatment.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Committee for Protection of Human Subjects (CPHS) by the Office for Protection of Human Subjects (OPHS) at the University of California, Berkeley. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the work presented in this manuscript. AA and AF designed the experiment and subsequent analyses. LS helped with hypotheses and data organization. AA ran the statistical analyses, analyzed the output data, created the tables and figures, and contributed to the writing of the manuscript.

FUNDING

Publication made possible in part by support from the Berkeley Research Impact Initiative (BRII) sponsored by the UC Berkeley Library.

REFERENCES

Bablas, V., Yap, K., Cunnington, D., Swieca, J., and Greenwood, K. M. (2016). Mindfulness-based stress reduction for restless legs syndrome: a proof of concept trial. *Mindfulness* 7, 396–408. doi: 10.1007/s12671-015-0457-9

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., et al. (2011). *Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide*. New York, NY: Oxford University Press.

Beck, A. T., and Steer, R. A. (1991). Relationship between the Beck anxiety inventory and the Hamilton anxiety rating scale with anxious outpatients. *J. Anxiety Disord.* 5, 213–223. doi: 10.1016/0887-6185(91)90002-B

Beesdo-Baum, K., and Knappe, S. (2012). Developmental epidemiology of anxiety disorders. *Child Adolesc. Psychiatr. Clin. North Am.* 21, 457–478. doi: 10.1016/j.chc.2012.05.001

Brown, C. L., Bosley, H. G., Kenyon, A. D., Chen, K. H., and Levenson, R. W. (2019). An idographic statistical approach to clinical hypothesis testing for routine psychotherapy: a case study. *Behav. Res. Therapy* 118, 43–53. doi: 10.1016/j.brat.2019.03.014

Brown, T. A., and Barlow, D. H. (2014). *Anxiety and Related Disorders Interview Schedule for DSM-5: Clinician Manual*. Oxford: Oxford University Press.

Bruss, G. S., Gruenberg, A. M., Goldstein, R. D., and Barber, J. P. (1994). Hamilton anxiety raing scale interview guide: joint interview and test-retest methods for interrater reliability. *Psychiatry Res.* 53, 191–202. doi: 10.1016/0165-1781(94)90110-4

Cincotta, A. L., Gehrman, P., Gooneratne, N. S., and Baime, M. J. (2010). The effects of a mindfulness-based stress reduction programme on pre-sleep cognitive arousal and insomnia symptoms: a pilot study. *Stress Health* 27, 299–305.

Cohen-Katz, J., Wiley, S. D., Capuano, T., Baker, D. M., and Shapiro, S. (2004). The effects of mindfulness-based stress reduction on nurse stress and burnout:

a quantitative and qualitative study. *Holist. Nurs. Pract.* 18, 302–308. doi: 10.1097/00004650-200411000-00006

DiNardo, P. A., Barlow, D. H., and Brown, T. A. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime Version (ADIS-IV-L)*. New York, NY: Oxford University Press.

Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Thompson-Hollands, J., Carl, J., et al. (2012). Unified protocol for transdiagnostic treatment of emotional disorders: a randomized controlled trial. *Behav. ther.* 43, 666–678. doi: 10.1016/j.beth.2012.01.001

Fernandez, K. C., Fisher, A. J., and Chi, C. (2017). Development and initial implementation of the dynamic assessment treatment algorithm (DATA). *PLoS One* 12:e0178806. doi: 10.1371/journal.pone.0178806

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: implications for personalized care. *J. Consult. Clin. Psychol.* 83, 825–836. doi: 10.1037/ccp0000026

Fisher, A. J., Bosley, H. G., Fernandez, K., Reeves, J. W., Diamond, A. E., Soyster, P. D., et al. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behav. Res. Therapy* 116, 67–79. doi: 10.1016/j.brat.2019.01.010

Fisher, A. J., Medaglia, J. D., and Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *PNAS* 115, E6106–E6115. doi: 10.1073/pnas.1711978115

Foa, E. B., and Kozak, M. J. (1986). Emotional processing of fear: exposure to corrective information. *Psychol. Bull.* 99, 20–35. doi: 10.1037/0033-2909.99.1.20

Frank, J. D., and Frank, J. B. (1991). *Persuasion and Healing: A Comparative Study of Psychotherapy*, 3rd Edn. Baltimore: Johns Hopkins University Press.

Galatzer-Levy, I. R., and Bryant, R. A. (2013). 636,120 Ways to Have Posttraumatic Stress Disorder. *Perspect. Psychol. Sci.* 8, 651–662. doi: 10.1177/1745691613504115

Hamilton, M. (1959). The assessment of anxiety states by rating. *Br. J. Med. Psychol.* 32, 50–55. doi: 10.1111/j.2044-8341.1959.tb00467.x

Hamilton, M. (1960). A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23:56.

Hayes, A. M., Beevers, C., Feldman, G., Laurenceau, J. P., and Perlman, C. A. (2005). Avoidance and emotional processing as predictors of symptom change and positive growth in an integrative therapy for depression. *Int. J. Behav. Med.* 12, 111–122. doi: 10.1207/s15327558ijbm1202_9

Hilsenroth, M. J., Ackerman, S. J., and Blagys, M. D. (2001). Evaluating the phase model of change during short-term psychodynamic psychotherapy. *Psychotherapy Res.* 11, 29–47. doi: 10.1080/713663851

Howard, K. I., Lueger, R. J., Maling, M. S., and Martinovich, Z. (1993). A phase model of psychotherapy outcome: causal mediation of change. *J. Consult. Clin. Psychol.* 61, 678–685. doi: 10.1037/0022-006x.61.4.678

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167, 748–751. doi: 10.1176/appi.ajp.2010.09091379

Jain, S., Shapiro, S. L., Swanick, S., Roesch, S. C., Mills, P. J., Bell, I., et al. (2007). A randomized controlled trial of mindfulness meditation versus relaxation training: effects on distress, positive states of mind, rumination, and distraction. *Ann. Behav. Med.* 33, 11–21. doi: 10.1207/s15324796abm3301_2

Johnsen, T. J., and Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. *Psychol. Bull.* 141, 747–768. doi: 10.1037/bul0000015

Luborsky, L., Singer, B., and Luborsky, L. (1975). Comparative studies of psychotherapies: is it true that "everyone has won and all must have prizes"? *Arch. Gen. Psychiatry* 32, 995–1008.

Moore, S. A., Zoellner, L. A., and Mollenholt, N. (2008). Are expressive suppression and cognitive reappraisal associated with stress-related symptoms? *Behav. Res. Therapy* 46, 993–1000. doi: 10.1016/j.brat.2008.05.001

Moras, K., Di Nardo, P. A., and Barlow, D. H. (1992). Distinguishing anxiety and depression: reexamination of the reconstructed Hamilton scales. *Psychol. Assess.* 4:224. doi: 10.1037/1040-3590.4.2.224

Persons, J. B. (2012). *The Case Formulation Approach to Cognitive-Behavior Therapy*. New York, NY: The Guilford Press.

Prochaska, J. O., and DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *J. Consult. Clin. Psychol.* 51, 390–395. doi: 10.1037/0022-006x.51.3.390

Schwartz, R. M. (1997). Consider the simple screw: cognitive science, quality improvement, and psychotherapy. *J. Consult. Clin. Psychol.* 65, 970–983. doi: 10.1037/0022-006x.65.6.970

Spates, C. R., and Kanfer, F. H. (1977). Self-monitoring, self-evaluation, and self-reinforcement in children's learning: a test of a multistage self-regulation model. *Behav. Therapy* 8, 9–16. doi: 10.1016/s0005-7894(77)80115-9

Steer, R. A., Beck, A. T., Riskind, J. H., and Brown, G. (1987). Relationships between the Beck depression inventory and the Hamilton psychiatric rating scale for depression in depressed outpatients. *J. Psychopathol. Behav. Assess.* 9, 327–339. doi: 10.1007/bf00964561

Steer, R. A., McElroy, M. G., and Beck, A. T. (1983). Correlates of self-reported and clinically assessed depression in outpatient alcoholics. *J. Clin. Psychol.* 39, 144–149. doi: 10.1002/1097-4679(198301)39:1<144::aid-jclp2270390128>3.0.co;2-m

Stricker, G., and Gold, J. R. (1996). Psychotherapy integration: an assimilative, psychodynamic approach. *Clin. Psychol.* 3, 47–58. doi: 10.1111/j.1468-2850.1996.tb00057.x

Svendsen, J. L., Kvernenes, K. V., Wiker, A. S., and Dundas, I. (2017). Mechanisms of mindfulness: rumination and self-compassion. *Nordic Psychol.* 69, 71–82. doi: 10.1080/19012276.2016.1171730

Wampold, B. E. (2001). *The Great Psychotherapy Debate: Model, Methods, and Findings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wampold, B. E. (2007). Psychotherapy: the humanistic (and effective) treatment. *Am. Psychol.* 62, 857–873. doi: 10.1037/0003-066x.62.8.857

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

OPINION
published: 29 April 2020
doi: 10.3389/fpubh.2020.00144

The Need for Personalized Approaches to Microbiome Modulation

*Nita Jain**

Independent Researcher, Lilburn, GA, United States

Keywords: microbiome, personalized medicine, gut ecology, metagenomics, FMT, host-microbe interactions

INTRODUCTION

The human microbiome has been a topic of interest for both research and clinical applications in recent decades. However, the considerable gut microbial variation observed across human populations poses a challenge in terms of targeted interventions. Diet (1), exercise (2), age (3), ancestry (4), and geographic latitude (5) all influence the composition of the gut microbiome. The individualized nature of microbiome compositions makes responses to modulatory interventions, including probiotics, prebiotics, and fecal microbiota transplantation, subject-specific (6).

MICROBIOME VARIABILITY AND HOST CHARACTERISTICS

Host Characteristics Influence Individual Variation in Gut Microbiota Composition

Host features govern the types of niches available for occupation such that only microbes adapted to host ecological conditions can successfully colonize. For this reason, autochthonous strains are more likely to possess the traits necessary to successfully persist in gut ecosystems, accounting for the failure of most allochthonous probiotic strains to colonize. Genes that encode for traits such as mucosal adherence and acid resistance can confer greater ecological fitness in a host environment (6). Habitat filters are influenced by various factors, including a host's genetics, metabolism, diet, and environment, and select for microbes with common traits, leading to phylogenetic underdispersion. Out of the hundreds of phyla encountered in terrestrial and aquatic ecosystems, the human gut is dominated by only five, illustrating the impact of selection through habitat filters (7). Certain gene polymorphisms can differentially impact intestinal microbiota composition through provision of adhesion sites and growth substrates such as secreted glycans (6).

The *FUT2* gene encodes for fucosyltransferase, which is responsible for the synthesis of the H antigen that serves as the precursor to the ABH histo-blood group antigens in mucus and other bodily secretions. Individuals that are homozygous for any non-functional *FUT2* allele are known as non-secretors and will not present ABH antigens on epithelial cell surfaces whereas individuals carrying at least one functional *FUT2* allele will express ABH antigens on intestinal mucosal surfaces (8). Secretor status determines the expression of fucosylated glycan epitopes in the human intestine, and the *FUT2* non-secretor phenotype has been linked to alterations in the gut microbiome in the form of reduced bifidobacterial diversity, richness, and abundance (9). However, large-scale studies have not been able to replicate these reports (10, 11). Collecting and analyzing additional metadata on diet and lifestyle habits may help resolve some of the discrepancies observed between different studies. A murine study, for example, showed that *FUT2* secretor status-associated changes in intestinal microbiota composition are diet-dependent (12). Citizen science initiatives, such as the American Gut Project, can help evaluate the effects of interactions among host genetics, diet, and environment on microbiome composition on larger

OPEN ACCESS

Edited by:
Eric Jay Daza,
Clarify Health, United States

Reviewed by:
Justine W. Debelius,
Karolinska Institutet (KI), Sweden
George Weinstock,
Jackson Laboratory for Genomic Medicine, United States

***Correspondence:**
Nita Jain
nitajain8@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *31 January 2020*
Accepted: *06 April 2020*
Published: *29 April 2020*

Citation:
Jain N (2020) The Need for Personalized Approaches to Microbiome Modulation. Front. Public Health 8:144. doi: 10.3389/fpubh.2020.00144

scales. Analyzing patient sequencing data along with self-reported metadata may also help elucidate possible associations between the non-secretor phenotype and increased risk of certain diseases, including Crohn's disease, type 1 diabetes, vaginal candidiasis, and urinary tract infections (13).

Efficacy of Microbiome Modulatory Interventions Depend on Baseline Host Characteristics

An individual's baseline microbiota composition determines the types of dietary fibers that may be fermented to produce short-chain fatty acids (SCFAs) such as butyrate. The microbiomes of some individuals may be capable of fermenting pectin to produce SCFAs while the microbiomes of others may require inulin to achieve the same effect (14). The inherent heterogeneity among the gut microbiota of healthy humans differentially affects functional degradation of fibers and the SCFAs produced in response (15). Dietary fiber may also improve glucose homeostasis in a subset of patients through colonic production of SCFAs; acetate and butyrate have been shown to stimulate production of glucagon-like peptide-1 (GLP-1) and peptide YY (PYY), which in turn stimulate insulin secretion. A study examining dietary fiber interventions in type 2 diabetes patients found that the microbiomes of positive responders possessed more genes for plant fiber utilization while the microbiomes of negative responders were more enriched in genes for utilization of animal carbohydrates derived from mucin (16). Data on a patient's microbiome composition may therefore be able to inform personalized dietary intervention strategies targeted toward increased SCFA production in order to ameliorate disease phenotypes.

Before targeting colonic SCFA production, it may be judicious to first evaluate a patient's immune system activity in order to prevent potential adverse effects. A study conducted by a research group from the Massachusetts Institute of Technology used microphysiological systems to demonstrate that SCFAs can either ameliorate or exacerbate ulcerative colitis disease severity depending on the activation state of CD4 T cells. In the setting of T cell-mediated acute inflammation, SCFAs led to further gut barrier disruption and hepatobiliary damage (17). Observations from a randomized controlled trial (RCT) on the use of fecal microbiota transplantation (FMT) in ulcerative colitis patients further elucidate the interface between immune system activation and host response to the resident gut microbiota, as individuals on immunosuppressive therapy were more likely to benefit from FMT compared to patients who were not on immunosuppressive therapy (46 vs. 15%) (18). Thus, immune-modulating strategies may theoretically help facilitate successful strain engraftment.

Individualized Microbiome Features Govern Strain Engraftment Efficacy

Exogenous species are more likely to successfully engraft through FMT when related species are already present (19). In light of this knowledge, reducing recipient microbial load with antibiotics may hinder successful engraftment of the donor microbiota (20). A study evaluating the effects of Rifaximin pre-treatment compared to FMT alone for the treatment of ulcerative colitis reported no significant difference between groups in terms of disease activity (21). Changes to gut microbiota composition caused by colonic lavage or laxative use may also have unintended effects on FMT efficacy (22). While Li et al. reported that new strains transfer more easily than new species, Stecher et al. similarly described a "like will to like" principle, suggesting that successful colonization of both pathogenic and commensal strains is dictated by prior establishment of related species (23). A research group from the University of Milan observed a significant increase in Proteobacteria abundance and a significant decrease in Firmicutes abundance at the phylum level immediately after colon cleansing (24). Considering the implication of high Proteobacteria abundance in various human diseases (25) and its possible utility as a marker for dysbiosis (26), bowel preparation procedures prior to fecal transplantation may negatively impact FMT efficacy by preferentially facilitating the engraftment of potential pathobionts, but this possibility would warrant further research.

While conspecific strains exhibit greater colonization success than new species (19), ecology theory conversely predicts that competition among phylogenetically related strains will be greater as a result of trait similarity and niche overlap. Consequently, the presence of certain strains at baseline may prevent the colonization of other strains within the same species due to competitive exclusion, or phylogenetic limiting. A study examining strain engraftment in the human gut found that *B. longum* subsp. *longum* AH1206 was more likely to successfully engraft in hosts who did not already harbor native *B. longum* strains, suggesting niche availability as a limiting factor for persistence. However, while baseline *B. longum* abundance generally inversely correlated with AH1206 persistence, this pattern did not hold true for all subjects. Since traits that define niches are not always phylogenetically conserved within species, researchers also evaluated metagenomic data in order to assess differences in functional microbiome composition between persisters and non-persisters. Specifically, AH1206 was able to engraft in a subset of subjects whose microbiome lacked certain carbohydrate utilization genes characteristic of *B. longum* strains (27). In regards to niche availability, functional gene distinctions may be more predictive of exclusion effects than phylogenetic considerations under certain environmental conditions, as horizontal gene transfer can facilitate the emergence of functionally similar bacteria in phylogenetically distinct taxa (28). The factors that determine whether habitat filtering or competitive exclusion takes precedence will likely include contextual and taxonomic considerations (6).

Considerations of Host-Microbe Coevolution May Enhance Efficacy of FMT

Whereas sharing a joint evolutionary history is a characteristic of autochthony (6), colonization of strains that did not evolve with a given host may result in hologenomic disequilibrium and cause negative health effects in the form of certain increased disease risks. For example, the presence of a specific strain of

Helicobacter pylori in a host that did not coevolve with that microorganism was associated with an increased risk of gastric cancer (29). Additionally, the equilibrium of hunter-gatherer microbiota may be disrupted after exposure to a Western diet and switch to a state of dysbiosis. Thus, classifications of "commensal" and "pathogenic" may be relative and dependent on evolutionary history among other considerations, underlying the importance of stratifying stool donors by ethnogeographic and social factors (30). Furthermore, most metagenomic studies sample both study and control groups from the same population, which is exposed to similar environmental conditions common to urban lifestyles in developed countries, suggesting that many "healthy" subjects may simply be in a prodromal period. For these reasons, FMT donor screening on the basis of pathogen testing alone may be insufficient to prevent potential adverse outcomes in recipients, and more rigorous screening may include clinical laboratory data as well as metagenomic analyses.

Clinical Outcomes in Response to FMT May Be Donor-Dependent

Microbial communities that exhibit greater phylogenetic diversity and evenness are considered more resilient to invasion. As a result, fecal microbiota transplantation results in a higher degree of engraftment in patients exhibiting severe microbiome perturbation, such as that encountered in the setting of active *C. difficile* infection, compared to patients with metabolic syndrome (19). High levels of genetic diversity in an incoming community increase the chance of successful invasions, as some organisms will likely possess the adaptations necessary to thrive. In particular, high microbial richness has been shown to be one of the most important factors in determining FMT outcome (31). While most literature on FMT has focused on bacteria as the therapeutically active agent, recent research suggests that phages may play a more significant role in disease resolution than previously realized. Donor-derived phages may target indigenous species of the host microbiome, expanding niche availability for incoming microbes. Zuo et al. reported that treatment response to FMT was associated with bacteriophage transfer involving *Caudovirales* taxa (32).

Based on the condition being treated, donors with certain microbiota profiles may be more effective than others. For example, fecal microbiota enriched with *Bifidobacterium* has been shown to be a positive predictor for the efficacy of FMT in IBS patients. Donor material rich in *Bifidobacterium* may stimulate the growth and expansion of undetectable strains in recipient microbiota to match the level of diversity observed in donor microbiota (33). In this manner, the efficacy of FMT likely depends upon stimulation of recipient microbiota by donor material rather than literal "transplantation" of donor microbiota. Similarly, an RCT examining the effects of FMT in patients with ulcerative colitis found that remission among responders was associated with increases in bacterial abundances of *Clostridium* clusters IV and XIVa (34). Another RCT involving UC patients found that FMT treatment success in response to one particular donor, donor B, was 39% vs. 10% for other donors, providing further evidence that clinical outcomes may be donor-dependent. The two most commonly used donors in the study, donor A and donor B, displayed significant differences in taxonomic composition, including enrichment in the family Lachnospiraceae and the genus *Ruminococcus* in donor B and enrichment in the order Clostridiales and the genera *Escherichia* and *Streptococcus* in donor A (18). The donor-dependent nature of FMT efficacy may help explain the disparity in clinical results observed among different studies conducted on a specific condition.

CONCLUSION

Given the considerable amount of variation observed in human populations, bridging the gap between microbiome research and clinical applications may allow for more targeted, personalized recommendations based on diet, ancestry, geography, and physiology as well as microbial phylogenetic, metagenomic, and metabolic considerations.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

1. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* (2014) 505:559–63. doi: 10.1038/nature12820
2. Grosicki GJ, Durk RP, Bagley JR. Rapid gut microbiome changes in a world-class ultramarathon runner. *Physiol Rep.* (2019) 7:e14313. doi: 10.14814/phy2.14313
3. Salazar N, Arboleya S, Fernández-Navarro T, de Los Reyes-Gavilán CG, Gonzalez S, Gueimonde M. Age-associated changes in gut microbiota and dietary components related with the immune system in adulthood and old age: a cross-sectional study. *Nutrients.* (2019) 11:1765. doi: 10.3390/nu11081765
4. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell.* (2014) 159:789–99. doi: 10.1016/j.cell.2014.09.053
5. Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. *Biol Lett.* (2014) 10:20131037. doi: 10.1098/rsbl.2013.1037
6. Walter J, Maldonado-Gómez MX, Martínez I. To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. *Curr Opin Biotechnol.* (2018) 49:129–39. doi: 10.1016/j.copbio.2017.08.008
7. Catford JA, Jansson R, Nilsson C. Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Divers Distribut.* (2009) 15:22–40. doi: 10.1111/j.1472-4642.2008.00521.x
8. Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, et al. A natural history of FUT2 polymorphism in humans. *Mol Biol Evol.* (2009) 26:1993–2003. doi: 10.1093/molbev/msp108
9. Wacklin P, Mäkivuokko H, Alakulppi N, Nikkil,ä J, Tenkanen H, Räbinä J, et al. Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. *PLoS ONE.* (2011) 6:e20113. doi: 10.1371/journal.pone.0020113
10. Davenport ER, Goodrich JK, Bell JT, Spector TD, Ley RE, Clark AG. ABO antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. *BMC Genomics.* (2016) 17:941. doi: 10.1186/s12864-016-3290-1

11. Turpin W, Bedrani L, Espin-Garcia O, Xu W, Silverberg MS, Smith MI, et al. FUT2 genotype and secretory status are not associated with fecal microbial composition and inferred function in healthy subjects. *Gut Microbes.* (2018) 9:357–68. doi: 10.1080/19490976.2018.1445956
12. Kashyap PC, Marcobal A, Ursell LK, Smits SA, Sonnenburg ED, Costello EK, et al. Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc Natl Acad Sci USA.* (2013) 110:17059–64. doi: 10.1073/pnas.1306070110
13. Wacklin P, Tuimala J, Nikkilä J, Sebastian T, Mäkivuokko H, Alakulppi N, et al. Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. *PLoS ONE.* (2014) 9:e94863. doi: 10.1371/journal.pone.0094863
14. Gurry T, HST Microbiome Consortium, Gibbons SM, Nguyen LTT, Kearney SM, Ananthakrishnan A, et al. Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort. *Sci Rep.* (2018) 8:12699. doi: 10.1038/s41598-018-30783-1
15. Gurry T, Nguyen LTT, Yu X, Alm EJ. Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota. *bioRxiv [preprint].* (2020). doi: 10.1101/2020.01.17.910638
16. Zhao L, Zhang F, Ding X, Wu G, Lam YY, Wang X, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science.* (2018) 359:1151–56. doi: 10.1126/science.aao5774
17. Trapecar M, Communal C, Velazquez J, Maass CA, Huang YJ, et al. Gut-Liver physiomimetics reveal paradoxical modulation of IBD-related inflammation by short-chain fatty acids. *bioRxiv [preprint].* (2019). doi: 10.1101/706812
18. Moayyedi P, Surette MG, Kim PT, Libertucci J, Wolfe M, Onischi C, et al. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology.* (2015) 149:102–9.e6. doi: 10.1053/j.gastro.2015.04.001
19. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science.* (2016) 352:586–9. doi: 10.1126/science.aad8852
20. Manichanh C, Reeder J, Gibert P, Varela E, Llopis M, Antolin M, et al. Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res.* (2010) 20:1411–9. doi: 10.1101/gr.107987.110
21. El-Nachef N, Kassam Z, Piceno YM, Ablaza, A.-J., Zydek M, et al. Does Rifaximin prior to fecal microbiota transplantation improve clinical outcomes compared to microbiome restoration alone in ulcerative colitis? A cohort study evaluating the impact of non-absorbable antibiotic pretreatment. *Gastroenterology.* (2017) 152:S1008–9. doi: 10.1016/S0016-5085(17)33421-2
22. Tropini C, Moss EL, Merrill BD, Ng KM, Higginbottom SK, Casavant EP, et al. Transient osmotic perturbation causes long-term alteration to the gut microbiota. *Cell.* (2018) 173:1742–54.e17. doi: 10.1016/j.cell.2018.05.008
23. Stecher B, Chaffron S, Käppeli R, Hapfelmeier S, Freedrich S, Weber TC, et al. Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathogens.* (2010) 6:e1000711. doi: 10.1371/journal.ppat.1000711
24. Drago L, Toscano M, de grandi R, Casini V, Pace F. Persisting changes of intestinal microbiota after bowel lavage and colonoscopy. *Eur J Gastroenterol Hepatol.* (2016) 28:532–7. doi: 10.1097/MEG.0000000000000581
25. Rizzatti G, Lopetuso L, Gibiino G, Binda C, Gasbarrini A. Proteobacteria: a common factor in human diseases. *Bio Med Res Int.* (2017) 2017:9351507. doi: 10.1155/2017/9351507
26. Shin NR, Whon T, Bae JW. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* (2015) 33:496–503. doi: 10.1016/j.tibtech.2015.06.011
27. Maldonado-Gómez MX, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van Sinderen D, et al. Stable engraftment of Bifidobacterium longum AH1206 in the human gut depends on individualized features of the resident microbiome. *Cell Host Microbe.* (2016) 20:515–26. doi: 10.1016/j.chom.2016.09.001
28. Boto L. Horizontal gene transfer in evolution: facts and challenges. *Proc R Soc B Biol Sci.* (2010) 277:819–27. doi: 10.1098/rspb.2009.1679
29. Kodaman N, Pazos A, Schneider BG, Piazuelo MB, Mera R, Sobota RS, et al. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci USA.* (2014) 111:1455–60. doi: 10.1073/pnas.1318093111
30. Tyakht AV, Alexeev DG, Popenko AS, Kostryukova ES, Govorun VM. Rural and urban microbiota: to be or not to be? *Gut microbes.* (2014) 5:351–6. doi: 10.4161/gmic.28685
31. Kump P, Wurm P, Gröchenig HP, Wenzl H, Petritsch W, Halwachs B, et al. The taxonomic composition of the donor intestinal microbiota is a major factor influencing the efficacy of faecal microbiota transplantation in therapy refractory ulcerative colitis. *Aliment Pharmacol Ther.* (2018) 47:67–77. doi: 10.1111/apt.14387
32. Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, et al. Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut.* (2018) 67:634–43. doi: 10.1136/gutjnl-2017-313952
33. Mizuno S, Masaoka T, Naganuma M, Kishimoto T, Kitazawa M, Kurokawa S, et al. Bifidobacterium-rich fecal donor may be a positive predictor for successful fecal microbiota transplantation in patients with irritable bowel syndrome. *Digestion.* (2017) 96:29–38. doi: 10.1159/000471919
34. Rossen NG, Fuentes S, van der Spek MJ, Tijssen JG, Hartman JHA, Duflou A, et al. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. *Gastroenterology.* (2015) 149:110–8.e4. doi: 10.1053/j.gastro.2015.03.045

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

PERSPECTIVE
published: 22 May 2020
doi: 10.3389/fdgth.2020.00003

The Importance of Starting With Goals in N-of-1 Studies

Sean A. Munson[1], Jessica Schroeder[2], Ravi Karkar[2], Julie A. Kientz[1], Chia-Fang Chung[3] and James Fogarty[2]*

[1] Human Centered Design and Engineering, DUB Group, University of Washington, Seattle, WA, United States, [2] Computer Science and Engineering, DUB Group, University of Washington, Seattle, WA, United States, [3] Informatics, Indiana University Bloomington, Bloomington, IN, United States

N-of-1 tools offer the potential to support people in monitoring health and identifying individualized health management strategies. We argue that elicitation of individualized goals and customization of tracking to support those goals are a critical yet under-studied and under-supported aspect of self-tracking. We review examples of self-tracking from across a range of chronic conditions and self-tracking designs (e.g., self-monitoring, correlation analyses, self-experimentation). Together, these examples show how failure to elicit goals can lead to ineffective tracking routines, breakdowns in collaboration (e.g., between patients and providers, among families), increased burdens, and even designs that encourage behaviors counter to a person's goals. We discuss potential techniques for eliciting and refining goals, scaffolding an appropriate tracking routine based on those goals, and presenting results in ways that advance individual goals while preserving individual agency. We then describe open challenges, including how to reconcile competing goals and support evolution of goals over time.

Keywords: N-of-1, goals, patient-centered, self-tracking, self-monitoring, self-experiment

OPEN ACCESS

Edited by:
Jane Nikles, University of Queensland, Australia

Reviewed by:
Unai Diaz-Orueta, Maynooth University, Ireland
Maria Vittoria Bulgheroni, Ab.Acus, Italy

****Correspondence:***
Sean A. Munson
smunson@uw.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *20 December 2019*
Accepted: *28 April 2020*
Published: *22 May 2020*

Citation:
Munson SA, Schroeder J, Karkar R, Kientz JA, Chung C-F and Fogarty J (2020) The Importance of Starting With Goals in N-of-1 Studies. Front. Digit. Health 2:3. doi: 10.3389/fdgth.2020.00003

INTRODUCTION

N-of-1 designs have received attention for their potential to facilitate understanding and management of health conditions that require individualized insights and approaches (1, 2). N-of-1 studies come in a variety of designs [e.g., Heyvaert and Onghena (3) and Daskalova et al., (4)], including self-monitoring, in which people track data related to their condition to examine progress toward goals or changes over time [e.g., Mishra et al., (5), Ayobi et al., (6), and Consolvo et al., (7)]; correlational analyses, in which people investigate what factors may affect their symptoms [e.g., behavioral, environmental, medical; (8)]; and self-experiments, in which people determine causality between factors and symptoms [e.g., Karkar et al., (9) and Riggare et al., (10)]. Each of these designs can support a range of goals. Self-monitoring can support tracking and tuning behaviors and understanding whether a condition is worsening (providing cues to take action) or improving (providing motivation to continue one's current management plan). Correlational analyses can support diagnosis and formation of hypotheses among possible contextual and behavioral factors and resulting outcomes. Self-experiments can provide additional rigor in testing relationships between individualized contributors and symptoms and in evaluating whether a management plan is effective.

Many people, individually or with support and encouragement from their healthcare providers, begin n-of-1 studies but struggle to achieve their goals (11–13). Drawing on research on using technology to collect, integrate, and reflect on data about oneself [self-tracking or personal informatics; (14)], we first examine how designers of n-of-1 tools often take a "data-first" perspective that does not place sufficient emphasis on understanding and supporting personalized

goals. We then illustrate how this perspective leads to misalignments between people's goals and the tools they use, or among people collaborating to understand and manage a health condition, as well as emergent design techniques for addressing these challenges. Finally, we discuss unaddressed challenges for researchers and designers of tools that support n-of-1 studies.

ARTICLE CONTEXT: OUR RESEARCH IN N-OF-1 STUDIES AND TOOLS

In this viewpoint article, we examine and reflect on findings from our n-of-1 research. Across this research, we have surveyed 1,396 people with chronic conditions, and conducted interviews, participatory design sessions, or focus groups with 108 people with chronic conditions and 32 health providers (9, 15–21). We also draw on three field deployments of novel prototype systems with 48 people (22, 23). Although most these studies were grounded in specific conditions (irritable bowel syndrome, migraine, juvenile idiopathic arthritis) or health behaviors (sleep, healthy eating), we anticipate the implications of this research apply broadly to n-of-1 studies. All studies were approved by the Human Subjects Division at the University of Washington.

Hypothesis Formation and Hypothesis Testing in Irritable Bowel Syndrome Management

Irritable bowel syndrome (IBS) is characterized by episodic gastrointestinal symptoms that are often caused by individualized dietary factors [i.e., different nutrients can be "triggers" for different individuals; (24)]. Providers often advise their patients record their foods and symptoms in a journal to attempt to identify these triggers, but both patients and providers struggle to interpret the data (16). We explored how the design of n-of-1 tools can better address these challenges. We then examined how interactive, exploratory visualizations can help people and their health providers better interpret their data and collaborate with each other (17). We further developed Foodprint, a photo-based food journaling system that reduces burden and explicitly elicits the patient's goals to better support personalized, actionable, collaborative review (23). Finally, we examined how self-experimentation could help people determine causality between a symptom and trigger (22). We designed, developed, and evaluated a system to support such self-experimentation, and investigated how Bayesian analyses could better answer the questions people want to answer via self-experiments (20). Together, these studies provide insights around how n-of-1 tools can help people form and test hypotheses about their personal IBS triggers.

Supporting Distinct Personalized Goals in Migraine Management

Migraine is characterized by unpredictable, intermittent, and poorly understood symptoms. Similar to IBS, providers often recommend their patients with migraine self-track to better understand and manage their migraines, but both again struggle to find value in the resulting data (18). We investigated how to better support individualized migraine management. We first investigated challenges and pitfalls people currently face, characterizing distinct types of migraine tracking goals people would like to pursue. We then developed and investigated *goal-directed self-tracking*, a new method that scaffolds the process of deciding what, when, and how to track toward a specific goal, and analyzes and visualizes the resulting data to support that goal (20).

SOURCES AND CONSEQUENCES OF GOAL MISALIGNMENTS

Both the tools used for n-of-1 studies and the people involved in planning, interpreting, and acting on those studies can be sources of goal misalignments. We draw on our results, as well as related research, to illustrate these misalignments.

Tools as Source of Goal Misalignment

People conducting n-of-1 studies for insights into their health often adopt tools that are misaligned with their personal goals. These misalignments generally fall into three categories: (1) designs that operationalize a broad goal in ways that are inconsistent with an individual's operationalization of that goal; (2) assumptions that a tracking goal implies other long-term goals; (3) data-first views that fail to scaffold use of that data to support individualized goals.

Designers often make assumptions about how people pursuing their own n-of-1 studies operationalize their goals in their daily lives. For example, many applications designed to support healthy eating promote calorie-counting goals (15). People pursue healthy eating goals, and adopt tools in support of those goals, for much more varied reasons (e.g., improving their energy levels, adopting a diet that they believe has health benefits, trying to reduce certain foods, managing an eating disorder) (15, 25). Tools that operationalize every goal as calorie counting both fail to support people's true goals and can lead to negative feelings and counterproductive behaviors (15).

Similarly, designers often assume that people track to pursue certain long-term goals. For example, most commercial menstrual tracking apps embed an assumption that people track to become or avoid becoming pregnant (26). This assumption can lead to features that are irrelevant or hurtful (e.g., an annoyance for people who are not having sex that could result in conception, a painful reminder for people who cannot conceive). We have been happy to see a trend toward making such features optional and disabled by default (e.g., in Apple's new cycle tracking application), but more work is needed to apply such design principles consistently across self-tracking tools.

An approach of creating general-purpose tools that allow people to collect a large range of data in various ways and to run analyses on that data may seem promising; a flexible tool could support a range of goals (27). However, this approach also leads to problems. Some self-tracking applications *do* enable collection and integration of large amounts of data, with the idea that supporting flexibility is the same as supporting individualized goals. However, *flexibility* is not the same as *support*. Added

flexibility for configuration requires a system to also support understanding *how* to configure for one's goals. This problem is particularly salient in n-of-1 tools, where people may not know what goals are achievable or reasonable (20) or how to translate their goals into tracking plans (28).

Even when people bring their own well-defined, achievable goals to n-of-1 tools, they face burdens when tracking and analyzing resulting data. They may reach incorrect conclusions or abandon a tool after considerable effort but without reaching their goals (11, 29). Tracking tools designed with a data-first view may also prioritize collection of as much data as possible, even when lower-burden tracking would also support a person's goals.

People as a Source of Goal Misalignment

Whether people conducting n-of-1 studies initiate those studies themselves or under the advice of a health provider, they frequently turn to others for support (e.g., family, peers, health experts). However, collaborators sometimes assume certain goals for both for *why* the person is tracking (e.g., *management goals* regarding what they want to address in their health and *tracking goals* regarding the information can help them achieve those management goals) and *how* they should track. Such assumptions can introduce misalignment in configuring, interpreting, and acting on self-tracking data (19).

Misaligned Management Goals

A person's goals for managing their health sometimes differs from their health provider's (16). For example, when reviewing food and symptom diaries, providers often try to identify potential contributors to a patient's digestive symptoms and suggest they eliminate those potential contributors. However, due to personal preferences and priorities, patients may instead choose to continue eating certain foods, tolerating resulting symptoms and planning for how those symptoms will affect their lives. Other patients may initially restrict their diet as suggested to control their symptoms, then collect food and symptom data toward a goal of re-diversifying their diet, which a provider may not expect if not explicitly told (17). Patients also sometimes use food and symptom diaries to elicit emotional support, such as seeking recognition of their effort in managing symptoms or showing the data as evidence of how symptoms affect their life. Although providers may primarily expect to use data for diagnosis and the design of treatment plans, acknowledging these other potential patient goals is also important throughout the collaboration.

People with migraine and health providers also encounter tensions when their management goals do not align (18). For example, prescription medications can prevent symptoms for some people with migraine. As many providers assume their patient's primary management goal is symptom prevention, a common first step in migraine treatment is to prescribe medications. However, some people with migraine have management goals of preventing symptoms *without* medications. One patient described this misalignment: "[My doctor's] approach was much more like, 'Let me figure out what drugs I can give you to have you stop having these headaches', rather than figuring out why I'm having them. I'm much more like, 'I want to know why this is happening'."

Misaligned Tracking Goals

Even when management goals do align, misalignment in tracking goals can still be detrimental in collaborations (e.g., within a family, between patients and providers). When providers encourage patients to track what they eat and relevant health indicators, they sometimes review the tracked data with the patient once and then expect the patient to continue independently reviewing data. However, provider disengagement with data can dissuade people from continuing tracking or suggest that self-tracking is no longer useful. We found similar misalignments in migraine, where some providers assumed patients would be able to analyze their data to identify trends.

Providers can also be removed from the lived experience of self-tracking, leading them to recommend burdensome tracking routines (16, 20). For example, providers might assume patients want the most validated answers possible and recommend rigorous but high-burden tracking (e.g., paper diaries detailing every consumed food). Patients may instead sacrifice some rigor to find a tracking regime that fits better in their life (e.g., a photo-based food diary that loses some detail but retains a reasonable record with less effort). In migraine, providers do not always recognize the burden of daily tracking, so they might recommend it given potential value of having more data [e.g., "obviously I like my patients to track every day," (20)]. However, people might prefer to reduce their tracking burden by building in breaks or tracking only when they experience symptoms.

Misalignments in management and tracking goals also interact to create further problems. For example, many people with migraine track with a goal of predicting the likelihood of symptoms so they can prepare for or attempt to prevent those symptoms. They often focus on tracking contributors to ensure they will notice contributor accumulation, which can result in symptoms. However, providers generally focus on overall symptom frequency, rather than the consequences of symptoms on a particular day. They therefore often want patients to focus on tracking treatments and symptoms. A patient's desired tracking routine may therefore differ from their provider recommendations.

BETTER SUPPORT FOR GOALS IN N-OF-1 TOOLS AND PROCESSES

Emerging design patterns can support explicit goal alignment and pursuit. These patterns include supporting patients and providers in aligning goals, configuring tracking routines to support goals, and analyzing and presenting resulting data to provide actionable insights that advance goals.

Eliciting and Aligning Patient and Provider Goals

Systems and health experts can prompt patients to articulate their goals, which can help people plan their tracking and subsequent actions.

In our research to support n-of-1 studies in migraine, we designed an interface to elicit a person's tracking goals (**Figure 1**). We first asked people to define their migraine tracking goal in their own words. Participants sometimes struggled to express their goals, but presenting explicit examples helped them reconcile their management and tracking goals. After expressing their goal, our design prompted people to categorize that goal as one of three distinct migraine tracking goal categories we had previously characterized (18). All participants reported being able to select a goal category that articulated why they wanted to track migraine-related data. The explicit categories also helped participants hypothesize about goals they might want to pursue in the future and helped differentiate those goals from their current goals. For example, one participant wanted to focus on learning about her migraines, but thought that, once she understood more, she would want to transition to a monitoring or predicting goal. Another ultimately wanted to learn about her migraines but decided to first focus on a lower-burden monitoring goal before committing to a goal that might require longer or more frequent tracking routines. Explicit goal categories therefore helped participants to navigate the critical path between their management and tracking goals and to reason about what goals would be most feasible and helpful to them at present and in the future.

Designs can also support communication about goals between patients and providers. In our food tracking research, we designed a pre-visit note to support explicit patient-provider communication about goals (23). The note elicited patient goals for the visit, their own summary of their data, and their questions for health providers. Providers could view this note at the start of a visit with that patient. Both patients and providers paid more attention to the patient's goals and questions during visits with the note. Having these explicit goals also helped providers tailor their advice to patient priorities. For example, one provider saw that a patient valued eating certain foods that could contribute to symptoms. Rather than urging that patient to eliminate those foods, they instead talked about alternative ways to prepare them that might mitigate symptoms. Another patient-provider pair also chose to focus on stress management instead of food elimination, because the patient had a goal of maintaining dietary diversity. Having awareness of a patient's goal allowed providers to better develop individualized management plans.

Scaffolding the Right N-of-1 Study Design Based on a Person's Goals

After a person's goals are understood, designs can scaffold n-of-1 studies that support those goals with the least burden by either matching people with the right tool among many or by changing how a tool is configured.

To support healthy eating and IBS management, Foodprint supported configurations specific to different goals (23) (**Figure 2**, left). For example, when individuals expressed a healthy eating goal of "eating more balanced meals," we configured their app to support annotating food groups (fruits, vegetables, grains, protein, dairy, oils). Individuals who wanted to understand relationships between food and mood or stress could instead report stress level and mood. Finally, for people tracking to understand potential IBS symptom contributors, we configured Foodprint to record common contributors and symptoms. During onboarding, researchers elicited patient goals to configure the tool, but we envision the design of onboarding processes that elicit goals and configure an appropriate n-of-1 tracking tool.

In our work on goal-directed self-tracking for migraine, we designed and evaluated low-fidelity prototypes for such an interface (20). Transforming goals into tracking regimes helped people avoid common tracking pitfalls. The system could prompt people to track all of the data they would need to support their goal, avoiding a breakdown in which people do not track all the data needed to meet their goals. It also could guide them away from tracking *too much* data, avoiding a breakdown in which people track too much, become fatigued, and abandon tracking. For example, when a person selects a goal of *identifying contributors to their migraines*, the system walks them through selecting symptoms and contributors they want to investigate. When a person selects a goal of *monitoring* their migraines, which typically does not require tracking contributors, the system encourages a focus on symptoms.

Given the variety of possible management and tracking goals, no single tool can realistically support every goal a person might have. Tool *selection*, and communication of a tool's limits, is therefore as important as tool *configuration*. Consider a person working to understand what factors contribute to their gastrointestinal symptoms. They might use Foodprint for preliminary analyses that suggest caffeine or lactose may be a trigger. However, that person might only consume caffeine when stressed, and might only consume caffeine in lattes. Each of these factors (i.e., caffeine, lactose, and stress) is a potential contributor, but Foodprint's correlational approach cannot untangle their confounds. Doing so requires a rigorous self-experiment, which is challenging for people to design and conduct due to the need for expertise in health, experimental design, and appropriate tracking burden (11). A system designed to scaffold such self-experiments (**Figure 2**, right) can design an appropriate experiment and explicitly support a corresponding tracking routine and analyses of results (9, 22). Guiding people to the right tool for their goal is therefore necessary: a person with a specific hypothesis would likely prefer a self-experimentation app, whereas a person who wants to learn about potential contributors would likely prefer a tool that supports correlation-based analyses of a broader range of factors.

Each n-of-1 design has a range of possible analysis approaches. In our correlational analysis of food photos, we explored both visual analysis of photos grouped by symptom severity and quantitative analysis graphing correlations between nutrients and symptoms (17, 23). Both approaches had advantages. Photos facilitated conversation and better supported action planning; the quantitative analysis supported understanding more complex nutrient-symptom interactions. Our work in self-experimentation revealed other tradeoffs. Our initial analysis presented a graph and a summary of a statistical analysis, including a p-value (22). This familiar (although flawed) statistical detail contributed to a sense of validity and trust for participants. We have since shown that Bayesian analyses can better support the questions people ask from n-of-1 studies and the decisions they want to make using those answers (18).

FIGURE 1 | Our goal-directed self-tracking prototype for migraine elicited people's tracking and management goals.

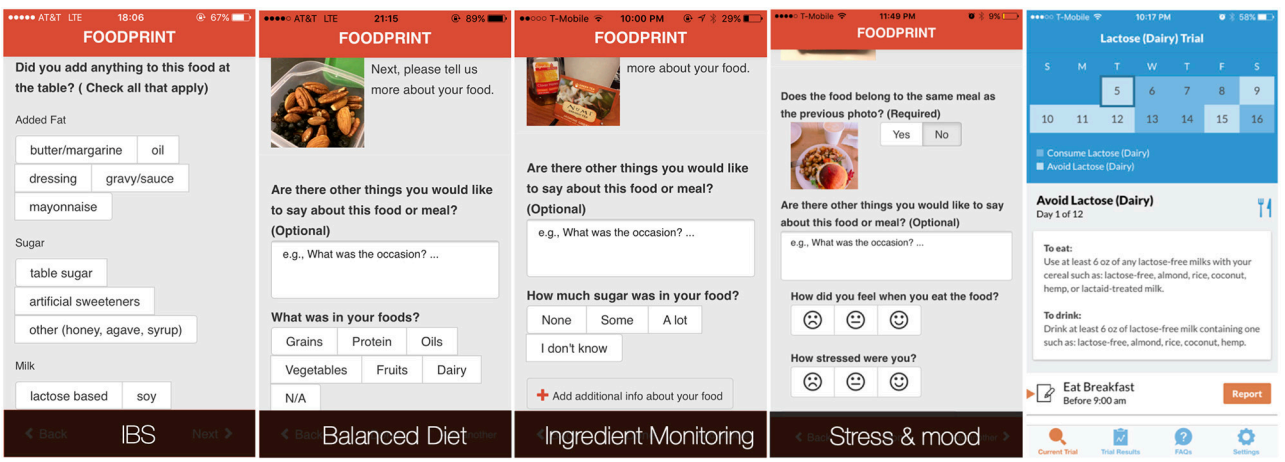


FIGURE 2 | By eliciting people's tracking goals, we could configure the Foodprint food diary application to better support those goals. People who wanted to test the relationship between a specific potential contributor and symptoms, however, benefited more from using our self-experimentation application, TummyTrials (rightmost).

Across all study designs and analyses, grounding results in examples from a person's data (e.g., particular foods, days when symptoms were severe) facilitated understanding and helped them determine next steps.

FUTURE CHALLENGES

Our research shows that eliciting goals and configuring systems to support them is often less straightforward than it sounds. Participants often approach tracking with underspecified [e.g., "I don't know [what my goal is]. I just want to know how to get rid of them faster," (20)] or unachievable goals. Techniques outlined above can help, but goal elicitation and specification remain challenging (28, 30).

Even when people can articulate and fulfill tracking goals, knowing the answer to a question can be far from acting on it. Research should develop techniques for providing actionable guidance tailored to a person's goals and their context, such by using explicit goal elicitation alongside context-aware computing [e.g., Rabbi et al., (31) and Lee et al., (32)]. Similar to our

scaffolding for migraine tracking (20), others have proposed interactive instructional materials, such as for effective planning (32). Designers might also develop techniques for helping people anticipate possible answers to a range of possible questions, so they could decide whether they would want to act on any of those answers *before* they begin tracking. This information would allow people to exclude n-of-1 studies designed to provide answers on which they would not want to act or that would provide insufficient evidence for them to act.

Fully supporting goal-directed self-tracking also requires supporting goal evolution, both between and within goals. People often change their goals as their understanding, experiences, and needs change (19, 33–35). For example, a person with migraine may initially want to learn about their migraines (e.g., understand what causes their symptoms), then switch to monitoring. Tools should support explicitly making such changes.

We have thus far designed and evaluated n-of-1 systems that focus on one person's goals and what their health providers believe those goals are or should be. However, many health behaviors are influenced by others, especially the people with whom one cohabitates, such as family members (36, 37). In such situations, we might instead think of the unit of analysis as a family. Within that family, people might have shared goals, compatible goals, or conflicting goals (21). Such uses will likely require new n-of-1 designs and approaches.

CONCLUSION

New technologies for collecting, integrating, and analyzing data promise to make n-of-1 studies more feasible than ever before. This trend offers important opportunities for understanding and managing personal health. However, we caution against assumptions, and especially implicit assumptions, about why and how people use tracking tools. Such assumptions often lead to frustrating goal misalignments and n-of-1 studies that provide the wrong answers or no answers. Instead we urge researchers and designers to start with people's goals, then provide scaffolding to support selection and configuration of tools to meet those goals.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Washington Human Subjects Division. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SM, JS, RK, JK, C-FC, and JF each led parts of the research described in this article and contributed to outlining, writing, and editing the manuscript.

FUNDING

This research was supported in part by the University of Washington under a Research Innovation Award; the National Science Foundation under awards IIS-1813675, IIS-1553167, and DGE 1256082; the National Institutes of Health under award R01 LM012810; the Agency for Healthcare Research and Quality under award R21 HS023654, and the National Center For Advancing Translational Sciences of the National Institutes of Health under award UL1 TR002319. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

REFERENCES

1. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personal Med.* (2011) 8:161–73. doi: 10.2217/pme.11.7
2. Kravitz RL, Duan N (editors). The DEcIDE Methods Center N-of-1 Guidance Panel, Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S. *Design and Implementation of N-of-1 Trials: A User's Guide.* AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality (2014). Available online at: www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm
3. Heyvaert M, Onghena P. Randomization tests for single-case experiments: state of the art, state of the science, and state of the application. *J Contextual Behav Sci.* (2014) 3:51–64. doi: 10.1016/j.jcbs.2013.10.002
4. Daskalova N, Desingh K, Papoutsaki A, Schulze D, Sha H, Huang J. Lessons learned from two cohorts of personal informatics self-experiments. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* New York, NY: Association for Computing Machinery (2017). 1:22. doi: 10.1145/3130911
5. Mishra SR, Klasnja PJ, MacDuffie Woodburn, Hekler EB, Omberg L, Kellen M, et al. Supporting coping with Parkins'n's disease through self tracking. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* (2019). p. 1–16. doi: 10.1145/3290605.3300337
6. Ayobi A, Marshall P, Cox AL, Chen Y. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In: *Proceedings of the (2017). CHI Conference on Human Factors in Computing Systems.* New York, NY: Association for Computing Machinery (2017). p. 6889–901. doi: 10.1145/3025453.3025869
7. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of UbiFit garden. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* (2008). p. 1797–806. doi: 10.1145/1357054.1357335
8. Mamykina L, Mynatt ED, Kaufman DR. Investigating health management practices of individuals with diabetes. *Proc SIGCHI Conf HumFactors Comput Sys.* (2006) p. 927–936. doi: 10.1145/1124772.1124910
9. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. *J Am Med Inform Assoc.* (2015) 23:440–8. doi: 10.1093/jamia/ocv150
10. Riggare S, Unruh KT, Sturr KJ, Domingos J, Stamford JA, Svenningsson P, Hägglund M. Patient-driven N-of-1 in Parkinson's disease. *Methods Inform Med.* (2017) 56(S.1): e123–8. doi: 10.3414/ME16-02-0040
11. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collecting and exploring personal data. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* New York, NY: Association for Computing Machinery (2014). p. 1143–52. doi: 10.1145/2556288.2557372
12. Chung CF, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *J Med Internet Res.* (2015) 17:e203. doi: 10.2196/jmir.4364

13. West P, Van Kleek M, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. In: *Proceedings of the (2018). CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2018). p. 1–13. doi: 10.1145/3173574.3174058
14. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. In: *Proceedings of the SIGCHI conference on human factors in computing systems.* New York, NY: Association for Computing Machinery (2010). p. 557−66.
15. Cordeiro F, Epstein DA, Thomaz E, Bales E, Jagannathan AK, Abowd GD, et al. Barriers and negative nudges: exploring challenges in food journaling. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2015). p. 1159–62. doi: 10.1145/2702123.2702155
16. Chung CF, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*. New York, NY: Association for Computing Machinery (2016). p. 770–86. doi: 10.1145/2818048.2819926
17. Schroeder J, Hoffswell J, Chung C, Fogarty J, Munson SA, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In: *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing.* New York, NY: Association for Computing Machinery (2017). doi: 10.1145/2998181.2998276
18. Schroeder J, Karkar R, Fogarty J, Kientz JA, Munson SA, Kay M. A patient-centered proposal for bayesian analysis of self-experiments for health. *J Health Informatics Res.* (2019). 3:124–55. doi: 10.1007/s41666-018-0033-x
19. Schroeder J, Chung C, Epstein DA, Karkar R, Parsons A, Murinova N, et al. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. In: *Proceedings of ACM Designing Interactive Systems.* New York, NY: Association for Computing Machinery (2018). doi: 10.1145/3196709.3196738
20. Schroeder J, Karkar R, Murinova N, Fogarty J, Munson SA. Examining opportunities for goal-directed self-tracking to support chronic condition management. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* New York, NY: Association for Computing Machinery (2019). 3:151. doi: 10.1145/3369809
21. Pina LR, Sien S, Ward T, Yip JC, Munson SA, Fogarty J, et al. From personal informatics to family informatics: understanding family practices around health monitoring. In: *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW 2017).* New York, NY: Association for Computing Machinery (2017). doi: 10.1145/2998181.2998362
22. Karkar R, Schroeder J, Epstein DA, Pina LR, Scofield J, Fogarty J, et al. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* New York, NY: Association for Computing Machinery (2017). doi: 10.1145/3025453.3025480
23. Chung CF, Wang Q, Schroeder J, Cole A, Zia J, Fogarty J, et al. Identifying and planning for individualized change: patient provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *PACM Interactive Mobile Wearable Ubiquitous Technologies. ACM.* (2019). doi: 10.1145/3314394
24. Eswaran S, Tack J, Chey WD. Food: the forgotten factor in the irritable bowel syndrome. *Gastroenterol Clin.* (2011) 40:141–62. doi: 10.1016/j.gtc.2010.12.012
25. Graham AK, Wildes JE, Reddy M, Munson SA, Barr Taylor C, Mohr DC. User-centered design for technology-enabled services for eating disorders. *Int J Eating Disord*. (2019) 52:1095–1107. doi: 10.1002/eat.23130
26. Epstein DA, Lee NB, Kang JH, Agapie E, Schroeder J, Pina LR, et al. Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the (2017). CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2017). p. 6876–88. doi: 10.1145/3025453.3025635
27. Kim YH, Jeon JH, Lee B, Choe EK, Seo J. OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking. *Proc ACM on Interactive Mobile Wearable Ubiquitous Technol.* (2017) 1:1–28. doi: 10.1145/3130930
28. Niess J, Wozniak PW. Supporting meaningful personal fitness: the tracker goal evolution model. In *Proceedings of the (2018). CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2018). p. 171. doi: 10.1145/3173574.3173745
29. Epstein DA, Pin A, Fogarty J, Munson SA. A lived informatics model of personal informatics. In *Proceedings of the (2015). ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: Association for Computing Machinery (2015). p. 731–42. doi: 10.1145/2750858.2804250
30. Rapp A, Cena F. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *Int J Hum Comp Stud.* (2016) 94:1–17. doi: 10.1016/j.ijhcs.2016.05.006
31. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In: *Proceedings of the (2015). ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: Association for Computing Machinery (2015). p. 707–18. doi: 10.1145/2750858.2805840
32. Lee J, Walker E, Burleson W, Kay MJ. S., Buman M, Hekler EB. Self-experimentation for behavior change: Design and formative evaluation of two approaches. In: *Proceedings of the (2017). CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2017). p. 6837–49. doi: 10.1145/3025453.3026038
33. Epstein DA, Kang JH, Pina LR, Fogarty J, Munson SA. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. In *Proceedings of the (2016). ACM International Joint Conference on Pervasive and Ubiquitous Computing*. (2016). p. 829–40. doi: 10.1145/2971648.2971656
34. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In: *Proceedings of the (2017). CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2017). p. 629–41. doi: 10.1145/3025453.3025750
35. Kim SI, Jo E, Ryu R, Cha I, Kim YH, Yoo H, et al. Toward becoming a better self: understanding self-tracking experiences of adolescents with autism spectrum disorder using custom trackers. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare.* Trento (2019). p. 169–78.
36. Saksono H, Castaneda-Sceppa C, Hoffman J, Seif El-Nasr M, Morris V, Parker AG. Family health promotion in low-SES neighborhoods: a two-month study of wearable activity tracking. In: *Proceedings of the (2018). CHI Conference on Human Factors in Computing Systems*. (2018). p. 1–13. doi: 10.1145/3173574.3173883
37. Lukoff K, Li T, Zhuang Y, Lim BY. TableChat: mobile food journaling to facilitate family support for healthy eating. *Proc ACM Hum Comp Interaction*. (2018) 2:1–28. doi: 10.1145/3274383

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 04 June 2020
doi: 10.3389/fdgth.2020.00002

Using Self-Study and Peer-to-Peer Support to Change "Sick" Care to "Health" Care: The Patient Perspective

Camille Nebeker[1,2,3]*, *Bethany Weisberg*[1], *Eric Hekler*[1,2,3] *and Michael Kurisu*[1,2,3]

[1] Department of Family Medicine and Public Health, School of Medicine, University of California, San Diego, La Jolla, CA, United States, [2] Center for Wireless and Population Health Systems, UC San Diego, La Jolla, CA, United States, [3] The Design Lab, UC San Diego, La Jolla, CA, United States

Background: Access to digital health technologies is contributing to a paradigm shift where *sick*care may become authentic *health*care. Individuals can now access personal health data through wearable sensors, affordable lab screenings, genetic and genomic sequencing, and real-time health tracking apps. Personal health data access creates opportunities to study health indicators 24/7 and in real time. This is especially useful for patients with hard-to-diagnose or treat diseases, which led to a self-formed patient group called Project Apollo. Project Apollo is composed of highly motivated patients with common experiences of undiagnosed conditions, a lack of clear treatment options, and shared frustrations with navigating the U.S. healthcare system. These experiences have led the Apollo cohort to supplement their health knowledge through self-study research.

Objective: To qualify the experience and expectations of patients affiliated with Project Apollo.

Methods: A qualitative approach involved record review and semi-structured interviews. One-hour semi-structured interviews were conducted to solicit motivations, expectations, and potential barriers and facilitators to self-study followed by a brief survey on digital tool use. Interviews were digitally recorded, transcribed, and analyzed to identify themes and patterns.

Results: Participants included six females and three males ranging in age from 30 to 70+ years. Responses were organized under five key themes including: frustration with healthcare system; community support; self-study/N-of-1 research; access to experts; moving from sick to healthcare. Facilitators include motivation, albeit stemming from frustration, a safe community where patients derive support, and access to experts for guidance. Increasing awareness of clinicians about the potential value of partnering with patients who are advancing health knowledge through self-study is critical.

Conclusions: N-of-1 self-study research, coupled with community support and digital health tools, appears to be one plausible pathway to shifting the paradigm from *sick*care toward patient-partnered *health*care.

Keywords: citizen science, N-of-1, digital health, self-tracking, participant-led research, peer-to-peer support, Research Ethics

OPEN ACCESS

Edited by:
Constantinos S. Pattichis, University of Cyprus, Cyprus

Reviewed by:
Elena Cardillo, National Research Council (Cnr), Italy
Abu Saleh Mohammad Mosa, University of Missouri, United States

***Correspondence:**
Camille Nebeker
nebeker@eng.ucsd.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *01 December 2019*
Accepted: *21 April 2020*
Published: *04 June 2020*

Citation:
Nebeker C, Weisberg B, Hekler E and Kurisu M (2020) Using Self-Study and Peer-to-Peer Support to Change "Sick" Care to "Health" Care: The Patient Perspective. Front. Digit. Health 2:2. doi: 10.3389/fdgth.2020.00002

INTRODUCTION

The structure and operations of healthcare in the United States (US) is grounded in prioritizing acute care over individual health promotion and disease prevention as well as public health (1). This *sick*care system was classically created to enable people to receive expert, evidence-based advice and support to help diagnose and treat diseases (1). A dominant paradigm in the United States and the United Kingdom (UK), among other countries, is to provide evidence-based medicine to ensure high quality support (2). (2) defined evidence-based medicine as:

> *"… the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research."*

As this definition implies, decisions on diagnosing and treating diseases involves balancing the information and wisdom between what is learned from prior scientific studies and the clinical training, experience, and expertise of clinicians. Increasingly, there is movement toward more patient-centered care and support (3). This includes recognizing and honoring the knowledge, preferences, and abilities of patients as an essential part of care and prioritizing the prevention of disease, or what is called *health*care (1).

This shift toward patient-centered care and challenges to assumptions on what is evidence-based are being further influenced by digital technologies. In particular, access to digital health technologies enables individuals to gather personal health data through wearable sensors, affordable lab screenings, genetic and genomic sequencing, and real-time health tracking apps. Personal health data access creates opportunities to study health indicators 24/7, in context and in real time. These new technologies are affording new forms of information and evidence to be incorporated into the provision of care. This is especially useful for patients with hard-to-diagnose or treat diseases, for whom classic external evidence from prior clinical trials or the training and expertise of clinicians providing the support do not have sufficient information to provide an accurate diagnosis and offer actionable care. These new technologies are resulting in a growing number of informed and empowered patients (4–6). Greater access to personal health data has enabled patients to document their individual health trends and status, which contributes to their health-related decisions and interactions with their healthcare providers (4). Indeed, obtaining personal health data can provide evidentiary support in the medical diagnosing and treatment of diseases (7).

From this context, the Project Apollo cohort emerged and was organized as a non-profit entity. The Precision Healthcare Ecosystem is a nonprofit corporation registered in California with the vision that "The Doctor of the Future is One's Self." Its inaugural program, Project Apollo, utilizes a multi-disciplinary, collaborative, and integrative care model, the "Study of Me," to educate, enable, and empower participants to lead a personalized health journey, guided by their own quantified, evidence-based data. Project Apollo is a patient-initiated effort with a goal helping people learn to "self-study" to better understand factors that influence their health. Project Apollo provides people with access to education and experts who can facilitate increased knowledge of how to conduct self-tracking and self-experiments. The genesis of Project Apollo began with Dr. Michael Kurisu, an osteopathic physician who actively integrates digital health data and information in pursuit of more holistic care. His idea was to form a community of patients to foster active self-tracking to learn about and be better health advocates for themselves and others. This community was inspired by one of Dr. Kurisu's more prominent patients, Dr. Larry Smarr, a well-known "Quantified Self" individual who is modeling what a patient may be in the future (8). As the community has evolved, it has also incorporated other roles, including researchers who can provide support on issues such as the ethical conduct of research or conducting rigorous N-of-1 self-study and other clinicians who can provide holistic care and support in alignment with the desires and self-study results of patients (e.g., QiGong).

The purpose of this paper is to report ethnographic research on the genesis of the Project Apollo Cohort. In particular, the Project Apollo Cohort represents a concrete, real-world, patient-initiated effort that aligns with more general aspirations of patient-centered care. The results of this qualitative inquiry shed light on the motivations, benefits, and challenges experienced by this cohort, which could be instructive for understanding efforts in participant-led research.

METHODS

Between February and May 2019, we conducted an ethnography of Project Apollo and its parent organization the Precision Healthcare Ecosystem, a 501c3 umbrella organization formed by the patients to advance the goals of Project Apollo. Qualitative data were collected through a 1-h semi-structured interview and a short survey with Apollo members to capture individual motivations, challenges, and goals. These data were augmented with meeting minutes and documents describing the formation and evolution of Project Apollo. Participants gave their informed consent prior to being interviewed and the study was verified as exempt from the Common Rule by the UC San Diego Institutional Review Board. Throughout the data collection period, the research team attended multiple Project Apollo meetings and participated in conference calls with the group.

Data Collection and Management

An inductive ethnographic approach was used to review documents that included recorded presentations, meeting minutes and organizational mission/vision statements. Semi-structured interviews were conducted with eight Project Apollo patients and the group's founding clinician. The interview questions were developed to better understand motivations and expectations as well as potential barriers and facilitators to self-study. Interviews included open-ended questions, for example:

"What role do you feel Project Apollo will play in your healthcare journey?"

"Describe for me, in your own words, N-of-1 (self-study) research."

"What guiding principles should be upheld in participant-led research?" and,

"What steps will you take to ensure the validity of your research results?". Interviews were approximately 60 min in duration and were digitally recorded, professionally transcribed, and inductively analyzed to identify themes and patterns as they emerged. All transcriptions were de-identified to protect confidentiality and stored in a password-protected file accessible to the research team members involved with data collection and analysis. The transcribed interviews, participant-observation field notes, and Project Apollo records were uploaded into a qualitative data analysis software program (9).

From July to August 2019, we asked interview participants to respond to a four-question survey and we received responses from six ($n = 6$) individuals. The survey was designed to contextualize the process of self-study research Project Apollo members are conducting. The survey included open-ended questions so as to not limit participant responses, including:

"What data are you collecting (e.g., sleep, pain, function, etc.)?"

"How do you collect your data (e.g., Oura ring, daily blood pressure device, self-assessment, etc.)?" "How do you record your data (e.g., spreadsheet like Excel, journal, app, etc.)?" and,

"Any additional information about your use of digital tools to support your self-study project?".

The survey responses supported the analysis of how Project Apollo members choose to conduct self-study research and preferred methods of tracking and storing their research data.

Analysis

All transcripts were de-identified and each participant was assigned an identification number. Interview data, including analytic memos and meeting and field notes, were imported into Dedoose and inductively analyzed. Data analysis involved an iterative process of reviewing all transcripts and supporting data by two of us (BW and CN) and then applying inductive coding to extrapolate the predominant themes (10). Initial codes were developed independently after reviewing two transcripts and then discussed to identify final codes. All transcripts were then coded by BW and further organized by major themes. To further contextualize the data, a brief anonymous survey was sent to participants to gauge experience and usage of digital technologies and mobile health apps. The results of the survey responses were analyzed and reported as descriptive statistics.

RESULTS

Participant Demographics

A total of nine ($N = 9$) individuals participated in an interview, with six responding to a follow-up survey. Participants were all adults over 18 years old and consisted of six females and three males. The estimated age range was 25–75 years of age with all reporting having complete college with the majority having a graduate degree. Participants included eight ($N = 8$) patients and one ($N = 1$) clinician associated with Project Apollo.

Major Themes

Responses were organized under the five key themes identified during data analysis: healthcare system frustration; community support; self-study/N-of-1; need for access to experts; moving from *sick*care toward a *health*care system.

What has led patients to Project Apollo is their shared frustration with the healthcare system. They receive community support that, along with advances in technology and access to health and research experts, fosters their motivation to study their health conditions through observational self-tracking and N-of-1 studies. A common theme of community support is in empowering their decisions to go forward with self-studies to supplement their healthcare decision-making. In addition to the peer-to-peer community support, they expressed the need for access to health experts and researchers in the process of their self-study research. Ultimately, their shared hope is that through the self-study research combined with advances in technologies, they will facilitate the transformation of the broken *sick*care system to a patient-centered, precision health ecosystem. Each theme is presented below and augmented with quotes from participant interviews.

Healthcare System Frustration

This theme is characterized by experiences with hard-to-diagnose diseases, which is a key attraction of this Project Apollo Cohort and why Project Apollo was formed. One of the most common points of contention among Apollo patients was their shared frustrations with the U.S. healthcare system.

Frustrations with the current healthcare system included how difficult it is to navigate, receiving unsatisfactory diagnoses, undiagnosed health issues, piecemeal care, high costs ("these financial burdens, they're not fair to patients" P05), and being brushed aside. A common frustration expressed was the difficult path many faced in obtaining diagnoses of their various health conditions, illustrated by this participant:

> *"It's painful. It's frustrating. That journey was so difficult for me, and I am a very strong person, but those were dark, dark times... because I was in pain, things were happening to my body, and no one could tell me what was wrong, or how to fix it" (P01).*

> *"So, it's been a long journey, and it doesn't look like it's getting... like there appear to be no solutions. It's incredibly frustrating.*

> *You know, people look at me and think I'm fine," (P17).*

Relatedly, several participants felt they were not listened to and their symptoms glossed over by health providers. For instance,

> *"It's sort of been a frustration for me for a lot of years to be kind of not in sync with my providers, where I actually have – this sounds ridiculous – to tell them which tests I need them to run. They'll question that and say, 'Well, how can you justify this, blah, blah?' I read a lot. I read studies and I read methodology and it's frustrating for me to run up against" (P02),* and

"If you don't know and the scientists don't know, or the scientists say, 'Well, it's all in your head,' which is one of the things that enraged so many of us" (P04).

Regarding feeling brushed aside by health providers, another participant (P05) simply stated,

"I didn't want to be doubted as a person or a patient."

P01 continued to describe how the common thread of frustrating healthcare inspired the formation of Project Apollo,

"But basically, the founding cohort of Project Apollo came together because we have managed to navigate the healthcare system to begin to get the kind of care we needed, and we unanimously felt that it shouldn't be this hard."

"P02: At least speaking for myself, and I know with some of the others, we find it difficult that the various specialists that we're seeing don't seem to be communicating together and/or don't seem to be, how can I say, not yet quite comfortable with the concept of collaborative care, where [it's] the patient them self who really has the most in depth understanding of how their body works."

Another participant analogized the broken healthcare system to that of a storybook character:

"I was thinking about medicine as Humpty Dumpty, in that the current healthcare system has broken the patient into "parts care" via specialists, and that only through an integrated patient-centered whole person approach can we put Humpty Dumpty back together again to help patients heal and become whole again. All the king's horses and men cannot do it...we must involve the patient." (P05).

Community Support

This theme of Community Support is perhaps the strongest predictor of how successful Apollo may be in the future. Resonating across all Apollo participants is the close community support system they have created. For patients who have experienced serious hardships in their healthcare journey, Project Apollo was often described by patients as place of solace and support.

While many participants expressed dissatisfaction with the healthcare they received, Project Apollo was explained to be a group that provides a place of support and guidance where patients can express their health desires and seek answers to the questions they possess.

"I just don't know what the answers are for me, and I need some community to help me figure that out" P17.

Explaining what draws the group together, one participant expressed that Project Apollo is:

"a community that brings a lot of support to one another in navigating this often-broken healthcare system, as well as deep diving into our own health and I guess promoting wellness," (P06).

The community support and group dynamic were also discussed as providing a healthy impact on participants' wellness journey, as illustrated by P15,

"Just being a part of this group has really helped me on my health journey," and

P07, *"we had this kind of group meditation and a check-in and the patients got to know each-other and I started noticing that aspects of their health got better just from that intervention."*

A strong social support system has been demonstrated to improve health outcomes (11) as well as provide meaning in life. Participant P05 spoke of how a community bond is a crucial aspect of overall health:

"I think the community is really important. And I think just empowering. I mean we all want to live a rich, fulfilled life and it doesn't have to be with a perfect body and perfect mind but a rich, fulfilled life. So, I think that's been a huge part."

Another participant declared that the key to a successful self-study lies in the community aspect,

"There's a lot of things I want to study and how would I like to study them? Well, no way better than a community of caring people who have their own self-study, with all these amazing researchers we have accessible through this project" (P01).

In a stark difference to other participants who highly regard the community Project Apollo brings together. One participant (P04) expressed this as being the weakest aspect of Project Apollo and needs to be strengthened.

"That community aspect is where we're weakest. Where there's a tight group of the original founding cohort, and then there's...if this is going to grow, the community has to be attended to. I'd say if anything, that's probably the place that needs the most work, in my mind" (P04).

Self-Study/N-of-1

The process of learning to self-track and carry out self-experiments plays a vital role in supplementing Apollo patients' healthcare experience. Several Apollo patients felt that without the tools currently available to assist them in conducting self-study research, they would not have been able to get this far in their health journey. In fact, our brief survey revealed that nearly all participants were using digital tools to facilitate their self-tracking process and progress. Of the nine participants who completed an interview, six individuals responded to our 4-item web-based survey. Respondents acknowledged tracking a diverse array of data, including symptoms, biomarkers, and/or physical attributes, using digital technologies including wearable sensors, mobile health apps, and real-time tracking. For example, several (five of the six respondents) had purchased an Oura ring[1] to track information on sleep quality and activity levels. All had begun to use applications and digital health technologies to

[1]https://ouraring.com/

assist in tracking different health variables to inform actionable health choices (**Figure 1**).

A core element of the Project Apollo Cohort is the opportunity to create and implement N-of-1, self-studies. Many attributed self-tracking and self-study to their ability to take control of their health journey in a substantive way.

> *"The folks at Project Apollo, I think many or most are very actively involved in their own healthcare. They've been doing a lot of their own tracking and a lot of their own finding providers that are most helpful" (P02).*

Similarly, participant P05 stated,

> *"I think giving any individual the tools to gather data in a meaningful manner that can help them make – well, the side benefit is it will help them make decisions about their own health."*

Self-tracking and N-of-1 research has flourished and continues to expand due to the ubiquity of wearable sensors, real-time tracking technologies, and affordable lab screenings. For example, P01 stated:

> *"I want the data, I don't just want to wear the watch and see the app, I want the data. Cause I want to link that data to my day, and to the stress, to the food, and to the exercise, to see what's going on and to see if I can learn something about why my blood pressure has been up for the last couple of years."*

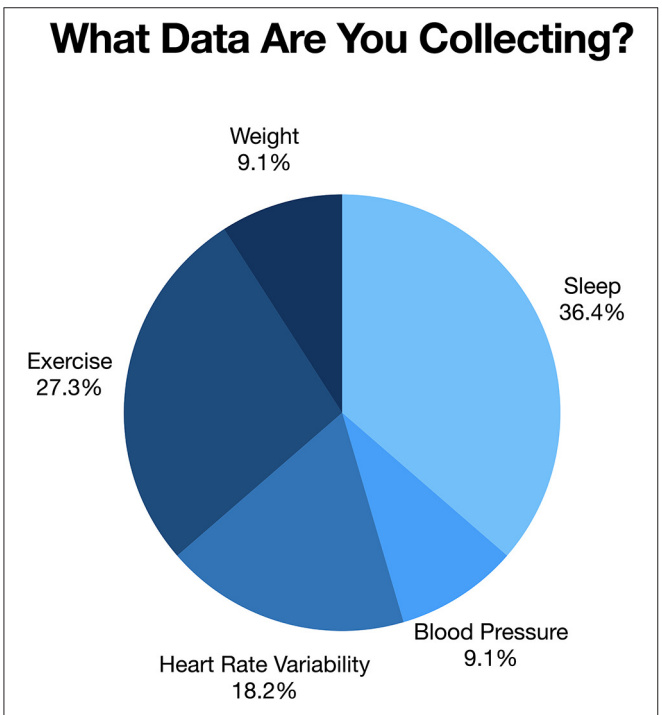


FIGURE 1 | Shows health domains that Apollo participants were tracking at the time of this study.

Participant P15 explained how technology can help track data for self-studies,

> *"I'm trying to start in the basics and I really like the Oura Ring because it just does it for me. Like even when I'm thinking about, I would like to maybe do a study on radiation-induced fatigue because it just knocked me on my butt and I'm just really curious about it... Is there a way that technology can do it for me, you know?"*

Along with advances in health-based technologies, participant-led research is growing because of a *sick*care system, which historically has disregarded patient input. Armed with shared frustrations of the healthcare they received and access to tools for self-study research, patients are empowered to act on their health conditions. Describing the conjunction of these factors, P07 stated:

> *"It has morphed into this idea that this group can become much more empowered rather than the medical system doing something to them, that then they have the power to act on it. I have all this data about myself, what do I do with it? And so, part of it is well what do you want to do with it? Let's create studies, let's create personalized plans for each individual."*

Participant P01 expressed that if they were to get the health outcomes they want, they would have to take matters into their own hands,

> *"So, until I got it that the only person that was going to drive my care was me and I'm not taking no for an answer, and if somebody is scratching their head, I'll find somebody that will dig deeper with me."*

Need for Access to Experts

For some in the Project Apollo Cohort, this is their first exposure to learning and applying the scientific method. Many were unfamiliar with the process of forming a hypothesis and research question and the steps of designing a study that could provide meaningful data. Moreover, the process of collecting data and skills necessary to analyze data and draw conclusions from that process is not trivial. As such, many emphasized the need to be walked through the scientific method by research experts to develop the foundational knowledge needed to do self-tracking and/or self-experimentation safely and ethically.

Apollo participants also felt it was vital when in the process of learning and doing participant-led research to receive feedback from experts including researchers and clinicians. For example, as P01 stated:

> *"Oh, God. I want to be handheld and walked through it every inch of the way... there's a lot of things in the digital universe that people do better than me. And I just, I know the limitations of my experience and my capabilities. And I can do things, but I just need step-by-step instructions."*

Project Apollo has added several researchers and clinical experts to support the self-study pursuits of its members, which

participants agreed is key for successful N-of-1 research. Areas participants expressed as important for access to experts included the protocol design *("It requires...input from other valuable support people, researchers, about what you could potentially encounter" P06*), data collection *("What are some good objective measures and what are some ways to track them?" P01*), and data analysis *("I could look at some patterns, but I immediately need that feedback." P05*) in their N-of-1 research studies.

To provide consistent foundational instruction about the scientific methods and responsible research practices, educational modules were adapted from the Building Research Integrity and Capacity (BRIC) curriculum developed by Dr. Camille Nebeker and made available to the Project Apollo Cohort (12). The adapted BRIC educational modules were made available online for the Cohort members to review in advance of planned face-to-face training sessions, which were designed to apply the concepts introduced via BRIC. Two face-to-face training sessions were convened to discuss the modules and begin the process of developing a research question, identifying measurement strategies, and creating a data collection and recording plan. Specific to the BRIC modules and group discussions, P04 exclaimed,

> *"Boy we need the training. I know that's where, after reading the threads in Slack, I know that's the push now, the BRIC [training modules], the realization that we've got to have training."*

Understanding of the scientific method takes time and applying the method to self-study takes practice and trial-and-error. Moreover, access to experts throughout the process was deemed critical.

> *"I'm not a statistician. I think it would be great to have people we could talk to with different expertise like that who could address, especially interaction between different factors. That, I'm not at all comfortable that I know how to do that" (P02).*

> And, *"at some point that question might be, 'how do I begin to answer this question?' and the answer to that might be, 'seek the insight of someone skilled in x, y, or z"* (P03).

The importance of access to researchers and clinicians includes how and when to share self-study results that may indicate the need to obtain medical attention. Participant P15 stated,

> *"But it's not like it's giving me information that's going to lead me to self-diagnosing and self-treating. Because I think that can be very dangerous, even with me as a nurse, like as Master's in Nursing, I don't feel comfortable doing anything without a doctor telling me to do it, especially with the cancer."*

Moving From *Sick*care Toward a *Health*care System

The transformational shift from *sick*care to healthcare involves integrative and personalized medicine supported by the patient's role in self-tracking and self-study. Both clinicians and patients must be actively involved to realize this paradigm shift.

A vision and mission of Project Apollo and its parent organization, the Precision Healthcare Ecosystem[2], is to create "a world of people empowered to realize optimal health" where the "doctor of the future is one's self" and subsequently, "transform healthcare through data-driven, patient-centered collaborative communities." The motivation to revolutionize the healthcare system such that it is tailored to the health experience of individuals through precision medicine and patient-led self-study. As stated by P07

> *"patient-led research can start driving us into a greater understanding by getting closer and closer to the unique lived and mysterious experience of each individual life."*

To realize this ambitious goal, medical education will need to change. As stated by P02,

> *"The medical education system hasn't yet changed sufficiently. I think it's changing with the existing model shifting to an individualized care model; but patients aren't in the middle of that equation; patients aren't even in the conversation. We're in a really exciting time given the technological advances, and although clinicians are experiencing a lot of burnout due to the current healthcare system - patients are experiencing patient burnout. Project Apollo provides a really great opportunity to move things forward and do what we all came here to do, which is promote health and live our best lives."*

From a clinician's perspective, the idea of individualized care may seem intuitive, as noted by P07

> *"It's not like I'm going use the same hands-on technique for every single person because their anatomy, their physiology, their life, everything is different. So, it has to be adapted. N-of-1 is the study of just one individual, and there's a lot of research right now being done on N-of-1 precision medicine and a lot of that is in the pharmaceutical grade, especially with designer drugs for cancer."*

Clearly, our current *sick*care system is not designed to support this level of individualized care and, as such, it will continue to take a toll on both patients and clinicians as the process of transformation takes place.

The impetus of Project Apollo was to explore whether patients who were already collecting data independently could be a collective force in shifting the health ecosystem. The idea of self-tracking and self-study maybe essential to transforming healthcare; however, while the concept may seem simple, in reality it is quite challenging. Independently, the patients who became the Project Apollo Cohort were navigating the complicated waters of the current *sick*care system and had developed expertise that collectively could help others avoid the frustrations they had experienced. Some of this expertise was in knowing what questions to ask and of whom, but it was also synthesizing the corpus of medical information amassed from various tests across a multitude of clinicians. Self-tracking, while

[2]https://precisionhealthcareecosystem.org/

it may not have been systematic or even labeled as such, was inherent to the Apollo Cohort members. As P02 recalled,

> *"this whole idea is a foundational part of our goal to generalize to large communities and perhaps people who know nothing about this or weren't aware of or have never maybe done a deep dive, reflective, introspective analysis of their own health and what could be better and things like that."*

The process of learning a more systematic approach to self-tracking involved learning new methods of collecting and making sense of personal health data, including new vocabulary. In P04's case, the phrase "*precision medicine*" was not familiar, and there was an excitement about taking control as expressed in this quote:

> *"Project Apollo, for us, represented an intriguing intellectually interesting endeavor... and a chance to break out of the limitations, get away from these predictions, get away from the statistics, and get into a level of medicine that's really about you, and not confined to a rushed 20-minute appointment."*

DISCUSSION

Digital health technologies and mobile health applications are integral to the success of Project Apollo self-studies and empowering patients in their health making decisions. In addition, peer-to-peer support and the creation and sense of being part of a community are also essential aspects of this work.

Implications for Patient-Centered Care

These results point to the possibility of patients not merely being "empowered" by professionals, but also taking leadership roles within their own care and, alongside professionals, advancing peer-to-peer support to one another. This has important implications both for understanding the role of patients in the health sciences and also on the future of care, particularly the active integration of peer-to-peer support.

With regard to the role of patients within health sciences, the Project Apollo Cohort could be viewed as a form of citizen science. Citizen science is an umbrella term with origins in the disciplines of ecology, ornithology, and astronomy that have involved the public in conservation and crowdsourcing (13). More recently, citizen science has moved into the health sector (14). As in other fields, within the realm of health, citizen science encompasses a very broad array of activities (15, 16). On one end of the spectrum, citizen scientists are involved in providing support to research efforts via volunteering their time and interest toward a well-specified and prescribed task established by researchers (17). For example, researchers have developed Fold-It[3], a "game" that enables people to work through the "puzzle" of finding different ways that amino acids/proteins can fold over on themselves to create different types of protein structures; it is a topic that is vital for understanding issues such as antibodies and care (18). On the opposite extreme are citizen-/patient-led efforts whereby the priorities, work, and efforts are completely driven by and for the persons experiencing the issue. For example, the #WeAreNotWaiting[4] community of patients with type I diabetes is a self-organized, highly networked, modular group of individuals with type I diabetes (or parents of children with type I diabetes) who found ways to drive advancements in their self-care (19). Some concrete examples of solutions that grew out of this community include Nightscout[5], an open source tool used to gain access to a patient's continuous glucose monitor data, and the Open Artificial Pancreas System (OpenAPS)[6], which, building on Nightscout, is a closed loop artificial pancreas system algorithm created via self-motivated patients and those who care for them (20). In between these two extremes are truly collaborative efforts in which power and agency is shared between traditional professionals and patient/citizen scientists. For example, the Robert Wood Johnson Foundation-funded Opening Pathways Project[7] was a research effort led by Principal Investigator (PI) Dana Lewis, a patient innovator who leads the OpenAPS community, with traditional professors playing roles of Co-PIs, Hekler & Johnston. The focus of that project was on advancing new pathways for non-traditional researchers to advance the care and health of patients. Across all of these domains, it is common for a community of individuals with shared interests, passions, or needs to come together to work toward a shared future vision of health.

Based on the wide range of ways in which citizen science manifests, from citizens supporting researchers to citizens running efforts without any traditional professional support, there are also a wide range of methods and tools used to advance these efforts. For example, Fold-It involves robust use of data informatics, human-centered and game design expertise, and robust knowledge on surfacing difficult and intractable challenges in understanding proteomics to be combined into a fun, engaging, challenging "puzzle" that any person interested in solving puzzles can engage in. The OpenAPS community, on the other hand, uses a mixture of techniques such as open source software development practices (e.g., robust use of GitHub), community "tuning" strategies[8] for iteratively and rigorously identifying and vetting assumptions related to any technologies developed by the OpenAPS community to ensure they are safe[9], and also open science practices[10] related to data sharing, data science best practices, and open data repositories, such as those supported on the Open Humans service (21).

Turning now to Project Apollo, the Apollo cohort are engaging in a wide range of hypothesis-driven "small data" approaches (22). There are a wide range of methods that fit into a small data paradigm. On one extreme, there are methods that are simple for most people to use and engage with, such as journaling and gathering of qualitative data. The value here is that most people can do it, but it may not necessarily produce as rigorous results in terms of inferring and predicting future responses of

[3]https://fold.it/portal/

[4]https://www.healthline.com/health/diabetesmine/innovation/we-are-not-waiting#1

[5]http://www.nightscout.info/

[6]https://openaps.org/

[7]http://openingpathways.org/

[8]http://openingpathways.org/communal-tuning

[9]http://openingpathways.org/is-it-safe

[10]https://opensource.guide/best-practices/

individuals. In the middle are quantitative self-tracking, more formalized hypothesis testing within an individual's time series, and non-randomized single case experimental studies used to glean insights on the impact of different decisions. These balance ease with rigor. At the more rigorous end, when very specific and concrete questions are being asked, are randomized N-of-1 cross-over designs meant to test the influence of various actions (e.g., taking a medication or not, choosing to eat a certain food or not) on targeted outcomes to use of system identification, which is a technique used by control systems engineers to identify computational models of complex, dynamic phenomena. The Apollo cohort appears to be engaging in almost all except the most extreme in terms of technical requirements (system ID) to advance understanding, as well as the quality of their own health.

The Project Apollo cohort, coupled with broader trends toward patient-centered health, points to the potential value and need for further advancing a small data paradigm to better support patient self-study. By self-study we mean the use of these small data methods by a person to help them better understand themselves toward achieving self-defined goals. Within a small data paradigm, success is defined for each person, such as reduction in symptoms, improved function, or increased self-understanding of one's own condition. By extension, this enables a clear alignment on the self-interests of patients/persons experiencing a condition, clinicians, and researchers. They also create space for different ways of knowing and understanding a person's health condition than is common from traditional evidence-based practice. Specifically, as the definition of evidence-based practice (provided at the beginning of the paper) suggests, it relies heavily on generalizable knowledge gleaned from the scientific literature and prior individuals coupled with the clinical expertise of clinicians to translate that wealth of knowledge into personalized recommendations and steps forward for each patient. This classic approach provides little structure or place for incorporating knowledge and insights from the person themselves experiencing a condition and their self-studies. A small data approach provides a structure for honoring the unique knowledge and insights self-study can bring into advancing decision-making around health issues.

As is likely obvious when looking at this spectrum of small data methods, the amount of training and prior knowledge needed to use the methods is one key tradeoff (e.g., journaling can be done by practically anyone; system ID requires deep specialized knowledge in mathematics, programming, and understanding of robust study designs to systematically test computational models). The complementary tradeoff, of course, is the capacity for these various methods to provide more rigorous insights from data for guiding thinking and decision-making related to complex phenomena (e.g., journaling has a higher risk of drawing spurious conclusions compared to N-of-1 cross-over trials or system ID studies). Based on this, a key implication from the work of the Project Apollo cohort is the need for a wide range and diversity of training materials (e.g. tailored education) and resources (e.g., health coach, professional researchers) that support the many different ways in which patients may engage in self-study, from basic journaling to rigorous predictive mathematical models designed for each person. This is not only the case for the patients themselves, but also points to the need for professionals, particularly clinicians, to learn how to understand, honor, and integrate this type of evidence into their clinical practice and support. It also may point to a new type of healthcare service, a self-study coach, who, alongside a health coach, physician, nurses, and others on the care team, could take the time to help individuals engage in appropriate self-study to facilitate self-learning and not over-generalize results, either to themselves and definitely not to others, as doing N-of-1 study does not, alone, produce transportable knowledge (22).

Moving now to the stated importance of community that emerged from the Project Apollo cohort, this work points to the possibility that "patient-centered" may, in fact, be too limiting of a concept. In particular, the Project Apollo members clearly highlight ways in which support and care can, and perhaps should, be offered that go well beyond the traditional dyadic relationship between patients and their providers, or even patients surrounded by providers. Indeed, the work points to the value patients receive when they can work and discuss their experiences with other peers. While "care" has always been identified within healthcare, in many ways the desires, interests, and active cultivation of a caring community of patients highlights that the professionalization of care may not be adequate, or even appropriate, compared to what people need. This fits with broader trends and interests in peer-to-peer support, such as the work of Susannah Fox[11] in supporting peer-to-peer advice online and Rajiv Mehta advancing an "Atlas of Caregiving[12]," whereby individuals learn to understand and cultivate care and caring within and across families and communities.

As with self-study though, peer-to-peer requires further reflection and training. For example, a key risk of peer-to-peer support involves a person translating personal history and beliefs on what was helpful for them into explicit recommendations of activities that others should engage in. As health sciences, writ large, demonstrates, it is no small task to determine if a recommendation is indeed an appropriate, safe, and effective recommendation for others to use. Peer-to-peer is not an appropriate venue for offering treatment recommendations and the like, as the underlying epistemology does not support that type of offering. With that said, peer-to-peer support does offer a place for care, warmth, and shared experiences to be communicated. Furthermore, peers can feasibly be excellent sounding boards for one another to help each other think through plausible pathways forward on a given condition, particularly when determining the right diagnostic, prevention, treatment, and health promotion options are unclear. In this domain, peers can share their stories and experiences and, grounded in a shared recognition of individual choice and agency, patients can then engage in balancing their understanding with what they might learn from self-study, their clinician, or the external evidence-base. It is these latter benefits

[11] https://susannahfox.com/
[12] www.atlasofcaregiving.org

of support, shared experience, and what not that the Project Apollo group is seeking from one another, not evidence-based recommendations. Based on this, future work points to the need for further understanding, reflection, and integration on how patient peers can and should interact with one another, their care team, and broader community members (e.g., health scientists and ethicists with expertise that could be valuable for patients to think through their self-study).

Overall, this work points to two ways of knowing and advancing ones' health that fall outside of the realm of the external evidence from the health sciences and clinical expertise. As such, the work of Apollo acknowledges potential limitations of the classic definition of evidence-based medicine. This work points to four plausible ways of knowing: external evidence, clinical expertise, self-study, and peer-to-peer support, as a possible foundation for a new type of evidence-based medicine, what might be thought of as evidence-based practice 2.0 (see **Figure 2**). Each of these ways of knowing have different strengths and limitations that, when combined, are highly complementary. For example, classical health science provides a robust "warm start" that enables people to quickly rule out different types of diagnoses and treatment options (22). Clinical expertise provides insights on patterns of responses across the many patients that clinicians see to further improve decision-making and rule out different diagnoses, treatments, and actions. Self-study provides a structure for enabling a person to identify and test assumptions around diagnosis and treatment specifically for themselves. Finally, peer-to-peer offers insights on plausible hypotheses, beliefs, and coping strategies that are not yet well-studied or understood in the scientific literature or part of clinical practice. When used together, these four references balance out the relative strengths and limitations of one another toward more robust, personalized decision-making.

There is a great opportunity to improve care if robust approaches to self-study (alone, in partnership between patients and clinicians, and even partnerships between patients, clinicians, and researchers) and peer-to-peer support can be defined. As alluded to, there is an opportunity to provide complementary knowledge and insights to that which is offered from traditional scientific methods and clinical expertise. There is also the opportunity for improving communication and understanding between the lived experience of persons and clinicians seeking to support them. This type of approach could also provide a foundation that enables patients to feel more capable of understanding themselves and finding solutions for their personal health needs. Finally, if these four references could be established as working synergistically together, they could enable new insights, ways of thinking about health and care, and strategies for improving health to emerge.

While the opportunities are great, there are also a myriad of challenges that this vision of care implies. At the most basic level, this type of approach, to the best of our knowledge, does not yet exist. Specifically, we are unaware of any health organization that actively and consciously balances knowledge, skills, and expertise across external scientific evidence, clinical expertise, self-study, and peer-to-peer support. This means that new skills, training, and even mindsets for all relevant stakeholders (e.g., patients, providers, researchers, and payers) are likely needed. Ideally, this training would focus on ensuring appropriate conclusions are drawn from each referent. For example, external scientific evidence produces "on average" insights, but that does not necessarily equate to an individual; in contrast, self-study produces insights that could be valuable for a person, but that does not mean those insights would be transportable to anyone else. Furthermore, traditional clinical training is focused on providing a clinician with a structure for calibrating between the scientific literature and patterns they have observed over time among their patients; this means that clinical expertise and intuitions could be inappropriate to be relied upon if either there is little research on something a person is experiencing or if a clinician has little prior personal exposure to other patients experiencing a given phenomenon. Finally, peer-to-peer provides a way for individuals to explore different ways of understanding themselves in context and develop alternative beliefs around health, but those alternative perspectives, even if appropriate for a given group of people, are not necessarily appropriate for others; thus, peer-to-peer can be thought of as a great venue for generating new hypotheses and strategies to move forward, but not as a venue for testing ideas or gaining reliable, rigorously vetted recommendations. New approaches that support this calibration are needed, along with appropriate training and guidance on this work.

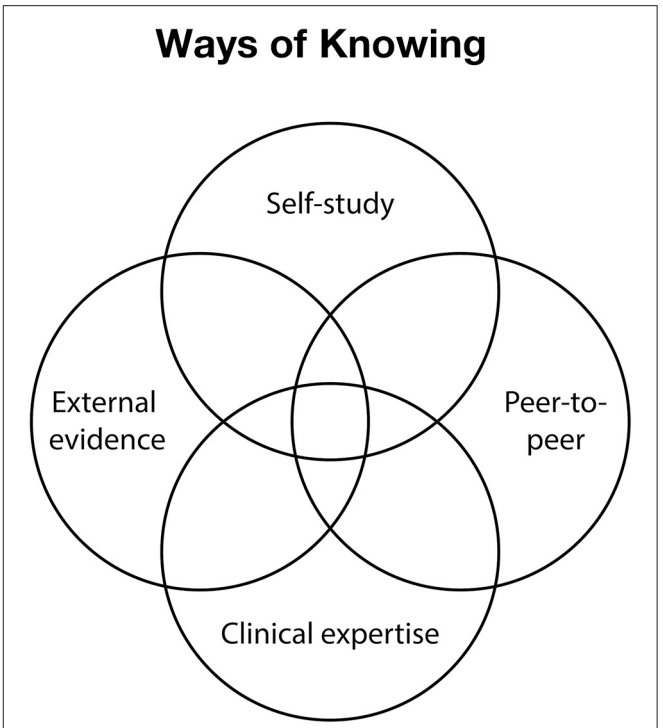


FIGURE 2 | Depicts four plausible ways of knowing presented as a Venn Diagram with external evidence, clinical expertise, self-study, and peer-to-peer support forming a foundation for a new type of evidence-based medicine.

Beyond basic use of the methods, this approach also introduces new challenges to traditional approaches for considering ethical practices and, by extension, appropriate

oversight and regulation. Traditional regulations largely rely on the implicit assumption that the professionals generating scientific evidence and clinical expertise have the requisite information needed to guide ethical decision-making. Moreover, federal regulations for the protection of human participants in research do not align with participant-led self-study forms of citizen science (4, 23). Based on this, the activities for regulating self-study and related ethical practices largely conform to the monitoring and regulation of professionals and their practices, leaving this novel form of citizen science unregulated. That is, individuals engaging in self-study and peer-to-peer support are by definition working outside the realm of existing regulation, and it is unclear the extent to which existing principles for the ethical conduct of research pertains (4). There is a clear need for strategies that enable an individual patient/person to conduct an ethical self-review, including assessment of the potential risks and benefits, to reduce the likelihood of negative unintended consequences either to oneself or to others. Furthermore, for peer-to-peer, there is a need for structures that enable checks on beliefs and also appropriate practices on what is appropriate vs. in appropriate in terms of peer-to-peer support. For example, in peer-to-peer circumstances, it is appropriate for people to share stories and current thinking on what they are doing and how they think their actions are resulting in positive changes. It is inappropriate for peers to engage in providing advice and recommendations, particularly if current beliefs are largely grounded in one's own experience. As we develop these practices, key lessons can be learned from how open source efforts, such as Wikipedia, function in terms of governance (24) as, almost by definition, an overly top-down regulatory process will not only be insufficient, it is likely inappropriate for this type of work. Interestingly, insights from philosophy of science, particularly on thinking through ways to develop trustworthy scientific consensus in ways that do not use top-down structures of regulation, could be another starting point for inspiration on different types of ethical practices (25).

OPEN QUESTIONS AND LIMITATIONS

We conclude our discussion with five provocative questions along with limitations to our study.

1. How can clinicians support patient agency? Clinical support for agency and autonomy for patients is fundamental in the practice of medicine but has been proven to be a very challenging and difficult task (26). The healthcare system that we are all a part of does not seem to be set up to allow this to happen naturally. Paternalism, whether good or bad, is a pervasive attribute that exists within the culture of practice of medicine. It is imperative that we look at this as we try to form a new paradigm of the doctor-patient relationship.
2. What are the barriers and facilitators to clinician support? It is essential that we consider a complete and holistic view of the patient. This includes all areas and aspects of patients' lives that impact health and wellness, including social determinants such as socioeconomic status, community, family dynamics, race, and culture. This gives a broader and more rich contextual relational understanding of the patient (27). Unfortunately, a holistic perspective is not emphasized and modeled enough in medical school, nor is the importance of the patient's role within the doctor-patient relationship included in the curriculum.
3. What is needed to develop clinician education to advance precision healthcare that involves patients throughout the process? Perhaps a more important factor regarding this topic is to take a broad holistic view of the healthcare system. The current system is multifaceted in complexity, and over time it seems to have undermined the doctor-patient relationship (28). There is over-reliance on information technology, medical devices, and procedures and less time spent nurturing empathy, compassion, and connection with the patient (29). The result is a growing distance in the doctor-patient relationship and a mirrored discontent amongst both participants (30).
4. How to foster autonomy in patients when one of the most common frustrations from the clinicians themselves is their own feeling of lack of autonomy within the healthcare system. Physician burnout is at an all-time high and its consequences are disastrous (31). On many measures, the actual clinicians have worse healthcare outcomes than the patients they are treating (32). This, in turn, creates a downward spiral negatively effecting the entire system (33). As designed, the healthcare system is not sustainable. It is time for change as the system needs to be designed to care for all involved.
5. What is the vision of Apollo moving forward? The essence of Project Apollo is a co-creative and collaborative nature of building community amongst patients, providers, and researchers. The cohort also prototyped different ways of experimenting with enhancing patient autonomy and agency. This will be accomplished through educational modules, self-tracking, retrospective analysis of healthy behavior changes, and utilizing technology not as a barrier, but as a tool to augment the connection the patient has with their physician. This can create a new paradigm empowering patients and physicians to have a more enhanced doctor-patient relationship. A co-creation of a new model of the way care is delivered can be designed benefiting all those within the healthcare system.

There are limitations to this study. The Apollo Cohort consists of a small group of about a dozen patients and, while those who agreed to be interviewed ($N = 9$) represent a majority of the group, it is not appropriate to view these data as representative of self-study as it relates to patient-centered care. In qualitative research, a goal is to reach saturation of the data to have confidence about the phenomena under study and, while our sample was a representation of the Apollo Cohort, the themes identified may not be generalizable to others involved with self-tracking or participant-led N-of-1 studies. Another consideration is that we, the authors, have been involved in discourse with the Project Apollo Cohort from the early days of its formation and it is feasible that we have influenced the community and their conceptualization, just as they have influenced us. To mitigate

the potential bias introduced, we followed best practices in qualitative research, such as including several data sources to inform our finding, involving two researchers coding the data, and engaging participants and peers in the review of our results. Given the novelty of patient/participant-led research and how it presents (e.g., DIY, lead innovators, Quantified Self), we believe the potential value in learning from the Project Apollo cohort experiences is noteworthy with respect to self-study and potential impact on the current sick/healthcare system.

CONCLUSION

Our study revealed how N-of-1 research, in the form of self-tracking and self-study plus communal support, can contribute to one's health journey and create a pathway for active collaboration in advancing precision healthcare. Facilitators to engaging patients in self-study include motivation, albeit stemming from frustration, and a safe community where support is derived from one another. Additional support in the form of access to experts who can help with important foundational knowledge necessary to conduct meaningful self-study is critical. Moreover, increasing awareness of healthcare professionals about the potential value of collaborating with patients who are advancing health knowledge through self-study will be a key factor in the success of patient-centered care and in shifting the paradigm from *sick*care to collaborative *health*care.

DATA AVAILABILITY STATEMENT

The datasets used for this study are qualitative and will be available via the Qualitative Data Repository (https://qdr.syr.edu/) once redacted to protect the identity of participants.

ETHICS STATEMENT

This study involved human participants and was reviewed and verified as exempt by the University of California San Diego, Institutional Review Board. Participants provided their verbal informed consent prior to being interviewed.

AUTHOR CONTRIBUTIONS

CN conceptualized the study and contributed to data analysis, writing, and editing of the manuscript. EH contributed to writing and editing the manuscript. BW contributed to data collection, analysis, and writing and editing the manuscript. MK contributed to writing and editing the manuscript.

FUNDING

This research was supported by intramural funding from the UC San Diego Institute of Public Health pilot grant program and the Walton Family Foundation (Grant index numbers: CFMMKN1 and 86N03A).

ACKNOWLEDGMENTS

We acknowledge the meaningful contributions of the Project Apollo Cohort and thank them for participating in our interviews, providing source documents, inviting our observation of group meetings and for attending a presentation of this research prior to our manuscript submission. We also recognize Mad Ball, Executive Director of Open Humans Foundation, who reviewed and provided valuable feedback on our manuscript.

REFERENCES

1. Marvasti FF, Stafford RS. From sick care to health care - reengineering prevention into the U.S. System. *New Engl J Med*. (2012) 367:889. doi: 10.1056/NEJMp1206230
2. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. 1996. *Clin Orthop Relat Res*. (2007) 455:3–5. doi: 10.1136/bmj.312.7023.71
3. Eklund JH, Holmström IK, Kumlin T, Kaminsky E, Skoglund K, Höglander J, et al. "Same same or different?" a review of reviews of person-centered and patient-centered care. *Patient Educ Couns*. (2019) 102:3–11. doi: 10.1016/j.pec.2018.08.029
4. Grant AD, Wolf GI, Nebeker C. Approaches to governance of participant-led research: a qualitative case study. *BMJ Open*. (2019) 9:e025633. doi: 10.1136/bmjopen-2018-025633
5. Lewis D, Leibrand S. Real-world use of open source artificial pancreas systems. *J Diabetes Sci Technol*. (2016) 10:1411. doi: 10.1177/1932296816665635
6. Pauwels E, Denton SW. *The Rise of the Bio-Citizen*. (2018). Available online at: https://www.wilsoncenter.org/sites/default/files/media/documents/article/rise_of_biocitizenfinal.pdf
7. Vizer LM, Eschler J, Koo BM, Ralston J, Pratt W, Munson S. "It's not just technology, it's people": constructing a conceptual model of shared health informatics for tracking in chronic illness management. *J Med Internet Res*. (2019) 21:e10830. doi: 10.2196/10830
8. Cohen J. *The Patient of the Future*. (2012). Available online at: http://joncohen.org/2012/02/21/the-patient-of-the-future/ (accessed November 21, 2019).
9. Dedoose Version 8.0.35. 2018. *Dedoose. 2018*. (2018). Available online https://www.dedoose.com/userguide/appendix (accessed August 25, 2019).
10. Saldana J. *The Coding Manual for Qualitative Researchers*. SAGE (2009).
11. Reblin M, Uchino BN. Social and emotional support and its implication for health. *Curr Opin Psychiatry*. (2008) 21:201–5. doi: 10.1097/YCO.0b013e3282f3ad89
12. Nebeker C, Lopez-Arenas A. Building research integrity and capacity (BRIC): introducing community health workers to the scientific method and ethical research practices. *J Microbiol Biol Educ*. (2016) 17:41–5. doi: 10.1128/jmbe.v17i1.1020
13. Wiggins A, Crowston K. From conservation to crowdsourcing: a typology of citizen science. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. Kauai, HI: IEEE (2011). p. 1–10.
14. Wiggins A, Wilbanks J. The rise of citizen science in health and biomedical research. *Am J Bioethics*. (2019) 19:3–14. doi: 10.1080/15265161.2019.1619859
15. Heigl F, Kieslinger B, Paul KT, Uhlik J, Dörler D. Toward an international definition of citizen science. *Proc Natl Acad Sci USA*. (2019) 116:8089–92. doi: 10.1073/pnas.1903393116
16. Mervis J. Citizen science needs to look more like society, report says. *Science*. (2018). doi: 10.1126/science.aav9305. Available online at: https://www.sciencemag.org/news/2018/11/citizen-science-needs-look-more-society-report-says

17. Haklay M. Citizen science and volunteered geographic information: overview and typology of participation. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice.* Netherlands: Springer (2013). p. 105–22.
18. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. *Nature.* (2010) 466:756–60. doi: 10.1038/nature09304
19. Omer T. Empowered citizen 'health hackers' who are not waiting. *BMC Med.* (2016) 14:118. doi: 10.1186/s12916-016-0670-y
20. Lewis D. History and perspective on diy closed looping. *J Diabetes Sci Technol.* (2019)13:790–3. doi: 10.1177/1932296818808307
21. Greshake Tzovaras B, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open humans: a platform for participant-centered research and personal data exploration. *GigaScience.* (2019) 8:giz076. doi: 10.1093/gigascience/giz076
22. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. *BMC Med.* (2019) 17:133. doi: 10.1186/s12916-019-1366-x
23. Vayena E, Blasimme A. Health research with big data: time for systemic oversight. *J Law Med Ethics.* (2018) 46:119–29. doi: 10.1177/1073110518766026
24. Forte A, Larco V, Bruckman A. Decentralization in wikipedia governance. *J Manag Inform Syst.* (2009) 26:49–72. doi: 10.2753/MIS0742-1222260103
25. Miller B. When is consensus knowledge based? Distinguishing shared knowledge from mere agreement. *Synthese.* (2013) 190:1293–316. doi: 10.1007/s11229-012-0225-5
26. Murgic L, Hébert PC, Sovic S, Pavlekovic G. Paternalism and autonomy: views of patients and providers in a transitional (post-communist) country. *BMC Med Ethics.* (2015) 16:65. doi: 10.1186/s12910-015-0059-z
27. Entwistle VA, Carter SM, Cribb A, McCaffery K. Supporting patient autonomy: the importance of clinician-patient relationships. *J Gen Int Med.* (2010) 25:741–5. doi: 10.1007/s11606-010-1292-2
28. Salvagioni DAJ, Melanda FN, Mesas AE, González AD, Gabani FL, de Andrade SM. Physical, psychological and occupational consequences of job burnout: a systematic review of prospective studies. *PLoS ONE.* (2017) 12:e0185781. doi: 10.1371/journal.pone.0185781
29. Weiner M, Biondich P. The influence of information technology on patient-physician relationships. *J Gen Int Med.* (2006) 21:35–9. doi: 10.1111/j.1525-1497.2006.00307.x
30. Katz J. *The Silent World of Doctor and Patient.* Baltimore, MD: Johns Hopkins University Press (2002).
31. Fred HL, Scheid MS. Physician burnout: causes, consequences, and (?) cures. *Texas Heart Inst J.* (2018) 45:198–202. doi: 10.14503/THIJ-18-6842
32. Wurm W, Vogel K, Holl A, Ebner C, Bayer D, Mörkl S, et al. Depression-burnout overlap in physicians. *PLoS ONE.* (2016) 11:e0149913. doi: 10.1371/journal.pone.0149913
33. Panagioti M, Geraghty K, Johnson J, Zhou A, Panagopoulou E, Chew-Graham C, et al. Association between physician burnout and patient safety, professionalism, and patient satisfaction. *JAMA Int Med.* (2018) 178:1317. doi: 10.1001/jamainternmed.2018.3713

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 25 June 2020
doi: 10.3389/fpubh.2020.00260

Feasibility, Acceptability, and Influence of mHealth-Supported N-of-1 Trials for Enhanced Cognitive and Emotional Well-Being in US Volunteers

Richard L. Kravitz[1†], Adrian Aguilera[2], Elaine J. Chen[3], Yong K. Choi[4], Eric Hekler[5], Chris Karr[6], Katherine K. Kim[4], Sayali Phatak[7], Sayantani Sarkar[4], Stephen M. Schueller[8], Ida Sim[9], Jiabei Yang[10] and Christopher H. Schmid[10]*

[1] Division of General Medicine, UC Davis Health, Sacramento, CA, United States, [2] School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States, [3] WNYC Public Radio, New York, NY, United States, [4] Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, United States, [5] Center for Wireless & Population Health Systems, Qualcomm Institute, Department of Family Medicine & Public Health, Design Lab, University of California, San Diego, San Diego, CA, United States, [6] Audacious Software, Inc., Chicago, IL, United States, [7] College of Health Solutions, Arizona State University, Tempe, AZ, United States, [8] Department of Psychological Science, University of California, Irvine, Irvine, CA, United States, [9] Department of Medicine, University of California, San Francisco, San Francisco, CA, United States, [10] Department of Biostatistics and Center for Evidence Synthesis in Health, School of Public Health, Brown University, Providence, RI, United States

OPEN ACCESS

Edited by:
Giuseppe Fico, Polytechnic University of Madrid, Spain

Reviewed by:
Rebeca Isabel García-Betances, Polytechnic University of Madrid, Spain
Francisco Lupiáñez-Villanueva, Open University of Catalonia, Spain

***Correspondence:**
Richard L. Kravitz
rlkravitz@ucdavis.edu

[†]Beyond the first and last author, all authors are listed alphabetically

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 24 December 2019
Accepted: 22 May 2020
Published: 25 June 2020

Citation:
Kravitz RL, Aguilera A, Chen EJ, Choi YK, Hekler E, Karr C, Kim KK, Phatak S, Sarkar S, Schueller SM, Sim I, Yang J and Schmid CH (2020) Feasibility, Acceptability, and Influence of mHealth-Supported N-of-1 Trials for Enhanced Cognitive and Emotional Well-Being in US Volunteers. Front. Public Health 8:260. doi: 10.3389/fpubh.2020.00260

Although group-level evidence supports the use of behavioral interventions to enhance cognitive and emotional well-being, different interventions may be more acceptable or effective for different people. N-of-1 trials are single-patient crossover trials designed to estimate treatment effectiveness in a single patient. We designed a mobile health (mHealth) supported N-of-1 trial platform permitting US adult volunteers to conduct their own 30-day self-experiments testing a behavioral intervention of their choice (deep breathing/meditation, gratitude journaling, physical activity, or helpful acts) on daily measurements of stress, focus, and happiness. We assessed uptake of the study, perceived usability of the N-of-1 trial system, and influence of results (both reported and perceived) on enthusiasm for the chosen intervention (defined as perceived helpfulness of the chosen intervention and intent to continue performing the intervention in the future). Following a social media and public radio campaign, 447 adults enrolled in the study and 259 completed the post-study survey. Most were highly educated. Perceived system usability was high (mean scale score 4.35/5.0, SD 0.57). Enthusiasm for the chosen intervention was greater among those with higher pre-study expectations that the activity would be beneficial for them ($p < 0.001$), those who *obtained* more positive N-of-1 results (as directly reported to participants) ($p < 0.001$), and those who *interpreted* their N-of-1 study results more positively ($p < 0.001$). However, reported results did not significantly influence enthusiasm after controlling for participants' interpretations. The interaction between pre-study expectation of benefit and N-of-1 results interpretation was significant ($p < 0.001$), such that those with the lowest starting pre-study expectations reported greater intervention enthusiasm when provided with results they interpreted as positive.

We conclude that N-of-1 behavioral trials can be appealing to a broad albeit highly educated and mostly female audience, that usability was acceptable, and that N-of-1 behavioral trials may have the greatest utility among those most skeptical of the intervention to begin with.

Keywords: N-of-1 trial, single patient trial, mobile health, digital health, behavioral health, psychological well-being

INTRODUCTION

Accumulating evidence supports the adoption of various habits and behaviors to improve cognitive and emotional well-being. For example, Americans are urged to be more physically active, reduce stress, and connect socially (1–5). One problem with the plethora of recommendations is that individuals may be confused about which behaviors to adopt first. They can turn to trusted experts, but most of the evidence upon which those experts rely is based on studies that generate average effects. Evidence derived from groups may not necessarily apply to the individual because of heterogeneity in person-level and contextual factors (e.g., age, gender, personal preferences, community resources, and fit with a person's life or workflow) (6, 7). Furthermore, the impact of any behavior is likely to yield modest benefits, potentially accumulating over time. More precise information on the likelihood of benefit *at the individual level* could help motivate long term behavior change.

Certainly, many people can and do assess the personal value of behavioral interventions informally through trial and error. Some, however, may be interested in a more rigorous approach. N-of-1 trials are multiple crossover trials conducted in a single individual (8). While sharing some characteristics with informal "trials of therapy," they lend rigor to the assessment and, along with parallel group randomized controlled trials, are ranked at the top of the so-called evidence hierarchy by experts (9, 10). They have been used extensively in clinical psychology and medicine (11–17). For fast-acting, short-lived behavioral interventions expected to influence near-term outcomes, N-of-1 trials are arguably the most direct method for inferring the effect of treatment on an individual. These trials may appeal to persons who wish to gain greater certainty that the behavioral intervention under consideration actually does (or does not) have benefit for them.

Despite their theoretical appeal, N-of-1 trials have gained limited traction among clinicians and the general public (18). Part of the reason may be that when implemented according to the highest scientific standards (which may include blinding, use of complex outcome measures, strict attention to adherence, etc.), many potential participants will decide that the likely benefits (in terms of insights and motivation) are simply not worth the trouble. However, we and others have demonstrated that the reach and feasibility of N-of-1 trials may be extended through use of mobile health (mHealth) technologies; in our own recent study of patients with chronic pain, 88% of patients starting an n-of-1 trial reported that the mobile app was "extremely or very helpful."(19) Another barrier may be the absence of scalable tools that allow non-scientists to conduct systematic evaluations of behavioral interventions on themselves (20).

We conducted this study to determine whether an mHealth supported N-of-1 trial assessing simple behavioral interventions for improving short-term cognitive and emotional well-being was feasible and perceived as beneficial. Specifically, we asked:

- Will members of the general adult population participate in an mHealth-facilitated behavioral N-of-1 trial?
- How do participants rate the usability of the mHealth N-of-1 trial system?
- To what extent are participant's attitudes toward the intervention and intentions to persist with it influenced by trial participation? Specifically,
  - Upon trial completion, how is enthusiasm for the chosen intervention related to expectations of benefit from the intervention, to the N-of-1 results themselves (as reported to the patient in terms of the difference in outcomes on days assigned to the intervention vs. days on their usual routine), and to the participant's interpretation of their N-of-1 results?

In addressing these questions, we sought to learn more about the utility of N-of-1 trials, the ways in which such trials affect participants' subsequent attitudes and behavioral intentions, and their prospects for broader adoption by the medical and behavioral community.

METHODS

Design overview

A national convenience sample of adult volunteers was recruited to engage in a 30-day single person (N-of-1) trial comparing the effects of one of four behavioral interventions on self-reported stress, focus, and happiness. Participants selected an intervention and were assigned for 30 days to randomly sequenced five-day periods performing the chosen activity and their "usual routine." Outcome measures were collected via secure text messaging. This report focuses on the 259 subjects who completed a post-study survey. Ethics approval was granted through the UC Davis Institutional Review Board (**IRB ID 1255435-4**).

Eligibility and Recruitment

We promoted the study through social media and The Brian Lehrer Show (WNYC Public Radio). Potentially interested

subjects were directed to the study website (studyofme.org), where they were given the opportunity to watch videos introducing the study and asked to select an activity of interest. Available activities included: (1) deep breathing meditation; (2) gratitude journaling; (3) physical activity; and 4) performing acts of kindness for strangers. Drawing on cognitive-behavioral techniques and positive psychology, activities were selected to be simple and easy to apply repeatedly (21–24). Volunteers were eligible if they were US adults > = 18 years, owned a smartphone or had regular access to the internet, and were interested in committing to a 30-day N-of-1 trial. In addition, subjects were encouraged to try an activity that they were not already doing, and if they were considering vigorous physical activity, they were advised to "first consult your doctor, especially if you have a chronic health problem, recurring injury, or are pregnant or nursing."

Baseline Survey

After confirmation of eligibility and provision of online informed consent, participants completed a baseline questionnaire asking for contact information (mobile phone number and valid email, both of which were deleted from the dataset prior to analysis); time zone (so that study reminders would go out at the right time of day); N-of-1 trial start date within the next 7 days; demographic information (ethnicity, race, gender, education level, and household size); and several questions concerning experience with self-tracking and interest in the chosen activity.

N-of-1 Trial Design and Conduct

All participants had the opportunity to read text and view a video providing detailed instructions on their chosen activity. Computer-generated N-of-1 trial sequences (e.g., UAUAAU, where U indicates 5 days performing usual routine and A indicates 5 days performing the chosen activity) were issued for each subject beginning on their chosen start date and continuing for 30 consecutive days. We used 5-day treatment periods as a compromise between the dictates of behavioral science (which would favor longer periods, to allow for adequate ramp-up and wash-out) (8) and statistical power (which would favor a greater number of switches between treatments). Participants received a text message through the HealthySMS system (25–27) each evening asking for ratings of stress, focus, and happiness for the day just finished and announcing tomorrow's activity.

Within 1 week of N-of-1 trial completion, HealthySMS sent participants a text message with a link to their personalized results (example provided in **Figure 1**).

Measures

Daily stress, focus, and happiness were each assessed with a single-item question sent each evening by text messaging: (1) On average, how stressed were you today? Please select a value from 0 (not at all stressed) to 10 (extremely stressed); (2) On average, how well were you able to focus? Please select a value from 0 (not able to focus at all) to 10 (extremely focused throughout the day); (3) On average, how happy were you today? Please select a value from 0 (not happy at all) to 10 (extremely happy throughout the day). Single-item measures of these constructs are typically used in studies that require daily responses by participants to reduce participant burden (28), and have demonstrated good reliability and validity when compared to longer measures (29–31).

At the end of the 30-day period, participants were sent a post-study questionnaire requesting completion of the System Usability Scale (32) (Cronbach's alpha in the sample, 0.83), and four questions assessing: (1) *pre-study expectations of benefit* of the chosen activity ("Before you started your personalized experiment, how confident were you that ACTIVITY would be beneficial for you?" 1 = not-at-all confident…5 = extremely confident); (2) *post-study* interpretation of results ("What is your best guess about what the RESULTS of your personalized experiment mean? 4 = highly beneficial, 3 = somewhat beneficial, 2 = minimally beneficial, 1 = not beneficial); (3) *post-study* perceptions of activity helpfulness ("Now that you have completed your personalized experiment, how helpful do you think ACTIVITY was for you? (1 = not at all helpful…5 = extremely helpful); and (4) *post-study* behavioral intentions ("Based on your personalized experiment, how likely are you to continue doing ACTIVITY on a regular basis over the next six months? 1 = not-at-all likely…5 = extremely likely).

We created an *Activity Enthusiasm Score* as the mean of perceived helpfulness of the activity (1–5 scale) and likelihood of continuing activity on a regular basis (1–5 scale), both measured after N-of-1 completion on a 1–5 scale. Cronbach's alpha for this 2-item index was 0.77, indicating acceptable to good internal consistency (33).

We summarized the *actual* results of each subject's N-of-1 trial in two ways. First, we directly evaluated differences in means for focus, stress, and happiness (each reported on a 0–10 scale) by taking the difference between the mean value during activity days and the mean value during the participant's usual routine, reversing the sign for stress, then summing across the three outcomes. The theoretical range of this scale was −30 to +30 and the actual range was −7.4 to 9.3. Second, we counted the number of outcomes (stress, focus, happiness) in which the mean value of the participant's responses while performing the chosen activity was better (more positive or less negative) than the mean value of the participant's response while performing their usual routine. Possible values of this count variable ranged from 0 (no outcome better, even by a small amount, during activity days) to 3 (all outcomes better during activity days). The difference variable accounts for the magnitude of benefit but does not consider precision (i.e., the metric does not take into account the within-individual variance in reported outcomes nor the number of measures reported by each participant. The count variable focuses on the number of outcome dimensions that were "improved," while ignoring the magnitude of the improvements. We chose to evaluate these metrics rather than more sophisticated alternatives because they more closely comport with the data actually supplied to participants as shown, for example, in **Figure 1**). Because the results using the two metrics were not materially different, we report only the difference measure.

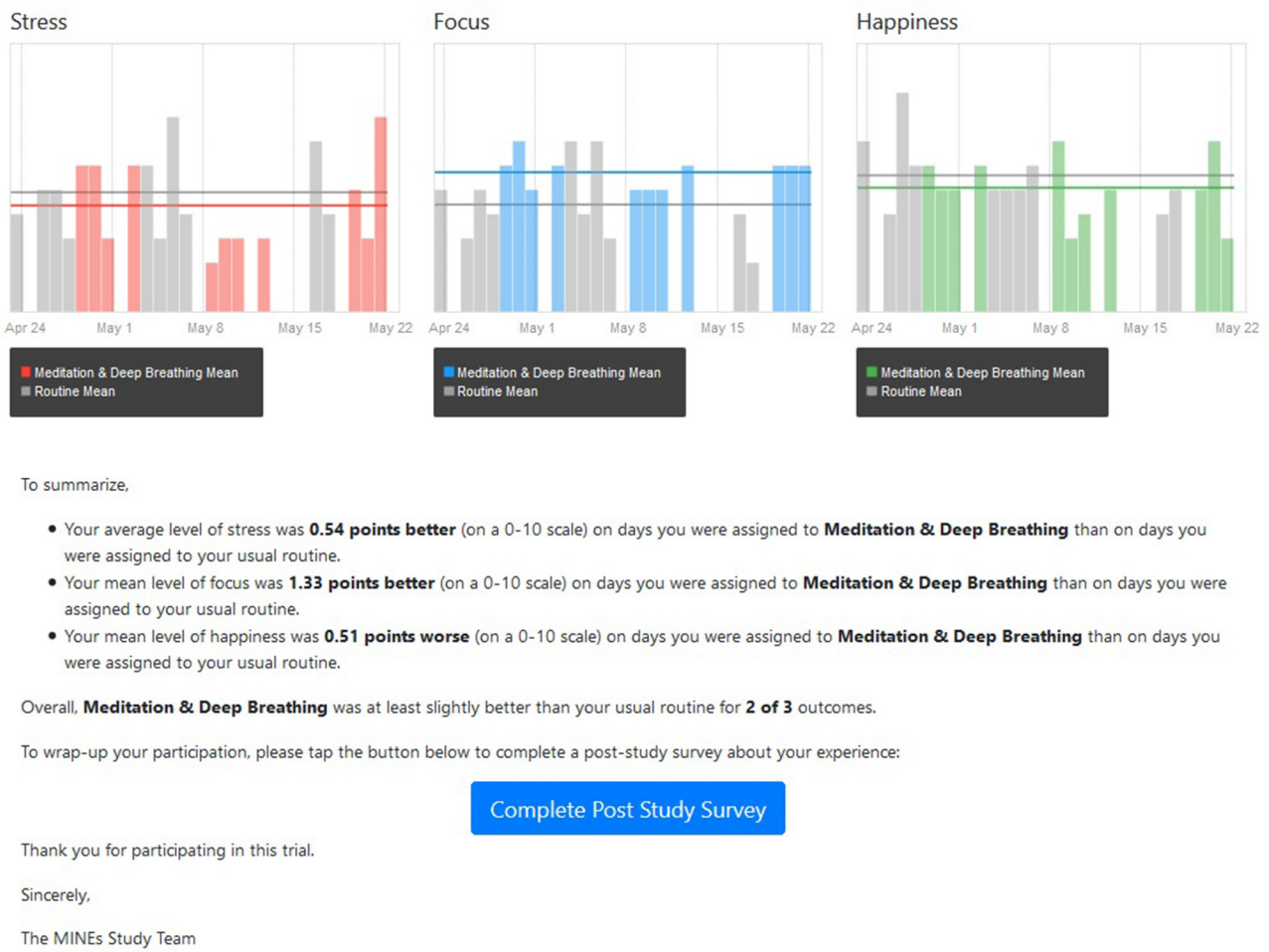


FIGURE 1 | Sample participant results report. Participants in the study were provided with a both a graphic and a written summary depicting their gains (or losses) on days assigned to the intervention compared with days assigned to usual routine.

Statistical Analysis

Values were expressed as means with standard deviations for continuous variables and counts with proportions for categorical variables. For characteristics of the analytic sample, analysis of variance (ANOVA) was performed to test whether there were differences in the means of continuous variables across different chosen activity groups. Chi-square test was performed to test for whether there was association between categorical variables and chosen activity groups if the minimum expected cell count was greater than 1; otherwise, Fisher's exact test was performed. The same procedure was also applied when comparing those who completed the post-study survey with those who did not.

Multiple linear regression was used to assess the relationship between Activity Enthusiasm Score and expectations of benefit from the intervention, interpretation of n-of-1 results, and actual reported results as represented by the summated score along with their pairwise interactions. Goodness-of-fit was expressed by the coefficient of determination, R^2. A significant relationship was determined by a $p < 0.05$. Stata software version 15 was used for regression modeling. R software version 3.6.1 was used to produce graphs.

RESULTS

Of 824 volunteers who *accessed* the online pre-study questionnaire (353 who selected deep breathing, 225 gratitude journaling, 191 physical activity, and 55 acts of kindness), 682

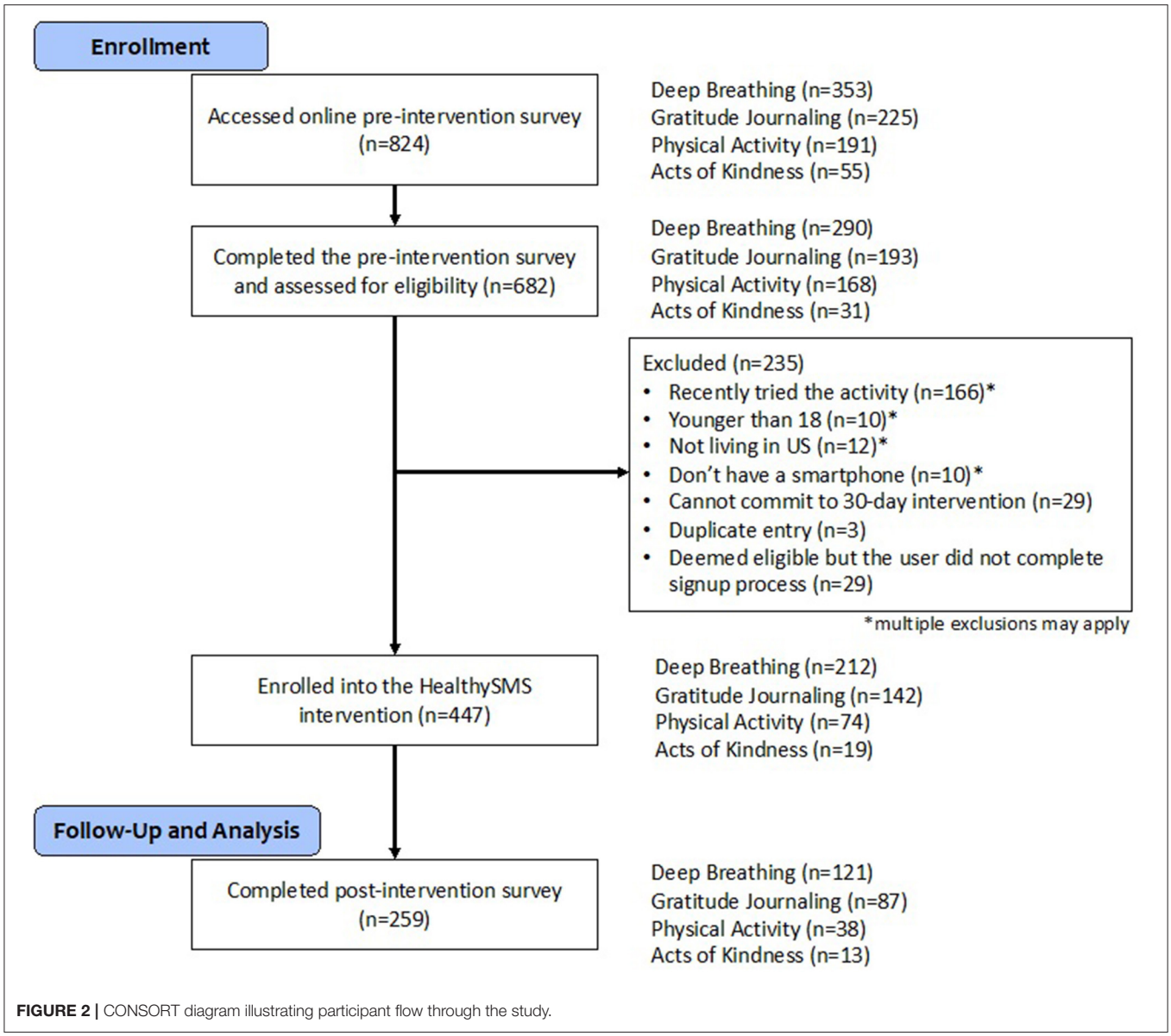


FIGURE 2 | CONSORT diagram illustrating participant flow through the study.

subjects completed the pre-study survey and were assessed for eligibility, 447 signed onto the HealthySMS platform, and 259 completed the post-study survey **Figure 2**. The mean proportion of daily assessments actually returned was 0.70 (SD 0.22).

Among the 259, the mean age was 51 and most were from the Eastern time zone, female, white, and very highly educated; a minority lived alone or had previously tried self-experimentation (**Table 1**) There were no significant associations between respondents' personal characteristics and the behavior intervention activity they chose to evaluate (**Table 1**). There were no meaningful or statistically significant demographic differences between the 259 participants included in the analytic sample and the remaining 188 participants who enrolled in the study but did not complete any part of the post-study questionnaire, except that completers were about 4 years older (**Supplemental Material Table 1**).

Most respondents strongly or somewhat agreed with positive statements about system usability and strongly or somewhat disagreed with negative statements (**Table 2**). The mean scale score was 4.35 (SD 0.57) (**Table 2**), corresponding to a percentage-based score of $4.35/5 \times 100 = 87.1$, well above the average of 68 previously reported and comparable to microwave ovens, which received a rating of 87 in a consumer survey of 1,058 participants (34, 35). System Usability Scale scores were lower, on average, among participants choosing Acts of Kindness compared to those choosing Deep Breathing or Physical Activity ($p = 0.008$, with pairwise comparisons between Acts of Kindness and both Deep Breathing and Physical Activity

TABLE 1 | Characteristics of analytic sample, overall, and by chosen activity.

Characteristic	Overall ($n = 259$)	Deep breathing ($n = 121$)	Gratitude journaling ($n = 87$)	Physical activity ($n = 38$)	Acts of kindness ($n = 13$)	*P*-value
Age, yrs. (SD)	50.5 (13.9)	51.8 (14.0)	48.4 (13.4)	50.7 (14.4)	51.5 (14.7)	0.36
Time zone, *n* (%)						0.44
Eastern	148 (57.1)	75 (62.0)	46 (52.9)	21 (55.3)	6 (46.2)	
Central	44 (17.0)	17 (14.0)	17 (19.5)	6 (15.8)	4 (30.8)	
Mountain	20 (7.7)	8 (6.6)	9 (10.3)	1 (2.6)	2 (15.4)	
Pacific	47 (18.1)	21 (17.4)	15 (17.2)	10 (26.3)	1 (7.7)	
Female, *n* (%)	219 (84.6)	107 (88.4)	69 (79.3)	33 (86.8)	10 (76.9)	0.27
Nonwhite (including mixed), *n* (%)	41 (15.8)	20 (16.5)	11 (12.6)	8 (21.1)	2 (15.4)	0.69
Education						0.21
<Bachelor's degree, refused, or other	52 (20.1)	25 (20.7)	11 (12.6)	11 (28.9)	5 (38.5)	
Bachelor's degree	66 (25.5)	31 (25.6)	26 (29.9)	7 (18.4)	2 (15.4)	
Advanced degree	141 (54.4)	65 (53.7)	50 (57.5)	20 (52.6)	6 (46.2)	
Lives alone, *n* (%)	44 (17.0)	14 (11.6)	16 (18.4)	10 (26.3)	4 (30.8)	0.08
Previously tried self-experimentation, *n* (%)	47 (18.1)	25 (20.7)	17 (19.5)	4 (10.5)	1 (7.7)	0.38

significant using the Bonferroni approach ($p < 0.01$ in each case, data not shown in tabular form). However, there were no significant differences in System Usability Scale scores by age, gender, race, or education (see **Supplemental Material Table 2** for details).

On the post-study questionnaire, 32% of respondents recalled that prior to starting the N-of-1 trial they were very or extremely confident that the chosen activity would be beneficial. At the same time, 27 (10%) interpreted their reported N-of-1 results as indicating that the activity was *not* beneficial for them, 77 (30%) that the activity was *minimally* beneficial, 119 (46%) that the activity was *somewhat* beneficial, and 25 (10%) that the activity was *highly* beneficial. Participants with positive expectations for intervention benefit (i.e., those who reported being "very" or "extremely" confident that the chosen intervention would be beneficial) were more likely than their more skeptical peers to interpret their results as showing that the activity was "somewhat" or "highly" beneficial (68 vs. 53%, $p = 0.036$, data not shown in tabular form).

Table 3 examines the effects of pre-study expectations for benefit, the participant's interpretation of their own N-of-1 results, and actual reported results (as represented by the difference metric; see Methods) on Activity Enthusiasm Score. Model 1 shows that both expectations for intervention benefit ($p < 0.001$) and interpretation of own results ($p < 0.001$) were significantly associated with enthusiasm for the chosen activity, accounting for 33% of the variance. Substituting *actual* results for *interpretation* of results resulted in a regression (Model 2) that explained only 14% of the variance in enthusiasm. Finally, actual results were not significantly associated with enthusiasm after adjusting for pre-study confidence and results interpretation (Model 3).

The relationship of pre-study expectations for benefit, post-study interpretation of results, and Activity Enthusiasm Score is further illustrated in **Figure 3**. Essentially, if at the outset respondents were highly confident that their chosen activity was beneficial (top row), enthusiasm remained moderate to high regardless of actual study results (plenty of orange and red dots, and very few blue dots, even among participants who interpreted their own n-of-1 results as showing that the activity delivered little to no benefit). On the other hand, if initial confidence for benefit was low-to-moderate (bottom row), enthusiasm was more strongly related to the participant's interpretation of their own results (mostly blue dots in the "not beneficial" column, mostly red dots in the "highly beneficial" column). These results are replicated in tabular form in **Supplemental Material Table 3**.

DISCUSSION

As the most direct approach to estimating individual treatment effects, N-of-1 trials have been called the holy grail of clinical investigation (36). The method's appeal may also extend to selected lay audiences, such as the quantified-self movement (37). Broader uptake of N-of-1 trials could help people with and without chronic diseases to more quickly identify treatments or lifestyle interventions that are both appealing and effective. However, logistical barriers, technical concerns, and simple lack of awareness have impeded dissemination and uptake. The main contribution of the current study is to demonstrate that N-of-1 trials of behavioral interventions can attract substantial interest from a relatively broad cross-section of US adults. However,

TABLE 2 | Respondents' experiences with N-of-1 trial system usability ($n = 252$)*.

Item	Strongly or somewhat agree, *n* (%)	Item mean* (SD)
I think I would like to use this system frequently	144 (55.6)	3.5 (1.1)
I found the StudyofMe system unnecessarily complex	10 (3.9)	1.5 (0.9)
I thought the StudyofMe system was easy to use	233 (90.0)	4.6 (0.8)
I think that I would need the support of a technical person to be able to use the system	6 (2.3)	1.2 (0.6)
I found the various functions in the StudyofMe system were well-integrated	156 (60.2)	3.8 (1.1)
I thought there was too much inconsistency in the StudyofMe system	26 (10.0)	1.8 (1.1)
I would imagine that most people would learn to use the StudyofMe system very quickly	237 (91.5)	4.6 (0.7)
I found the StudyofMe system very cumbersome to use	24 (9.3)	1.6 (1.1)
I felt very confident using the StudyofMe system	209 (80.7)	4.4 (1.0)
I needed to learn a lot of things before I could get going with the StudyofMe system	5 (1.9)	1.3 (0.7)
System usability scale	–	4.35 (0.57)⊥

*$N = 252$ rather than 259 because 7 subjects did not complete a majority of scale items.
⊥ In calculating the mean scale score, items with negative valence were reversed.

TABLE 3 | Influence of pre-study confidence, interpretation of own results, and actual (reported) results on participant's "enthusiasm" for the behavioral intervention£.

	Model 1 (pre-study confidence and interpretation of own results, without interaction) ($n = 248$)		Model 2 (pre-study confidence and actual results) ($n = 245$)		Model 3 (pre-study confidence, interpretation of own results, and actual results) ($n = 245$)	
Predictor variable	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI
Pre-study confidence (1–5 scale)	0.20	(0.10, 0.30)*	0.27	(0.16, 0.38)*	0.20	(0.10, 0.30)*
Interpretation of own results (1–4 scale)	0.54	(0.43, 0.65)*	–		0.48	(0.36, 0.60)*
Actual results†	–	–	0.10	(0.05, 0.14)*	0.03	(-0.01, 0.07)
R-squared	0.33		0.14		0.31	

£ *All models in this table use multiple linear regression to estimate "enthusiasm" as a function of various predictors. As described in Methods, "enthusiasm" is an index ranging from 1 (low) to 5 (high) combining perceived "helpfulness" of the intervention and likelihood of persisting with the intervention over the next 6 months. Model 1 examines the influence on enthusiasm for the intervention (1–5 scale) of pre-study confidence (1 = not-at-all confident…5 = extremely confident) and the participant's interpretation of their own n-of-1 results (1 = intervention not beneficial…4 = intervention extremely beneficial). In a variation of this model (not shown), the interaction of confidence and results interpretation was significant with a negative sign, indicating that interpretation of own results was a more potent predictor of enthusiasm among those with lower pre-study confidence. However, this model is not further considered for ease of interpretation. Model 2 uses multiple linear regression to estimate the effects on enthusiasm of confidence and actual n-of-1 study results as reported to the participant (using the "difference measure" as defined in Methods, actual range −7.4 to 9.3), and Model 3 evaluates confidence (1–5 scale), interpretation of results (1–4 scale), and actual results (difference measure). Interaction terms are not reported for Models 2 and 3 because preliminary analysis showed no significant contribution of any two-way interaction.*
†Average of the mean difference between intervention and control rating stress, focus, and happiness. Scores for stress were reversed so that a more positive difference between intervention and control consistently represents a better outcome.
*****$p < 0.001$.

participants were highly educated and tilted strongly female. There are several possible explanations for limited participation among men and those without a college degree, including relative indifference to the topic of "wellness;" competing demands from other responsibilities; or a persistent "digital divide" curtailing access or limiting comfort with mobile devices (38). Nevertheless, our mHealth platform supporting these trials was rated highly usable by participants. Finally, participants' *a priori* expectations for benefit of their chosen behavioral intervention (as measured by confidence that the activity would be beneficial) as well as their *a posteriori* interpretation of their N-of-1 trial results were both significant independent predictors of enthusiasm for the intervention going forward.

We recruited participants using social media (principally Facebook) plus an on-air interview with WNYC Public Radio host Brian Lehrer. Over a brief recruitment period, 824 individuals demonstrated interest by visiting the study website, but unsurprisingly, there was significant attrition at every stage thereafter. Among the 259 subjects in the analytic cohort, the modal participant was a white, middle-aged, highly educated woman. However, less than one in five had prior experience with self-experimentation, indicating both that our sample was

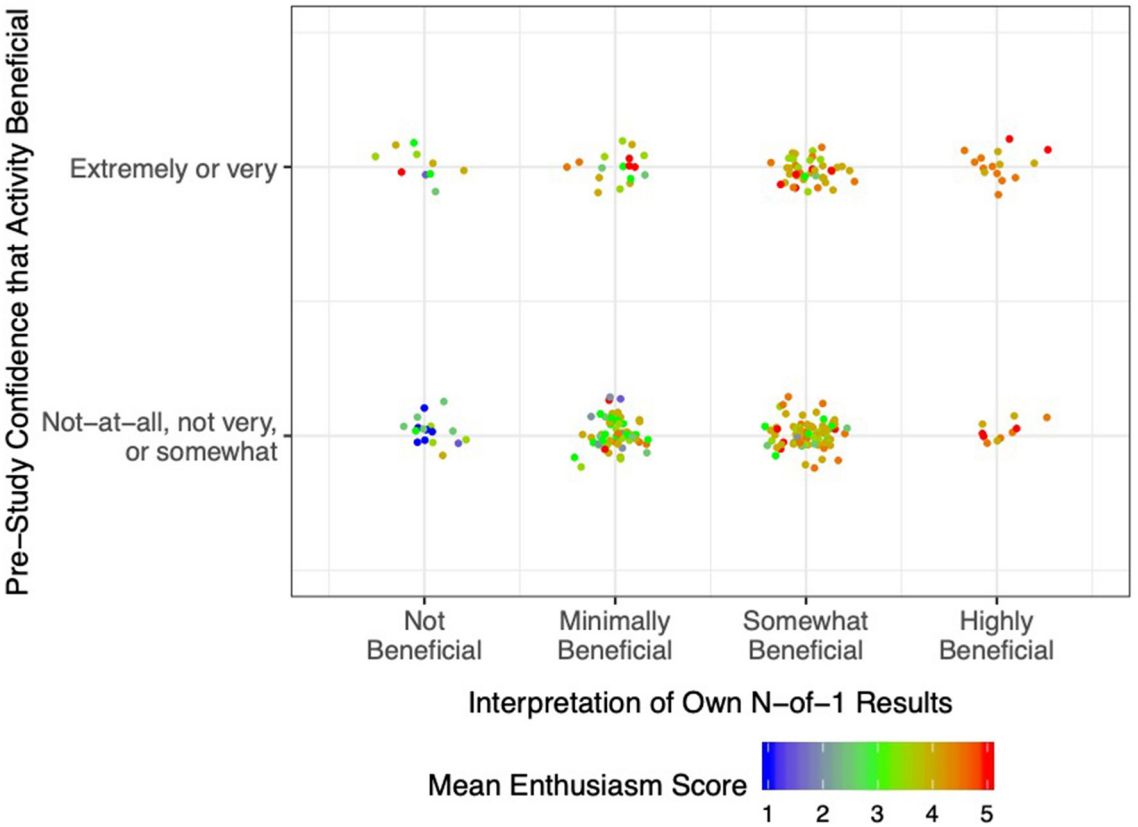


FIGURE 3 | Post-Study Enthusiasm Scores (mean of perceived intervention helpfulness and likelihood of persisting with the activity in future) as a function of pre-study confidence in the activity (reported retrospectively) and interpretation of own n-of-1 trial results. Each dot represents a single participant, with warmer colors indicating greater enthusiasm.

open to novel experiences and that N-of-1 trials may have appeal beyond the established self-tracking community.

Participants generally rated the HealthySMS platform as highly accessible and easy to use without technical support, despite modest misgivings about functional integration. These findings are especially remarkable in light of the native complexity of N-of-1 trials. For example, in our study, patients needed to become comfortable with a new behavioral intervention, switch off regularly between the intervention and their usual routine, and report daily ratings of stress, focus, and happiness.

Pre-study expectations for benefit from the chosen behavioral intervention was modestly associated with post-study enthusiasm for the intervention. At the same time, participants who interpreted their N-of-1 results as highly beneficial had much greater enthusiasm than those who interpreted their results as indicating that the intervention was not beneficial. However, the effect of participants' interpretations of their own results on enthusiasm for the intervention was greater among those with the least confidence in the intervention to begin with.

One interpretation of these findings is that N-of-1 trials had greater information value for participants who were more skeptical at the outset; in Bayesian terms, those with weak or negative priors relied more on the incoming data (39). Although considerable work in cognitive psychology indicates that humans are poor Bayesians (40), our results suggest that in the context of a self-experiment in which they are personally vested, participants may form conclusions based on a weighted average of pre-trial expectations and post-trial results. A plausible implication is that investigators should explicitly account for participants' prior beliefs in the context of N-of-1 experiments and, indeed, use them in constructing posterior probabilities that are returned to patients. Another possible explanation, drawing on expectation disconfirmation theory (41), is that participants who were pleasantly surprised by a positive result (despite expecting a negative outcome) were more likely to be enthused about the activity going forward.

Although participants' actual results (as conveyed by a graphical interface supported by text, as in **Figure 1**) were moderately correlated with their subjective interpretations, the former did not significantly predict intervention enthusiasm after adjusting for the latter, suggesting that actual results are mediated through participants' interpretations. Furthermore, participants' interpretations may not fully and accurately incorporate actual results—even among the highly educated. More work is needed on ways to accurately convey n-of-1 results to participants,

especially in real-world, non-clinical settings where clinicians and investigators are unavailable to help.

The strengths of this study include attention to several novel questions and the use of innovative methods to attract participants and to support them in conducting their own single-patient trials. However, as with all studies, the findings must be evaluated in light of certain limitations. First, there was substantial attrition between expressing initial interest and completing a minimum number of study procedures. Second, the analytic sample was demographically narrow, limiting generalizability (42). This likely reflected some combination of our outreach methods (social media and public radio, which may appeal to a more socio-economically advantaged cohort); the "digital divide;" and the intrinsic appeal of "wellness" interventions and self-monitoring to certain demographic groups (e.g., women). Third, in measuring daily outcomes with single items, we likely sacrificed reliability in the interest of minimizing respondent burden. Fourth, measuring pre-study confidence in the benefits of the intervention after participants completed their N-of-1 trial could have introduced recall or "hindsight" bias. In retrospect, it would have been preferable to measure expectations prospectively, and future studies should do this. Hindsight bias would tend to narrow the gap between participants' expectations and *post-hoc* enthusiasm for the intervention. Had we measured expectations prospectively, we might have seen a more consistent gap in enthusiasm between those with high and low expectations. Finally, we made no attempt to measure either behavior change or psychological outcomes beyond 4–6 weeks after the start of each participant's N-of-1 trial.

In summary, this study demonstrates that N-of-1 trials can be disseminated to a broad, if demographically slanted, subset of the general US population in the interest of enhancing psychological well-being. Subjects appear to learn from their own N-of-1 experiences, although their learnings are tempered by prior beliefs. Our finding of increased influence of trial results among those with the lowest a priori expectations of benefit suggests that mHealth-supported, behavioral N-of-1 trials may have the greatest value for those with the lowest outcomes expectations; these individuals may be exactly those with more health problems and higher need. Further research is needed to clarify who can benefit from such trials, under what circumstances, and with respect to which medium and long-term outcomes.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be accessed by writing to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the UC Davis Institutional Review Board (IRB ID 1255435-4). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RK conceived of the study, obtained funding, performed selected analyses, and wrote the first draft of the manuscript. AA, CK, KK, and SS supported the technical implementation of the study, participated in data collection, and edited the manuscript. EC, EH, SP, SMS, and IS conceived of the study, contributed to design and implementation, and edited the manuscript. YC, JY, and CS performed and/or supervised the statistical analysis and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant number UL1 TR001860. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00260/full#supplementary-material

REFERENCES

1. Clarke TC, Barnes PM, Black LI, Stussman BJ, Nahin RL. Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. *NCHS Data Brief.* (2018) 1–8.
2. Warburton DE, Bredin SS. Reflections on physical activity and health: what should we recommend? *Can J Cardiol.* (2016) 32:495–504. doi: 10.1016/j.cjca.2016.01.024
3. Holt-Lunstad J, Robles TF, Sbarra DA. Advancing social connection as a public health priority in the United States. *Am Psychol.* (2017) 72:517. doi: 10.1037/amp0000103
4. Christian H, Bauman A, Epping JN, Levine GN, McCormack G, Rhodes RE, et al. Encouraging dog walking for health promotion and disease prevention. *Am J Lifestyle Med.* (2018) 12:233–43. doi: 10.1177/1559827616643686
5. Oman D. Elephant in the room: why spirituality and religion matter for public health. In: Oman D, editor. *Why Religion and Spirituality Matter for Public Health. Religion, Spirituality and Health: A Social Scientific Approach, Vol. 2.* (Cham: Springer) (2018).
6. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q.* (2004) 82:661–87. doi: 10.1111/j.0887-378X.2004.00327.x
7. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol.* (2016) 45:2075–88. doi: 10.1093/ije/dyw118
8. Kravitz R, Duan N, Eslick I, Gabler N, Kaplan HC, Larson E, et al. *Design and Implementation of N-of-1 Trials: A User's Guide.* Rockville, MD: Agency for Healthcare Research and Quality. (2014).

9. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015. *Statement. J Clin Epidemiol.* (2016) 76:9–17. doi: 10.1016/j.jclinepi.2015.05.004
10. OCEMB. *The Oxford Levels of Evidence 2 Oxford, England: Oxford Center for Evidence Based Medicine*. (2016). Available online at: https://www.cebm.net/index.aspx?o = 5653 (accessed May 1, 2020).
11. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care.* (2011) 49:761–8. doi: 10.1097/MLR.0b013e318215d90d
12. Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. *J Rheumatol.* (1991) 18:447–51.
13. Li J, Gao W, Punja S, Ma B, Vohra S, Duan N, et al. Reporting quality of N-of-1 trials published between 1985 and 2013: a systematic review. *J Clin Epidemiol.* (2016) 76:57–64. doi: 10.1016/j.jclinepi.2015.11.016
14. Nikles CJ, Mitchell GK, Del Mar CB, Clavarino A, McNairn N. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. *Pediatrics.* (2006) 117:2040–6. doi: 10.1542/peds.2005-1328
15. Nikles CJ, Mitchell GK, Del Mar CB, McNairn N, Clavarino A. Long-term changes in management following n-of-1 trials of stimulants in attention-deficit/hyperactivity disorder. *Eur J Clin Pharmacol.* (2007) 63:985–9. doi: 10.1007/s00228-007-0361-x
16. Nikles CJ, Yelland M, Del Mar C, Wilkinson D. The role of paracetamol in chronic pain: an evidence-based approach. *Am J Ther.* (2005) 12:80–91. doi: 10.1097/00045391-200501000-00011
17. Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI, et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. *J Rheumatol.* (2006) 33:2069–77.
18. Kravitz RL, Paterniti DA, Hay MC, Subramanian S, Dean DE, Weisner T, et al. Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. *Contemp Clin Trials.* (2009) 30:436–45. doi: 10.1016/j.cct.2009.04.001
19. Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, et al. Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Int Med.* (2018) 178:1368–77. doi: 10.1001/jamainternmed.2018.3981
20. Chen C, Haddad D, Selsky J, Hoffman JE, Kravitz RL, Estrin DE, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. *J Med Internet Res.* (2012) 14:e112. doi: 10.2196/jmir.2152
21. Curry OS, Rowland LA, Van Lissa CJ, Zlotowitz S, McAlaney J, Whitehouse H. Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. *J Exp Soc Psychol.* (2018) 76:320–9. doi: 10.1016/j.jesp.2018.02.014
22. Davis DE, Choe E, Meyers J, Wade N, Varjas K, Gifford A, et al. Thankful for the little things: a meta-analysis of gratitude interventions. *J Couns Psychol.* (2016) 63:20–31. doi: 10.1037/cou0000107
23. Johnson S, Gur RM, David Z, Currier E. One-session mindfulness meditation: a randomized controlled study of effects on cognition and mood. *Mindfulness.* (2015) 6:88–98. doi: 10.1007/s12671-013-0234-6
24. Yeung RR. The acute effects of exercise on mood state. *J Psychosom Res.* (1996) 40:123–41. doi: 10.1016/0022-3999(95)00554-4
25. Avila-Garcia P, Hernandez-Ramos R, Nouri SS, Cemballi A, Sarkar U, Lyles CR, et al. Engaging users in the design of an mHealth, text message-based intervention to increase physical activity at a safety-net health care system. *JAMIA Open.* (2019) 2:489–97. doi: 10.1093/jamiaopen/ooz049
26. García Y, Ferrás C, Aguilera A, Ávila P. Usability and feasibility study of a remote cognitive behavioral therapy system with long-term unemployed women. *J Technol Hum Serv.* (2017) 35:219–30. doi: 10.1080/15228835.2017.1345672
27. Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P. Automated text messaging as an adjunct to cognitive behavioral therapy for depression: a clinical trial. *J Med Inter Res.* (2017) 19:e148. doi: 10.2196/jmir.6914
28. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. *J Pain.* (2018) 19:699–716. doi: 10.1016/j.jpain.2018.01.006
29. Abdel-Khalek AM. Measuring happiness with a single-item scale. *Soc Behav Personal.* (2006) 34:139–49. doi: 10.2224/sbp.2006.34.2.139
30. Elo AL, Leppanen A, Jahkola A. Validity of a single-item measure of stress symptoms. *Scand J Work Env Hea.* (2003) 29:444–51. doi: 10.5271/sjweh.752
31. Gardner DG, Cummings LL. Single-item versus multiple-item measurement scales: an empirical comparison. *Educ Psychol Meas.* (1998) 58:898–915. doi: 10.1177/0013164498058006003
32. Brooke J. SUS-A quick and dirty usability scale. *Usabil Eval Indus.* (1996) 189:4–7.
33. Gliem JA, Gliem RR (editors). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. In: *Midwest Research to Practice Conference in Adult, Continuing, and Community Education Voices on the Web*. Columbus, OH (2003).
34. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Intl J Hum Comput Inter.* (2008) 24:574–94. doi: 10.1080/10447310802205776
35. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. *Int J Hum Comput Inter.* (2013) 29:67–76. doi: 10.1080/10447318.2012.681221
36. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personal Med.* (2011) 8:161–73. doi: 10.2217/pme.11.7
37. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. *Big data.* (2013) 1:85–99. doi: 10.1089/big.2012.0002
38. Gonzales A. The contemporary US digital divide: from initial access to technology maintenance. *Inform Commun Soc.* (2016) 19:234–48. doi: 10.1080/1369118X.2015.1050438
39. Duan NH, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *J Clin Epidemiol.* (2013) 66:S21–8. doi: 10.1016/j.jclinepi.2013.04.006
40. Kahneman D, Tversky A. On the psychology of prediction. *Psychol Rev.* (1973) 80:237. doi: 10.1037/h0034747
41. Serrano CI, Shah V, Abràmoff MD. Use of expectation disconfirmation theory to test patient satisfaction with asynchronous telemedicine for diabetic retinopathy detection. *Int J Telemedicine Appl.* (2018) 2018:7015272. doi: 10.1155/2018/7015272
42. Bethlehem J. Selection bias in web surveys. *Int Stat Rev.* (2010) 78:161–88. doi: 10.1111/j.1751-5823.2010.00112.x

Conflict of Interest: CK is the founder of Audacious Software LLC and provided software development, design, and support services for research efforts. AA is (together with University of California, Berkeley) the owner of the HealthySMS license. Fees are paid to UC Berkeley for its use. EH is scientific advisor to Fitabase. KK is an advisor to Bluenote Therapeutics. SS is a consultant with Otsuka Pharmaceutical. IS is a member of the Medical Advisory Board for 98point6; has stock options in Myia Labs, Inc.; is a scientific advisor and consultant for Myovant; and is a Board member and consultant at Vivli.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

BRIEF RESEARCH REPORT
published: 26 June 2020
doi: 10.3389/fdgth.2020.00007

PERSONAL: Feasibility Study Protocol for Placebo-Controlled, Randomized n-of-1 Trials of Tamsulosin for Lower Urinary Tract Symptoms

Scott R. Bauer[1,2,3], Benjamin N. Breyer[2,4], Akinyemi Oni-Orisan[5], Michael A. Steinman[3,6], Ida Sim[1], Charles E. McCulloch[4] and Stacey A. Kenfield[2,4]*

[1] Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States, [2] Department of Urology, University of California, San Francisco, San Francisco, CA, United States, [3] Veterans Affairs Medical Center, San Francisco, San Francisco, CA, United States, [4] Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States, [5] Department of Clinical Pharmacy and Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, United States, [6] Division of Geriatrics, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

OPEN ACCESS

Edited by:
Eric Jay Daza, Clarify Health, United States

Reviewed by:
Chenyun Dai, Fudan University, China
Iraklis Paraskakis, South East European Research Centre, Greece

***Correspondence:**
Scott R. Bauer
Scott.Bauer@ucsf.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *30 January 2020*
Accepted: *18 May 2020*
Published: *26 June 2020*

Citation:
Bauer SR, Breyer BN, Oni-Orisan A, Steinman MA, Sim I, McCulloch CE and Kenfield SA (2020) PERSONAL: Feasibility Study Protocol for Placebo-Controlled, Randomized n-of-1 Trials of Tamsulosin for Lower Urinary Tract Symptoms. Front. Digit. Health 2:7. doi: 10.3389/fdgth.2020.00007

Background: Lower urinary tract symptoms (LUTS) affect more than half of men over age 70 and contribute to both poor health-related quality of life and polypharmacy. Tamsulosin hydrochloride, a selective α_1-blocker, is the most common medication used to treat LUTS due to presumed benign prostatic hyperplasia and is often prescribed indefinitely, although not all men benefit from long-term therapy. N-of-1 trials allow for individualized estimates of benefit and harm and could facilitate decisions regarding chronic tamsulosin therapy for LUTS, particularly among older men. Our team developed the PERSONAL (PlacEbo-controlled, Randomized, patient-Selected Outcomes, N-of-1 triALs) app to track daily urinary symptoms and medication side effects for n-of-1 trials among older men with LUTS.

Materials and Methods: We will conduct a feasibility study of 20 individual randomized n-of-1 trials using the PERSONAL app to compare tamsulosin (0.4 or 0.8 mg) vs. placebo among older men taking tamsulosin for LUTS. We will include men over age 65 with a smartphone for whom temporary discontinuation of tamsulosin is safe, (e.g., no history of acute retention). Participants will work with research staff to prospectively identify the most important urinary symptoms and medication side effects that they would like to digitally track. Men will then be randomized to 2-week treatment periods of tamsulosin or placebo followed by a 1-week wash-out with placebo, for 4 distinct treatment periods and 3 wash-out periods, totaling 11 weeks. Study medications will be blinded using over-encapsulation of tamsulosin pills and matching placebo. Our primary outcomes for this study will be recruitment and retention of eligible men, completion rates of n-of-1 trials and daily questionnaires using the PERSONAL app, and participants' perceived usefulness of their n-of-1 trial for determining whether tamsulosin is effective for them. Linear mixed effects models with individual-specific intercepts and intervention effects will also be used to estimate within-individual effects of tamsulosin.

Discussion: The goal of this innovative study is to establish feasibility and acceptability of using a mobile health app and n-of-1 trials to provide older men with individualized estimates of benefits and harms of chronic tamsulosin therapy for LUTS.

Keywords: randomized clinical trial design, personalized medicine, patient-reported outcomes, medication side effects, benign prostatic hyperplasia, deprescribing, α-antagonist

INTRODUCTION

Lower urinary tract symptoms (LUTS), such as nocturia, urinary urgency, and weak stream, affect more than half of men over age 70 (1). LUTS are associated with increased risk of polypharmacy, falls, and psychological distress, all of which contribute to poor health-related quality of life (2–4). Guidelines recommend treating LUTS due to presumed benign prostatic hyperplasia (BPH) with α_1-blockers (5, 6), which inhibit smooth muscle contraction in the prostate and bladder neck and are increasingly prescribed globally (7–9). Although large randomized controlled trials have demonstrated modest efficacy of α_1-blockers for improving LUTS severity scores [2.1 to 3.7 point mean difference in the International Prostate Symptom Score (10)], average effect sizes in most individual trials and meta-analyses do not reach the accepted minimally important difference (3 points) (11, 12). These trials also use mean change in LUTS severity scores as the primary outcome, which assumes that all patients use overall symptom severity rather than specific bothersome symptoms to make LUTS treatment decisions (13). Harms of α_1-blockers, such as orthostatic hypotension and dizziness which lead to falls and fractures, have led to recommendations that they be used with caution in older men (14–16). Unfortunately, alternative LUTS medications, such as 5α-reductase inhibitors, anti-muscarinics, and most recently desmopressin, are also problematic for older men (14, 15, 17). In the setting of modest benefits and known harms, a more personalized and patient-centric approach is needed to ensure that only older men in whom benefits outweigh the harms continue to receive chronic tamsulosin therapy for LUTS.

N-of-1 trials, or multiple crossover trials conducted within a single individual, are a powerful yet underused tool that could be used to optimize prescriptions for symptomatic conditions such as LUTS (18). This study design is particularly well-suited to address a major barrier of deprescribing for both patients and prescribers: the fear of worsening symptoms or complications after stopping a medication that may have provided benefit initially (19, 20). N-of-1 trials carry an additional benefit to older adults given their lack of representation in most rigorous randomized controlled trials, greater heterogeneity in causes of urinary symptoms and response to treatments, and potential for harms from medication side effects and polypharmacy (e.g., adverse drug events and drug-drug interactions) (21). Whereas the trials evaluating efficacy of tamsulosin for LUTS due to BPH were conducted predominantly among relatively healthy white men <65 years old (11), the majority of men who currently receiving chronic tamsulosin therapy are over age 65, have multiple comorbidities, and match the racial diversity in the United States (9, 22). N-of-1 trials can be implemented with or without the involvement of clinicians and are able to accommodate patient-selected outcomes that may be more influential in treatment decisions than overall LUTS severity scores. By leveraging mobile health technology to implement a more personalized approach to prescribing and deprescribing, n-of-1 trials could potentially replace current recommendations to treat bothersome LUTS due to BPH with indefinite α_1-blocker therapy.

The goal of this study is to establish feasibility and acceptability of using the PERSONAL (PlacEbo-controlled, Randomized, patient-Selected Outcomes, N-of-1 triALs) mobile health app to conduct placebo-controlled n-of-1 trials among older men receiving chronic tamsulosin therapy for LUTS to facilitate deprescribing decisions. This study will include a total of 20 men who will undergo individualized n-of-1 trials in order to collect and report the parameters necessary to plan an optimal and adequately powered full study of drug effectiveness.

MATERIALS AND METHODS

Study Design

Our research team is following a mixed methods approach (23) to develop and evaluate digital health interventions as advocated by the World Health Organization (24). While we plan to conduct focus groups and semi-structured interviews of study participants to further refine the PERSONAL app and study design, the following protocol focuses on the feasibility of conducting placebo-controlled n-of-1 trials among older men with LUTS using a mobile health app.

Study Setting

The proposed study will be located within the San Francisco Bay Area with recruitment occurring at multiple clinical sites within the University of California, San Francisco (UCSF) Medical Center. Participants' UCSF clinicians will not be routinely informed of their participation in this study.

Study Hypothesis

The primary study hypothesis is that it is feasible to conduct a series of individual placebo-controlled n-of-1 trials among older men receiving tamsulosin for LUTS using PERSONAL app. Specifically, we hypothesize that it is possible to recruit and retain 20 eligible men from a single healthcare system within a reasonable timeframe (e.g., 3–6 months), >70% of enrolled participants will complete n-of-1 trials, participants will complete >50% of daily questionnaires, and >50% of participants will describe the PERSONAL app as "extremely helpful" or "very helpful" for deciding whether tamsulosin is an effective medication for them to continue or discontinue.

TABLE 1 | Participant inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Android or iPhone smartphone or tablet with an active data plan and/or connected to a home WiFi network	History of urinary incontinence, acute urinary retention, recurrent urinary tract infections, nephrolithiasis, obstructive kidney disease, urethral stent
Age ≥65 years	Active cancer treatment or medical condition that would limit the patient's life expectancy to <6 months
Diagnosis code for BPH or other micturition problem based on ICD-10 (N40.1 or R39198)	Dementia, bipolar disorder, schizophrenia, active suicidality, active substance use disorder
Taking tamsulosin 0.4 to 0.8 mg daily for ≥12 months	Current participation in another smartphone app-based clinical study
LURN 10-Item Symptom Index (SI-10) ≤10 (none/mild to moderate symptoms)	Planning to relocate from study area within 6 months
Downloaded an app from the Google Play or App Store within the past year	Impaired vision that limits the use of smartphone apps
Ability to speak and read English	Unwilling to temporarily stop tamsulosin

BPH, benign prostatic hyperplasia; ICD, International Statistical Classification of Disease and Related Health Problems; LURN, Lower Urinary Tract Dysfunction Research Network.

Secondary and tertiary outcomes will include standardized measures of LUTS severity, global urinary bother, satisfaction with LUTS treatment, attitude toward deprescribing, medication adherence, and health-related quality of life.

Eligibility Criteria

Study participants include a broad diversity of patients recruited from the UCSF Medical Center electronic health record (EHR) at 3 UCSF-affiliated clinical sites: Mission Bay Campus, Parnassus Heights Campus, and Zuckerberg San Francisco General Hospital. Participants must meet the following eligibility criteria (**Table 1**): English speaking men over age 65 with a diagnosis of LUTS or BPH based on International Statistical Classification of Disease and Related Health Problems (ICD-10 codes; N40.1 or R39198); currently prescribed daily tamsulosin therapy [(0.4 or 0.8 mg) for at least the past 12 months; owns an eligible iOS or Android smartphone or tablet; have a Lower Urinary Tract Dysfunction Research Network 10-Item Symptom Index (LURN SI-10(25)] ≤10 (corresponds to none/mild to moderate symptoms on a scale of 0 [no symptoms] to 38 [most severe symptoms]). Participants will be excluded if they have urinary incontinence or a condition that requires continuous tamsulosin treatment, such as history of acute urinary retention, recurrent urinary infections, obstructive kidney disease, or ureteral stent. Participants will also be excluded if they have medical conditions, such as dementia or active substance use disorder, that will interfere with their participation in the study.

Recruitment

The participant flow diagram is shown in **Figure 1**. Patients who meet inclusion criteria based on data available in the EHR and who have previously agreed to be contacted by UCSF Research Participant Services will receive a secure EHR message informing them about the study and inviting them to contact research staff if interested in participating. Patients who have not enrolled in secure EHR messaging will receive a paper-based letter with the same information. Recruitment materials will offer eligible participants $100 for completing the study.

Screening

Eligibility will be determined via telephone screening. Research staff will explain the study and ask questions to determine which inclusion and exclusion criteria are met, including whether the patient has an eligible phone device. Once a patient is deemed eligible, they will be asked to schedule a study orientation visit. They will then receive a confirmatory email with a link to complete the baseline questionnaire using Research Electronic Data Capture (RedCap), a secure online portal. They will also receive instructions to download the PERSONAL app (available for free on Google Play and Apple's App Store) on their smartphone prior to the orientation visit.

Study Orientation Visit

The orientation visit will be conducted by trained research staff who will first obtain informed consent to participate in the study and a Health Insurance Portability and Accountability Act (HIPPA) authorization form. Research staff will first describe the n-of-1 study design, daily questionnaires, and real-time data visualization and will then ensure successful installation of the PERSONAL app. Each participant will then receive a unique study ID and login credentials for the PERSONAL app. Participants and research staff will customize the app together by selecting symptoms and side effects to track and setting reminder notifications based on the preferences of the participant (e.g., morning reminder if tracking nighttime symptoms or evening reminder if tracking daytime symptoms). The participants will then select a start date for their n-of-1 trial which will be entered directly into the PERSONAL app. Participants will receive a bubble pack with 11 weeks of tamsulosin (at their previously prescribed dose) or matching placebo and will be instructed to start taking the study medications after successfully completing the run-in period. At the end of the orientation visit, research staff will assess understanding of the app features and provide verbal and written instructions for the rest of the study. Research staff contact information will be provided for reporting severe or concerning symptoms or technical app support for the duration of the study.

Patient-Selected Outcomes

Participants will be asked at baseline to identify their most bothersome urinary symptoms and perceived medication side effects. Using responses from the baseline questionnaire as a guide, research staff will guide participants to select up to 2

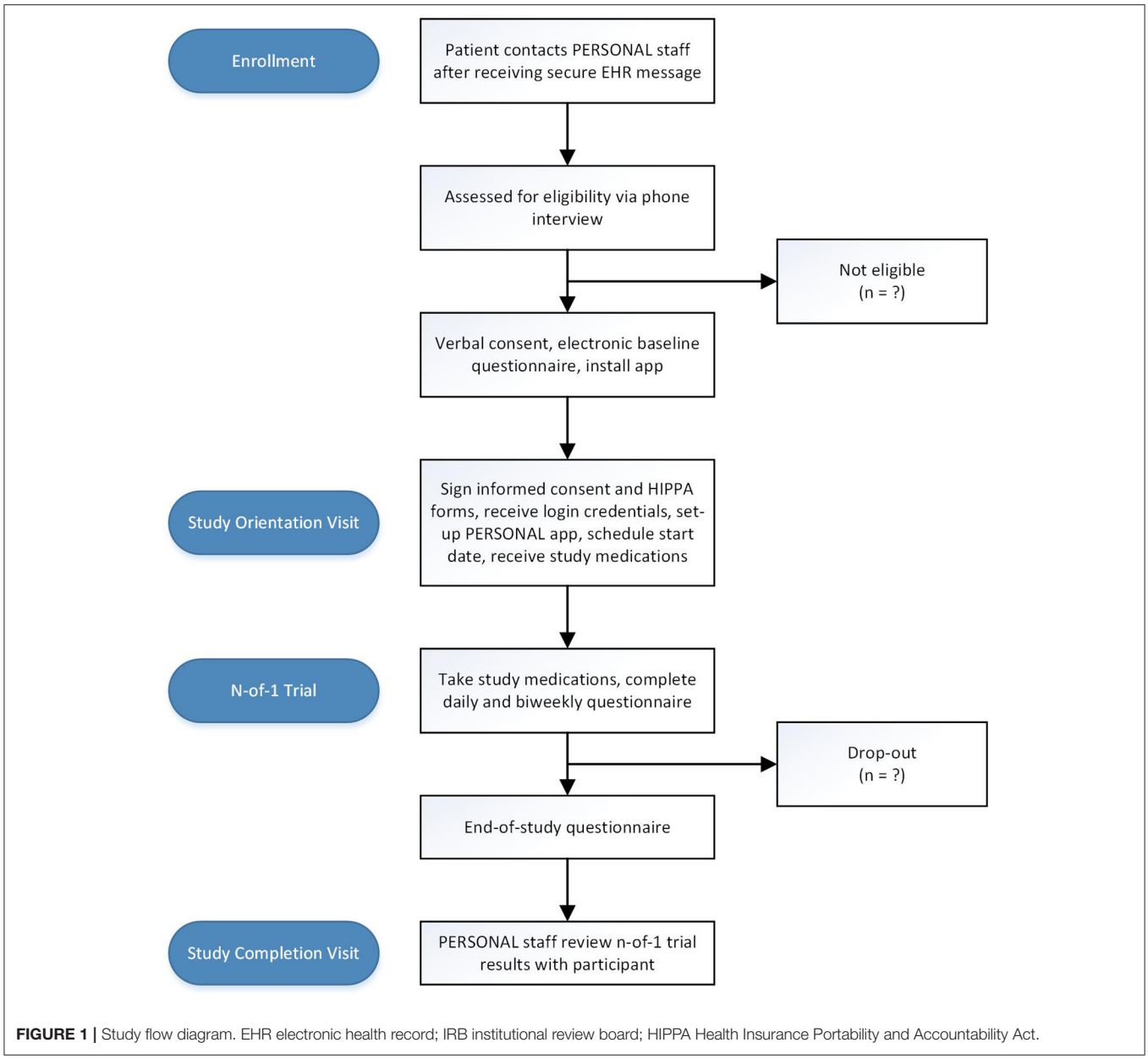


FIGURE 1 | Study flow diagram. EHR electronic health record; IRB institutional review board; HIPPA Health Insurance Portability and Accountability Act.

symptoms and up to 2 side effects to track during their n-of-1 study (see Assessments and Outcome Measures below). We will consider allowing participants to track additional symptoms and side effects if they perceive the burden of additional daily questions to be low. Once selected, these outcomes will be entered directly into the PERSONAL app to create a personalized n-of-1 trial focused on the specific outcomes of interest for each participant. If participants cannot identify a preferred symptom or side effect for tracking, this will be recorded, and the default urinary symptom will be their most frequent symptom identified on the baseline questionnaire. The default medication side effect will be dizziness/lightheadedness because it is one of the most common (15% to 17%) and serious side effects of tamsulosin among older men (26).

N-of-1 Trial Description

Participants will start with 1-week open label run-in period where they will use the PERSONAL app to track daily symptoms and side effects while not taking their tamsulosin or any study pills. Based on the pharmacokinetics and expected timeframe of symptomatic relief from tamsulosin (half-life = 14 to 15 h; steady state by the 5th day of daily dosing) (26), all n-of-1 trials will have a total duration of 11 weeks during which participants will complete 2 cycles consisting of a pair of 2-week treatment periods (taking tamsulosin or placebo) separated by 1 week of wash-out on placebo (**Figure 2**). The order of treatment periods within a cycle will be random (e.g., ABAB, BABA, ABBA, or BAAB) according to pre-filled bubble packs given to participants during their orientation visit. Participants will receive placebo

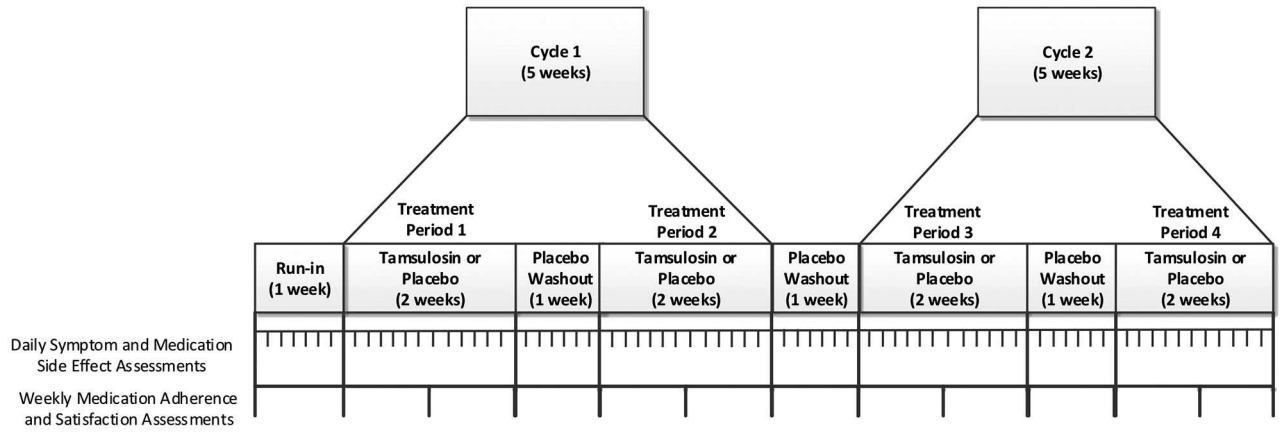


FIGURE 2 | Individual n-of-1 trial.

during wash-out periods between treatment periods and cycles, but they will not be aware of the order or duration of treatment periods or cycles to prevent them from self-correlating symptoms to specific treatments.

Our team, in collaboration with Overlap Health (https://www.overlaphealth.com/), has developed and tested the PERSONAL app among older men with LUTS due to presumed BPH. The PERSONAL app presents participants with a daily questionnaire to track their individually selected urinary symptoms and medication side effects (**Figure 3**). All participants will also be presented a global urinary symptom bother question. Depending on how many symptoms and side effects they desire to track, participants will be asked a minimum of 3 and maximum of 5 daily questions for the duration of their n-of-1 trial. At the end of each week, participants will receive additional medication adherence and treatment satisfaction questionnaires administered via the PERSONAL app as well as motivational messages summarizing their progress in the trial. Participants will be able to view a graphical representation of their responses summarized in chronological order for the prior day, week, or month. To maximize adherence to daily questionnaires, participants will be contacted via email or phone if they have completed fewer than 4 daily questionnaires in any week during their n-of-1 trial.

Study Completion Visit

Upon completion of their n-of-1 trial, participants will be sent an end-of-study RedCap questionnaire via secure email along with an invitation to schedule a study completion visit with research staff. Research staff will access the participants results using the PERSONAL desktop interface and review results with them during the study completion visit. N-of-1 trial results will be displayed in a series of graphs and text output showing the mean daily urinary symptom, global urinary bother, and medication side effect scores while taking tamsulosin vs. placebo. Within-person treatment effects and confidence limits based on linear mixed effects models (see section Data Analysis below for more detail) will also be shared using patient-friendly graphs and text. We will obtain qualitative feedback on the orientation, PERSONAL app, and usefulness of the data and visualization at the end of study completion visit.

Assessments and Outcome Measures

Feasibility and Acceptability Outcomes

We will evaluate 4 primary outcomes that will be used to determine feasibility and acceptability of a larger future n-of-1 study:

1) Recruitment and retention of 20 eligible men within reasonable timeframe (goal 3–6 months)
2) N-of-1 trial completion rate (goal >70%)
3) Daily questionnaire completion rate (goal >50%)
4) Percentage of participants who describe the PERSONAL app as "extremely helpful" or "very helpful" for deciding whether tamsulosin is an effective medication for them to continue or discontinue (goal >50%).

To calculate the recruitment timeframe, we will log the start date of each individual n-of-1 trial. N-of-1 trials will be considered complete when participants complete the end-of-study questionnaire. Completion of daily questionnaires is tracked by the PERSONAL app along with session duration and distribution of time spent with each component of the app to further quantify user engagement. To characterize the experience of participants using the PERSONAL app, we will ask them at the end of the study to rate the helpfulness of the app from 1 (extremely helpful) to 5 (not at all helpful) across multiple domains based on prior mobile health studies (27). We will also administer an adapted System Usability Scale to characterize usability and functionality of the PERSONAL app (28).

Patient-Selected Outcomes

Daily urinary symptom severity

Participants will select from daily urinary symptom measures that were adapted from the LURN SI-10 (25). Specifically, we will use 7 of the 10 questions in the LURN SI-10 (including frequency, nocturia, urgency, voiding, and post-micturition symptoms) which have been previously

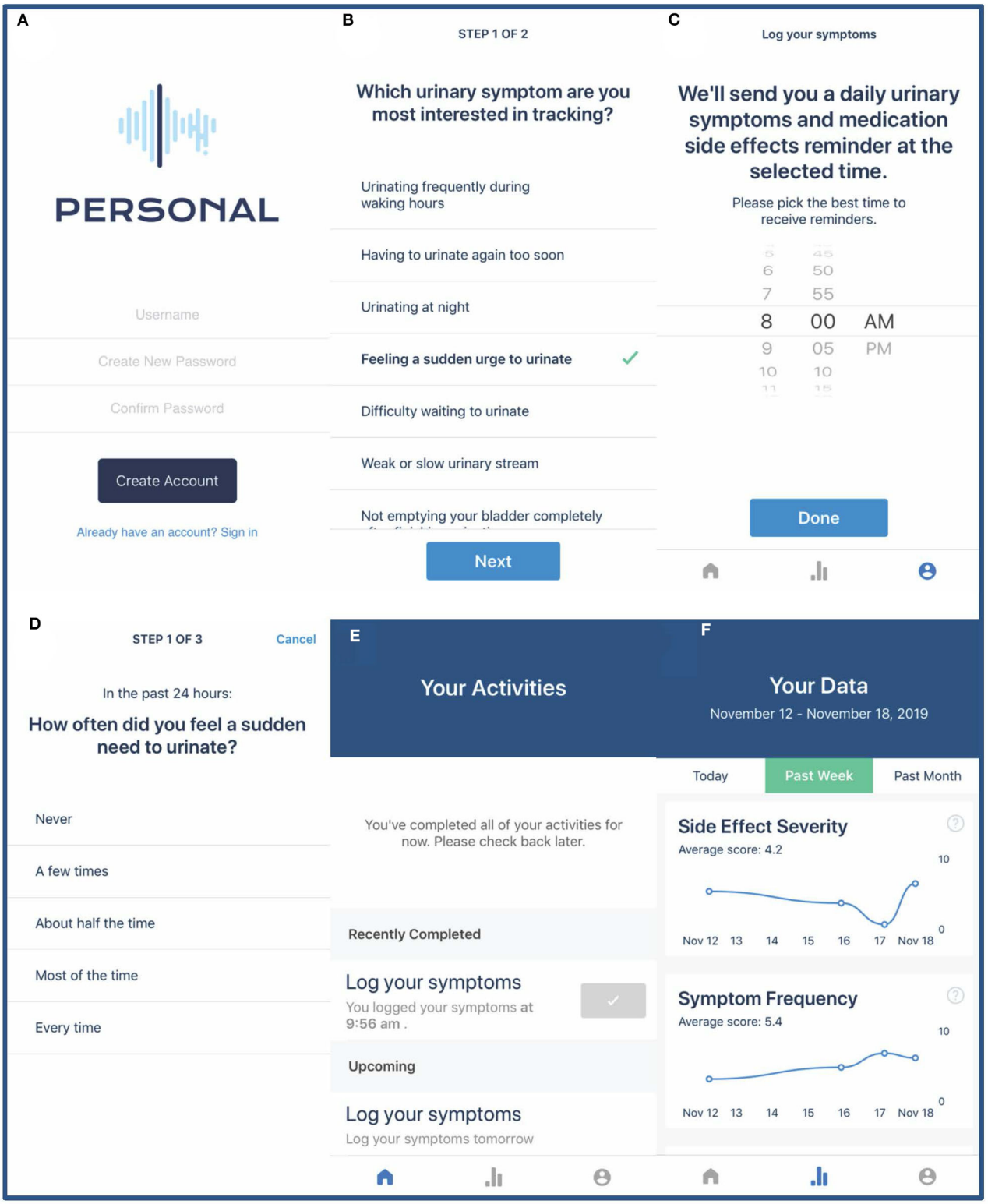


FIGURE 3 | PERSONAL app screenshots. Screenshots demonstrate the PERSONAL app login page **(A)**, selecting urinary symptoms for daily tracking **(B)**, setting a daily questionnaire reminder **(C)**, completing the daily questionnaire **(D)**, the PERSONAL app home page with current and future tasks **(E)**, and graphical representation of participant data **(F)**.

published using 24-h recall periods and were integrated in the PERSONAL app (29). The LURN SI-10 questions excluded from this study refer to urinary incontinence, an exclusion criteria, and bladder pain, a type of LUTS that is not commonly treated with tamsulosin. Specific wording and response options for each question are listed in **Supplemental Table 1**. For visualization purposes, all responses will be normalized to a scale of 0 (minimal severity or bother) to 10 (maximum severity or bother) to ensure uniform graphical representation.

Daily urinary symptom bother

We adapted the American Urologic Association Symptom Index global bother question and changed the recall period from the past month to the past 24 h (30). All participants will be asked "Over the past 24 h, how bothered were you by urinary symptoms?" regardless of which urinary symptoms they are tracking. Responses include "Not at all bothered," "Somewhat bothered," "Very bothered," and "Extremely bothered."

Daily medication side effect bother

To evaluate side effects of tamsulosin, we further adapted the global bother question and will ask participants to quantify how bothered they were by specific perceived side effects. Participants will be asked "Over the past 24 h, how bothered were you by [side effect]?" for each of the medication side effects selected to track during the study orientation visit. Responses rangeD from "Not at all bothered" to "Extremely bothered." Specific wording and response options for each question are listed in **Supplemental Table 1**.

Investigator-Selected Outcomes

At baseline, we will collect demographic data on age, marital status, race, ethnicity, employment, income, and educational attainment via questionnaire. We will ask participants about their smoking history, alcohol and caffeine intake, and physical activity as well as history of medical conditions, including cardiovascular disease, diabetes, prostate cancer, and prostatitis. At baseline and the end of study, participants attitudes toward deprescribing will be assessed using Revised Patients' Attitudes Toward Deprescribing (rPATD) (31) and health-related quality of life will be assessed using the NIH PROMIS 29+2 Profile (32). At baseline and each week, we will assess self-reported medication adherence and reasons for non-adherence (33) as well as overall satisfaction with current LUTS treatment regimen.

Data Analysis

Sample Size

Based on prior mobile health studies (27, 34–36), we expect to observe a failed primary outcome, such as inability to recruit sufficient participants or reach goal questionnaire and n-of-1 trial completion rates, at least 10% of the time. Therefore, with a sample size of 20 participants, we will have 90% power to observe at least one failed primary outcome during this feasibility study (37, 38).

Analytic Plan

Primary outcomes will be assessed as binary variables (e.g., did or did not successfully recruit goal sample size within appropriate timeframe) and the feasibility study will be considered successful if all 4 primary outcome objectives are met. We will describe the primary outcomes as well as baseline demographic and clinical characteristics using percentages, means ± standard deviations for normally distributed variables, and medians with interquartile ranges for skewed variables. The change in secondary and tertiary outcomes from baseline to end-of-study, including LUTS treatment satisfaction, attitudes toward deprescribing, and health-related quality of life, will be evaluated using paired samples *t*-test for continuous measures and McNemar tests for binary variables. We will use multivariable-adjusted linear mixed effects models, with random intercepts and slopes and an unstructured variance-covariance matrix, to estimate variation in daily urinary symptoms or medications side effects. We are aware that there is unlikely to be a sufficient sample size of men who tracked the same urinary symptom or medication side effect needed to calculate valid between-person estimates in this feasibility study, however, participants may have sufficient data to calculate valid within-person variability for each treatment group as well as within-person treatment effects. These estimates will be more accurate and precise in a larger future trial where data from other participants tracking the same symptoms or side effects is incorporated into the linear mixed effects models and contributes to within-person estimates.

Safety

Participant Confidentiality

Data entered into the PERSONAL app will be hosted on the Overlap Health secure environment and will contain no personal health information. To protect participant confidentiality, Overlap Health will only have access to participant study ID numbers. The raw data collected by the PERSONAL app will not be available to other applications. Data transfers will use HIPPA-compliant file encryption (at rest and in transit), secure file transfer (SFTP), Secure Sockets Layer (SSL) for interface data transfers, predefined authentication routes, and a role-based permission system. Questionnaires will be collected electronically via RedCap surveys managed in secure environments behind institutional firewalls. All study staff will be trained in good clinical practice, HIPPA procedures, and participant confidentiality.

Data Monitoring

A unblinded Safety Monitoring Committee will be established to review unanticipated or serious adverse events for the duration of the study and will report directly to the University of California, San Francisco Institutional Review Board.

Ethics Approval

Ethical approval was granted to our team by the University of California, San Francisco Institutional Review Board for a PERSONAL app pilot study (#19-28557). The feasibility study protocol will build off this prior work and will be

submitted for approval by the University of California, San Francisco Institutional Review Board as well as registered on ClinicalTrials.gov once finalized.

DISCUSSION

The PERSONAL study leverages mobile health technology, n-of-1 trials, patient-selected outcomes, and placebo controls to provide older men with personalized information regarding the benefits and harms of continuing or discontinuing chronic tamsulosin therapy. This study protocol seeks to evaluate the feasibility and acceptability of using the PERSONAL app to conduct a series of n-of-trials. We will collect the data needed to plan a larger future n-of-1 study to provide individualized estimates of patient-selected benefits and harms of chronic tamsulosin therapy among older men with LUTS.

There is likely both undertreatment and overtreatment of LUTS with chronic tamsulosin therapy. Potential contributors to overtreatment include situations where there is minimal or no benefit of long-term treatment (e.g., overestimation of symptomatic relief due to placebo effects or regression to the mean, waning efficacy with longer-term treatment, symptoms refractory to tamsulosin) and situations where harms exceed benefit (e.g., medication side effects, polypharmacy, adverse drug events, drug-drug interactions). Harms from chronic tamsulosin therapy are often insidious; tamsulosin or polypharmacy-related side effects may be attributed to other medications and comorbidities or inappropriately tolerated as a "normal process of aging." Conversely, men who would benefit from chronic tamsulosin therapy may prematurely discontinue due to misattributed harm (e.g., a mechanical fall in the absence of symptomatic orthostatic hypotension) or perceived lack of benefit due to overlapping conditions (e.g., improved nocturia but persistent insomnia). Since there is currently no recommended minimum or maximum duration of α_1-blocker therapy for LUTS, n-of-1 trials could be used to personalize LUTS treatments by quantifying both benefits and harms of continuing or discontinuing chronic tamsulosin therapy.

Several observational and small open-label randomized clinical trials provide evidence that discontinuation of chronic tamsulosin therapy will not lead to worsening symptoms in many men. Among 33 men who initially experienced symptomatic improvement with α_1-blocker monotherapy, mean symptom severity were not increased up to 6 months after unblinded discontinuation and more than two-thirds of men remained off medication (39). Another unblinded study among 75 men with symptomatic improvement after α_1-blocker monotherapy demonstrated stable symptoms for up to 12 months after discontinuation with only 30% of men requesting re-initiation of treatment (40). Even among men with more severe LUTS who are treated with combined α_1-blockers plus 5α-reductase inhibitor therapy, both observational and randomized studies have demonstrated no symptomatic progression in the majority of men who discontinue α_1-blockers but continue 5α-reductase inhibitor monotherapy (41–44). In the largest randomized clinical trial of 230 men receiving combined therapy who were assigned to discontinue either 5α-reductase inhibitor or α_1-blocker, 74% of men in both groups had no worsening of symptoms after 12 months (45). Despite preliminary evidence that α_1-blockers can be safely discontinued in men with a wide range of LUTS severity without significant worsening of symptoms, the effects of chronic tamsulosin discontinuation remain unknown and rigorous placebo-controlled studies of α_1-blocker discontinuation are lacking. We also know that the effects averaged over large numbers of participants in traditional placebo-controlled RCTs do not translate directly to individuals, particularly older adults (18, 21). New approaches to personalized prescribing and deprescribing, such as n-of-1 trials, are needed to determine whether an individual man is receiving more benefit than harm from chronic tamsulosin therapy.

N-of-1 trials have the potential to greatly increase the accuracy and precision with which urologic medications are prescribed for symptomatic conditions such as LUTS. Mobile health technology has lowered many of the barriers to implementing this powerful study design in research, clinical, and non-clinical settings by decreasing the burden of frequent data collection and facilitating data interpretation through instantaneous visualization, however it remains unknown if placebo-controlled n-of-1 trials using patient-selected outcomes are feasible. To address this gap, we will establish the feasibility and acceptability of placebo-controlled n-of-1 trials using the PERSONAL app and patient-selected outcomes among older men with LUTS.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SB, BB, AO-O, MS, IS, and SK conceptualized the study and contributed to the study design. SB and CM developed the analytic plan. BB, AO-O, MS, IS, and SK provided administrative, technical, or material support. SB wrote the manuscript and obtained funding. All authors contributed to the refinement of the study protocol and the manuscript. All authors read and approved the final manuscript.

FUNDING

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the NIH under Award Number KL2TR001870. SB was further supported by Grant No. 1K12DK111028 from the National Institute of Diabetes, Digestive, and Kidney Disorders. MS was supported by grants from the National Institute on Aging (1R24AG064025, K24AG049057, P30AG044281). SK is supported by the Helen Diller Family Chair in Population Science for Urologic Cancer at the University of California, San Francisco.

ACKNOWLEDGMENTS

We thank Anthony Enriquez for his support coordinating with medical center administration, obtaining IRB approval, reviewing study documentation and protocols, and facilitating recruitment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2020.00007/full#supplementary-material

REFERENCES

1. Wei JT, Calhoun E, Jacobsen SJ. Urologic diseases in america project: benign prostatic hyperplasia. *J Urol*. (2008) 179(Suppl. 5):S75–80. doi: 10.1016/j.juro.2008.03.141
2. Coyne KS, Wein AJ, Tubaro A, Sexton CC, Thompson CL, Kopp ZS, et al. The burden of lower urinary tract symptoms: evaluating the effect ofLUTS on health-related quality of life, anxiety and depression:epiLUTS. *BJU Int*. (2009) 103(Suppl. 3):4–11. doi: 10.1111/j.1464-410X.2009.08371.x
3. Kupelian V, Wei JT, O'Leary MP, Norgaard JP, Rosen RC, McKinlay JB. Nocturia and quality of life: results from theboston area community health survey. *Eur Urol*. (2012) 61:78–84. doi: 10.1016/j.eururo.2011.05.065
4. Taylor BC, Wilt TJ, Fink HA, Lambert LC, Marshall LM, Hoffman AR, et al. Prevalence, severity, and health correlates of lower urinary tract symptoms among older men: theMrOS study. *Urology*. (2006) 68:804–9. doi: 10.1016/j.urology.2006.04.019
5. McVary KT, Roehrborn CG, Avins AL, Barry MJ, Bruskewitz RC, Donnell RF, et al. Update onAUA guideline on the management of benign prostatic hyperplasia. *J Urol*. (2011) 185:1793–803. doi: 10.1016/j.juro.2011.01.074
6. Gratzke C, Bachmann A, Descazeaud A, Drake MJ, Madersbacher S, Mamoulakis C, et al. EAU guidelines on the assessment of non-neurogenic male lower urinary tract symptoms including benign prostatic obstruction. *Eur Urol*. (2015) 67:1099–109. doi: 10.1016/j.eururo.2014.12.038
7. Verhamme KM, Dieleman JP, Bleumink GS, Bosch JL, Stricker BH, Sturkenboom MC. Treatment strategies, patterns of drug use and treatment discontinuation in men withLUTS suggestive of benign prostatic hyperplasia: theTriumph project. *Eur Urol*. (2003) 44:539–45. doi: 10.1016/S0302-2838(03)00376-2
8. Cornu JN, Cussenot O, Haab F, Lukacs B. A widespread population study of actual medical management of lower urinary tract symptoms related to benign prostatic hyperplasia acrossEurope and beyond official clinical guidelines. *Eur Urol*. (2010) 58:450–6. doi: 10.1016/j.eururo.2010.05.045
9. Erickson BA, Lu X, Vaughan-Sarrazin M, Kreder KJ, Breyer BN, Cram P. Initial treatment of men with newly diagnosed lower urinary tract dysfunction in the veterans health administration. *Urology*. (2014) 83:304–9. doi: 10.1016/j.urology.2013.09.042
10. Barry MJ, Fowler FJ Jr, O'Leary M P, Bruskewitz RC, Holtgrewe HL, et al. The American urological association symptom index for benign prostatic hyperplasia. *J Urol*. (2017) 197:S189–s97. doi: 10.1016/j.juro.2016.10.071
11. Yuan JQ, Mao C, Wong SY, Yang ZY, Fu XH, Dai XY, et al. Comparative effectiveness and safety of monodrug therapies for lower urinary tract symptoms associated with benign prostatic hyperplasia: a network meta-analysis. *Medicine*. (2015) 94:e974. doi: 10.1097/MD.0000000000000974
12. Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker-Corkery E, et al. Benign prostatic hyperplasia specific health status measures in clinical research: how much change in theAmerican urological association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients *? J Urol*. (1995) 154:1770–4. doi: 10.1016/S0022-5347(01)66780-6
13. Agarwal A, Eryuzlu LN, Cartwright R, Thorlund K, Tammela TL, Guyatt GH, et al. What is the most bothersome lower urinary tract symptom?Individual- and population-level perspectives for both men and women. *Eur Urol*. (2014) 65:1211–7. doi: 10.1016/j.eururo.2014.01.019
14. Oelke M, Becher K, Castro-Diaz D, Chartier-Kastler E, Kirby M, Wagg A, et al. Appropriateness of oral drugs for long-term treatment of lower urinary tract symptoms in older persons: results of a systematic literature review and international consensus validation process (LUTS-FORTA 2014). *Age Ageing*. (2015) 44:745–55. doi: 10.1093/ageing/afv077
15. American Geriatrics Society.Updated AGS Beers Criteria(R) for Potentially Inappropriate Medication Use in Older Adults. *J Am Geriatr Soc*. (2019) 67:674–94 doi: 10.1111/jgs.15767
16. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing*. (2015) 44:213–8. doi: 10.1093/ageing/afu145
17. Sarkar RR, Parsons JK, Bryant AK, Ryan ST, Kader AK, McKay RR, et al. Association of treatment with 5alpha-reductase inhibitors with time to diagnosis and mortality in prostate cancer. *JAMA Intern Med*. (2019) 179:812–9. doi: 10.1001/jamainternmed.2019.0280
18. Price JD, Grimley Evans J. N-of-1 randomized controlled trials ('N-of-1 trials'): singularly useful in geriatric medicine. *Age Ageing*. (2002) 31:227–32. doi: 10.1093/ageing/31.4.227
19. Reeve E, To J, Hendrix I, Shakib S, Roberts MS, Wiese MD. Patient barriers to and enablers of deprescribing: a systematic review. *Drugs Aging*. (2013) 30:793–807. doi: 10.1007/s40266-013-0106-8
20. Anderson K, Stowasser D, Freeman C, Scott I. Prescriber barriers and enablers to minimising potentially inappropriate medications in adults: a systematic review and thematic synthesis. *BMJ open*. (2014) 4:e006544. doi: 10.1136/bmjopen-2014-006544
21. Davidson KW, Cheung YK, McGinn T, Wang YC. *Expanding the Role of N-of-1 Trials in the Precision Medicine Era: Action Priorities and Practical Considerations*. Washington, DC: NAM Perspectives. Commentary, National Academy of Medicine. (2018). doi: 10.31478/201812d
22. Nichol MB, Knight TK, Wu J, Barron R, Penson DF. Evaluating use patterns of and adherence to medications for benign prostatic hyperplasia. *J Urol*. (2009) 181:2214–21; discussion 21–2. doi: 10.1016/j.juro.2009.01.033
23. *Designing and Conducting Mixed Methods Research. 3rd ed.* Thousand Oaks, CA: SAGE Publications, Inc. (2017).
24. WHO. *Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment*. Geneva: World Health Organization; (2016). Licence: CC BY-NC-SA 3.0 IGO.
25. Cella D, Smith AR, Griffith JW, Kirkali Z, Flynn KE, Bradley CS, et al. A new brief clinical assessment of lower urinary symptoms for women and men:LURNSI-10. *J Urol*. (2019) 203:164–70. doi: 10.1097/JU.0000000000000465
26. Astellas Pharma Inc. *Flomax® (tamsulosin hydrochloride) [package insert]*. U.S. Food and Drug Administration website. Available online at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2007/020579s020lbl.pdf. Revised July. (2006) (accessed November 17, 2019).
27. Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, et al. Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. *JAMA Intern Med*. (2018) 178:1368–77. doi: 10.1001/jamainternmed.2018.3981
28. Brooke J.SUS:a "quick and dirty" usability scale. InJordan PW, Thomas BA, Weerdmeester I, McClelland IL, editors. *Usability Evaluation in Industry*. London: Taylor & Francis. (1996).
29. Flynn KE, Mansfield SA, Smith AR, Gillespie BW, Bradley CS, Cella D, et al. Can 7 or 30-Day recall questions capture self-reported lower urinary tract symptoms accurately? *J Urol*. (2019) 202:770–8. doi: 10.1097/JU.0000000000000310
30. Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holtgrewe HL, et al. The American urological association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol*. (1992) 148:1549–57. doi: 10.1016/S0022-5347(17)36966-5
31. Reeve E, Low LF, Shakib S, Hilmer SN. Development and validation of the revised patients' attitudes towards deprescribing (rpatd) questionnaire:

versions for older adults and caregivers. *Drugs Aging*. (2016) 33:913–28. doi: 10.1007/s40266-016-0410-1
32. Hays RD, Spritzer KL, Schalet BD, Cella D.PROMIS((R))-29 v2.0 profile physical and mental health summary scores. *Qual Life Res*. (2018) 27:1885–91. doi: 10.1007/s11136-018-1842-3
33. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. *Med Care*. (2012) 50:1013–9. doi: 10.1097/MLR.0b013e318269e121
34. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among USadults. *J Med Syst*. (2016) 40:153. doi: 10.1007/s10916-016-0492-7
35. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. *Rheum Dis Clin*. (2019) 45:159–72. doi: 10.1016/j.rdc.2019.01.004
36. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR mHealth uHealth*. (2017) 5:e37. doi: 10.2196/mhealth.6496
37. Moore CG, Carter RE, Nietert PJ, Stewart PW. Recommendations for planning pilot studies in clinical and translational research. *Clin Transl Sci*. (2011) 4:332–7. doi: 10.1111/j.1752-8062.2011.00347.x
38. Viechtbauer W, Smits L, Kotz D, Bude L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. *J Clin Epidemiol*. (2015) 68:1375–9. doi: 10.1016/j.jclinepi.2015.04.014
39. Kobayashi M, Tokue A, Morita T. Discontinuation of tamsulosin treatment in men with lower urinary tract symptoms: a pilot study. *Urol Int*. (2006) 76:304–8. doi: 10.1159/000092052
40. Yokoyama T, Watanabe T, Saika T, Nasu Y, Kumon H, Miyaji Y, et al. Natural course of lower urinary tract symptoms following discontinuation of alpha-1-adrenergic blockers in patients with benign prostatic *hyperplasia*. *Int J Urol*. (2007) 14:598–601. doi: 10.1111/j.1442-2042.2007.01784.x
41. Son H, Cho SY, Park S, Kang JY, Kim CS, Kim HG. A retrospective study of clinical outcomes of alpha-blocker or finasteride monotherapy followed by combination therapy: determination of the period of combination therapy of alpha-blocker and *finasteride*. *Int J Clin Pract*. (2013) 67:351–5. doi: 10.1111/ijcp.12045
42. Lee JY, Kang DH, Park SY, Lee SW, Kim YT, Choi HY, et al. Effect ofdiscontinuation of tamsulosin in korean men with benign prostatic hyperplasia taking tamsulosin and dutasteride: an open-label, prospective, randomized pilot study. *Lower Urinary Tract Symptoms*. (2012) 4:35–40. doi: 10.1111/j.1757-5672.2011.00109.x
43. Chung JH, Lee JY, Kang DH, Jo JK, Lee JW, Lee SH, et al. Evaluation of patient outcome after discontinuation of alfuzosin treatment for benign prostatic hyperplasia: a multicentre, prospective study. *Int J Clin Pract*. (2013) 67:870–5. doi: 10.1111/ijcp.12108
44. Nickel JC, Barkin J, Koch C, Dupont C, Elhilali M. Finasteride monotherapy maintains stable lower urinary tract symptoms in men with benign prostatic hyperplasia following cessation of alpha blockers. *CanUrolAssoc J*. (2008) 2:16–21. doi: 10.5489/cuaj.520
45. Lin VC, Liao CH, Kuo HC. Progression of lower urinary tract symptoms after discontinuation of 1 medication from 2-year combined alpha-blocker and 5-alpha-reductase inhibitor therapy for benign prostatic hyperplasia in men–a randomized multicenter study. *Urology*. (2014) 83:416–21. doi: 10.1016/j.urology.2013.09.036

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

PERSPECTIVE
published: 30 June 2020
doi: 10.3389/fcomp.2020.00021

A Conceptual Framework for Personal Science

Gary Isaac Wolf[1*†] *and Martijn De Groot*[2†]

[1] *Article 27 Foundation, Berkeley, CA, United States,* [2] *Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands*

This paper introduces a conceptual framework to guide research and education into the practice of personal science, which we define as using empirical methods to pursue personal health questions. Personal science consists of five activities: questioning, designing, observing, reasoning, and discovering. These activities are conceptual abstractions derived from review of self-tracking practices in the Quantified Self community. These practices have been enabled by digital tools to collect personal real-world data. Similarities and differences between personal science, citizen science and single subject (N-of-1) research in medicine are described. Finally, barriers that constrain widespread adoption of personal science and limit the potential benefits to individual well-being and clinical and public health discovery are briefly discussed, with perspectives for overcoming these barriers.

Keywords: quantified self, self-tracking, N-of-1, citizen science, digital health, personal science

OPEN ACCESS

Edited by:
Suzanne McDonald, The University of Queensland, Australia

Reviewed by:
Andrew McNeill, Northumbria University, United Kingdom
Shihao Yang, Harvard Medical School, United States

****Correspondence:***
Gary Isaac Wolf
gary@article-27.org

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *14 October 2020*
Accepted: *28 May 2020*
Published: *30 June 2020*

Citation:
Wolf GI and De Groot M (2020) A Conceptual Framework for Personal Science. Front. Comput. Sci. 2:21. doi: 10.3389/fcomp.2020.00021

INTRODUCTION

The use of empirical methods for learning and discovery is a valuable human capacity. Our interest in the use of empirical methods for personal exploration was stimulated by a decade of work in the Quantified Self community. The term Quantified Self was introduced as a journalistic description of the emerging practice of using technology for self-tracking (Wolf, 2010). Prior to its first publication, however, the term was in use for several years as the name for a hobbyist group of users and makers of self-tracking tools whose participants shared an interest in "self-knowledge through numbers." The first meeting of the Quantified Self was held in October, 2008 in Pacifica, California. Participants in the original Quantified Self group met frequently and in growing numbers during 2008 and 2009. Over time, these meetings came to focus mainly on first person accounts of self-tracking projects and experiments. This focus was encouraged by "three prime questions" addressed in each talk: What Did You Do? How Did You Do It? What Did You Learn?[1] In the period between 2009 through 2019, the network of Quantified Self meetings grew to ∼110 meetings in 30 countries.

The Quantified Self community has become a focus of academic research interest in the cultural effects of new technologies (see, for instance: Lee, 2013; Morozov, 2013; Swan, 2013; Choe et al., 2014; Lupton, 2016; Neff and Nafus, 2016; Ajana, 2017; Sharon, 2017; Heyen, 2020). A key set of scholarly questions addressed by sociological researchers into the Quantified Self community relates to how the self-research activities seen in the meetings and documented in its archive of presentations can be characterized (Almalki et al., 2015; Lupton, 2016; Heyen, 2020). As one of these scholars, Nils Heyen, has recently written:

> From the perspective of science (and technology) studies and public understanding of science, these self-tracking research practices are an interesting empirical phenomenon that deserves further scrutiny. It seems that, far away from institutional science, some lay people or citizens, at least no professional scientists, use methods and procedures known from science such as research design, data collection or data analysis in order to produce knowledge for self-use in their daily lives. How can the relationship between science and society, or science and the public... be characterized here?

[1] https://quantifiedself.com/blog/our-three-prime-questions/

Building on earlier work, Heyen uses the term "personal science" to describe these self-research practices. Here, we introduce a conceptual framework for personal science, in order to guide research into this practice and support its acquisition as a skill. The features we describe as typical of personal science are analytical abstractions grounded in self-tracking practices common in the Quantified Self community.We hope our framework will be useful to scholars working on characterizing this kind of self-research, as well as to practitioners and tool makers who share our interest in advancing everybody's capacity to take advantage of empirical methods to address personal questions.

PERSONAL SCIENCE: DEFINITION, EXAMPLES, AND ACTIVITIES

We define personal science as the practice of using empirical methods to explore personal questions. Copious material exemplifying this practice can be reviewed in the public archive of the Quantified Self community, which currently contains records of 1,093 presentations, of which 508 have been transcribed and 385 have been published online.

For instance, in September, 2011, Lindsey Meyer suffered an ontologic emergency: the loss of all hearing in one ear. Later that year, she tracked the partial return of her hearing through treatment with oral prednisone and intratympanic injections of dexamethasone. She was able to both plot the improvement and, in advance of her physician, determine when the benefit was leveling off (Meyer, 2011).

In March 2012, Sara Riggare used a smartphone app for finger tapping to explore daily variations of the effects of her medications for Parkinson's disease. She was thereby able to glean important insights about her disease and treatment that could not easily have been discovered in any other way (Riggare, 2012).

In the spring of 2018, Thomas Blomseth Christiansen used a one-button instrument for recording the times he felt an itching in his nose. Using this data in combination with pictures from a GoPro camera, he was able to determine which plants caused his allergies (Christiansen, 2018).

These examples, among many other, have five kinds of activities in common as described in the following paragraphs and depicted in **Figure 1**.

Questioning

Personal science specifically addresses personal questions. We use the word personal in its ordinary language sense, pointing to questions directly relevant to the individual asking the questions, and concerning their private life, experiences, and emotions. It is the self-reflexive quality of the questions that makes personal science personal; that is, the researcher's own life is the research domain. Personal science always involves the deliberate choice of the individual about what questions to ask, what methods to use and what observations to make.

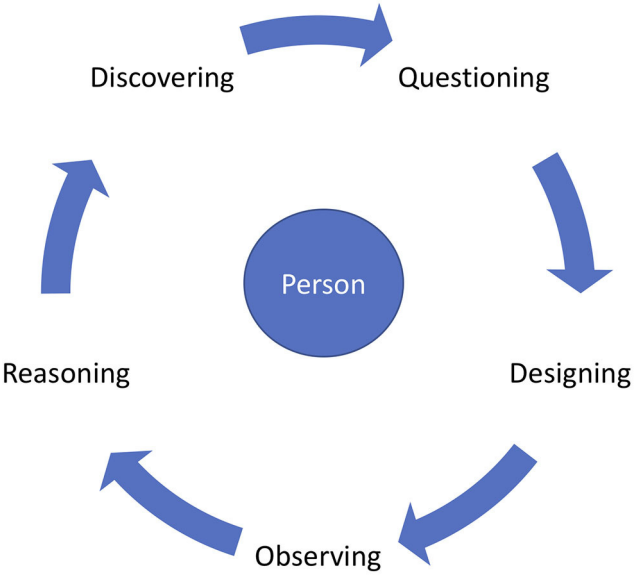


FIGURE 1 | Personal science is defined as the practice of using empirical methods to explore personal questions. This practice is characterized by five kinds of activities and begins with questioning.

Designing

Empirical approaches vary widely in formality and complexity. Simple self-observation of a single variable across time allows self-reflection on patterns of change. More complex self-experiments with alternating conditions or interventions can reveal cause-and-effect relationships when important questions justify the effort. The design activities in personal science involve exploring, applying, and adapting empirical methods to suit the aims, needs and skills of the person involved.

Observing

Personal science uses observations made through self-tracking, which is a process of deliberately collecting and structuring observations about one's own life. Widely varying types of instrumentation can be chosen to make these observations, including medical assays, digital technology, and pen and paper. The data can represent physical, emotional, social, or environmental phenomena. Observations can be gathered actively, as when a self-tracker presses a button on a device or creates handwritten notes; or, passively, using sensors on the body or "data exhaust" from digital records associated with other devices such as mobile phones and computers. The development of personal science has been supported by digital technologies that allow the collection of detailed real-world data about individuals. However, the most important distinguishing feature does not lie in the instrumentation used for making observations, but in the control of the self-tracking process by individuals exploring their own questions.

Reasoning

The reasoning process in personal science is carried out mainly by the individual who is its main subject. Individual control of reasoning means determining what questions to ask, what methods to use, who to consult for help, and balancing risks and benefits. Approaches to reasoning with one's own data may include data exploration, analysis, and visualization, development of formal experiments, and active reflection on one's own and with the help of peers and healthcare practitioners. Records of self-observation in a project of personal science often take the form of tabular real-world data, and the reasoning process typically involves common methods of digital data analysis, such as creating time-series graphs.

Discovering

In personal science, learning occurs throughout the whole cycle. We use the word discovering to refer to specific steps of consolidation and sharing. By consolidation we mean practical actions to improve daily life and by sharing we mean public discussion and dissemination. Insights gained during this process often lead to further questions that can initiate a new cycle of exploration. Consolidation and sharing of discovery depend on having time, freedom, and power to act on what's been learned, as well as support from peers and professionals.

PERSONAL SCIENCE IN HEALTH

Motives for developing a question that lead to self-tracking projects include curiosity, the need for problem solving, the quest for improvement to health and well-being, and pleasure in tinkering and creative expression. Self-tracking projects are often inspired by an encounter with the insights others have gained through similar projects presented at Quantified Self meetings or shared on the internet. Although the skills and activities associated with empirical self-study and encompassed by the concept of personal science are widely applicable, here we focus on health for three reasons: first, health is a dominant theme in Quantified Self practices; second, many individuals attempting to address consequential personal questions using their own self-collected data depend on collaboration from health practitioners, who require clear definitions and methodological context for evidence-based practice; and, third, the greatest promise of widespread benefit from access to the tools and methods of personal science lie in the health domain.

The specific goal of a personal science project may be as simple as checking intermittent clinical measurement of key biometrics with more frequent home measurements, or as complex as assessing the effect of highly structured self-administered trials of interventions. Along with the examples described above, self-trackers in the Quantified Self community have presented projects relating to, for instance: adult-onset acne, atrial fibrillation, Crohn's disease, IBS, asthma, and headache. The value of these projects to participants are not limited to cases where medical treatment is the main concern. Self-trackers have presented personal explorations on effects of diet on sleep; progress toward sports and fitness goals; patterns in mood and stress; along with many other distinct topics.

Self-trackers presenting at Quantified Self meetings often report that their projects have been useful even when their initial question was not resolved. Ancillary benefits include deeper learning about a health topic; generation of new ideas for improving their own care; productive engagement with clinicians; and providing a sense of agency while dealing with the stress of disease and treatment. The sense of agency is noted by self-trackers across a wide range of expertise. Even individuals who are clinicians themselves, or otherwise highly trained as researchers, have described the discouraging passivity induced by the experience of illness. Reasoning about one's own conditioning counteracts that passivity. As Dr. Larry Smarr, who suffers from Crohn's disease, put it in a QS community "Show&Tell" talk in 2011: "When you get a sense of knowledge, and, if not control then at least a sense that you can understand what's happening to you—then there's hope."

PERSONAL SCIENCE AND N-OF-1 TRIALS IN MEDICINE

In the health domain, interest in addressing highly individual questions has a long history. The personal science framework reaches back to the tradition of single subject research design in applied fields of psychology, education, and human behavior, where it has benefited from extensive methodical research and practical guidance for practitioners (Kratochwill and Levin, 2010; Kazdin, 2011). However, the personal science framework and single subject research design are not the same.

Single subject research design is a scientific method in which an individual person serves as the research subject. Where single subject research takes the form of a rigorously conducted N-of-1 trial, it has high evidential value in medicine (Vohra et al., 2016). The limitations of N-of-1 trials in medicine are well-understood (Mirza et al., 2017). The urgency of acute illness offers limited time for rigorous trials, while more slowly progressing or chronic conditions require lengthy commitment to individually-focused discovery that clinical practice does not normally support (Kravitz and Duan, 2014). The aim of clinical studies of N-of-trials in medicine is to deliver results that simultaneously provide personal benefit, are clinically practical, and create generalizable knowledge that can be broadly applicable. Proponents of N-of-1 trials in medicine have the burden of showing that they improve patient outcomes in comparison to standard treatment (Kravitz et al., 2018; Mirza and Guyatt, 2018; McDonald et al., 2019).

The aims of personal science are different. As the extensive examples in the Quantified Self archive demonstrate, personal science typically involves reasoning about problems medicine often does not address: highly individual, often long term personal challenges, and questions such as finding the triggers of intermittent conditions in everyday life, understanding the effects of changes in diet and daily activities on physical and mental health, or using regular measurements to guide day-to-day decision making about sports, travel, work, and management of chronic disease. Individuals spend time and effort figuring out for themselves what they should measure that can give them insight into their question, choosing measurements that are personally

relevant rather than clinically defined. For these reasons, personal science should not be understood to be identical to N-of-1 trials in medicine. Although self-trackers benefit from clinical knowledge and use it, where relevant, to guide their projects, personal science is a distinct form of empirical reasoning focused on health questions arising in daily life.

CITIZEN SCIENCE AND PERSONAL SCIENCE

In 1991, Martin and Brouwer (1993) introduced the term personal science to describe an approach to characterizing scientific practice for young students, an approach that emphasized that "science is not simply rational and objective but that the inquiring person is an integral part of the enquiry." Personal science as a concept was derived from work of noted physical chemist and philosopher of science Michael Polanyi, whose work emphasized the tacit and subjective dimensions of mainstream scientific practice (Polanyi, 1958). Although in its origin the term personal science has wide scope, it has more recently been used to refer to the kinds of self-research described here (Roberts, 2014; de Groot et al., 2017; Heyen, 2017). In a well-informed and sensitive analysis that draws extensively on research conducted in the Quantified Self community, the sociologist Nils Heyen proposed "personal science" as the specific term labeling the practice of exploring one's own personal questions using empirical methods (Heyen, 2020). The distinctness of personal science relates both to the methods of its practice and to the domain of its application. Personal science shares the same overarching empirical framework as science generally. However, the research is motivated by personal questions salient in everyday life, it's methods are typically simple, and the discoveries are applicable directly by the person doing the research.

The practice of using empirical observations to address personal questions has clear similarities to citizen science. In the mid-1990s, Alan Irwin proposed the term "citizen science" to describe active collaboration between scientists and the public to understand complex ecological challenges and develop new approaches (Irwin, 1995). In published scientific literature, most projects described as citizen science involve volunteer contributions of observations and classification of data in ornithology, astronomy, meteorology and microbiology (Kullenberg and Kasperowski, 2016). However, in recent years, the term citizen science has come to encompass a wide range of diverse approaches to involving non-professionals in research. In its most general definition, citizen science describes research in which non-professionals play an active role in funding, data collection/generation, analysis, interpretation, application, dissemination, or evaluation (Mueller et al., 2011; Prainsack, 2014).

In the practice of personal science non-professionals occupy most, if not all, of the significant roles in research. However, this approach is distinct from citizen science in significant ways. Personal science is self-directed: that is, the subject of the research is also the primary investigator. This feature is not present in the majority of citizen science projects (NAS - National Academies of Sciences, 2018). The selection of topics and questions in personal science are determined by the researcher's personal motive alone; in citizen science the questions are typically determined by the research agenda of a scientific discipline. Where personal science addresses a health question, it typically aims at a specific personal question, where citizen science aims at creating generalizable knowledge. Despite these key differences, personal science and citizen science are aligned around a common commitment to democratic participation in science (Vayena and Tasioulas, 2015).

PROSPECTS

We envision a world of personal scientists, in which everybody has access to the support that they need to address their own questions about health and well-being using the tools and methods of science. Broadening participation in personal science is needed to catalyze new discoveries at both the individual and collective level. Widespread participation in personal science primarily benefits the individuals who make personal discoveries, but it can also make important contributions to the next generation of clinical and public health studies, which depend on data gathered in daily life which, when aggregated, may allow for population-based effect estimates. Engagement in personal science strengthens the empirical foundation for discovery generally.

Three developments in digital technology have been important in encouraging the spread of personal science: access to science on the web; online community and peer support; and, digital tools for sensing and tracking. However, to make personal science widely accessible, significant barriers must be overcome. First, methods for personal science are underdeveloped. Today, practical approaches to formulating good questions, setting up a self-tracking project, and visualizing and learning from one's own personal data tend to be handcrafted by individuals as they work on their projects. Translating common features of these handcrafted methods into designs that can be easily shared and adapted for personal use by many people will lower the barrier to participation. Second, personal science depends on self-tracking tools. However, today's commercial digital self-tracking tools are not appropriate for many types of personal questions. Digital data is often not accessible to users, who face important privacy and security threats, while instrumentation for many kinds of personal questions is expensive, inflexible, or lacking altogether. Third, all learning requires social support. Support for personal science is especially needed from those health professionals most directly concerned with individual care, including nurses, physical therapists, and specialists in rehabilitation and elder care. Broad recognition of the value of personal science is needed so that individuals can find encouragement, inspiration, and education both in the healthcare system and in other domains where consequential questions arise. Finally, people doing self-research require time to make discoveries and power to act on them. In this respect, support for personal science inevitably touches on broader social issues of democratization and social change. We hope our framework is helpful in furthering this change.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://quantifiedself.com/show-and-tell/.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Ajana, B. (2017). Digital health and the biopolitics of the Quantified Self. *Digital Health* 3:2055207616689509. doi: 10.1177/2055207616689509

Almalki, M., Gray, K., and Sanchez, F.M. (2015). The use of self-quantification systems for personal health information: big data management activities and prospects. *Health Inf. Sci. Syst.* 3 (Suppl. 1):S1. doi: 10.1186/2047-2501-3-S1-S1

Choe, E.K., Lee, N.B., Lee, B., Pratt, W., and Kientz, J.A. (2014). "Understanding quantified-selfers' practices in collecting and exploring personal data," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON), 1143–1152. doi: 10.1145/2556288.2557372

Christiansen, T.B. (2018). *Which Grasses Aggravate My Allergies?* Quantified Self Show&Tell Library. Available online at: https://quantifiedself.com/show-and-tell/?project=1122 (accessed February 13, 2020).

de Groot, M., Drangsholt, M., Martin-Sanchez, F. J., and Wolf, G. (2017). Single subject (N-of-1) research design, data processing, and personal science. *Methods Inform. Med.* 56, 416–418. doi: 10.3414/ME17-03-0001

Heyen, N.B. (2017). *Quantified Self as Personal (Citizen) Science*. Havard Bill of Health. Available online at: http://blogs.harvard.edu/billofhealth/2017/05/11/quantified-self-as-personal-citizen-science/ (accessed February 12, 2020).

Heyen, N. B. (2020). From self-tracking to self-expertise: the production of self-related knowledge by doing personal science. *Public Underst Sci.* 29, 24–138. doi: 10.1177/0963662519888757

Irwin, A. (1995). *Citizen Science: A Study of People, Expertise, and Sustainable Development*. London: Routledge.

Kazdin, A. E. (2011). *Single-Case Research Designs: Methods for Clinical and Applied Settings (2nd ed.)*. New York, NY: Oxford University Press.

Kratochwill, T.R., and Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychol Methods* 15, 124–144. doi: 10.1037/a0017736

Kravitz, R. L., and Duan, N. (2014). *Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF*. Rockville, MD: Agency for Healthcare Research and Quality. Available online at: www.effectivehealthcare.ahrq.gov/N-~1-Trials.cfm

Kravitz, R. L., Schmid, C. H., Marois, M., Wilsey, B., Ward, D., and Hays, R. D. (2018). Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain. *Eur PMC.* 178, 1368–1377. doi: 10.1001/jamainternmed.2018.3981

Kullenberg, C., and Kasperowski, D. (2016). What is citizen science? – A scientometric meta-analysis. *PLoS ONE* 11:e0147152. doi: 10.1371/journal.pone.0147152

Lee, V. R. (2013). The Quantified Self (QS) movement and some emerging opportunities for the educational technology field. *Educ. Technol.* 39–42. doi: 10.2190/ET.42.1.d

Lupton, D. (2016). *The Quantified Self*. Malden, MA: Polity Press.

Martin, B., and Brouwer, W. (1993). Exploring personal science. *Sci. Educ.* 77, 441–459. doi: 10.1002/sce.3730770407

McDonald, S., and McGree, J., and Bazzano, L. (2019). Finding benefit in n-of-1 trials. *JAMA Internal Med.* 179, 454–455. doi: 10.1001/jamainternmed.2018.8382

Meyer, L. (2011). *Tracking My Hearing Loss, Quantified Self Show&Tell Library*. Available online at: https://quantifiedself.com/show-and-tell/?project=424 (accessed February 13, 2020).

Mirza, R. D., and Guyatt, G. (2018). A randomized clinical trial of n-of-1 trials—tribulations of a trial. *JAMA Internal Med.* 178, 1378–1379. doi: 10.1001/jamainternmed.2018.3979

Mirza, R. D., Punja, S., Vohra, S., and Guyatt, G. (2017). The history and development of N-of-1 trials. *J. R. Soc. Med.* 110, 330–340. doi: 10.1177/0141076817721131

Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. Chester, UK: Public Affairs.

Mueller, M. P., Tippins, D., and Bryan, L.A. (2011). The future of citizen science. *Democr. Educ.* 20, 1–12. Available online at: https://democracyeducationjournal.org/home/vol20/iss1/2.

NAS - National Academies of Sciences, Engineering, and Medicine (2018). *Learning Through Citizen Science: Enhancing Opportunities by Design*. Washington, DC: National Academies Press.

Neff, G., and Nafus, D. (2016). Self-tracking. Cambridge: MIT Press. doi: 10.7551/mitpress/10421.001.0001

Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: University of Chicago Press

Prainsack, B. (2014). "Understanding participation: the 'citizen science' of genetics," in *Genetics as Social Practice*, eds B. Prainsack, G. Werner-Felmayer, G. Schicktanz (Farnham: Ashgate), 147–164.

Riggare, S (2012). *Tracking Parkinsons and Medication.* Quantified Self Show&Tell Library. Available online at: https://quantifiedself.com/show-and-tell/?project=479 (accessed February 13, 2020).

Roberts, S. (2014). *Seth's Blog - Personal Science, Self Experimentation, Scientific Method.* Available online at: https://sethroberts.net/category/personal-science/ (accessed February 12, 2020).

Sharon, T. (2017). Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. *Phil. Technol.* 30, 93–121. doi: 10.1007/s13347-016-0215-5

Swan, M. (2013). The quantified self: fundamental disruption in big data science and biological discovery. *Big data* 1, 85–99. doi: 10.1089/big.2012.0002

Vayena, E., and Tasioulas, J. (2015). We the scientists: a human right to citizen science. *Phil. Technol.* 28, 479–485. doi: 10.1007/s13347-015-0204-0

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., et al. (2016). CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *J. Clin. Epidemiol.* 76, 9–17. doi: 10.1016/j.jclinepi.2015.05.004

Wolf, G. (2010). The data-driven life. New York, NY: The New York Times, 28.

Conflict of Interest: GW is co-founding director of Quantified Self Labs (LCC, San Francisco), a social enterprise that supports the Quantified Self community worldwide since 2009. MD was (2012–2017) co-founding director of Quantified Self Institute at Hanze University of Applied Sciences Groningen. Both authors are currently associated with the Article 27 Foundation, an international non-profit network organization which aims to support personal science.

frontiers in Computer Science

OPINION
published: 27 August 2020
doi: 10.3389/fcomp.2020.00038

The Quantified Self-in-Place: Opportunities and Challenges for Place-Based N-of-1 Datasets

*Benjamin W. Chrisinger**

Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

Keywords: N-of-1, wearable sensors, neighborhood, built environment and health, mHealth, citizen science, evidence based policy

Causal mechanisms connecting place and health have long vexed public health researchers and epidemiologists. While some relationships occur along readily measurable pathways, many linkages are less clear. This is especially true of non-communicable or "lifestyle" diseases, which often exist in a conceptual "black box"—wherein multiple, possibly interacting and interconnected, mechanisms complicate population-level generalizations about exactly how places affect health (Macintyre et al., 2002). Without observations over a variety of potential pathways, time periods, and individual-level characteristics and behaviors, researchers are limited to relatively high-level observations of associations between the characteristics of people and the places where they live, work, and play.

When paired with geocoordinates, data from self-tracking technologies can offer new opportunities for researchers and participants to explore these causal pathways. This paper describes how geolocated N-of-1 datasets could contribute to inquiries about place and health, and improve upon common limitations of place-based research. It also identifies several significant logistical, methodological, and ethical issues that could present barriers to these kinds of projects. Overall, the paper offers a vision for situating the quantified self *in place*, where researchers could support and amplify the creativity of self-tracking communities, and build testable hypotheses from rich, multidimensional datasets. With appropriate attention to the inherent challenges and limitations, researchers and self-trackers can meaningfully expand our knowledge of place effects on human health.

OPEN ACCESS

Edited by:
Eric Hekler,
University of California, San Diego, United States

Reviewed by:
Marta Jankowska,
University of California, San Diego, United States
Jochen Meyer,
Institute for Information Technology (OFFIS), Germany

***Correspondence:**
Benjamin W. Chrisinger
benjamin.chrisinger@spi.ox.ac.uk

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *29 November 2019*
Accepted: *23 July 2020*
Published: *27 August 2020*

Citation:
Chrisinger BW (2020) The Quantified Self-in-Place: Opportunities and Challenges for Place-Based N-of-1 Datasets. Front. Comput. Sci. 2:38. doi: 10.3389/fcomp.2020.00038

TECHNOLOGICAL OPPORTUNITIES IN SELF-TRACKING

A new era of place-based study has been ushered in by the development of relatively low-cost, portable geographic positioning system (GPS) technologies (Pendyala and Bhat, 2012). Researchers can now more precisely describe daily exposures within *activity spaces*, rather rely on static, administrative, and often residentially-focused representations of place, which suffer from a variety of validity issues (Cummins et al., 2007; Juarez et al., 2014). "Modifiable unit" problems can arise out of the relative arbitrariness of certain cut-offs and levels of aggregation (e.g., time points or "spatial boundaries") (Dark and Bram, 2007; Cheng and Adepeju, 2014). For instance, estimates of an individual's exposure to tobacco retailers could vary substantially if outlets are summarized as counts within streets, blocks, or counties. Kwan famously extended these challenges by describing the *uncertain geographic context problem* (UGCoP), which highlighted the potentially significant influence of spatial boundary definitions on constructs of exposure and behavior (Kwan, 2012a). High-resolution GPS data can alleviate some of these concerns (Kwan, 2012b), and the development of wearable, environmental sensing technologies, such as portable air pollution monitors (Koehler and Peters, 2015; Seto et al.), has also enabled new kinds of individual-level

datasets to be generated. Thus, individual-level, geolocated datasets could allow researchers to improve upon common limitations of GIS approaches, including issues with spatiotemporal joining of layers and datasets.

Opportunities for Place-Based N-OF-1

The proliferation of GPS and GIS technologies present particular opportunities to those interested in single-subject or N-of-1 investigations (De Groot et al., 2017). As identified by a recent systematic review of GPS measures in built environment-physical activity studies by Yi and colleagues, individual-level, geolocated datasets might also help address selection bias in place-based research by helping account for potentially influential individual characteristics, and enabling improved experimental or quasi-experimental research designs (Yi et al., 2019). Analytical frameworks, such as those proposed by Jankowska, Schipperijn, and Kerr for physical activity research (Jankowska et al., 2015), also provide guidance for integrating participant-level GPS and behavior data with existing GIS layers, and conceptualizing the different temporal, spatial, and behavioral exposures.

By placing N-of-1 investigations in environmental contexts, researchers might better understand how treatments and interventions work, for whom, and under which conditions. In patient-provider settings, the place-health nexus often presents a challenge to prescribing appropriate behavioral solutions to health problems, even in situations supported by self-tracking. For example, a physician might recommend that a patient increase their daily minutes of physical activity and use a smartphone app to track their progress. Absent contextual information, these activity data say little about how and where the provider's suggestions were put into practice. If the patient achieves the prescribed targets, important insights could be gained from learning exactly how they did it, and where. Did the patient pursue physical activity in their neighborhood? At a gym or at work? Using community parks or sidewalks? Contextual, and especially environmental information, present new dimensions of considering health behavior change, and, with appropriate supports for data collection and interpretation, could offer providers clues about the generalizability of their recommendations.

Place-based N-of-1 datasets might also help patients and providers retest significant observations to understand if and how findings depend on contextual factors. Many smartphone owners have already collected baseline datasets that, if paired with location data, could be used prospectively or retrospectively to investigate changes and relationships between environment, behavior, and health. These might include actively recorded information (e.g., meal records logged in a diet app), as well as those collected passively, such as daily step counts. For instance, a patient might observe a correlation between their daily step counts and the average walkability of the environments where they spent time. However, if they base their observation on data from warm summer months, and fail to adjust for average outside temperature, they could wrongly conclude that the place effect is not generalizable to colder times or environments. The outside temperature is likely to moderate the place effect of walkable street networks, but not wholly determine it. With observations across multiple spatial and temporal settings, place-based N-of-1 datasets open new avenues for considering the relative influence of different contextual factors.

Everyday "Natural Experiments"

Together, researchers and long-term self-tracking participants could also conceptualize possible areas for inquiry in advance of changes to the built environment or policy, or in retrospective studies that leverage geolocated tracking data across numerous dimensions (Fox and Duggan, 2013). These investigations could include "natural experiments" that emulate the methodological rigor of randomized controlled trials (RCTs), while also providing feedback to stakeholders about the effectiveness of a program or policy (Sampson, 2010). Natural experiments can pose challenges for researchers, including how to conceptualize and adequately measure treatment (e.g., the dose, duration, or intensity of the intervention), and develop rigorous baseline datasets with appropriate counterfactuals. These issues can sometimes prove to be prohibitive, especially in terms of time and resources required to prospectively collect baseline data, or in cases where the intervention could not have been anticipated (e.g., disruptions to a transport network, or shelter-in-place orders related to COVID-19).

Natural experiment designs can also face logistical or ethical challenges when applied in community settings (Sampson, 2010). For example, while random assignment to a treatment or control group is a cornerstone of RCTs that helps guard against selection bias, such assignments can be impossible or impractical. Quasi-experimental evaluation of how a new neighborhood park influences residents' physical activity demands that a reasonable control group be identified (i.e., a neighborhood that could have received the treatment, but did not). However, the determination of where and when a new park is built is, in reality, far from random, and subject to influence from unobserved or unmeasured political and community factors. As Sampson observes in his critique of the "experimental turn" in evaluation research, "the hard truth is that we have little choice but to adapt in creative ways to the limitations that confront all social science inquiry" (Sampson, 2010). The place-based N-of-1 dataset offers one such creative adaptation to the random assignment problem by making available potential counterfactuals from within-subject baseline data. This could strengthen both the validity of the treatment or exposure variable, as well as our confidence in the comparability of the control units.

Quantified Self-in-Place

The opportunities for place-based N-of-1 studies are complemented by a broadening public use of wearable and self-tracking technologies, including a growing "quantified self" movement of individuals who use self-measurement to improve or optimize aspects of their lives, like health, happiness, or productivity (Fox and Duggan, 2013; Lupton, 2016; De Moya and Pallud, 2020). These avid self-trackers might be willing to volunteer long-term "baseline" data and may also be tracking across multiple devices or applications. Additionally, those in the quantified self-community could be interested in developing and testing new self-tracking technologies in

collaboration with researchers. Thus, by working with quantified self-participants, researchers not only gain access to unique and potentially geolocated datasets, but also draw upon the community's ingenuity and curiosity to develop new tools or hypotheses.

Thinking of place-based N-of-1 research in a community-engaged or participatory research frameworks also has epistemological and ontological benefits. Citizen science approaches, which might include the quantified self-movement, can help integrate participants' environmental perceptions into otherwise "objective" data collection methodologies (Pykett et al., 2020). While more controlled research settings use standardized ecological momentary assessments (EMA), citizen science approaches may instead follow more participant-driven protocols, though they might still employ standardized tools (e.g., photo-taking, neighborhood audits). This allows researchers to further engage participants in formulating hypotheses and interpreting outcomes, which could be especially relevant in exploratory settings without clear expectations of cause-and-effect relationships.

These engagement-focused approaches respond to calls for self-tracking researchers to leverage both quantitative and qualitative methods (Gilmore, 2016), such as the "citizen social science" described by Pykett and colleagues, whereby individual-level data are measured via wearables and also elicited through surveys and interviews (Pykett et al., 2020). Communities of self-trackers might also share insights to help one another optimize a behavioral intervention, or collectively assess the impact of an environmental change. Examples of quantified self-communities organizing for mutual support, learning, and advocacy are also evocative of the empowerment and engagement potential identified by De Groot et al. (2017), King et al. (2019), and De Moya and Pallud (2020). Thus, a complete vision for "quantified self-in-place" projects should include possible hypotheses of place-health relationships, and also make room for participants to suggest new directions, tools, or variables.

CHALLENGES

These opportunities are not without logistical, methodological, and ethical challenges. While some of these may be addressed with future technological improvements, it is important to recognize both the current limits to our capabilities and methods, as well as the potential risks that the imagined high-resolution, individual-level datasets might introduce.

Logistical Issues

Collecting and interpreting place-based N-of-1 datasets is no small task. Geospatial researchers have increased the internal validity of exposure measures with high-frequency GPS tracking, though conceptual (e.g., how do we operationalize exposure to a neighborhood park?) and logistical (e.g., how often should location be recorded to capture exposure?) questions remain. Furthermore, broad heterogeneity rooted in device-specific particularities and individual motivations for participation are likely to complicate or preclude between-participant comparisons from crowdsourced data. Crowdsourcing also requires that participants know how to collect and extract high-resolution, geolocated data from mobile applications or wearable devices. Even among tech-savvy quantified self-communities, self-trackers are sometimes limited in their ability to extract and analyze data from tracking devices.

When GPS, GIS, and biometric data are collected with the express purpose of integration, analysts can anticipate some of these challenges by setting data formatting standards, conducting sensitivity analyses, or making adjustments to statistical models. Examples of web-based dashboards that integrate specific kinds of geodata, such as Patrick and Kerr's Personal Activity Location Measurement System (PALMS), may provide inspiration for future open-source platforms that could guide users through the various steps and stages of collecting, curating, and interpreting their own multidimensional datasets (Kerr et al., 2011). Additionally, advanced computing technologies like machine learning could provide future opportunities for automating data cleaning and harmonization, as well as uncovering relationships between spatial, temporal, and biometric variables.

Representation and Inclusion

Social determinants of health exert strong influences on both health behaviors and outcomes, but these constructs may not be well-represented in place-based N-of-1 datasets (Gabriels and Moerenhout, 2018). The degree to which these variables are integrated from self-tracking sources depends both on whether they are valued and collected by researchers or data collection/integration platforms, and whether participant populations are distributed across these socioeconomic gradients. As others have noted, disadvantaged populations face barriers in accessing tracking technologies and responding to insights gleaned from self-tracking data (Lupton, 2016; Gabriels and Moerenhout, 2018; Lupton). Well-documented mistrust of academic, medical, and research communities among many marginalized and exploited groups, stemming from decades of real and perceived harms perpetrated against them, could also limit the applicability and acceptability of place-based N-of-1 projects in certain settings (George et al., 2013; Bonevski et al., 2014). Furthermore, participation through geolocated personal data could elevate concerns about the independence of researchers from other potentially mistrusted and surveillance-interested parties, such as financial institutions, police, immigration officials, or case workers. Clear delineation of data protection, processing, and sharing protocols are necessary to make N-of-1 studies accessible to all communities, including legally-informed procedures regarding data requests from outside parties.

Without addressing these concerns about inclusion, place-based N-of-1 studies may thus be limited to a subset of "worried well," relatively healthy and high-income individuals with time and resources to devote to self-study (Gabriels and Moerenhout, 2018; Lupton). While researchers might still leverage learnings from pilot testing among this specific community for broader applications, more inclusive thinking is needed to avoid replicating inequality in N-of-1 research. Novel participatory approaches which aim to reduce power imbalances

between researchers and participants potentially enable new collaborations that would not be possible in conventional patient research paradigms (English et al., 2018; King et al., 2019).

Construct Validity

Researchers may seek to operationalize relatively ambiguous constructs in place-based N-of-1 studies. In some domains, validated measurement standards provide a strong start, so that, for instance, a daily step count is derived from an assemblage of high-resolution accelerometer data. Environment may complicate this relationship; as in the step count example, local terrain or topography could be markedly different than the validated standard (Huang et al., 2016). Other constructs are the result of more complex physiological measurements (e.g., accelerometer data to gauge movement and intensity), and still others aim to represent psychological or social constructs (e.g., excitement, fear, stress) (Pykett et al., 2020). While these measures may have a biological basis derived from laboratory settings, important questions emerge about the validity of these measures when used in the field, especially measures that indicate response to external stimuli (Chrisinger and King, 2018; Pykett et al., 2020). These questions become even more complex when comparisons between individual datasets are desired, but data have been collected with different kinds of applications or devices, and/or individuals' motivations for contributing their data are unclear.

Additionally, researchers must be aware of the contested nature of place terminologies themselves, including UGCoP and other challenges (Kwan, 2012a). Still, recent projects recognizing the "personal and subjective" nature of spatial perception (e.g., two neighbors may define their neighborhood quite differently) provide examples for how these uncertainties might be conceptualized and addressed (Pykett et al., 2020; Meier and Glinka). Furthermore, the moderating influences of environmental perceptions are also important to consider and could be measured with complementary methods, such as EMA (Yi et al., 2019). Ultimately, N-of-1 researchers should be cognizant of the uncertainties that surround the different constructs invoked by their analyses, and how interpretation of these constructs might vary between places and people.

Participant Privacy

Finally, as it becomes easier to volunteer and merge discrete streams of personal information into geolocated datasets, we must bear in mind the risks to participant privacy. For example, the "digital fingerprints" corporations assemble for individual consumers using multiple sources of online activity data illuminates just one of the risks posed by the "exploited" self-tracking described by Lupton (2016), and emerging "surveillance capitalism" identified by Zuboff (2015). Potential ethical concerns about the intrusiveness of continuous GPS monitoring are possibly ameliorated in a quantified self-paradigm, where researchers and participants often agree to collect far more data than might otherwise be deemed necessary.

De Moya and Pallud describe possible benefits to self-tracking as "self-surveillance," observing how empowerment can come through visibility and accountability in data-sharing communities, and the ability to integrate data across multiple platforms and types (De Moya and Pallud, 2020). Still, the "consented self-surveillance" they describe relies on transparency of data integration and sharing, providing users opportunities to (dis)allow personal data from different sources to be integrated across platforms (De Moya and Pallud, 2020). Quantified self-participants might be more willing to accept more radical transparency in terms of data-sharing, though the exact privacy risks of geolocated datasets may not be entirely clear until they are created. Given the size and scale of place-based N-of-1 datasets, it may be difficult to ask participants to fully review and understand their contributions before volunteering them.

CONCLUSION

As new technologies expand our conceptualization of human and environmental variables, and the instruments used to measure them become more accessible to the general public in terms of cost, size, and skills required, the datasets available to health and built environment researchers will become increasingly large and multi-dimensional. For each the hundreds of potentially observable data points available to researchers, still greater numbers of linkages could be made to existing or simultaneously collected environmental data, enabling innovative observational and mechanistic studies that describe and predict the effects of place on human health. These place-based N-of-1 datasets also create new opportunities for engagement and collaboration outside of traditional researcher-participant paradigms, and may draw inspiration from flourishing quantified self and citizen science communities. While encouraging patients and citizen scientists to collect, analyze, and share their own data, researchers can also help educate participants about the challenges and opportunities inherent in place-based research. By developing a higher-resolution understanding of how place and health are connected for different individuals, the contours of etiological black boxes will become more legible, including the contextual and conditional dynamics that so often exist within them.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., et al. (2014). Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Med. Res. Methodol.* 14:42. doi: 10.1186/1471-2288-14-42

Cheng, T., and Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PLoS ONE* 9:e100465. doi: 10.1371/journal.pone.0100465

Chrisinger, B. W., and King, A. C. (2018). Stress experiences in neighborhood and social environments (SENSE): a pilot study to integrate the quantified self with citizen science to improve the built environment and health. *Int. J. Health Geogr.* 17:17. doi: 10.1186/s12942-018-0140-1

Cummins, S., Curtis, S., Diez-Roux, A. V., and Macintyre, S. (2007). Understanding and representing 'place' in health research: a relational approach. *Soc. Sci. Med.* 65, 1825–1838. doi: 10.1016/j.socscimed.2007.05.036

Dark, S. J., and Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress Phys. Geogr. Earth Environ.* 31, 471–479. doi: 10.1177/0309133307083294

De Groot, M., Drangsholt, M., Martin-Sanchez, F., and Wolf, G. (2017). Single subject (N-of-1) research design, data processing, and personal science. *Methods Inf. Med.* 56, 416–418. doi: 10.3414/ME17-03-0001

De Moya, J.-F., and Pallud, J. (2020). From panopticon to heautopticon: a new form of surveillance introduced by quantified-self practices. *Information Syst. J.* doi: 10.1111/isj.12284. [Epub ahead of print].

English, P. B., Richardson, M. J., and Garzón-Galvis, C. (2018). From crowdsourcing to extreme citizen science: participatory research for environmental health. *Annu. Rev. Public Health* 39, 335–350. doi: 10.1146/annurev-publhealth-040617-013702

Fox, S., and Duggan, M. (2013). *Tracking for Health*. Washington, DC: Pew Research Center. Available online at: https://www.pewresearch.org/internet/2013/01/28/tracking-for-health/ (accessed May 15, 2020).

Gabriels, K., and Moerenhout, T. (2018). Exploring entertainment medicine and professionalization of self-care: interview study among doctors on the potential effects of digital self-tracking. *J. Med. Internet Res.* 20:e10. doi: 10.2196/jmir.8040

George, S., Duran, N., and Norris, K. (2013). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *Am. J. Public Health* 104, e16–e31. doi: 10.2105/AJPH.2013.301706

Gilmore, J. N. (2016). Everywear: the quantified self and wearable fitness technologies. *New Media Soc.* 18, 2524–2539. doi: 10.1177/1461444815588768

Huang, Y., Xu, J., Yu, B., and Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait Posture* 48, 36–41. doi: 10.1016/j.gaitpost.2016.04.025

Jankowska, M. M., Schipperijn, J., and Kerr, J. (2015). A framework for using GPS data in physical activity and sedentary behavior studies. *Exerc. Sport Sci. Rev.* 43, 48–56. doi: 10.1249/JES.0000000000000035

Juarez, P. D., Matthews-Juarez, P., Hood, D. B., Im, W., Levine, R. S., Kilbourne, B. J., et al. (2014). The public health exposome: a population-based, exposure science approach to health disparities research. *Int. J. Environ. Res. Public Health* 11, 12866–12895. doi: 10.3390/ijerph111212866

Kerr, J., Duncan, S., Schipperijn, J., and Schipperjin, J. (2011). Using global positioning systems in health research: a practical approach to data collection and processing. *Am. J. Prev. Med.* 41, 532–540. doi: 10.1016/j.amepre.2011.07.017

King, A. C., Winter, S. J., Chrisinger, B. W., Hua, J., and Banchoff, A. W. (2019). Maximizing the promise of citizen science to advance health and prevent disease. *Prev. Med.* 119, 44–47. doi: 10.1016/j.ypmed.2018.12.016

Koehler, K. A., and Peters, T. M. (2015). New methods for personal exposure monitoring for airborne particles. *Curr. Envir. Health Rpt.* 2, 399–411. doi: 10.1007/s40572-015-0070-z

Kwan, M.-P. (2012a). The uncertain geographic context problem. *Ann. Assoc. Am. Geograph.* 102, 958–968. doi: 10.1080/00045608.2012.687349

Kwan, M.-P. (2012b). How GIS can help address the uncertain geographic context problem in social science research. *Ann. GIS* 18, 245–255. doi: 10.1080/19475683.2012.727867

Lupton, D. (2014). "Self-tracking cultures: towards a sociology of personal informatics," in *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures the Future of Design - OzCHI'14* (Sydney, NSW: ACM Press), 77–86.

Lupton, D. (2016). The diverse domains of quantified selves: self-tracking modes and dataveillance. *Econ. Soc.* 45, 101–122. doi: 10.1080/03085147.2016.1143726

Macintyre, S., Ellaway, A., and Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Soc. Sci. Med.* 55, 125–139. doi: 10.1016/S0277-9536(01)00214-3

Meier, S., and Glinka, K. (2017). Psychogeography in the age of the quantified self—mental map modelling with georeferenced personal activity data," in *Advances in Cartography and GIScience Lecture Notes in Geoinformation and Cartography*, ed M. P. Peterson (Cham: Springer International Publishing), 507–522.

Pendyala, R. M., and Bhat, C. R. (2012). *Travel Behaviour Research in an Evolving World*. Available online at: Lulu.com

Pykett, J., Chrisinger, B., Kyriakou, K., Osborne, T., Resch, B., Stathi, A., et al. (2020). Developing a Citizen Social Science approach to understand urban stress and promote wellbeing in urban communities. *Palgrave Commun.* 6, 1–11. doi: 10.1057/s41599-020-0460-1

Sampson, R. J. (2010). Gold standard myths: observations on the experimental turn in quantitative criminology. *J. Quant. Criminol.* 26, 489–500. doi: 10.1007/s10940-010-9117-3

Seto, E. Y. W., Giani, A., Shia, V., Wang, C., Yan, P., Yang, A. Y., et al. (2009). "Wireless body sensor network for the prevention and management of asthma," in *2009 IEEE International Symposium on Industrial Embedded Systems* (Lausanne), 120–123. doi: 10.1109/SIES.2009.5196203

Yi, L., Wilson, J. P., Mason, T. B., Habre, R., Wang, S., and Dunton, G. F. (2019). Methodologies for assessing contextual exposure to the built environment in physical activity studies: a systematic review. *Health Place* 60:102226. doi: 10.1016/j.healthplace.2019.102226

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* 30, 75–89. doi: 10.1057/jit.2015.5

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 28 August 2020
doi: 10.3389/fdgth.2020.00013

Optimizing Aggregated N-Of-1 Trial Designs for Predictive Biomarker Validation: Statistical Methods and Theoretical Findings

Rebecca C. Hendrickson[1,2]*, *Ronald G. Thomas*[3], *Nicholas J. Schork*[4,5] and *Murray A. Raskind*[1,2]

[1] VISN 20 Northwest Network Mental Illness Research, Education and Clinical Center (MIRECC), VA Puget Sound Health Care System, Seattle, WA, United States, [2] Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States, [3] Department of Biostatistics, University of California, San Diego, San Diego, CA, United States, [4] Quantitative Medicine and Systems Biology, The Translational Genomics Research Institute (TGen), Phoenix, AZ, United States, [5] The Joint City of Hope/TGen IMPACT Center (NJS), City of Hope National Medical Center, Duarte, CA, United States

OPEN ACCESS

Edited by:
Raghav Sundar, National University Health System, Singapore

Reviewed by:
Stephen John Senn, Retired, Edinburgh, United Kingdom
Michael Daniele, North Carolina State University, United States

***Correspondence:**
Rebecca C. Hendrickson
Rebecca.Hendrickson@va.gov

Specialty section:
This article was submitted to Personalized Medicine, a section of the journal Frontiers in Digital Health

Received: 31 March 2020
Accepted: 06 July 2020
Published: 28 August 2020

Citation:
Hendrickson RC, Thomas RG, Schork NJ and Raskind MA (2020) Optimizing Aggregated N-Of-1 Trial Designs for Predictive Biomarker Validation: Statistical Methods and Theoretical Findings. Front. Digit. Health 2:13. doi: 10.3389/fdgth.2020.00013

Background and Significance: Parallel-group randomized controlled trials (PG-RCTs) are the gold standard for detecting differences in mean improvement across treatment conditions. However, PG-RCTs provide limited information about individuals, making them poorly optimized for quantifying the relationship of a biomarker measured at baseline with treatment response. In N-of-1 trials, an individual subject moves between treatment conditions to determine their specific response to each treatment. Aggregated N-of-1 trials analyze a cohort of such participants, and can be designed to optimize both statistical power and clinical or logistical constraints, such as allowing all participants to begin with an open-label stabilization phase to facilitate the enrollment of more acutely symptomatic participants. Here, we describe a set of statistical simulation studies comparing the power of four different trial designs to detect a relationship between a predictive biomarker measured at baseline and subjects' specific response to the PTSD pharmacotherapeutic agent prazosin.

Methods: Data was simulated from 4 trial designs: (1) open-label; (2) open-label + blinded discontinuation; (3) traditional crossover; and (4) open label + blinded discontinuation + brief crossover (the N-of-1 design). Designs were matched in length and assessments. The primary outcome, analyzed with a linear mixed effects model, was whether a statistically significant association between biomarker value and response to prazosin was detected with 5% Type I error. Simulations were repeated 1,000 times to determine power and bias, with varied parameters.

Results: Trial designs 2 & 4 had substantially higher power with fewer subjects than open label design. Trial design 4 also had higher power than trial design 2. Trial design 4 had slightly lower power than the traditional crossover design, although power declined much more rapidly as carryover was introduced.

Conclusions: These results suggest that an aggregated N-of-1 trial design beginning with an open label titration phase may provide superior power over open label or open label and blinded discontinuation designs, and similar power to a traditional crossover design, in detecting an association between a predictive biomarker and the clinical response to the PTSD pharmacotherapeutic prazosin. This is achieved while allowing all participants to spend the first 8 weeks of the trial on open-label active treatment.

Keywords: N-of-1 trials, crossover trials, posttraumatic stress disorder (PTSD), prazosin, biomarkers, personalized medicine

INTRODUCTION

Parallel-group randomized controlled trials (PG-RCTs) are the gold standard for detecting differences in mean improvement across treatment conditions (1). However, PG-RCTs provide limited information about the response of individuals to treatment, as they provide no information about the potential response to active treatment for those in the placebo group, and for those who do receive active treatment and experience clinical improvement, it is not possible to distinguish whether this improvement is treatment-specific, or whether the individual would have responded similarly to placebo. This makes PG-RCTs poorly optimized for quantifying the relationship of a biomarker measured at baseline to a treatment-specific response, or identifying subgroups of treatment-specific response (2).

These limitations also affect the utility of trial participation for participants, who receive limited information about whether they have a treatment-specific response (1, 3). Additionally, this trial design requires that many participants spend the full duration of the study on placebo, which may limit the enrollment of patients with particularly acute symptoms. The risk of under enrolling acutely symptomatic patients in a PG-RCT may be particularly high in contexts where the treatment in question or treatments very similar to it are already in active clinical use (4), as is often the case in clinical trials designed to address questions in the realm of personalized medicine.

In N-of-1 trials, an individual subject experiences several treatment conditions, such as active treatment and placebo, in order to assess the individual's specific response to each treatment (1). In aggregated N-of-1 trials, a cohort of individuals moves through this same type of trial design, and their outcomes are analyzed to answer questions about e.g., patterns of treatment response (5). Aggregated N-of-1 trials can be designed to optimize both statistical power and clinical or logistical constraints, such as allowing all participants to begin with an open-label stabilization phase to facilitate the enrollment of more acutely symptomatic participants. They can also mix elements that facilitate standardized assessment of change across all participants with evaluative elements that are individualized to address symptoms that are specific or important to individual participants (1, 6). These features suggest that aggregated N-of-1 clinical trial designs may have significant advantages over PG-RCTs in addressing hypotheses related to personalized medicine (2, 7).

Despite these potential advantages, N-of-1 trials have been slow to gain traction in the biomedical research community. One reason may be that N-of-1 trials have statistical complexities that are different from those encountered in PG-RCTs, and their design and utilization has been limited by the availability of statistical methods to validate and interpret the results (1). Not only do standard methods of power calculations not apply to an aggregated N-of-1 clinical trial, but the breadth of trial designs that are possible using an aggregated N-of-1 approach mean that the questions a researcher would like to ask when computing power calculations may differ from those asked when designing PG-RCTs. For example, the power of an aggregated N-of-1 clinical trial generally increases with increasing repetitions of each treatment condition (5). This effect is limited, however, by the fact that the shorter the period of time an individual is on a given treatment before the effect of that treatment is measured and the treatment condition changed, the larger any carry-over effects from the previous treatment blocks are likely to be (5). The relative cost- vs. benefit of longer but fewer total blocks of treatment, vs. shorter but a larger number of blocks of treatment, then, will depend on the researcher's estimate of carry-over in their particular experimental context (8)–and it is important that methods for power calculations for aggregated N-of-1 trial designs take this type of a factor into account.

Another area in which the assessment of power in an aggregated N-of-1 trial may be more complex is in the area of drop-outs. In traditional power calculation methods, it is often hopefully assumed that dropouts will be unbiased with respect to the effects being measured (9); when the risk of biased drop outs is addressed, this is usually done during analysis by using last-measure-carried-forward, multiple-imputation, or similar strategies (10). In a clinical trial design where participants will at some point move from active treatment to placebo, or from one treatment condition to another treatment condition, it becomes harder to ignore the likelihood that those who have the strongest response to one particular treatment condition may be the most likely to drop out when the move from that treatment condition to one that is less effective for them (11). At the same time, the increased flexibility of the trial design means that it is may be possible to explicitly structure a clinical trial to both minimize dropouts and maximize the ability to obtain the most critical information from each participant prior to periods where the likelihood of dropout increases, if these factors can be quantified and compared across potential clinical trial designs.

Finally, while the option to include both open-label and blinded treatment blocks into an aggregated N-of-1 trial design has the potential to significantly increase the representation of acutely symptomatic patients in a clinical trial, it also makes assessing the impact of a participant's expectation of benefit on their outcome more complex than in a purely double-blinded RCT (5). For example, if a participant begins on open-label treatment, it is expected that their change in outcome measurements during this period of time would constitute the combined effect of both treatment-specific effects and non-treatment-specific effects, which includes the impact of just knowing that they are receiving active treatment. The question arises, then, as to what is expected to happen when they transition from this period of open-label treatment to a treatment block when they are on blinded but active treatment. How does the impact of knowing they *may* be on treatment compare to the impact of knowing they *are* on treatment?

The increased relevance of factors such as biased dropouts and expectancy related effects to statistical power means that wider adoption of aggregated N-of-1 clinical trial designs will require the development of statistical methods that allow clinical trialists to compare the statistical power of different potential trial designs in answering their particular research questions, and given their best estimates of the extent to which effects such as carry-over or biased dropout rates will impact their study population. As many of these factors do have non-trivial relevance even to PG-RCTs, however (12, 13), it is also possible that the development and such methods may eventually improve our understanding and interpretation of more traditional clinical trial designs, as well.

Although the simplest form of an N-of-1 trial, the crossover trial, is one of the earliest forms of clinical trial and has been studied extensively (14–17), most work addressing the statistical properties of more complex N-of-1 clinical trial designs has been done in the past decade (5, 8). In 2014, Chen and Chen compared both simple (paired *t*-test) and more complex (mixed effects models) approaches for conducting tests of treatment efficacy using aggregated N-of-1 trial results, and found that in their examples, mixed effects models were inferior in the absence of carryover effects but superior when these were included (18). This work was critiqued by Araujo et al. who point out that the models evaluated by Chen and Chen do not include a treatment by patient interaction (19), an interaction that has been advocated for in the meta-analysis literature; the relevance of Chen and Chen's approach may also be limited by the assumption of compound symmetry and auto-regressive covariance structure. More recently, Percha et al. implemented a stochastic time-series model to simulate individual N-of-1 studies, and characterized the impact of the number of treatment blocks, the ordering of treatments within blocks, the duration of each treatment, and the sampling frequency on both the statistical power to detect a difference in efficacy and in the accuracy of the estimated effect size (20). However, little work thus far has explicitly attempted to model the impact of expectancy and biased dropout on statistical power in aggregated N-of-1 clinical trial design, or to incorporate the possibility of non-traditional combinations of treatment conditions, such as trials that include both open label and blinded conditions, or blinded discontinuation blocks.

Finally, although it is expected to be an important application of this type of trial design (2, 7), there is extremely little that has been published addressing methods for the validation of predictive biomarkers in aggregated N-of-1 trials. A publication by Grenet et al. earlier this year provides a statistical framework for comparing the power of crossover vs. parallel-group clinical trials to detect a relationship between a binary predictive biomarker and treatment effect (21). However, we are not aware of any published methods for analyzing more complex aggregated N-of-1 clinical trials to test for the relationship between a putative predictive biomarker and treatment response, nor for calculating a trial's power to test this type of a hypothesis.

Here, we provide an initial set of tools designed to address a number of the above statistical challenges in the design and analysis of aggregated N-of-1 trials. Specifically, we describe a set of statistical simulation studies that were used to compare the expected statistical power of different potential clinical trial designs, the aim of which was to detect a relationship between a biomarker measured at entry into the study and the efficacy of a specific treatment. Importantly, then, the power that is being calculated in this set of examples does not address whether the treatment is effective as compared with placebo, but rather whether the biomarker measured at baseline is able to predict which individuals will respond to the treatment and which will not.

This sample application is based on work conducted by the authors to plan a randomized clinical trial to test the relationship between standing systolic blood pressure measured at baseline to the decrease in PTSD symptoms produced by the α_1 adrenoceptor antagonist prazosin. A relationship between this simple, clinically-accessible biomarker and treatment response that is large enough to be potentially relevant to treatment selection has been found in a *post hoc* analysis of a PG-RCT of prazosin for PTSD conducted in a primarily young, male population (22), but the relationship has not yet been validated in a prospective trial, or in a trial with a less homogeneous population. Further, the potential to conduct further PG-RCTs of prazosin for PTSD is believed to be limited by already wide utilization of prazosin for PTSD, such that the acutely symptomatic patients thought to be most responsive to PTSD are unlikely to be referred to trials where they may be placed on placebo, rather than simply treated (4).

The use of this real example of computing power calculations for what is now an ongoing aggregated N-of-1 clinical trial allows us to demonstrate how estimates of population means and variances were extracted from extant data sets when possible, while variables that could not be estimated based on existing data were allowed to vary so that the dependence of the power calculations on these estimates could be assessed. However, it is hoped that the methods described will be of general utility. To this end, the functions used to generate and run these simulations are also provided in a publicly available github repository.

METHODS

Approach

Conceptually, the work can be broken down into three broad steps, which are detailed in **Figure 1**: the statistical simulation of a single clinical trial, including a simulated data set and the estimated effect size and p-value that result from the analysis of that trial (**Figure 1A**); the repetition of the individual clinical trial simulation 1,000 times, producing an estimate of statistical power and the distribution of bias in the effect size estimate (**Figure 1B**); and a repetition of this entire process while systematically varying the parameter space and the clinical trial design, in order to quantify the relative power and bias distributions for the different clinical trial designs, and the sensitivity of these results to variable parameters such as carryover effects of dropout patterns (**Figure 1C**). All work was done using R (23) and RStudio (24). The R functions and vignettes documenting the steps used to generate these results are available as a package at https://github.com/rchendrickson/pmsimstats.

Selection of Clinical Trial Designs for Comparison

Potential clinical trial designs were selected to allow the comparison of statistical power and bias across the four most plausible trial designs for testing the relationship of a baseline biomarker to treatment response: (1) a single-group open label trial, (2) a single-group open label trial followed by a blinded discontinuation block, (3) a traditional crossover trial, and (4) the proposed N-of-1 trial design, consisting of a single-group open label trial followed by first a blinded discontinuation block and then two crossover blocks (**Figure 2**). In each design, the titration period for prazosin is expected to be 2.5 weeks each time it is initiated. In blinded discontinuation blocks, the transition from blinded but active treatment to placebo can occur at only two points, either after 1 week or after 2 weeks in the block. However, this aspect of the design is not revealed to participants, who are told only that during this block they may be on either active drug or placebo, and that this may change during the course of the block.

The inclusion of an open-label period at the beginning of designs 2 and 4 was selected to address concerns that highly symptomatic individuals would be less likely to be referred to or enroll in a clinical trial where they may be initially assigned to a placebo group. The inclusion of a blinded discontinuation period in two of the trial designs was designed to allow a higher intensity of data capture, including of personalized assessment measures, during the period of discontinuation after the open-label portion. A traditional PG-RCT was not included among the tested designs because, in the specific example being explored, existing data had already demonstrated that there was a negligible chance that the biomarker predicted response to placebo (22), which meant that minimal information would be obtained from the ∼50% of participants randomized to placebo.

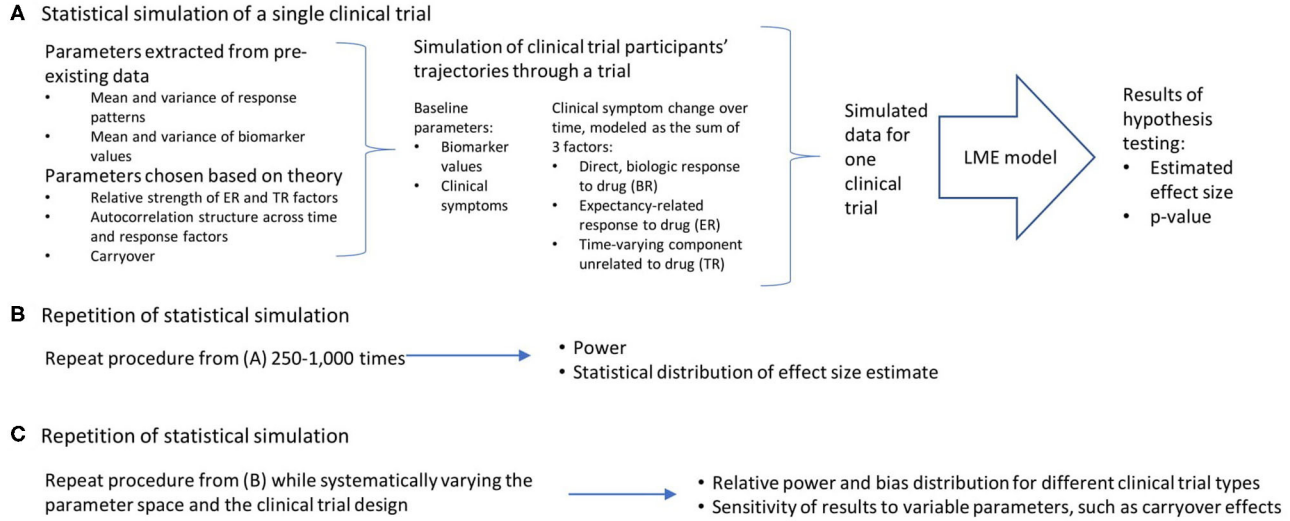


FIGURE 1 | Schematic overview of approach to simulating and analyzing clinical trials data. **(A)** A multi-step process was used to simulate and analyze the results of a single simulated clinical trial. Parameters used to generate the simulated data were derived in part from existing data sets, but also involved the selection of some parameters that could not be estimated directly from existing data. The generation of data was done using a model that presumed there were three basic factors that linearly combine to describe the trajectory of participants' symptoms over time (the direct, biologic response to drug (BR), the expectancy-related response to drug (ER), and the time-varying component unrelated to drug (TR). These results were then analyzed using a linear mixed effects model, as is proposed for the analysis of the actual clinical trial results. This analysis is structured to test the hypothesis that the biomarker measured at baseline will be significantly associated with the degree of clinical response a given participant has to the intervention. The analysis produces a *p*-value describing the statistical significance of the results if they were being analyzed as a single extant clinical trial, and an estimate of effect size. **(B)** This simulation process is repeated 1,000 times using the same clinical trial design and parameter selection, allowing an estimate of the power of this trial design to detect the proposed relationship, and the distribution of bias in the effect size estimated. **(C)** This entire process can then be repeated with (a) different clinical trial designs, and (b) different parameter selection, in order to determine how statistical power and bias in effect size estimation vary as a function of trial design, response parameters, and model assumptions.

A

	Baseline	Open-label prazosin							
		20 weeks							
Treatment received		A							
Participant blinded		N							
Study staff blinded		N							
Assessment points	X	X	X	X	X	X	X	X	X

B

	Baseline	Open-label prazosin				Blinded discotinuation			
		16 weeks				4 weeks			
Treatment received		A				A	A/P	P	P
Participant blinded		N				Y	Y	Y	Y
Study staff blinded		N				N	Y	N	N
Assessment points	X	X	X	X	X	X	X	X	X

C

	Baseline	Crossover period 1				Crossover period 2			
		10 weeks				10 weeks			
Treatment received		A/P				A/P			
Participant blinded		Y				Y			
Study staff blinded		Y				Y			
Assessment points	X	X	X	X	X	X	X	X	X

D

	Baseline	Open-label prazosin	Blinded discotinuation				Crossover period 1	Crossover period 2
		8 weeks	4 weeks				4 weeks	4 weeks
Treatment received		A	A	A/P	P	P	A/P	A/P
Participant blinded		N	Y	Y	Y	Y	Y	Y
Study staff blinded		N	N	Y	N	N	Y	Y
Assessment points	X	X	X	X	X	X	X	X

FIGURE 2 | Four potential clinical trial designs that were compared on their ability to detect a relationship between a biomarker measured at baseline and response to treatment with the pharmacotherapeutic agent prazosin. Trial designs were matched in duration and the number of evaluation points. **(A)** Open-label trial design: All participants receive open-label prazosin throughout the trial. **(B)** Open-label followed by blinded discontinuation design: All participants receive active drug for 16 weeks, then enter a 4 week blinded discontinuation block. During the blinded discontinuation block, all participants receive active drug during the first week and placebo during the last 2 weeks, such that only the participant is blinded to treatment condition during these weeks; the treatment condition during the second week is randomly assigned and a double blind is maintained during this week. **(C)** Traditional crossover trial design: Participants are randomized to 10 weeks on active drug followed by 10 weeks on placebo, or the reverse. **(D)** Proposed N-of-1 trial design: all participants begin with an 8 week open label period, then enter a 4 week blinded discontinuation period, then complete 8 weeks of crossover. There are two independent randomization points—whether the participant is on active drug (A) or placebo (P) during the second week of the blinded discontinuation block, and whether the participant's crossover blocks are active drug then placebo or the reverse. A = Active drug (prazosin); P = placebo. X indications the timing of assessment points for clinical outcome measure.

By the most general definition of an aggregated N-of-1 clinical trial, all but the open label and PG-RCT designs can be considered to be a form of N-of-1 trial, because each participant spends time on both treatment conditions (active drug and control). However, it is primarily the fourth trial design that takes advantage of the opportunity for multiple periods of treatment in each treatment condition.

Statistical Simulation of Data

The expected trajectory of clinical symptoms over time was modeled as the linear sum of 3 factors (**Figure 1A**), each of which describes one aspect of how symptoms change over time from their baseline values: (1) a direct, biologic response to a pharmacologic agent (the biologic response, or BR); (2) an expectancy-related response to taking a medication that is either known to be or know to possibly be an active treatment (the

expectancy-related response, or ER); and (3) a component that is a function of time since study entry, but which is not related to the actual or expected presence of active treatment (the time-dependent response, or TR). The time-dependent response is presumed to include both regression-to-the-mean effects and the impact of the structure, attention and regular interaction with staff involved in study participation.

A function describing the expected mean and variance of each factor as a function of time and study design was fit using a three-parameter gompertz function, allowing a non-linear monotonic trajectory over time with a maximum asymptote. The three parameters characterize: the maximum response, the displacement, and the rate. Initial estimates for these variables were based on fits to existing data from a parallel group randomized controlled trial of prazosin for PTSD in active duty service members (25), utilizing the following assumptions: the trajectory of the BR factor was taken to be the difference between the trajectory of the prazosin group and the placebo group; the trajectory of the placebo group was taken to represent the sum of the TR and ER factors; in the absence of any data to separate the trajectory of the placebo group into the TR and ER components, the maximum response, rate and variance of the TR and ER factors were assumed to be equal [tabula rasa (TaRa) parameter set]. In further sensitivity analyses, however, these values were varied to assess the impact of these parameters on simulated clinical trial performance. The means and variances of the baseline symptom intensity [as measured by the clinician administered PTSD scale for DSM-IV, or CAPS-IV (26)] and baseline biomarker values (systolic blood pressure 2 min after standing) were based directly on the baseline measurements from the existing data.

The ER factor was presumed to be scaled directly by participant expectancy regarding whether they were taking an active medication or not. For open label trial components, the expectancy was set to 1, while for blinded portions where the participant had been informed there was an equal chance they were taking active drug vs. placebo the expectancy was set to 0.5. The BR factor was set to zero at times when participants had never been on active drug; however, a carryover effect was built in such that when a participant moved from active drug to being off active drug, the value of the BR at the last timepoint on active drug was exponentially decayed, with the half-life of this decay being maintained as a model parameter.

Using the above factor parameterizations to provide the expected mean and variance for each factor at each time point, simulated data with the specified covariance structure, coerced to be positive definite, was generated using the function *mvrnorm* from the R package MASS (27). This simulated data consisted of baseline symptom intensity, baseline biomarker value, and the value of each of the three factors at each timepoint within the trial for a variable number (N) of participants (**Figure 3**). The *mvrnorm* function takes as input a vector specifying the means of each variable, as well as a covariance matrix. The covariance matrix was assembled based on a set of modifiable parameters defining the correlation between the baseline biomarker and the BR components, the autocorrelation over time (relating the value of a factor for one participant at one timepoint to the value of that factor for that participant at subsequent timepoints), the correlation at a single time point between the 3 factors, and the variance of each component. Once the factor values at each time point were generated, the sums of the three factors BR, ER and TR were subtracted from the baseline values for each simulated participant to produce a full set of results for the simulated clinical trial, consisting of the baseline biomarker measurement and symptom intensity measurement at each timepoint.

In some simulations, a censoring filter was applied following the production of the stimulated trajectories, in order to assess the effects of participant dropout. The probability of a simulated participant dropping out per unit time was calculated as the sum of a flat hazard function (β_0,) and a probability scaled by the square of the change in symptoms since baseline (shifted by 100 so that all values are positive; β_1). Thus, depending on the parameters β_0 and β_1, this function produces a probability of dropping out that is higher for participants experiencing worsening or high continued levels of symptoms and lower for participants who are experiencing benefit.

Analysis of Simulated Data

Each simulated trial data set was analyzed using a MMRM (mixed effect model with repeated measures) to assess the significance of the biomarker-vs.-drug exposure interaction. Consistent with the recommendations of Barr et al. (28), models were initially run with maximal random effects structure justified by the design, which was then limited based on empirical success with model fits. An unstructured variance/covariance matrix was assumed.

For trial designs that include timepoints both on and off active drug excluding baseline, fixed effects in the model were time, drug-exposure, baseline biomarker and an interaction term between drug-exposure and baseline biomarker. Individual subject was included with a random intercept. The inclusion of expectancy as a fixed effect was found to increase the frequency of collinearity leading to poor model fits while changing the results minimally, and thus was not included in any of the results presented. Thus, the model implemented for these designs was: $S_{i,t} = \beta_{i,0} + \beta_1 \cdot bm + \beta_2 \cdot Db + \beta_3 \cdot t + \beta_4 \cdot bm \cdot Db$. A non-zero coefficient for the interaction term, β_4, serves as indication of a significant effect of baseline biomarker on drug response.

For trial designs where, excluding baseline, each participant only experiences a single treatment condition, the above model was poorly fit, and produced a significantly inflated type I error rate (data not shown). Instead, consistent with the *post hoc* analysis of a parallel group RCT's results that served as the preliminary data for this work (22), trial designs of this type (primarily OL) were analyzed with a model that included time, baseline biomarker and an interaction term between time and baseline biomarker: $S_{i,t} = \beta_{i,0} + \beta_1 \cdot bm + \beta_2 \cdot t + \beta_3 \cdot bm \cdot t$, with a non-zero interaction term (this time represented by β_3) again signifying a significant effect of biomarker on treatment response.

In each case, the model was fit using the *lmer* function from the R package *lme4* (29). For each simulated trial analysis, the *p*-value for the biomarker-vs.-drug interaction was evaluated for significance at the alpha = 0.05 level by examining the *p*-value corresponding to the interaction terms described above as calculated by *lmer*. The rate of significant interactions provided

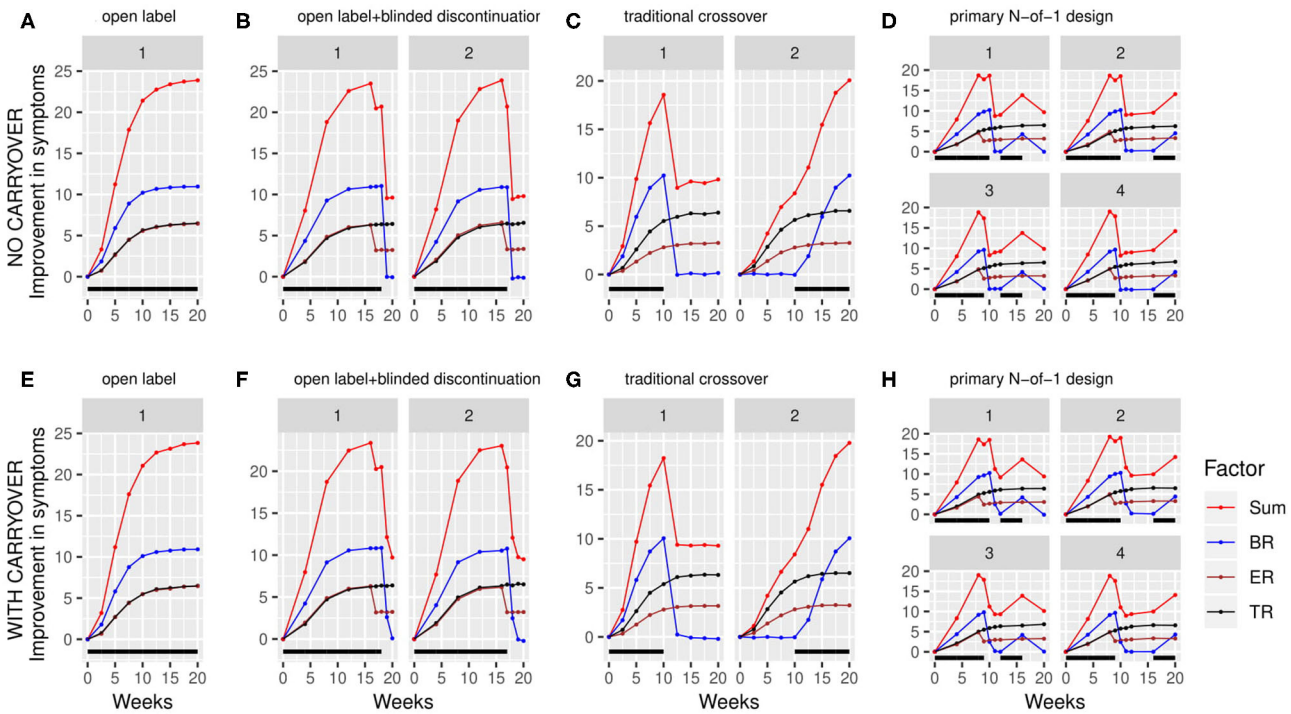


FIGURE 3 | Simulated clinical trials data plotted as change in baseline symptom score as a function of time, broken down by clinical trial design **(A–H)** and randomization path (numbered facets within each trial design). **(A)** through **(D)** show results with carryover set to 0; **(E)** through **(H)** show results with carryover set to 0.1 weeks. Individual simulated factors (BR = biologic response to drug, ER = expectancy-related response to drug, and TR = time-varying component unrelated to drug) are shown along with their summed effect. Plotted data represents the averaged output of 500 replicates for each clinical trial design, divided across the number of randomization paths, and was generated using the tabula rasa response parameters. OL = open label, OL+BDC = open label followed by blinded discontinuation, CO = cross over, N-of-1 = proposed N-of-1 trial design. Black bar represents times active drug was scheduled to be present.

an estimate of the power of each design to detect the simulated interaction signal for each combination of parameters. The distribution bias in the estimate of the association between the biomarker and the response to active drug was quantified for each trial design and censoring pattern as the differences between the β from each replicate and the β when the analysis was run across all replicates with that parameter set but with no censoring.

RESULTS

The statistical power to detect a relationship between the baseline biomarker and the response to prazosin treatment was significantly different among the four clinical trial designs. When simulations were run using tabula rasa parameter set, assuming equal magnitude and variance for the TR and ER factors, and without a carryover effect, the proposed N-of-1 trial design demonstrated superior power to detect a true relationship between the baseline biomarker and response to drug compared to the open label and open label + blinded discontinuation designs (**Figure 4A**). The N-of-1 design had lower power than the traditional crossover design. Increased censoring lowered power across all trial designs, but, consistent with the increased vulnerability of the open-label plus blinded discontinuation design to participant loss prior to the blinded component, this design's power dropped more rapidly.

Impact of Carryover on Statistical Power

When a non-zero carryover term was added to describe the persistence of improvement related to the biologic effect of the drug even after the active drug is discontinued, described as an exponential decay with $t_{1/2}$ measured in weeks, the presence of even a short (0.1 weeks) carryover component resulted in a precipitous decline in power in both the N-of-1 and, to a lesser but still very significant degree, the open label + blinded discontinuation design (**Figure 4B**). A decrease in power in the cross over design was also seen, but this was significantly less severe.

Impact of Response Trajectory Parameters on Statistical Power

The impact of changes in the parameters used to define the trajectories of the three response factors (BR, TR, and ER) were explored by systematically varying either the maximum values and standard deviations (set equal to the maxima) of each factor while retaining the tabula rasa values for the rates (**Figure 5A**) or by systematically varying the rates while retaining the tabula rasa values for the maximums

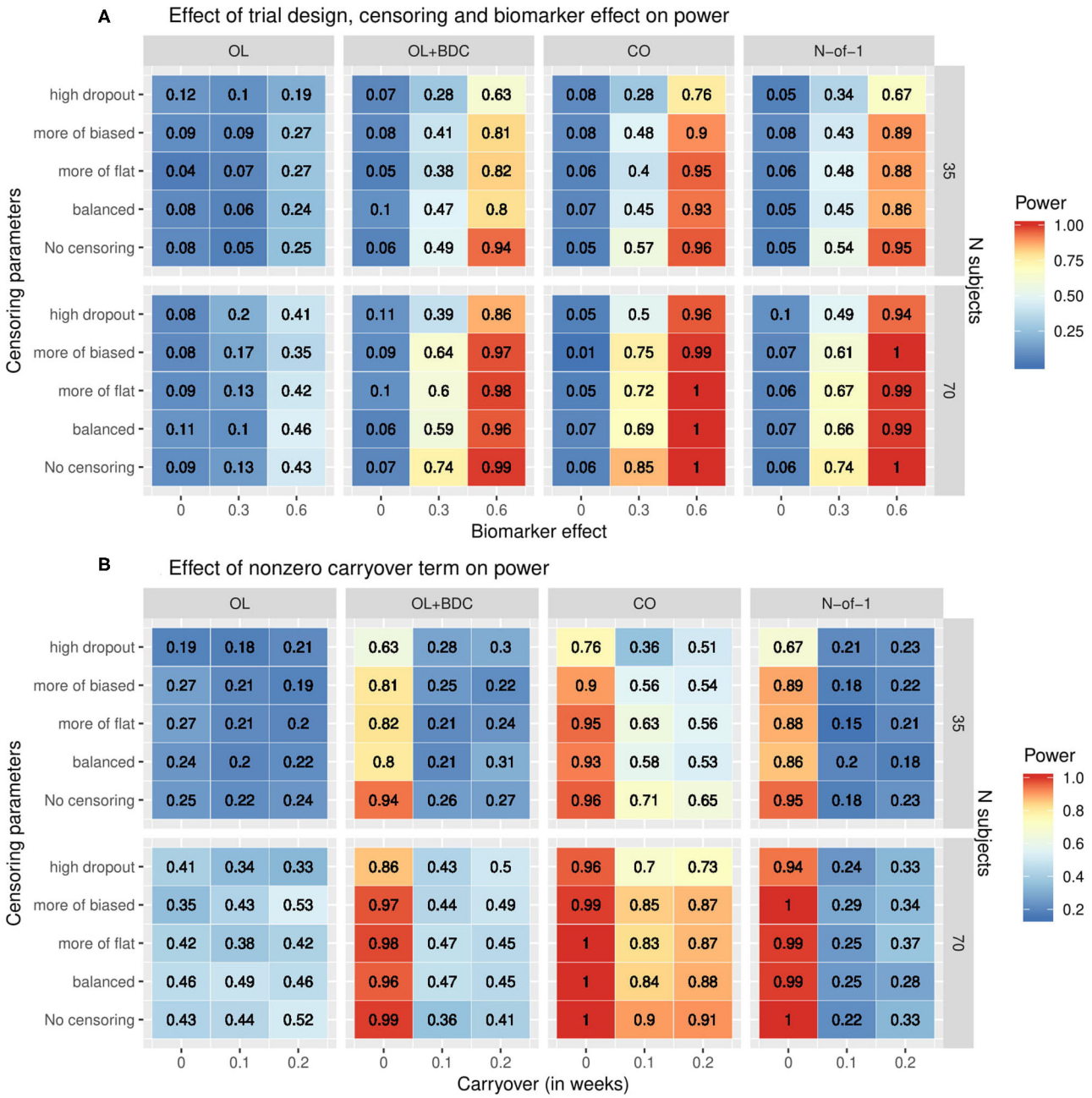


FIGURE 4 | Heat map showing statistical power as a function of **(A)** clinical trial design, the number of subjects, the correlation coefficient relating the biomarker to the biologic response to drug, and the censoring parameters describing dropout patterns, or **(B)** clinical trial design, the number of subjects, the timecourse of the carryover effect of the intervention ($t_{1/2}$ in weeks), and the censoring parameters describing dropout patterns, for each of the clinical trial designs described in **figure 1**. In **(A)** the carryover effect is set to zero; in **(B)** the correlation between baseline biomarker and the biologic response to drug is set to 0.6. OL = open label, OL+BDC = open label followed by blinded discontinuation, CO = cross over, N-of-1 = proposed N-of-1 trial design.

(**Figure 5B**). Consistent with expectation, increased maximal response of the BR factor improved power across all trial designs. Increasing the maximal response of the ER factor decreased power across all trial designs, but with a greater decrease in power in the two trial designs (OL+BDC and N-of-1) where the expectancy values changes across the trial. Increasing the maximal response of the TR factor decreased power across all trial designs. Increasing the maximal response of the ER factor decreased power more substantially for the trial design where expectancy changes.

The impact of changes in rate parameters were less consistent across trial designs. In the two trial designs with blinded

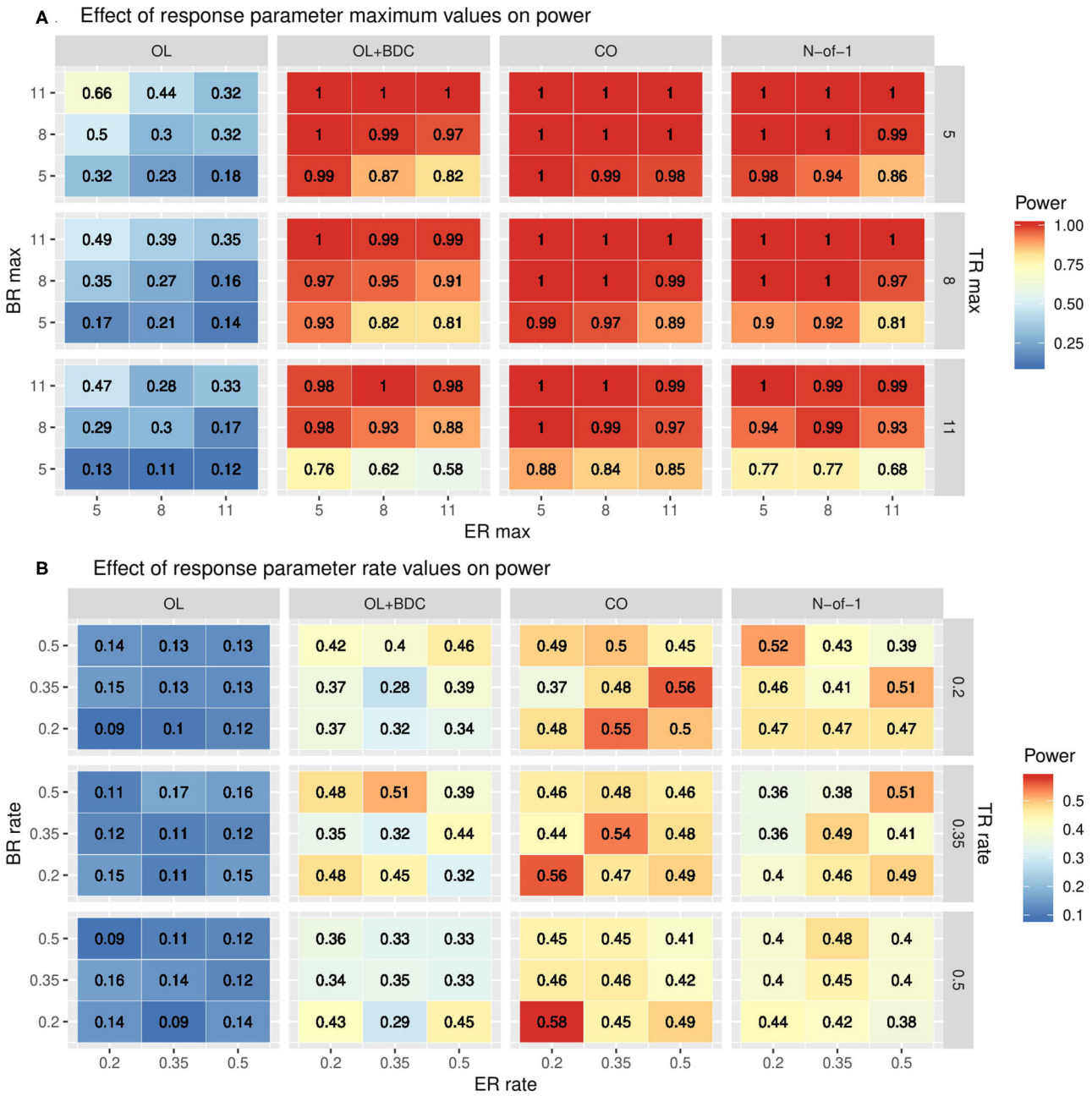


FIGURE 5 | Heat maps showing statistical power as a function of trial design and the parameters used to define **(A)** the maximum response or **(B)** the rate of change in the modified gompertz function defining the trajectories of each of the three response factors (BR = biologic response to drug, ER = expectancy-related response to drug, and TR = time-varying component unrelated to drug). In panel both panels, N is set to 35 and carryover is set to zero. In **(A)** the correlation between biomarker and the BR factor is set to 0.6, while in **(B)** it is set to 0.3. OL = open label, OL+BDC = open label followed by blinded discontinuation, CO = cross over, N-of-1 = proposed N-of-1 trial design.

discontinuation portions, an increased rate for the BR factor did generally correspond with increased power across most of the parameter space; however, for the crossover design, increased BR rate was associated with decreased power across most of the parameter space.

Variability and Bias in Effect Size Estimates

The variability and bias in the effect size estimates as a function of trial design and parameters was also explored. The mean across replicates of the estimated standard error in the coefficient for the interaction term used to carry out the hypothesis

testing increased across all trial designs with increasing censoring, but with a greater effect for the OL+BDC trial design (**Figure 6**).

Bias was assessed in two ways. First, the estimate of the coefficient for the interaction term used to carry out hypothesis testing (β) was extracted for all replicates for a given trial design and parameter set but with the correlation between the biomarker and BR set to zero, and both the mean β and the *p*-value applying a one-sample two-sided *t*-test with $\mu = 0$ to the distribution of β were examined (**Figure 7A**). The β-values were for most censoring patterns for the OL+BDC trial design and several censoring parameters of the CO design significantly biased toward a negative non-zero effect ($p < 0.0001$), while for the N-of-1 design, the non-censored condition showed a significant bias toward a positive non-zero effect (i.e., in the direction opposite the expected effect of a biomarker that predicts a decrease in symptoms; $p < 0.0001$).

Second, looking this time at simulations where the correlation between the biomarker and the BR response factor was set to 0.3 or 0.6, the β from each replicate was compared to the β obtained when the model was applied to all simulated participants across all replicates in the absence of censoring (**Figure 7B**). This analysis method allows the impact of different censoring patterns on effect size estimates to be assessed. For the open label design, for a larger true effect size, censoring was seen to result in a larger estimated effect size across all types of censoring parameters utilized ($p < 0.0001$), and in the high dropout case even with the lower true effect size. Censoring parameters did not have a significant effect on the other three trial designs.

DISCUSSION

These results suggest that an aggregated N-of-1 trial design beginning with an open label titration phase may provide superior power compared to an open label or open label followed

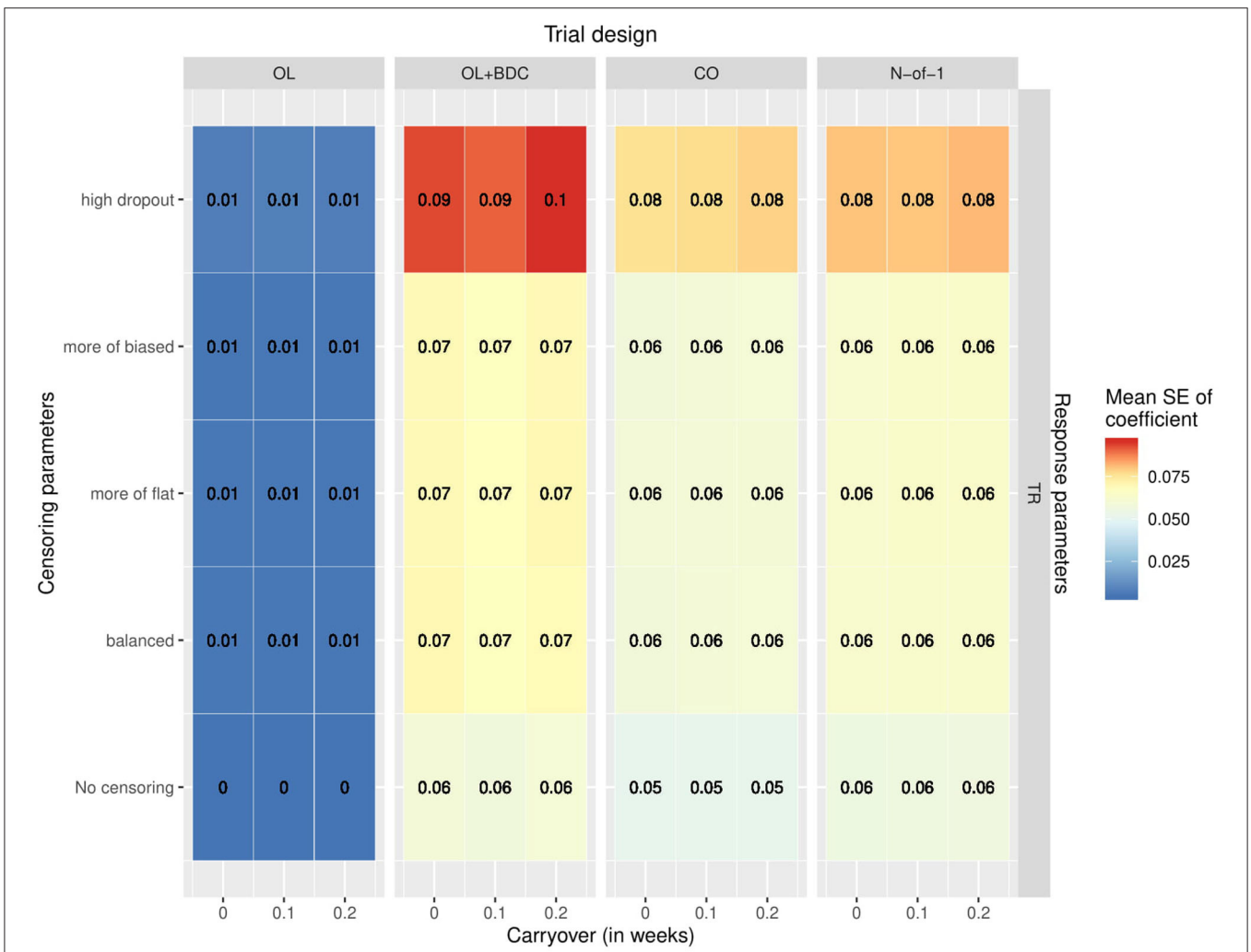


FIGURE 6 | Mean standard error in the coefficient for the interaction term used for hypothesis testing across simulated replicates, as a function of trial design, response parameters, carryover, and censoring parameters. OL = open label, OL+BDC = open label followed by blinded discontinuation, CO = cross over, N-of-1 = proposed N-of-1 trial design.

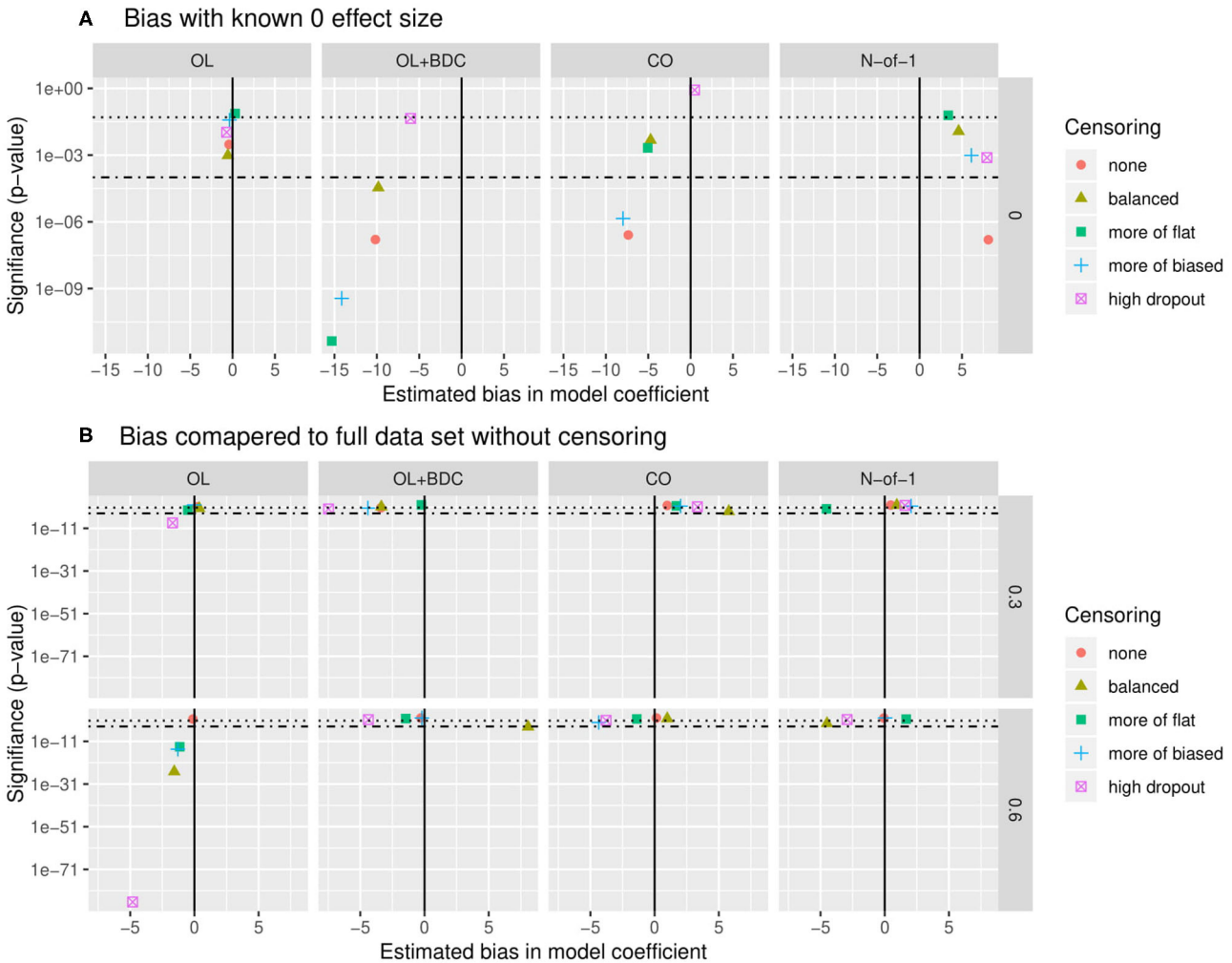


FIGURE 7 | Quantification of bias in effect size estimate as a function of trial design and censoring parameters. **(A)** Bias in model coefficient for interaction term being used for hypothesis testing (β) quantified as the mean coefficient across simulated replicates when the true effect size was set to zero (estimated bias in model coefficient). *P*-value (y axis) indicates results of a two-tailed, one-sample *t*-test comparing the coefficients across the set of replicates to $\mu = 0$. **(B)** Bias in model coefficient for interaction term being used for hypothesis testing quantified as the mean difference (Δβ) between the coefficient for a single replicate (β) and the "gold standard" coefficient for that parameter set (β_t), with the "gold standard" defined as the coefficient calculated across all simulated participants from all replicates with no censoring. *P*-value (y-axis) indicates results of a two-tailed, one-sample *t*-test comparing the Δβ across the set of replicates to $\mu = 0$. Dotted line indicates $p = 0.05$, dot-dash line indicates $p = 0.0001$. β and Δβ values multiplied by 1,000 for ease of visualization. OL = open label, OL+BDC = open label followed by blinded discontinuation, CO = cross over, N-of-1 = proposed N-of-1 trial design.

by blinded discontinuation trial design, and similar but slightly decreased power compared to a traditional crossover trial design, in detecting an association between a predictive biomarker and the clinical response to the PTSD pharmacotherapeutic prazosin. In contrast to the traditional crossover design, this increased power is achieved in a clinical trial design that allows all participants to start on open-label active treatment, a significant advantage in allowing the recruitment of a symptomatic study population.

The increased statistical power seen in the N-of-1 trial design as compared with the purely open-label trial design is consistent with the information-theoretic expectation that any clinical trial design that provides minimal or no information about who in a purely active treatment group is showing a response that is specific to the intervention provided, vs. who is showing a response to treatment that is not dependent on the specific biologic treatment provided, will have an associated decrease in statistical power when used to assess the relationship of a baseline biomarker to treatment response. A significant decrease in statistical power was across all tested trial designs except the open label design when a carryover term was introduced; this effect was particularly large for the N-of-1 and open-label followed by blinded discontinuation trial designs. Although this decrease in power is consistent with expectation, the magnitude of this drop with even short half-life carryover effects underscored the critical nature of this parameter in determining the appropriateness of

an N-of-1 trial design; it also suggests that the development of analysis methods that incorporate and expectation of carryover may significantly improve the power and utility of N-of-1 clinical trial designs in personalized medicine applications.

These results support both the use of aggregated N-of-1 clinical trial designs to optimize both statistical power to detect a relationship between predictive biomarkers and treatment response and clinical-logistical constraints, such as a need to allow patients to begin with active treatment. They also support the use of statistical simulation to quantitatively compare alternative clinical trial designs in such a context.

In addition to providing guidance for the design and selection of aggregated N-of-1 clinical trial designs, these types of results can help to quantify the extent to which the outcomes measured in clinical trials depend on factors such as drug carryover effects, the impact of expectancy on outcome measures, and biased dropout patterns—each of which has the potential to be highly relevant to more traditional clinical trial designs, as well.

Implications of Carryover Effect

The impact of carryover on the design and analysis of clinical trials where participants cross from one treatment condition to another has been considered for over 50 years, and extensively researched (16, 17). In these models, we assume an exponential decay analogous to a pharmacokinetic half-life, although the simulations could be easily adapted to incorporate an alternative model. We do not assume that the half-life of the carryover effect should equal the half-life of the pharmacotherapeutic agent, however. Instead, the carry-over effect is presumed to reflect a combination of factors that includes the pharmacokinetic half-life, the time lag that may be involved in participants becoming aware of changes in their symptoms or in reporting changes on assessment tools that may have a longer lookback period, and the impact of physiologic or behavioral changes that may have resulted from changes in primary symptoms but may also serve to sustain positive changes even after the intervention has ceased. While empirical data describing the magnitude and relevance of these factors is limited for most treatments of interest, this gap in our knowledge base regarding even our relatively well-studied interventions will decrease as N-of-1 trial designs become more common. Increased characterization of the effects of discontinuing treatments has the potential to provide important clinically relevant information well-beyond its utility in the design of N-of-1 clinical trials.

Potential for Biased Dropout

The impact of both treatment response and side effect burden on how likely different participants are to withdraw early from a trial, and at what points, is of particular importance in estimating the relative power of different N-of-1 type clinical trial designs. As is illustrated by these results, the impact on a trial of dropout rates that are biased by a participant's response to treatment has the potential to be both positive and negative. For example, the expense of running a clinical trial in which participants are enrolled for many months is significant—and if participants for whom either (a) no significant response to treatment at all, or (b) a clear response to treatment that is lost when the participant transitions back to placebo are the most likely to withdraw prior to later crossover blocks, this actually allows the additional expense of offering these extended blocks to be preferentially spent on participants for whom participation in the initial phases was inconclusive, and for whom additional blocks are the most important scientifically. This "happy accident" is of course not truly coincidental—rather, it can be seen as the result of aligning the participants' goals for trial participation (determining whether this treatment works for them, and if so, whether they need to continue to take it to maintain the effect) with the scientific goals of the trial (determining who has a specific response to the treatment that is not present with a placebo intervention).

At the same time, the potential that participants who are particularly likely to have strong placebo responses may also be particularly anxious about and likely to avoid entering discontinuation blocks is one that would decrease the power and potentially increase bias in aggregated N-of-1 trials, particularly ones that begin with an open-label titration and stabilization phase. Although the current statistical simulations do not incorporate an estimate of this type of an effect, it would be a straightforward extension do so. Additionally, as trials such as the one described here begin to be run, additional information will become available about the extent to which non-treatment-specific changes in symptoms may be associated with transitions in what the participant knows about what condition they are in (such as from open-label to blinded active treatment). This type of additional information will have the potential both to further inform N-of-1 trial design, and also help elucidate the different mechanisms and implications of non-specific treatment response. Similarly, these types of statistical models can easily incorporate the possibility of a confounding relationship between biomarkers that are putative predictors of treatment response, side effects, and actual treatment response, thus allowing researchers to assess the potential magnitude of bias in their estimates of biomarker-based predictions of treatment response as a function of drop outs biased by patterns of side effect emergences.

Implications Regarding Placebo Response and Expectancy

One of the most complicated factors to emerge when seeking to statistically model the response patterns of participants moving between treatment conditions, and particularly between open label and blinded phases of treatment, is the expected patterns of non-treatment-specific aspects of clinical responses. By non-treatment-specific responses, we mean changes in symptoms over the course of trial enrollment that are not a result of the direct biologic action of the treatment itself—i.e., are not specific to the presence of a particular active treatment. In PG-RCTs, such effects are often grouped together under the term "placebo response," which is used to describe all factors that together lead to changes in symptoms in the group receiving a placebo (30). This terminology is inconsistent, however, with the definition of placebo response that is used in research on the pathophysiology of the placebo effect (31), where the term is most commonly

reserved for the changes in symptoms and/or physiology that are the result of a patient's expectation that they are or may be receiving an active treatment.

Here, we have considered at least four primary factors likely to contribute to the overall course of symptom change in participants: (1) the direct biologic action of the drug; (2) the average trajectory of symptoms seen as a function of time following the point at which a participant is recruited to participate in a trial (generally a regression to the mean effect, for a trial seeking to recruit acutely symptomatic, treatment-seeking participants); (3) the change in symptoms related to general factors involved in clinical trial participation, including regular contact with warm, supportive staff and ongoing monitoring of symptoms and behavioral patterns such as substance use; and (4) the change in symptoms related to the participant's belief that they are taking a medication that is likely to help them. In a typical PG-RCT, factors 2–4 are generally grouped together as the "placebo response," and presumed to be present in both groups, with the additional impact of the direct biologic action of the medication presumed to be additive, such that it can be obtained by look at the difference between the response in the placebo group and the active treatment group (30). In the N-of-1 trial design discussed here, however, the expected timecourse of factor 4 can no longer be presumed to be static over the course of the trial, and must be modeled separately. Although this introduces additional complexity into the interpretation of the clinical trials data, it also introduces interesting additional potential analyses.

One potential benefit may be the ability to better understand the relationship of traditional PG-RCT results to the treatment effects observed in routine clinical care or in open-label trials (13). Most concretely, it has been observed that open-label contexts may result in more positive outcomes than blinded treatment conditions (32). In addition to factors such as patient selection or contact frequency, one contributing factor could be that the placebo effect is lessoned in the case of blinded treatment condition vs. open-label treatment. Additional experience with how patients' clinical outcomes differ across blinded and open-label treatment conditions may thus improve our ability to understand the relationship between the results of PG-RCTs and our clinical care contexts.

Importantly, although it is often assumed that such an effect would be linear and separable from other aspects of treatment effects, it is increasingly accepted this assumption is frequently in error, particularly for central nervous system (CNS) clinical trials (30). For example, the observation effect size and the frequency of positive clinical trial outcomes have trended downward over time as the magnitude of placebo effect in these trials has increased is frequently interpreted in the field as being due to a large placebo effect "masking" or interfering with the possibility of measuring a statistically significant treatment-specific effect (31, 33). In other words, it is being attributed to a presumed non-linearity in how treatment-specific and non-specific treatment responses combine, specifically a subadditivity—which is, in fact, consistent with emerging work on the additivity of treatment-specific and non-specific effects in clinical trials (31, 34).

There is also evidence to support the presence of interaction effects beyond subadditivity, as well. For example, in studies of two different analgesic medications operating via two different mechanisms, the treatment-specific effect was found to be either dependent on (35) or bi-directionally modulated by (36) the presence of an expected result of the intervention. In fact, such interactive effects between biologic response, non-treatment-specific effects, and even augmenting treatments are often explicitly hoped for and pursued in the context of routine clinical care (37), where a psychiatrist may e.g., remind a patient with PTSD whose treatment goals include increased behavioral activation and acclimating to attendance at anxiety-producing events that one of the expected mechanisms of action of a treatment is to allow increased ability to tolerate and learn from being present at such events. In this case, the clinician is hoping that not only does increased exposure to these activities have the potential to improve the patient's outcome both by itself and in combination with the pharmacologic treatment, but that the patient's knowledge that he is taking a medication that he expects to increase his ability to tolerate and benefit from this experience will increase his willingness to engage in the recommended activity. Although such interactive effects may significantly complicate the design and interpretation of N-of-1 clinical trials, additional experience throughout our field exploring and understanding how such factors affect patient outcomes holds the potential to make our research results more relevant to and effective for the optimization of actual clinical care practices.

The potential for complex interactive effects may also come into play in new ways as we increase the role of precision medicine methods in research and clinical care. In tests of biomarker guided treatment selection or decision making, it will be necessary to keep in mind the possibility that biomarker results may be associated not just with treatment-specific outcomes, but also with placebo response or the interaction between placebo and treatment-specific responses. For example, genetic variations in the Catechol-O-methyltransferase (COMT) gene, a key enzyme in catecholamine catabolism, has been found to be associated with the magnitude of placebo response in a variety of treatment trials (38–40). For a clinical trial such as is being modeled here, where the primary disease state (PTSD) has itself been suggested to be associated with COMT function (41, 42) and the primary hypothesis being tested is whether biomarkers of catecholamine signaling at baseline are predictive of treatment response, this suggests that the potential for interactive effects between biologic variation in placebo response, disease state, and relevant biomarkers may not be simply theoretical. Increasing use of study designs that allow increased independent assessment of expectancy-related and other non-treatment specific components of symptom change may thus become increasingly important as we seek to move toward personalized medicine models of care.

Potential for Biased Enrollment, and Early Experience With Currently Enrolling Clinical Trial

One concern that is sometimes raised in this context is whether patients with highly distressing symptoms will be willing to enroll

in a trial that involves discontinuing a what may have already been demonstrated to be an effective treatment, explicitly to see if symptoms return. Although the impact of such an effect is expected to vary significantly based on the specifics of each trial, in our experience the likelihood of this concern affecting trial enrollment is often significantly overestimated. First, those without clinical experience may underestimate the frequency with which patients in routine clinical care discontinue effective treatments to see if they still need them, with or without the awareness of their treating physician(s) (43, 44). Particularly when a treatment may require indefinite use, patients are often very interested in finding out whether any improvement they may have experienced when starting the intervention truly requires its continuation. In contexts such as antidepressant or pain management trials, where the placebo effect can be substantial, this is often a very rational question for patients to ask.

It is also possible to actively shape the likelihood that participants concerned about this possibility will avoid enrollment or not by shaping the way expectations for the duration of trial participation are conveyed. For example, in the currently running trial, ensuring a full representation of the spectrum of patients with PTSD who present for clinical care was of high priority. Thus, when the trial was described to patients, it was emphasized not only that participation was at all times voluntary, but that it was understood that at all times, the participant would need to do what was best for their own well-being—and that at times, this might mean discontinuing participation, if it turned out that symptoms exhibited substantial return during periods of discontinuation. It was emphasized that even if the potential participant were not sure if they would be able to participate in the entire trial, we would appreciate their participation for as long as it worked for them to participate. As was incorporated mathematically into the statistical modeling, it was expected that those choosing to terminate participation prior to completion of the full trial would more commonly be those for whom response to prazosin was either clearly significant or clearly minimal—while those who elected to continue for the entire trial would more commonly be those for whom it remained unclear to both participant and researchers alike whether the participant had had a significant, specific positive response to treatment or not.

Currently, the authors (RCH and MAR) are just over 1 year in to recruitment for the clinical trial (NCT03539614) that was designed based on the statistical simulation work presented here. Consistent with the concern discussed above that acutely symptomatic patients would be less likely to enroll in a trial of a widely available treatment if there were the potential that they would be initially randomized to a placebo group, and the finding that statistical power is only minimally worse for the N-of-1 design beginning with an open label period as compared with a traditional crossover design, the proposed N-of-1 design from these models was selected as the basis of the currently running clinical trial. The understanding that the trial would begin with active treatment for all participants, but that later treatment blocks would include both blinded drug and placebo, was clearly conveyed to all participants as part of informed consent. In addition to the types of outcome measures described in this simulation study, participants also completed daily symptom logs for a subset of weeks during the trial; these symptom logs included both items that were common to all participants, and items designed by the participants to reflect issues of particular importance to them in understanding how the treatment did or did not benefit them. The participants were informed at the beginning of the trial that they would be provided the data describing how their symptom reports changed during treatment with active drug and placebo at the end of the trial, and that one of our goals in the trial was to provide them as well as us as much information as possible about the ways in which the treatment did vs. did not help them, and whether they need to continue to take the medication in order to maintain any benefit that was achieved.

We found that recruitment for this type of a trial design was unexpectedly rapid, and in fact outpaced the resources we had allocated for the trial; we were eventually awarded a significant supplemental budget increase to accommodate the larger than expected recruitment interest. This experience is in contrast to multiple other PTSD treatment trials that have been run by our research group and others at our research site. The two factors that have been cited by participants and by those referring to the trial have been (1) the fact that everyone starts on active treatment, and (2) the fact that the trial is designed to provide participants with personalized information regarding their own individual response to treatment, including to what extent their symptoms were found to return when they transitioned from active treatment to placebo.

Relevance to Clinical Trial Analysis Methods and the Development of Predictive Models of Response

One of the primary goals of clinical research into predictive biomarkers is to allow biomarker guided treatment selection. For example, if the current running trial, described above, is found to support a significant association between noradrenergic biomarkers measured at baseline and response to treatment with prazosin, the next step in testing the clinical relevance of this finding would be a clinical trial where all participants are treated with one of two active treatments, but that randomizes participants to a biomarker-guided treatment selection group vs. a non-biomarker-guided treatment selection group, and compares outcomes across the group where biomarkers are used and the group where biomarkers are not used in treatment selection.

To accomplish this, one needs to use the results of the current clinical trial to inform the development a treatment selection algorithm, which can in turn be used to guide treatment selection for individual patients. Although the focus of work presented here was on the use of statistical simulations to guide clinical trial design, the methods implemented can also be applied directly to the results of an actual clinical trial. Because the measurement of treatment response used here is continuous rather than binary, the results do not by default take the form of a classifier of individuals predicted to be treatment responsive

vs. treatment non-responsive; instead, they produce a predictive model of expected mean change in total symptom severity, over a given window of time on treatment, for someone with a given combination of baseline symptom severity and biomarker measurements (an example of how to implement this analysis using existing clinical trial data is provided in vignette three in the R package accompanying this publication). This predicted response curve can then be combined with information about the expected chance of benefit and degree of benefit from an alternative treatment, along with information regarding the relative risk, cost, and convenience of both treatment options, in order to create a treatment selection algorithm for a biomarker-guided decision making trial, or for use in clinical care.

Limitations

This work has a number of important limitations. First, the statistical simulations of clinical trials necessarily makes several simplifying assumptions, such as the presumption of linearity in combining treatment effects, the adherence of carryover effects to an exponential decay curve, the constriction of the direct, biologic response, the expectancy-related response, and non-treatment dependent effects each to a single time course, and many others. The addition of further complexity to the models has the potential for both risk and benefit. Here, where the primary goal of our statistical simulations was to guide in the selection of and power calculations for as specific predictive-biomarker clinical trial, our goal in statistical design choices was to have known oversimplifications in model implementation be at least unbiased with respect to impact on clinical trial design, and for the impact to be small or comparable relative to the degree of oversimplification in traditional power analyses. In other contexts, however, the relative cost vs. benefit of adding in or leaving out explicit modeling of different factors may be quite different.

There are also potential benefits to aggregated N-of-1 clinical trial designs that are not directly addressed in this particular set of models. For example, based on our experiences with previous clinical trial enrollment patterns, we expect there to be a significant likelihood of differential enrollment of higher acuity patients and those with a higher likelihood of being treatment responders between trial designs that begin with an open label phase and those that are entirely placebo-controlled. Although this type of differential enrollment would directly affect the power for our primary outcome, it was not explicitly included in the model.

It is also our experience from the first year of enrolling participants in this clinical trial that the opportunity to receive one's own data addressing the extent to which one (a) responded to a particular intervention, and (b) needs to remain on that intervention to maintain any observed benefit is perceived as a significant benefit by many participants, and has helped to increase not only participant recruitment but also participant engagement throughout the trial. For example, participants in the current trial have completed both medication logs and daily symptom logs at higher rates than has been observed by the authors in similar studies using PG-RCT designs (unpublished observations). This experience is consistent with previously reported assessments of patient experiences in n-of-1 trial designs (45). Such an effect might well-influence such factors as dropout and adherence, which could in turn be explicitly included in the model so as to capture their potential effect on statistical power and effect size estimation. In addition, however, these factors appear to reflect the perception by patients that participation in this type of a clinical trial design simply provides them increased personal benefit compared with participation in a traditional PG-RCT—a factors that may not directly affect power or bias calculations, but which we believe to be meaningful in and relevant to clinical trial design in and of its own right.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The R code and parameters extracted from pilot data that were used for this study, along with the analysis code used to generate the results presented, can be found as a GitHub package at https://github.com/rchendrickson/pmsimstats.

DISCLOSURE

The views expressed are those of the authors and do not reflect the official policy of the Department of Veterans Affairs or the U.S. Government.

AUTHOR CONTRIBUTIONS

RH and MR contributed to the conception of the study. RH, NS, and MR contributed to the initial design of the study. RH implemented the statistical simulations. RT validated this implementation. RH wrote the first draft of the manuscript. RT wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by the Department of Veterans Affairs (VA) Clinical Sciences Research and Development Service Career Development Award IK2CX001774 (RH); VA Northwest Network MIRECC (RH, MR); the VA Office of Academic Affiliations Advanced Fellowship Program in Mental Illness Research and Treatment (RH); NIH Grants Nos. UH2 AG064706, U19 AG023122, U24 AG051129, and U24 AG051129-04S1 (NS); Dell Inc (NHS); and the Ivy and Ottosen Foundations (NS).

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Steven P. Millard and Jane B. Shofer for significant assistance with early development of the statistical models eventually implemented.

REFERENCES

1. Kravitsz R, Duan N, Vohra S, Li J, The DEcIDE Methods Center N-of-1 Guidance Panel Duan N, Eslick I. Introduction to N-of-1 trials: indications and barriers. In: Kravitz R, Duan N, editors. *Design and Implementation of N-of-1 Trials: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality (2014). p. 1–11.
2. Schork NJ. Personalized medicine: time for one-person trials. *Nature.* (2015) 520:609–11. doi: 10.1038/520609a
3. Punja S, Eslick I, Duan N, Vohra S, The DEcIDE Methods Center N-of-1 Guidance Panel Duan N, Eslick I. An ethical framework for N-of-1 trials: clinical care, quality improvement or human subjects research? In: Kravitz R, Duan N, editors. *Design and Implementation of N-of-1 Trials: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality (2014). p. 13–22.
4. Raskind MA, Peskind ER, Chow B, Harris C, Davis-Karim A, Holmes HA, et al. Trial of prazosin for post-traumatic stress disorder in military veterans. *N Engl J Med.* (2018) 378:507–17. doi: 10.1056/NEJMoa1507598
5. Schmid C, Duan N, The DEcIDE Methods Center N-of-1 Guidance Panel (Duan N EI. Statistical design and analytic considerations for N-of-1 trials. In: Kravitz R, Duan N, editors. *Design and Implementation of N-of-1 Trials: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality. (2014) p. 33–53.
6. Ness PH Van, Murphy TE, Ali A. Attention to individuals : mixed methods for N-of-1 health care interventions. *J Mix Methods Res.* (2017) 11:342–54. doi: 10.1177/1558689815623685
7. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ, et al. The N-of-1 clinical trial: the ultimate strategy for individualzing medicine? *Per Med.* (2011) 8:161–73. doi: 10.2217/pme.11.7
8. Wang Y, Schork NJ. Power and design issues in crossover-based N-of-1 clinical trials with fixed data collection periods. *Healthcare.* (2019) 7:84. doi: 10.3390/healthcare7030084
9. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clin Trials.* (2007) 4:125–39. doi: 10.1177/1740774507077849
10. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals. *Clin Trials.* (2004) 1:368–76. doi: 10.1191/1740774504cn032oa
11. Taylor FFB, Martin P, Thompson C, Williams J, Mellman TA, Gross C, et al. Prazosin effects on objective sleep measures and clinical symptoms in civilian trauma posttraumatic stress disorder: a placebo-controlled study. *Biol Psychiatry.* (2008) 63:629–32. doi: 10.1016/j.biopsych.2007.07.001
12. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ.* (2013) 346:e8668. doi: 10.1136/bmj.e8668
13. Rief W, Bingel U, Schedlowski M, Enck P. Mechanisms involved in placebo and nocebo responses and implications for drug trials. *Clin Pharmacol Ther.* (2009) 90:722–6. doi: 10.1038/clpt.2011.204
14. Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol.* (1979) 8:7–20. doi: 10.1111/j.1365-2125.1979.tb05903.x
15. Sheiner LB, Hashimoto Y, Beal SL. A simulation study comparing designs for dose ranging. *Stat Med.* (1991) 10:303–21. doi: 10.1002/sim.4780100303
16. Senn S. *Cross-Over Trials in Clinical Research. Second*. West Sussex: John Wiley & Sons Ltd (2002).
17. Senn S. Cross-over trials in statistics in medicine: the first "25" years. *Stat Med.* (2006) 25:3430–42. doi: 10.1002/sim.2706
18. Chen X, Chen P. A comparison of four methods for the analysis of n-of-1 trials. *PLoS ONE.* (2014) 9:e87752. doi: 10.1371/journal.pone.0087752
19. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. *PLoS ONE.* (2016) 11:e0167167. doi: 10.1371/journal.pone.0167167
20. Percha B, Baskerville EB, Johnson M, Dudley JT, Zimmerman N. Designing robust N-of-1 studies for precision medicine. *J Med Internet Res.* (2019) 21:1–12. doi: 10.2196/12641
21. Grenet G, Blanc C, Bardel C, Gueyffier F, Roy P. Comparison of crossover and parallel-group designs for the identification of a binary predictive biomarker of the treatment effect. *Basic Clin Pharmacol Toxicol.* (2020) 126:59–64. doi: 10.1111/bcpt.13293
22. Raskind MA, Millard SP, Petrie EC, Peterson K, Williams T, Hoff DJ, et al. Higher pretreatment blood pressure is associated with greater ptsd symptom reduction in soldiers treated with prazosin. *Biol Psychiatry.* (2016) 80:736–42. doi: 10.1016/j.biopsych.2016.03.2108
23. R Development Core Team. *R: A Language end Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. (2017). Available online at: http://www.r-project.org
24. R Studio Team. *R Studio: Integrated Development for R*. Boston, MA: RStudio Inc (2018). Available online at: http://www.rstudio.com/
25. Raskind MA, Peterson K, Williams T, Hoff DJ, Hart K, Holmes H, et al. A trial of prazosin for combat trauma PTSD with nightmares in active-duty soldiers returned from Iraq and Afghanistan. *Am J Psychiatry.* (2013) 170:1003–10. doi: 10.1176/appi.ajp.2013.12081133
26. Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS, et al. The development of a clinician-Administered PTSD scale. *J Trauma Stress.* (1995) 8:75–90.
27. Ripley B, Venables B, Bates DM, Hornik K, Gebbhardt A, Firth D. Support functions and datasets for Venables and Ripley's MASS. *CRAN.* (2019)
28. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang.* (2013) 68:255–78. doi: 10.1016/j.jml.2012.11.001
29. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* (2015) 67:1–48. doi: 10.18637/jss.v067.i01
30. Kube T, Rief W. Are placebo and drug-specific effects additive? questioning basic assumptions of double-blinded randomized clinical trials and presenting novel study designs. *Drug Discov Today.* (2017) 22:729–35. doi: 10.1016/j.drudis.2016.11.022
31. Vase L, Amanzio M, Price DD. Nocebo vs. placebo : the challenges of trial design in analgesia research. *Clin Pharmacol Ther.* (2015) 97:143–50. doi: 10.1002/cpt.31
32. Rochon PA, Binns MA, Litner JA, Litner GM, Fischbach MS, Eisenberg D, et al. Are randomized control trial outcomes influenced by the inclusion of a placebo group? a systematic review of nonsteroidal antiinflammatory drug trials for arthritis treatment. *J Clin Epidemiol.* (1999) 52:113–22. doi: 10.1016/S0895-4356(98)00149-8
33. Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain relationship to study characteristics. *Neurology.* (2008) 70:263–72. doi: 10.1212/01.wnl.0000275528.01263.6c
34. Benedetti F, Carlino E, Piedimonte A. Increasing uncertainty in CNS clinical trials: the role of placebo, nocebo, and Hawthorne effects. *Lancet Neurol.* (2016) 15:736–47. doi: 10.1016/S1474-4422(16)00066-1
35. Benedetti F, Amanzio M, Maggi G. Potentiation of placebo analgesia by proglumide. *Lancet.* (1995) 346:1231. doi: 10.1016/S0140-6736(95)92938-X
36. Bingel U, Wanigasekera V, Wiech K, Mhuircheartaigh RN, Lee MC, Ploner M, et al. The effect of treatment expectation on drug efficacy : imaging the analgesic benefit of the opioid remifentanil. *Sci Transl Med.* (2011) 3:70ra14. doi: 10.1126/scitranslmed.3001244
37. Enck P, Bingel U, Schedlowski M, Rief W. The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov.* (2013) 12:191–204. doi: 10.1038/nrd3923
38. Colloca L, Wang Y, Martinez PE, Chang YC, Ryan KA, Hodgkinson C. rs324420 genes interact with placebo procedures to induce hypoalgesia. *Pain.* (2019) 160:1824–34. doi: 10.1097/j.pain.0000000000001578
39. Leuchter AF, Mccracken Þ*JT, Hunter Þ*AM, Cook IA, Alpert JE. Monoamine oxidase a and catechol- O -methyltransferase functional polymorphisms and the placebo response in major depressive disorder. *J Clin Psychopharmacol.* (2009) 29:372–7. doi: 10.1097/JCP.0b013e3181ac4aaf

40. Hall KT, Lembo AJ, Kirsch I, Ziogas DC, Douaiher J, Jensen KB, et al. Catechol-O-methyltransferase val158met polymorphism predicts placebo effect in irritable bowel syndrome. *PLoS ONE.* (2012) 7:1–6. doi: 10.1371/journal.pone.0048135
41. Deslauriers J, Toth M, Zhou X, Risbrough VB, Perrine SA. Heritable differences in catecholamine signaling modulate susceptibility to trauma and response to methylphenidate treatment : relevance for PTSD. *Front Behav Neurosci.* (2019) 13:111. doi: 10.3389/fnbeh.2019.00111
42. Haxhibeqiri V, Haxhibeqiri S, Topciu-shufta V, Agani F, Goci Uka A, Hoxha B, et al. The association of catechol-O-methyl-transferase and interleukin 6 gene polymorphisms with posttraumatic stress disorder. *Psychiatr Danub.* (2019) 31:241–8. doi: 10.24869/psyd.2019.241
43. Bulloch AGM, Patten SB. Non-adherence with psychotropic medications in the general population. *Soc Psychiatry Psychiatr Epidemiol.* (2010) 45:47–56. doi: 10.1007/s00127-009-0041-5
44. Cooper JK, Love DW, Raffoul PR. Intentional prescription nonadherence (Noncompliance) by the elderly. *J Am Geriatr Soc.* (1982) 30:329–33. doi: 10.1111/j.1532-5415.1982.tb05623.x
45. Nikles CJ, Clavarino AM, Del Mar CB. Using n-of-1 trials as a clinical tool to improve prescribing. *Br J Gen Pract.* (2005) 55:175–80.

Conflict of Interest: MR is a paid advisory board member of Pfizer Laboratories, Merck, and Takeda Pharmaceuticals. RH is principle investigator of a VA CSR&D funded N-of-1 clinical trial using the methods discussed in this publication. NS has received funding from Dell Inc. These funding sources were not involved in the study design, implementation, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychiatry

BRIEF RESEARCH REPORT
published: 04 September 2020
doi: 10.3389/fpsyt.2020.530995

Exploring the Potential for Collaborative Use of an App-Based Platform for n-of-1 Trials Among Healthcare Professionals That Treat Patients With Insomnia

Jason R. Bobe[1]*, *Jessica K. De Freitas*[1,2] *and Benjamin S. Glicksberg*[1,2]*

[1] Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [2] Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

OPEN ACCESS

Edited by:
Wulf Rössler, Charité – Universitätsmedizin Berlin, Germany

Reviewed by:
Scott Bauer, University of California, San Francisco, United States
Salima Punja, University of Alberta, Canada
Claire Helen McCallum, Northumbria University, United Kingdom

***Correspondence:**
Jason R. Bobe
jason.bobe@mssm.edu
Benjamin S. Glicksberg
benjamin.glicksberg@mssm.edu

Specialty section:
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

Received: *30 January 2020*
Accepted: *17 August 2020*
Published: *04 September 2020*

Citation:
Bobe JR, De Freitas JK and Glicksberg BS (2020) Exploring the Potential for Collaborative Use of an App-Based Platform for n-of-1 Trials Among Healthcare Professionals That Treat Patients With Insomnia. Front. Psychiatry 11:530995. doi: 10.3389/fpsyt.2020.530995

Background: N-of-1 trials are single patient, multiple crossover, and comparative effectiveness experiments. Despite their rating as "level 1" evidence, they are not routinely used in clinical medicine to evaluate the effectiveness of treatments.

Objective: We explored the potential for implementing a mobile app-based n-of-1 trial platform for collaborative use by clinicians and patients to support data-driven decisions around the treatment of insomnia.

Methods: A survey assessing awareness and utilization of n-of-1 trials was administered to healthcare professionals that frequently treat patients with insomnia at the Icahn School of Medicine at Mount Sinai in New York City.

Results: A total of 45 healthcare professionals completed the survey and were included in the analysis. We found that 64% (29/45) of healthcare professionals surveyed had not heard of n-of-1 trials. After a brief description of these methods, 75% (30/40) of healthcare professionals reported that they are likely or highly likely to use an app-based n-of-1 trial at least once in the next year if the service were free and easy to offer to their patients.

Conclusions: An app-based n-of-1 trials platform might be a valuable tool for clinicians and patients to identify the best treatments for insomnia. The lack of awareness of n-of-1 trials coupled with receptivity to their use suggests that educational interventions may address a current barrier to wider utilization of n-of-1 trials.

Keywords: n-of-1, sleep, clinical informatics, mHealth, RWE, crossover

INTRODUCTION

Healthcare professionals (HCPs) routinely practice individualized care. They design treatment plans based on unique patient characteristics and clinical presentation, consider various levels of evidence for treatment efficacy, and help patients weigh the risks of side effects and other potential treatment burdens and trade-offs. While there is widespread agreement that we should not expect most

treatments to work uniformly across most populations, the systematic evaluation of treatment effect remains a challenge for healthcare professionals and patients (1). Healthcare professionals who practice "evidence-based medicine" generally use comparisons of means and proportions between large groups of patients (e.g., from clinical trials) but are intuitively aware that there exists large heterogeneity of effects for many disease processes and interventions. N-of-1 trials create an opportunity in some contexts for healthcare professionals and patients to individualize treatment selection in a more systematic way. They are designed to help both parties make objective, data-driven treatment choices.

What Are n-of-1 Trials?

In clinical medicine, n-of-1 trials are used as a decision support tool to inform individualized treatment selection (2). The Oxford Centre for Evidence-Based Medicine ranks n-of-1 trials as "level 1" evidence for determining whether a treatment benefits a patient (3). N-of-1 trials are also viewed as a tool to enhance patient-centered care, insofar as the patient may be involved in the selection of treatments to compare, the selection of outcomes to measure, and the selection of the treatment to continue at the end of the trial (4). Typically, in an n-of-1 trial, a single patient completes a baseline period without any treatments, then alternates between two treatments in a sequence (i.e., "multiple crossover") (5, 6). Where feasible, treatments may be blinded or placebo-controlled. Outcomes are measured during baseline and each treatment period. At the end of the trial, outcome measurements for each treatment are compared and a treatment is selected. N-of-1 trials may also be deployed to answer other common treatment investigations, such as whether to begin a treatment, proper dosing (7), disease-related nutrition recommendations (8–10), assessing treatment response in people with characteristics (e.g., rare genetic variants) not studied in randomized controlled trials (RCTs) of approved medications (11–13), and deprescribing (14), among others.

To be sure, n-of-1 trials are not useful in every context. They work best in patients with chronic or stable conditions. Non-curative treatments with rapid onset and rapid washout are ideal candidates for n-of-1 trials, whereas treatments with cumulative effectiveness (e.g., some antidepressants) or treatments that disrupt the nature of the underlying condition (e.g., surgery) are not. N-of-1 trials are particularly relevant in contexts where evidence for treatment efficacy is weak or where treatment effects are known to be heterogeneous across populations and among individuals (5).

Chronic Insomnia: A Testbed for Implementation of n-of-1 Trials

There are many therapeutic contexts where n-of-1 trials are able to serve unmet patient needs. Chronic insomnia is a good target disorder because there is a high prevalence of affected individuals in the general population and also across distinct clinical populations, including Alzheimer's disease (AD), Parkinson's disease (PD), attention deficit hyperactivity disorder (ADHD), pregnancy, post-treatment Lyme disease (PTLD) patients (15), and many others.

N-of-1 trials have been implemented or are currently underway in several populations with insomnia. Coxeter et al. reported the effectiveness of valerian versus placebo in 24 patients with chronic insomnia (16). Punja et al. have described a protocol to assess the effectiveness of melatonin in children (ages 6–17) taking medications for ADHD (17). Nikles et al. have described a protocol to assess the effectiveness of melatonin in patients with PD (18).

Many patient populations have insomnia. Patients with insomnia have significant interest in finding the most effective treatment due to the negative impact on quality-of-life for patients, family, and caregivers (19, 20). Poor sleep may also interact with recovery or progression of some diseases, including cardiovascular disease, depression, diabetes, AD, among others (21–25). Estimates of the prevalence of insomnia in the general population vary in part due to the definition of insomnia used. The American Academy of Sleep Medicine (AASM) estimates that 33%–50% of adults experience symptoms of insomnia and 10%–15% of adults experience insomnia disorders that disrupt daily functioning (26).

There is limited efficacy data available for insomnia treatments in many patient populations, so decisional conflict is common. The American Academy of Physicians recommends cognitive behavioral therapy for insomnia (CBT-I) as the first line therapy in adults with chronic insomnia (27). In a systematic review of RCTs that compared CBT-I and medications for the treatment of insomnia, evidence supports the notion that CBT-I is better than medications in some contexts (28). However, among patients that choose to pursue CBT-I, around 20%–30% fail to respond (29). Furthermore, many people do not choose CBT-I because of several treatment-related burdens, including limited access to trained healthcare professionals, weekly therapy appointments, and out-of-pocket costs (30, 31). There is widespread use of pharmacological interventions and over-the-counter (OTC) sleep aids. Around 20% of U.S. adults use prescribed or OTC sleeping medications each month (32). While commonly used, many OTC sleep treatments have limited efficacy and safety data (33). Furthermore, some sleep medications that are commonly used in younger populations for sleep problems are, for example, potentially inappropriate for use in older populations (e.g., hypnotics or Z drugs) (34). For n-of-1 trials to be an effective tool for patients and healthcare professionals, the design of these trials should incorporate the precise needs of the populations they intend to serve.

Key Questions for the Current Study

Several common themes related to barriers to physician adoption of n-of-1 trials were reported in 2009, based on focus groups with 32 patients and 21 providers in California. First, some clinicians were unclear about the validity of n-of-1 trials. Common concerns raised about cross-over, "sample size" and statistical validity indicates that education in the scientific basis of such trials will be important for adoption. Second, some clinicians were concerned about the potential for n-of-1 trials to interfere with the patient-physician relationship. Third and most germane to the current study, some physicians voiced concerns about the potential time and resource burdens n-of-1 trials would introduce (35). An app-based n-of-1 trial service that is able to reduce some logistical burdens inherent to n-of-1 trials through

the automation of previously resource-intensive processes, such as the trial design, study administration and analysis of results, presents an opportunity to potentially enhance adoption of these methods (36, 37).

The purpose of this study is to collect preliminary evidence to inform potential routes for the implementation of an app-based n-of-1 trial platform, called the N1 app. While this platform may be adapted to inform the optimal selection of treatments for a variety of disorders and wellness-related goals, patients with insomnia have been identified as a potential population to target due to the prevalence of the condition, limited evidence for efficacy of treatments across some patient populations, the possibility to incorporate wearable devices for the passive collection of outcome measures (38), among others. With the recent emergence of several app-based tools to support n-of-1 trials generally (39) or self-experimentation with sleep improvement in particular (40, 41), questions about barriers and facilitators to implementation of similar tools for the improvement of patient care or wellness will be of broad interest to the field.

In this study, we aim to assess the familiarity and experience with n-of-1 trials in a convenience sample of healthcare workers that frequently care for patients with insomnia. Our hypothesis is that healthcare professionals are currently familiar with but do not regularly use n-of-1 trials. A secondary aim of the study is to see if there are associations between awareness or utilization of n-of-1 methods and other variables (e.g., age and years in clinical practice). For example, a greater number of years in clinical practice may facilitate exposure to a broader repertoire of clinical methods or lead to the adoption of more traditional clinical practices.

METHODS

Survey

Instrument

The Office for the Protection of Human Subjects at Mount Sinai approved a protocol to administer a voluntary, anonymous survey to healthcare professionals. The exploratory, cross-sectional survey administered has three sections (a) sociodemographics, (b) experience and satisfaction with the treatment of patients with insomnia, and (c) awareness and utilization of n-of-1 trials (see Datasheet S1). The survey was administered online through REDCap (Research Electronic Data Capture) hosted by the Icahn School of Medicine at Mount Sinai (42).

Sample and Recruitment

We chose to recruit a convenience sample of clinicians and nurse practitioners through the Department of Psychiatry due to the regularity with which these healthcare professionals treat patients with insomnia. We sent messages through the department mailing list and advertised at bimonthly Grand Rounds events during October and November of 2019.

Statistical Analysis

We summarize the survey responses in **Table 1**. We further assessed differences of participants' awareness of n-of-1 using two-sided Chi-squared test for categorical variables and a t-test for continuous variables which produced p-values and odds ratios. Using the same methods, we also assessed associations for willingness to use a n-of-1 digital service, which we discretized due to low sample size and imbalanced responses into "More likely" (consisting of "Strongly agree" and "Agree" responses) and "Less likely" (consisting of "Neutral", "Disagree", and Strongly disagree" responses).

Ethics Approval

This study has been approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai (IRB-18-00789).

RESULTS

Survey Results

Sample Characteristics

A total of 66 participants consented to the survey, 49 participants completed the survey and the responses from 45 participants were included in the analysis (see **Supplementary Figure 1**). This sample of 45 healthcare professionals from a large, urban hospital are mostly practicing clinicians (88.9%), affiliated with the Department of Psychiatry (88.9%), in their first or second decade of clinical

TABLE 1 | Sociodemographics of the sample of healthcare professionals surveyed.

Age (mean ± SD)	**48.4 ± 16.5**
Sex	
Female (n, %)	26 (57.8)
Male (n, %)	18 (40.0)
Not reported (n, %)	1 (2.2)
Race	
Asian (n, %)	3 (6.7)
Black or African-American (n, %)	2 (4.4)
White (n, %)	34 (75.6)
Multiple (n, %)	3 (6.7)
Unknown/Not reported (n, %)	3 (6.7)
Ethnicity	
Hispanic/Latino (n, %)	3 (6.7)
Not Hispanic/Latino (n, %)	41 (91)
Unknown/Not reported (n, %)	1 (2.2)
Clinical degree(s)	
MD (n, %)	39 (86.7)
NP (n, %)	5 (11.1)
MD/NP (n, %)	1 (2.2)
Years of clinical practice	
0–10 (n, %)	20 (44.4)
11–20 (n, %)	9 (20.0)
21–30 (n, %)	7 (15.6)
31–40 (n, %)	5 (11.1)
>40 (n, %)	3 (6.7)
Not reported (n, %)	1 (2.2)
Department(s)	
Psychiatry (n, %)	43 (95.6)
Psychiatry and Anesthesiology (n, %)	1 (2.2)
Not reported (n, %)	1 (2.2)
Primary Specialty	
Psychiatry and Neurology (n, %)	42 (93.3)
Other (n, %)	1 (2.2)
Not reported (n, %)	2 (4.4)

practice (64.4%), with a primary specialty of Psychiatry and Neurology (93.3%) (see **Table 1**).

Experience and Satisfaction of Sample With Current Treatments for Patients With Insomnia

Most healthcare professionals surveyed frequently see patients with insomnia in their clinical practice with 88.9% (40/45) reporting daily or weekly encounters. A majority of participants also expressed their dissatisfaction with available treatment options for their patients with insomnia, with 55.6% (25/45) disagreeing or strongly disagreeing with the statement: "I am satisfied with the available treatment options for my patients with insomnia." Participants also perceived their patients as being dissatisfied with available treatment options, with 62.2% (28/45) disagreeing or strongly disagreeing with the statement: "My patients are satisfied with available treatment options for insomnia" (see **Table 2**).

Awareness and Use of n-of-1 trials

Most participants surveyed were unfamiliar with the concept of n-of-1 trials, with 64.4% (29/45) reporting that they had never heard of them before. Following this survey question, participants were presented with a short description of n-of-1 trials that also included a screenshot of the N1 app, a smartphone based n-of-1 trial platform developed at the Icahn School of Medicine (see **Data sheet S1**) (36, 43). They were asked to consider the scenario that "a free service [existed] that made it easy for you to offer n-of-1 trials to select patients in your practice with insomnia. The patient would conduct the mobile app-based trial at home. At the conclusion of the trial, the analyzed results would be available to you and the patient to review together." Excluding five participants that did not respond to this question, 75% (30/40) of participants reported that they were either "likely" or "highly likely" to "use a service like this to make data-driven treatment choices at least once in the next year."

Of the 16 healthcare professionals that reported that they were aware of n-of-1 trials, three reported that they had previously used an n-of-1 trial in the treatment of their patients. For the remaining 11 participants that were aware of n-of-1 trials, had never used them in their clinical practice, and reported a primary reason for the lack of adoption, 45.5% (5/11) cited inadequate training in n-of-1 trial design.

We further assessed the relationship between various participant characteristics and having heard of n-of-1 trials (n = 45). We found no significant relationship between having heard of n-of-1 trials and age ($p = 0.54$, $t = -0.61$), sex ($p = 0.09$, χ-squared = 2.91), race ($p = 0.10$, χ-squared = 7.92), ethnicity ($p = 0.75$, χ-squared = 0.58), clinical degree ($p = 0.84$, χ-squared = 0.04; one individual with both degrees not included), number of years practiced ($p = 0.29$, χ-squared = 5.02), frequency of seeing patients with insomnia ($p = 0.50$, χ-squared = 2.36), satisfaction of current treatments for insomnia ($p = 0.23$, χ-squared = 5.65), or perceived patient satisfaction of current treatments for insomnia ($p = 0.80$, χ-squared = 0.99).

We also assessed the relationship between various participant characteristics and willingness to use a n-of-1 digital service app (n = 40; discretized response). We found no significant relationship between willingness to use a digital app service and age ($p = 0.22$, $t = -1.27$), sex ($p = 0.23$, χ-squared = 1.45), race ($p = 0.54$, χ-squared = 3.13), ethnicity ($p = 0.20$, χ-squared = 3.26), clinical degree ($p = 0.81$, χ-squared = 0.06; one individual with both degrees not included), number of years practiced ($p = 0.67$, χ-squared = 2.36), frequency of seeing patients with insomnia ($p = 0.64$, χ-squared = 1.68), satisfaction of current treatments for insomnia ($p = 0.21$, χ-squared = 5.87), or perceived patient satisfaction of current treatments for insomnia ($p = 0.82$, χ-squared = 0.93) (see **Table 2**).

TABLE 2 | Summary of survey results from a sample of healthcare professionals on their experience and satisfaction with current treatments for insomnia.

How often do you see patients in your practice with insomnia?	
Daily (n, %)	17 (37.8)
Weekly (n, %)	23 (51.1)
Monthly (n, %)	3 (6.7)
Quarterly (n, %)	2 (4.4)
Total (n)	45
I am satisfied with the available treatment options for my patients with insomnia.	
Strongly agree (n, %)	2 (4.4)
Agree (n, %)	7 (15.6)
Neutral (n, %)	11 (24.4)
Disagree (n, %)	21 (46.7)
Strongly disagree (n, %)	4 (8.9)
Total (n)	45
My patients are satisfied with available treatment options for insomnia.	
Strongly agree (n, %)	0 (0.0)
Agree (n, %)	8 (17.8)
Neutral (n, %)	9 (20.0)
Disagree (n, %)	24 (53.3)
Strongly disagree (n, %)	4 (8.9)
Total (n)	45
Have you ever heard of n-of-1 trials?	
No (n, %)	29 (64.4)
Yes (n, %)	16 (35.6)
Total (n)	45
How likely are you to use a free service that made it easy for you to offer n-of-1 trials to select patients in your practice with insomnia in order to make data-driven treatment choices at least once in the next year?	
Highly likely (n, %)	12 (30.0)
Likely (n, %)	18 (45.0)
Neutral (n, %)	6 (15.0)
Unlikely (n, %)	3 (7.5)
Highly unlikely (n, %)	1 (2.5)
Total (n)	40

DISCUSSION

The N1 app aims to facilitate the design, administration, and analysis of n-of-1 trials (36, 37). Individuals with insomnia or other chronic sleep disturbance issues are a population that may benefit from access to n-of-1 trials for data-driven treatment selection. While these multi-crossover, comparative effectiveness trials have been in use for decades, awareness, and adoption by healthcare professionals continues to face challenges, as our survey further indicates. Yet, we are encouraged that the

healthcare professionals sampled did also report substantial receptivity to future use of app-based n-of-1 trials that were free and enabled collaboration with patients with insomnia who could conduct trials from home.

The lack of awareness of n-of-1 trials coupled with receptivity to their use suggests that educational interventions may address a current barrier to wider utilization of n-of-1 trials. One limitation of this study is the lack of generalizability due to the small sample of healthcare professionals surveyed at a single healthcare system. An important area for future study is to understand how awareness and utilization of n-of-1 trials differs across health systems and other medical specialties that regularly treat patients with insomnia, such as primary care. The survey and analysis conducted here could be deployed in a larger and more diverse sample encompassing providers across multiple specialties and health systems in order to obtain more generalizable knowledge about awareness, utilization, and barriers to adoption of n-of-1 methods and receptivity to the use of an app-based service. The addition of more open-ended questions to the survey, for example, related to barriers to adoption, may also lead to new findings that may not be readily captured in the current instrument. The convenience sampling method used for this survey may also have biased our results due to the possibility that respondents that enrolled were more interested in the idea of app-based n-of-1 trials compared to individuals that did not enroll. A larger and more diverse sample of HCPs may also be able to provide insights on associations between awareness or utilization of n-of-1 trials and other relevant variables, such as years of clinical practice, medical specialty, or age. We found no significant association among the variables we assessed but were also limited by a small sample size. Any such association identified in future studies may help to identify implementation strategies that are tailored for specific subsets of potential end users.

This study also suggests additional work is needed to identify the key barriers and facilitators to implementation in the specific context of an app-based n-of-1 trial service. Prior focus groups identified several key themes among HCPs related to barriers to implementation of n-of-1 methods (35), but that study did not contemplate the potential efficiencies gained or the potential problems introduced through the use of an app-based n-of-1 trial service. The Theoretical Domains Framework (TDF) has been deployed to investigate implementation problems in many healthcare settings, as reviewed in Francis et al. (44). TDF could be used to design interview questions that more thoroughly explores implementation issues related to an app-based n-of-1 trial service with a sample of HCPs and patients in future qualitative research.

While the vast majority of participants expressed a willingness to use an app-based n-of-1 trial platform, we recognize the limitations of this exploratory survey. For example, we do not address the potential complexities involved in getting an app effectively incorporated into existing clinical workflows and training of health care professionals in appropriate use. Moreover, while an app-based platform may reduce some aspects of n-of-1 trials that were previously labor-intensive, such as automated data analysis, new burdens may also be introduced, such as remote end-user support. The engagement of potential end-users, including both patients and providers, to collect more involved feedback about key features and functionality of an app-based n-of-1 trial platform, along with usability testing are important future directions.

Although n-of-1 trials are a powerful tool for patient-centered care in some contexts, we were surprised to find low awareness among our sample of healthcare providers. There may be an important role for organizations such as the Patient-Centered Outcomes Research Institute (PCORI) in the development of guidance materials, or a rubric, that raises the awareness among patients and providers and facilitates implementation of methodologically sound n-of-1 trials.

Historically, several centers have been established to support the implementation of these trials for local clinicians as a service (7, 35, 45). One center currently operates at the University of Queensland in Australia with a focus on sleep (46). While these centers have documented many cases where patients and clinicians were aided in the selection of treatments, the centers are often experiments themselves that last as long as funding permits their operation, for example. For a time in the United States, there was a Current Procedural Terminology (CPT) code for "personalized medicine tests" that suggested a route to reimbursement for the effort required to design, administer, and analyze an n-of-1 trial (4). Writing in 2008, Kravitz et al. suggested that one alternative to the model where n-of-1 trials take root and gain traction primarily through academic clinical centers is the possibility that they "cast off some of their academic trappings and focus on appealing to what patients want and need" (4). Our survey did not address the needs of patients with sleep problems who may benefit from n-of-1 trials. An important future direction is to include a sample of patients with chronic insomnia in qualitative research exploring their perspectives about the use of an app-based n-of-1 trial service for the optimal selection of treatments, either in collaboration with their HCP, or in the case of OTC treatments, through self-guided experiments.

While healthcare practitioners take into consideration each patient's unique characteristics and strive to use the most up-to-date information to make informed therapeutic recommendations, there will always be some variability and uncertainty in outcomes. Leveraging n-of-1 trials as self-contained experiments can quantify individual outcomes and can better optimize treatment selection in some contexts. We believe growth in the adoption of n-of-1 trials will enhance the precision of treatment selection for many individuals. App-based platforms offer the potential to reduce some barriers, but other challenges still remain.

DATA AVAILABILITY STATEMENT

The raw survey dataset for this manuscript is not publicly available because of privacy considerations of clinical practitioners.

ETHICS STATEMENT

This study has been approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai (IRB-18-00789).

The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study is funded by the Icahn School of Medicine at Mount Sinai, a grant from the Steven and Alexandra Cohen Foundation, and a generous gift from Julian Salisbury.

ACKNOWLEDGMENTS

Thanks to Drs. Richard Kravitz MD (UC Davis) and Andrew Avins MD (Kaiser Permanente) for providing feedback on the survey instrument and the manuscript, and also thanks to Alex Charney MD PhD (Mount Sinai) for providing feedback on the survey instrument.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.530995/full#supplementary-material.

REFERENCES

1. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med* (2011) 8:161–73. doi: 10.2217/pme.11.7
2. Samuel J, Holder T, Molony D. N-of-1 Trials as a Decision Support Tool in Clinical Practice: A Protocol for a Systematic Literature Review and Narrative Synthesis. *Healthcare (Basel)* (2019) 7. doi: 10.3390/healthcare7040136
3. OCEBM Levels of Evidence - CEBM. CEBM [Internet]. 1 May 2016 [cited 8 Oct 2019].https://www.cebm.net/2016/05/ocebm-levels-of-evidence/.
4. Kravitz RL, Duan N, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. *Milbank Q* (2008) 86:533–55. doi: 10.1111/j.1468-0009.2008.00533.x
5. Mitchell G, Nikles J eds. *The Essential Guide to N-of-1 Trials in Health*. Springer Netherlands: Springer (2015).
6. the DEcIDE Methods Center N-of-1 Guidance Panel. *Design and Implementation of N-of-1 Trials: A User's Guide*. Kravitz RL, Duan N, Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S, editors. AHRQ: Rockville, MD (2014). Jan. Report No.: Publication No. 13(14)-EHC122-EF. Available at: https://effectivehealthcare.ahrq.gov/products/n-1-trials/research-2014-5.
7. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Ann Intern Med* (1990) 112:293–9. doi: 10.7326/0003-4819-112-4-293
8. Kaplan HC, Opipari-Arrigan L, Schmid CH, Schuler CL, Saeed S, Braly KL, et al. Evaluating the Comparative Effectiveness of Two Diets in Pediatric Inflammatory Bowel Disease: A Study Protocol for a Series of N-of-1 Trials. *Healthcare (Basel)* (2019) 7. doi: 10.3390/healthcare7040129
9. Liu L, Zhang Y, Wei J, Chen Z, Yu J. A Pilot Study of Amino Acids in Unresectable Non-Small-Cell Lung Cancer Patients During Chemotherapy: A Randomized Serial N-of-1 Trials Design. *Nutr Cancer* (2019) 71:399–408. doi: 10.1080/01635581.2018.1515962
10. Soldevila-Domenech N, Boronat A, Langohr K, de la Torre R. N-of-1 Clinical Trials in Nutritional Interventions Directed at Improving Cognitive Function. *Front Nutr* (2019) 6:110. doi: 10.3389/fnut.2019.00110
11. Magaret AS, Mayer-Hamblett N, VanDevanter D. Expanding access to CFTR modulators for rare mutations: The utility of n-of-1 trials. *J Cyst Fibros* (2019). 19(1):1–2. doi: 10.1016/j.jcf.2019.11.011
12. Nick JA, St Clair C, Jones MC, Lan L, Higgins M. VX12-770-113 Study Team. Ivacaftor in cystic fibrosis with residual function: Lung function results from an N-of-1 study. *J Cyst Fibros* (2019) 19(1):91–8. doi: 10.1016/j.jcf.2019.09.013
13. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab* (2009) 96:20–6. doi: 10.1016/j.ymgme.2008.10.003
14. Clough AJ, Hilmer SN, Naismith SL, Gnjidic D. The Feasibility of Using N-Of-1 Trials to Investigate Deprescribing in Older Adults with Dementia: A Pilot Study. *Healthcare (Basel)* (2019) 7. doi: 10.3390/healthcare7040161
15. Rebman AW, Aucott JN. Post-treatment Lyme Disease as a Model for Persistent Symptoms in Lyme Disease. *Front Med* (2020) 7:57. doi: 10.3389/fmed.2020.00057
16. Coxeter PD, Schluter PJ, Eastwood HL, Nikles CJ, Glasziou PP. Valerian does not appear to reduce symptoms for patients with chronic insomnia in general practice using a series of randomised n-of-1 trials. *Complement Ther Med* (2003) 11:215–22. doi: 10.1016/S0965-2299(03)00122-5
17. Punja S, Nikles CJ, Senior H, Mitchell G, Schmid CH, Heussler H, et al. Melatonin in Youth: N-of-1 trials in a stimulant-treated ADHD Population (MYNAP): study protocol for a randomized controlled trial. *Trials* (2016) 17:375. doi: 10.1186/s13063-016-1499-6
18. Nikles J, O'Sullivan JD, Mitchell GK, Smith SS, McGree JM, Senior H, et al. Protocol: Using N-of-1 tests to identify responders to melatonin for sleep disturbance in Parkinson's disease. *Contemp Clin Trials Commun* (2019) 15:100397. doi: 10.1016/j.conctc.2019.100397
19. Kyle SD, Morgan K, Espie CA. Insomnia and health-related quality of life. *Sleep Med Rev* (2010) 14:69–82. doi: 10.1016/j.smrv.2009.07.004
20. Edelson J, Byrnes J, Mitchell G, Heussler H, Melaku M, Nikles J. Protocol for a Longitudinal Study of Melatonin Therapy and Cost Effectiveness Analysis in Stimulant-Treated Children with ADHD and Insomnia: An N-of-1 Trial. *Contemp Clin Trials Comm* (2020) 100530. doi: 10.1016/j.conctc.2020.100530
21. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet* (2019) 51:394–403. doi: 10.1038/s41588-018-0333-3
22. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J Affect Disord* (2011) 135:10–9. doi: 10.1016/j.jad.2011.01.011
23. Lane JM, Jones SE, Dashti HS, Wood AR, Aragam KG, van Hees VT, et al. Biological and clinical insights from genetics of insomnia symptoms. *Nat Genet* (2019) 51:387–93. doi: 10.1038/s41588-019-0361-7
24. Ju Y-ES, Lucey BP, Holtzman DM. Sleep and Alzheimer disease pathology–a bidirectional relationship. *Nat Rev Neurol* (2014) 10:115–9. doi: 10.1038/nrneurol.2013.269
25. Fan M, Sun D, Zhou T, Heianza Y, Lv J, Li L, et al. Sleep patterns, genetic susceptibility, and incident cardiovascular disease: a prospective study of 385 292 UK biobank participants. *Eur Heart J* (2019). 41(11):1182–9. doi: 10.1093/eurheartj/ehz849
26. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. *J Clin Sleep Med* (2008) 4:487–504. doi: 10.5664/jcsm.27286
27. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD. Clinical Guidelines Committee of the American College of Physicians. Management of Chronic Insomnia Disorder in Adults: A Clinical Practice Guideline From the American College of Physicians. *Ann Intern Med* (2016) 165:125–33. doi: 10.7326/M15-2175
28. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. *BMC Fam Pract* (2012) 13:40. doi: 10.1186/1471-2296-13-40

29. Morin CM. Cognitive-behavioral approaches to the treatment of insomnia. *J Clin Psychiatry* (2004) 65 Suppl 16:33–40.
30. Koffel E, Bramoweth AD, Ulmer CS. Increasing access to and utilization of cognitive behavioral therapy for insomnia (CBT-I): a narrative review. *J Gen Intern Med* (2018) 33:955–62. doi: 10.1007/s11606-018-4390-1
31. Morin CM. Issues and challenges in implementing clinical practice guideline for the management of chronic insomnia. *J Sleep Res* (2017) 26:673–4. doi: 10.1111/jsr.12639
32. Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep*. (2014) 37:343–9. doi: 10.5665/sleep.3410
33. Meolie AL, Rosen C, Kristo D, Kohrman M, Gooneratne N, Aguillard RN, et al. Oral nonprescription treatment for insomnia: an evaluation of products with limited evidence. *J Clin Sleep Med* (2005) 1:173–87. doi: 10.5664/jcsm.26314
34. By the 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 Updated AGS Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults. *J Am Geriatr Soc* (2019) 67:674–94. doi: 10.1111/jgs.15767
35. Kravitz RL, Paterniti DA, Hay MC, Subramanian S, Dean DE, Weisner T, et al. Marketing therapeutic precision: Potential facilitators and barriers to adoption of n-of-1 trials. *Contemp Clin Trials* (2009) 30:436–45. doi: 10.1016/j.cct.2009.04.001
36. Bobe JR, Buros J, Golden E, Johnson M, Jones M, Percha B, et al. 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* (2020) 9:e16362. doi: 10.2196/16362
37. Golden E, Johnson M, Jones M, Viglizzo R, Bobe J, Zimmerman N. Measuring the Effects of Caffeine and L-Theanine on Cognitive Performance: A Protocol for Self-Directed, Mobile N-of-1 Studies. *Front Comput Sci* (2020) 2:4. doi: 10.3389/fcomp.2020.00004
38. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep* (2019) 42(12):1–19. doi: 10.1093/sleep/zsz180
39. Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, et al. Effect of Mobile Device–Supported Single-Patient Multi-crossover Trials on Treatment of Chronic Musculoskeletal Pain A Randomized Clinical Trial. *JAMA Intern Med* (2018) 178:1368–78. doi: 10.1001/jamainternmed.2018.3981
40. Daskalova N, Metaxa-Kakavouli D, Tran A, Nugent N, Boergers J, McGeay J, et al. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. *In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. New York, NY, USA: Association for Computing Machinery (2016) 347–58. doi: 10.1145/2984511.2984534
41. Daskalova N, Yoon J, Wang Y, Araujo C, Beltran G, Nugent N, et al. (2020). A Personalized Automated Self-Experimentation System for Sleep Recommendations. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. New York, NY, USA: Association for Computing Machinery (2020) 1–13. doi: 10.1145/2984511.2984534
42. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J BioMed Inform* (2009) 42:377–81. doi: 10.1016/j.jbi.2008.08.010
43. N1app.org. In: N1 app [Internet]. [cited 19 Jan 2020]. http://n1app.org.
44. Francis JJ, O'Connor D, Curran J. Theories of behaviour change synthesised into a set of theoretical groupings: introducing a thematic series on the theoretical domains framework. *Implement Sci* (2012) 7:35. doi: 10.1186/1748-5908-7-35
45. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. *JAMA*. (1993) 270:2708–12. doi: 10.1001/jama.270.22.2708
46. *N-of-1 trials and sleep research*. Centre for Clinical Research -University of Queensland. Available at: https://clinical-research.centre.uq.edu.au/nikles-group. [cited 20 Jan 2020].

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 September 2020
doi: 10.3389/fpsyg.2020.516485

Self-Regulatory Processes, Motivation to Conserve Resources and Activity Levels in People With Chronic Pain: A Series of Digital N-of-1 Observational Studies

Gail McMillan[1] and Diane Dixon[2]*

[1] Department of Psychology, Carleton University, Ottawa, ON, Canada, [2] Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

Objectives: Motivational and self-regulatory processes during goal pursuit may account for activity patterns in people with chronic pain. This article describes a series of N-of-1 observational studies designed to investigate the influence of goal-related factors on fluctuations in motivation to conserve resources and objectively measured activity levels.

Methods: Four participants with chronic pain who attended a formal pain management program (PMP; 41–59 years old; three female) were recruited and completed digital daily diaries for 11–12 weeks. The daily dairies, delivered via text message, measured self-regulatory fatigue, goal self-efficacy, goal striving, perceived demands, pain, and motivation to conserve resources. Continuously worn accelerometers measured physical activity and sedentary time. Analyses were conducted individually for each participant. The effects of self-regulatory fatigue, goal self-efficacy, goal striving, perceived demands, and pain on motivation to conserve resources, physical activity and sedentary time were assessed with dynamic regression modeling.

Results: Different patterns of associations between the predictors and outcomes were observed across participants. Most associations occurred concurrently (e.g., on the same day). Perceived demand was the only variable to predict motivation to conserve resources, physical activity, and sedentary time. Motivation to conserve resources and sedentary time were most frequently predicted by goal striving and perceived demand. Self-regulatory fatigue and pain intensity both predicted motivation to conserve resources in two participants and sedentary time in one participant. Motivation to conserve resources predicted sedentary time in two participants.

Conclusion: This study was the first to examine the impact of fluctuations in self-regulatory processes on motivation to conserve resources and objective activity levels within individuals with chronic pain. The results generally supported recent affective-motivational views of goal pursuit in chronic pain. This study demonstrated that N-of-1 observational studies can be conducted with patients during a PMP using digital technologies. The use of these approaches may facilitate the application of personalized medicine.

Keywords: chronic pain, N-of-1, digital health, self-regulation, self-regulatory fatigue, motivation

OPEN ACCESS

Edited by:
Suzanne McDonald, The University of Queensland, Australia

Reviewed by:
Jingguang Li, Dali University, China
Rute Vieira, University of Aberdeen, United Kingdom

****Correspondence:***
Gail McMillan
gail.mcmillan@carleton.ca;
gail_mcmillan@hotmail.co.uk

Specialty section:
This article was submitted to Health Psychology, a section of the journal Frontiers in Psychology

Received: *01 December 2019*
Accepted: *13 August 2020*
Published: *04 September 2020*

Citation:
McMillan G and Dixon D (2020) Self-Regulatory Processes, Motivation to Conserve Resources and Activity Levels in People With Chronic Pain: A Series of Digital N-of-1 Observational Studies. Front. Psychol. 11:516485. doi: 10.3389/fpsyg.2020.516485

INTRODUCTION

Both underactivity and overactivity patterns have previously been deemed maladaptive and implicated in the maintenance and exacerbation of chronic pain (Philips, 1987; Vlaeyen and Linton, 2000, 2012; Hasenbring and Verbunt, 2010; Hasenbring et al., 2012). Underactivity, also known as pain avoidance behavior, is defined as a decrease in general daily activity and physical activity (Vlaeyen and Crombez, 1999). Overactivity, also known as persistence or endurance behavior, can increase pain and lead to long, inactive recovery periods or a "yo-yo" pattern of activity (Fordyce, 1976; Nielson et al., 2013). More recently, the utility of describing underactivity or overactivity as maladaptive has been challenged. It seems that only a small subset of people with chronic pain reduce activity levels (Bousema et al., 2007; Pincus et al., 2010; van Weering et al., 2011). Meanwhile, there is still ambiguity as to when endurance behavior can be detrimental or advantageous (Kindermans et al., 2011; Andrews et al., 2012; Hasenbring et al., 2012).

The reasons *why* individuals with chronic pain engage in different activity patterns may be more important than the patterns themselves. The interruptive nature of pain is often considered a barrier to engaging in valued activities and goals in people with chronic pain (Affleck et al., 1998; Eccleston and Crombez, 1999; Karoly and Ruehlman, 2007; Bushnell et al., 2013). However, the psychosocial, motivational and affective context of pursuing valued goals and activities must be examined (Crombez et al., 2012; Murphy, 2015; Van Damme and Kindermans, 2015). That is, the adoption of different physical activity patterns depends on the individual context of goal pursuit.

The Goal Centered, Self-regulatory, Automated, Social Systems Psychology (GRASSP) model (Karoly, 2010, 2018) is an integrative motivational model that assumes that goal pursuit in people chronic pain is accounted for by day-to-day goal-guided self-regulatory processes, neurobiological factors, and the individual psychosocial, motivational and affective context. The experience of chronic pain determines motivation, or *goal directedness*, by impacting goal-related thoughts, feelings and striving, and capacity to engage in self-regulatory efforts and strategies (Karoly, 2018). According to the GRASSP model, motivation during goal pursuit episodes is impacted by altering the value of activities and the cost-benefit analysis of engaging in activities (Karoly, 2018). Given that perceived demands of activities are considered in a cost-benefit analysis of goal pursuit, perceived demands may be directly related to motivation to conserve resources and activity levels. In addition to factors which undermine motivational and self-regulatory processes, GRASSP considers motivational buffers which facilitate goal-striving (Karoly, 2018). Most notably, self-efficacy, confidence in one's ability to complete a task (Bandura, 1977), is implicated in the allocation and conservation of resources during goal pursuit in people with chronic pain, facilitating or inhibiting goal striving (Karoly, 2018).

The capacity to engage in self-regulatory effort and strategies in people with chronic pain is affected by self-regulatory fatigue (Solberg Nes et al., 2010; Karoly, 2018). Self-regulatory fatigue is a decrease in general self-regulatory capacity, meaning self-regulation in cognitive, emotional and behavioral domains are more taxing and less effective (Solberg Nes et al., 2010, 2011). Experimental methods have demonstrated that people with chronic pain have lower self-regulatory capacity than healthy controls, resulting in poorer self-regulatory performance (Solberg Nes et al., 2010, 2011). People with chronic pain also have lower heart rate variability, a physiological indicator of lower self-regulatory capacity, compared to healthy controls (Koenig et al., 2016; Rost et al., 2017). Self-regulatory fatigue impacts motivation in people with chronic pain by increasing motivation to conserve resources (Hobfoll, 1989; Muraven et al., 2006; Eisenlohr-Moul et al., 2013). Pain intensity has a dose dependent effect on self-regulatory performance, where higher pain was associated with poorer performance (Solberg Nes et al., 2010).

An examination of the role of fluctuations in self-regulatory processes including self-regulatory fatigue, pain, self-efficacy perceived demands, and goal striving on motivation to conserve resources and activity patterns in people with chronic pain will further our understanding of mechanisms of goal pursuit. Investigating the dynamic pursuit of valued personal goals and their determinants has been identified as an important line of research for understanding the effects of pain in the broader context of living a meaningful life (Winger et al., 2019). Yet, the majority of past research with clinical samples has relied on pre-post intervention assessments with retrospective self-report, which are subject to recall and error biases (Stone et al., 2003, 2004, 2005; Stone and Broderick, 2007; Broderick et al., 2008). These approaches have not captured the dynamic nature of motivational processes of pursuing goals in daily life while living with chronic pain (Karoly, 2018; Mun et al., 2019). Self-regulation is a dynamic process, which requires dynamic measurement (Neal et al., 2017). Therefore, using methods that observe dynamic fluctuations in pain, motivation, and self-regulatory processes over time within-person are needed.

N-of-1 designs, which involve intensive longitudinal repeated measurement within an individual, are one such method of assessing within-person variability (Johnston and Johnston, 2013). These designs allow conclusions to be drawn about intraindividual variation over time which will advance the science of pain dynamics (Karoly, 2018; Mun et al., 2019). It has been recommended that N-of-1 methods are used to test theory and interventions (Craig et al., 2008; McDonald et al., 2017a; Kwasnicka and Naughton, 2020). For example, N-of-1 methods have been used to assess whether social cognitive constructs predict physical activity within individuals (Hobbs et al., 2013; O'Brien et al., 2016; McDonald et al., 2017b; Smith et al., 2019; Kwasnicka and Naughton, 2020).

The capability to evaluate the dynamic processes of pain and motivation has been facilitated tremendously by developments in digital health methodologies. The ability of text messaging, mHealth applications (apps) and wearable devices to provide precise, real-time observations of physical (e.g., pain), psychological (e.g., self-efficacy), physiological (e.g., heart rate), and exogenous (e.g., day of the week and weather) variables provides real opportunity to reduce recall biases and burden for participants (Winger et al., 2019). Thus, digital health technologies facilitate the collection of more ecologically valid data. Moreover, the use of multiple digital health technologies

simultaneously (e.g., wearable accelerometers, heart rate devices, and recording of cognitions through a smartphone) allows for a holistic bio-psychosocial approach to be taken to data collection and the subsequent design of interventions (Marceglia and Conti, 2017; Mun et al., 2019; Winger et al., 2019).

Understanding dynamic motivational processes via digital health technologies can have a direct impact on treatment for people with chronic pain. An advantage of digital health technologies is that data-driven, individual treatment plans can be easily accessed by the majority of the population at low cost (Marceglia and Conti, 2017). mHealth apps accessible to patients via their smartphones provide the opportunity for patients to self-monitor and gain insights which facilitate behavior change and self-management (Aaron et al., 2005), which is the ultimate goal of treatment for chronic pain. Real-time recording through digital health technologies also provides both patients and healthcare providers with detailed reports of progress and obstacles (Winger et al., 2019). Furthermore, when designing interventions to increase physical activity, taking a personalized approach may yield better results (Noar et al., 2007; Hobbs et al., 2013). Particularly, personalized, data-driven pacing plans in people with chronic pain may be of particular benefit (Murphy et al., 2010; Murphy, 2015).

Therefore, using a combination of a digital daily diary method and wearable accelerometer devices, the aim of the present study was to examine the effect of variation in self-regulatory process during goal pursuit. The effects of self-regulatory fatigue, goal self-efficacy, pain, goal striving and perceived demands on motivation to conserve resources, physical activity and sedentary time during daily living were examined in individuals with chronic pain. Based on between-person group-level studies, it would be expected that self-regulatory fatigue, pain and perceived demands predict motivation to conserve resources and sedentary time, while negatively predicting physical activity. It is hypothesized that goal self-efficacy and goal striving would be negatively related to motivation to conserve resources and sedentary time while being positively related to physical activity.

METHODS

Design

A series of N-of-1 observational studies were conducted for approximately 84 days (12 weeks) over the duration of a Pain Management Program (PMP). A digital daily diary method was used to measure study variables by self-report twice daily, once in the morning (between 7 am and 10 am) and again 12 h later. Therefore, there were around 168 observations in total for each participant on each variable (84 in the morning and 84 in the evening).

Participants

Participants who were due to attend a National Health Service (NHS) based PMP in Scotland were recruited by clinician referral. Inclusion criteria for this study were that patients were between the age of 18 and 65 years old, experienced chronic pain (defined as persistent pain lasting longer that 3 months), fluent in the English language, not currently experiencing acute injury and that they were due to begin the PMP within 3 months. Patients who were interested in participation were provided a letter of invitation and information about the study. Patients who expressed an interest were given a 1-week consideration period, and were then contacted and invited to participate in the study. Seven participants (six female and one male) were invited to take part. Of those seven invitees, one decided not to take part prior to the baseline meeting and one participant had to withdraw as they could not commence the PMP until after the data collection period would end. Another participant began the study but withdrew less than half-way through the PMP and a technical issue compromised their evening data collection meaning the available data could not be examined. Therefore, four participants completed the study. The study was granted ethical approval by the NHS South West-Central Bristol Research Ethics Committee (reference number: 18/SW/0076).

Measures

Baseline

Demographics

Each participant provided their age and gender. Participants were asked to describe any physical or mental health conditions they were experiencing.

Pain

Participants provided the duration of their pain (years). Current and average pain (pain over the past 6 months) intensity was rated on an 11-point Likert scale from 0 (no pain) to 10 (pain as bad as can be). Measuring current pain intensity by numerical rating scale is a valid, reliable and sensitive method of assessing present pain level (Williamson and Hoggart, 2005; Ferreira-Valente et al., 2011).

Physical functioning

Physical functioning was assessed by self-report using the PROMIS Physical Function Short Form 8a (PROMIS PF-8a). The PROMIS PF-8a (Cella et al., 2010) is an eight item measure developed from the PROMIS items bank of 124 physical functioning items which measure mobility, dexterity, movement of neck and back, and instrumental activities. The PROMIS PF-8a assesses current ability to perform basic activities of daily living. Four items on the measure (e.g., "Are you able to go up and down stairs at a normal pace"; "Are you able to run errands and shop?") are rate on a 5-point Likert scale anchored by 5 ("Without any difficulty") to 1 ("Unable to do"). Four items (e.g., "Does your health now limit you from doing 2 h of physical labor?"; "Does your health now limit you in lifting and carrying groceries?") are measured on a 5-point Likert scale anchored by 5 ("Not at all") to 1 ("Cannot do"). All items are summed and the scale provides a score range of 8–40 where higher scores indicate better physical functioning.

Self-regulatory fatigue

The Self-regulatory Fatigue Scale (Solberg Nes et al., 2013) measures self-regulation fatigue, or a reduced capacity to self-regulate, in chronic multisymptom illness (e.g., "It is easy for me to set goals"). Each item is scored on a 5-point Likert scale from strongly disagree to strongly agree. The scale measures cognitive (6 items), emotional (7 items) and behavioral (5 items)

components of self-regulatory fatigue to produce an 18-item scale with a range of 18–90 where higher scores indicate higher self-regulatory fatigue.

Pain self-efficacy

The Pain Self-efficacy Questionnaire (Nicholas, 1989) measures confidence in ability to cope despite pain in a variety of situations (e.g., "I can enjoy things, despite the pain"). It is a 10-item instrument where items are scored on a range of 0 (not at all confident) to 6 (completely confident) for a total score range of 0–60 where higher levels indicate higher pain self-efficacy.

Mood

The Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983) was designed to screen for anxiety and depression in those with illness where symptoms may be conflated (e.g., aching muscles). The Hospital Anxiety and Depression Scale has a depression subscale and an anxiety subscale with 7 items each. Each item is scored on a scale of 0 to 3 relating to the frequency that a symptom has been experienced over the past 7 days, thus each subscale has a range of 0–21.

Fear of movement

The 13-item version of the Tampa Scale of Kinesiophobia (Miller et al., 1991) is a modified version of the original Tampa Scale of Kinesiophobia (TSK) where reverse-scored items were removed. The TSK was used to assess pain-related fear of movement. The TSK assesses pain-related fear beliefs (e.g., "Pain always means I have injured my body") and fear of movement (e.g., "No one should have to exercise when he/she is in pain") on a scale from 1 (strongly disagree) to 4 (strongly agree) resulting in a scale range from 13 to 52. Higher scores indicate higher fear of movement.

Daily Activity Levels

This study measured day-to-day minutes spent being active or sedentary. All the participants wore the accelerometers on their left wrist (this was the non-dominant hand for all but participant 1). A bout of physical activity was defined as 10 consecutive minutes of physical activity of any intensity. Given the study sample (i.e., people with chronic pain), the focus of this study was on measuring physical activity that occurred in daily life. Therefore, bouts of continuous physical activity of light (101–1,951 counts/minute), moderate (1,952–5,724 counts/minute) or vigorous (>5,725 counts/minute) intensity were included in the definition of physical activity, as calculated by the Freedson algorithm (Freedson et al., 1998). Sedentary bouts were defined as consecutive minutes (≥1 min) where there is <100 counts/min (Freedson et al., 1998). Physical activity in this study was operationalized as minutes spent in physical activity bouts and sedentary time was operationalized as minutes spent in sedentary bouts. Physical activity and sedentary time were treated as continuous variables.

Daily Diary Measures

Motivation to conserve resources

Motivation to conserve resources was measured with one item ("How important was it for you to conserve energy or strength today?"). This was measured on a scale from 1 (Not at all) to 5 (Very much).

Pain

Current pain intensity was rated on an 11-point Likert scale from 0 (no pain) to 10 (pain as bad as can be).

Self-regulatory fatigue

Self-regulatory fatigue was assessed by a three-item Self-regulatory Fatigue Scale short form (SRFS-3) developed in an unpublished PhD thesis (McMillan, 2019). The behavior, cognitive, and emotion facets of self-regulatory fatigue were measured by one item each from the behavior ("I have urges to hit, throw, break, or smash things"), cognitive ("I have no trouble making decisions") and emotion subscales ("I get easily upset"). The items were measured on a 5-point Likert scale from 1 (Strongly disagree) to 5 (Strongly Agree). The item scores were summed to form a scale range from 3 to 15 where higher scores reflected higher self-regulatory fatigue.

Goal selection

Participants were presented with an item to assess which goal they would pursue each day ("Which goal is most important to you today?"). Participants could respond by selecting the goal they chose at the baseline meeting (see section "Baseline" below) or by selecting "other" and providing their daily goal response within a free-text box.

Goal self-efficacy

Goal self-efficacy was measured by up to four personalized self-efficacy items (Francis et al., 2004). The self-efficacy items were specific to the participant's individual goal. One item assessed general confidence in the ability to achieve the goal ("I am confident I can pursue my goal today") in all participants. Then, further items assessed confidence in ability to achieve the goal in the face of barriers of increasing difficulty. The barriers were also personal to each participant. Goal self-efficacy was measured with three or four items for each participant (depending on number of identified barriers) on a 5-point Likert scale from 1 (Not at all confident) to 5 (Completely confident), providing a score range of 1–20. The full list of additional goal self-efficacy items for each participant can be found in **Supplementary Table 1**.

Goal striving

Goal striving was measured with two items. One item measured goal efficiency ("How efficiently have you worked on your goal today?") and was measured on a scale from 1 (Not at all) to 5 (Very much). One item measured goal pursuit frequency ("How often did you work on your goal today?") on a scale from 1 (Not time at all) to 5 (All the time). The two items were summed to generate a score range from 1 to 10 where higher scores indicated higher goal striving.

Perceived demand

Perceived demand was measured with one item ("Overall, how demanding was your day?") on a scale from 1 (Not at all) to 5 (Very much).

Apparatus

Physical activity and sedentary time were measured using ActiGraph GT3X wearable accelerometer devices (ActiGraph GT3X; ActiGraph LLC, Pensacola, FL, United States). The GT3X collects raw tri-axial accelerometry data and takes measurements of wear time, energy expenditure, bouts of physical activity including duration and intensity of activity bout, metabolic rates, sedentary bouts, heart rate, an inclinometer which determines whether subjects are standing, sitting or lying down or if the device has been removed, and sleep activity. Accelerometers have demonstrated good reliability and validity in measuring physical activity (Eyler et al., 2003; Kelly et al., 2013).

A link to the daily diary was delivered via automated SMS text message to participants' own smartphone (except in the case of participant one who did not have a smartphone and so was provided with one). Automated text messages were sent using a bulk SMS text message provider (Voodoo, 2020). Smartphones used in the study could be either Android or iOS operating systems. The smartphones were required to have 3G or 4G capability to ensure the diary could be completed without interruption. The text message provided a prompt to complete the diary.

Procedure

Baseline

A brief semi-structured interview was conducted with each participant to illicit their valued activities, and to identify a goal and barriers, which were used to construct the personalized self-efficacy items. These interviews were conducted at the PMP (participant 1), at the University of Strathclyde (participant 2), in a public place chosen by the participant (participant 3) and at the participant's home (participant 4). Participants then completed the baseline measures and were given a demonstration of how they would receive the daily diary and how to complete it. To reduce participant burden, measures of fear of movement and mood were not recorded by the researcher at the initial meeting as they were recorded at the first session of the pain management program by clinicians.

Pain Management Program

The PMP was delivered within a Scottish NHS secondary care setting by a multidisciplinary team (e.g., clinical psychologist, specialist nurse, and physiotherapist). The program was a weekly group intervention based on Acceptance and Commitment therapy (ACT) principles and included pain education, physiotherapy, pacing, acceptance, and mindfulness strategies as well as commitment to values and behavior change. Each participant engaged in the pain management program, which lasted either 10 or 12 weeks regardless of their participation in the research study.

Daily Diary Phase

The participants were provided the opportunity to complete the daily diary from the day following the baseline meeting, which was up to 1 week prior to the first day of the PMP. Completion of diary entries prior to the commencement of the PMP was to allow participants to get accustomed to the procedure, and so were not included in the analysis. The daily diary was completed online on the Qualtrics platform. A link to the diary was sent via a text message to participants' smartphones at the agreed morning time. The morning diary included measures of pain intensity, goal identification, self-regulatory fatigue, goal striving, and goal self-efficacy. The evening diary, which was prompted by text message 12 h after the morning diary measured pain intensity, self-regulatory fatigue, perceived demand, and motivation to conserve resources. Additional morning and evening diary variables measured included mood, goal motivation, expected demand, expected progress and expected fatigue but these are not examined in this study. Every 2 weeks after beginning the diary phase, a face-to-face meeting was conducted at the site of delivery of the PMP to discuss any issues with the study, to ensure continued consent to participate, and to provide them with a new fully charged accelerometer. Participants were also encouraged to contact the researcher if any problems arose throughout the diary phase. After the diary phase was complete, a final face-to-face meeting was arranged to debrief the participant and provide the remuneration (£50 GBP) for their participation.

DATA ANALYSIS

Data Processing

Raw data were downloaded from the accelerometers and participants' data files from each accelerometer were combined into one file for each participant. The downloaded raw data files were processed into epochs of 10 s using ActiLife software v6.13.3. Wear-time validation was conducted and a non-wear period was defined as 60 consecutive minutes of no activity using ActiLife software (Troiano et al., 2008). Bouts of physical activity and sedentary bouts were calculated by ActiLife software.

Statistical Analysis

The data were analyzed individually for each participant using R statistical software v3.4.4. Missingness maps were produced for each participant using the AMELIA II package v1.7.5 (Honaker et al., 2011). Missingness maps were inspected visually to determine patterns of missingness. Where there was a very small number of daily diary observations missing at random (e.g., $\leq$0.05%), the mean of prior and subsequent observations was input. Otherwise, missing data was handled with multiple imputation using the AMELIA II package. The AMELIA II package uses an expectation-maximization bootstrapping (EMB) algorithm to model missing observations, specifically designed for time series data (Honaker et al., 2011). Five imputed datasets were produced where missing observations were imputed. As a bout of physical activity is defined as continuous movement for 10 min, imputed values $<$10 on physical activity were recoded to 0. All analysis was conducted on each of the five datasets and statistic estimates were calculated by pooling the results from each imputed dataset using Rubin's rules (Rubin, 1996). Using Rubin's rules to calculate parameter estimates accounts for the within and between variance of the combined results and calculating estimates with this method provides 95% confidence in inference when using multiply imputed datasets.

Time plots were examined for trends in the data. Autocorrelation, the correlation between a variable at the current time-point in a time series (t_0) and the same variable at earlier time points or lags (e.g., where t-$_1$ denotes one observation previous and t-$_2$ denotes two observations previous), can arise when there are many repeated measurements of the same variables. Autocorrelograms were assessed for each variable to determine whether autocorrelation was present (Naughton and Johnston, 2014). Dynamic regression modeling was conducted to examine the relationship between the predictor variables and motivation to conserve resources, physical activity, and sedentary time. Using dynamic models to analyze N-of-1 data has been recommended because it is a flexible modeling approach (Vieira et al., 2017). Dynamic regressions can account for autocorrelation by including lags of the predictors and outcome variables as well as exogenous variables including trends in time and periodicity (e.g., morning and evening). Including lagged variables in the model which represent autocorrelation allows for independence between data points to be assumed. Dynamic regression models will not be formally described here as this has been done previously (Vieira et al., 2017).

Descriptive and multivariate analysis was conducted. As the purpose was to determine which variables had the most impact on motivation to conserve resources, physical activity and sedentary time, a stepwise approach was used to ascertain the model with the best model fit as determined by Akaike's Information Criterion. Based on examination of the time plots and autocorrelograms, lags of the outcome variables, week, and weekday (i.e., whether it was a workday or weekend) were included as control variables as needed prior to the inclusion of predictor variables. The model residuals were then assessed for normality using a histogram and Q–Q plots and autocorrelation using autocorrelation function (ACF) plots and partial autocorrelation function (PACF) plots.

RESULTS

Participant Characteristics

The participants' demographic information, description of physical health condition(s) and baseline recordings of pain, self-regulatory fatigue, pain self-efficacy, fear of movement, mood and personal goal are shown in **Table 1**. Questionnaire scores for fear of movement, anxiety and depression for participant three are missing as this was not recorded at the first PMP session. Additional goals pursued by participants over the course of the study can be found in **Supplementary Table 2**. It should be noted that all participants chose a goal related to improving their emotional or social wellbeing.

Descriptive Statistics

Compliance with diary completion was very high. Participants 2 and 3 completed 100% of diary entries and there were no missing observations. Participant 1 had one diary entry missing, meaning there was 0.006% of possible occasions and 0.05% of observations missing. Given the small amount of missing observations within the dataset for participant 1, the mean of the preceding and subsequent observations was inputted. Participant 3 had 1% observations missing as there was a technical issue with the accelerometer for the last 6 days of measurement. Participant 4 completed the diary on 97.5% of possible occasions and, overall, 4% of observations were missing. Evening observations were more likely to be missing than morning observations for participant 4. Therefore, multiple imputation was undertaken in participant 3 and 4's data to provide full datasets. The results for participants 3 and 4, reported below, are the product of pooled estimates from five imputed datasets. Time plots of motivation to conserve resources, physical activity, sedentary time, and the predictor variables are shown in **Figure 1**.

Figure 1 illustrates that there is evidence of variance across participants and within participants over time on all variables. There may have been ceiling effects for participants 3 and 4 on goal striving and for participant 4 on self-efficacy. The means and standard deviations for physical activity, sedentary time, motivation to conserve resources, pain, self-regulatory fatigue, goal self-efficacy, goal striving and perceived demand for each participant are displayed in **Table 2**.

Dynamic regression models were conducted individually for each participant. Within each model, the reference measurement (t_0) is either current morning or evening, depending on when the variable was measured. Pain intensity and self-regulatory fatigue were measured in both morning and evening diaries. The time of day of measurement is indicated in **Table 3**. Lag 1 (t-$_1$) is the observation prior to t_0, while lag 2 (t-$_2$) is the observation prior to t-$_1$. For example, lag 1 of variables measured in the evening (e.g., perceived demand) refers to the previous evening, while lag 1 of variables measured in the morning (e.g., goal self-efficacy) refers to the previous morning. We used autocorrelograms with ACF and PACF to guide the selection of the number of lags for predictors. It was unusual for there to be significant autocorrelation beyond lag 2. However, when significant autocorrelation of earlier lags (lag 3 onward) appeared to be present in autocorrelograms, these lags were included in models. When more recent lags were also accounted for within models (e.g., lag 0, lag 1, and lag 2), there was no effect of earlier lags (e.g., lag 3).

Dynamic Regression Modeling Results

An overview of individual dynamic regression models of the effect of the pain, self-regulatory fatigue, goal self-efficacy, goal striving, and perceived demand on motivation to conserve resources, physical activity and sedentary time is displayed in **Table 3**.

In participant 1, the small positive association between week and motivation to conserve resources suggests that motivation to conserve resources increased slightly across the course of the study. In participant 1, motivation to conserve resources was higher on days when perceived demands, evening pain intensity, and previous morning self-regulatory fatigue were higher, and goal striving was lower. For participant 3, motivation to conserve resources was higher on days when goal striving, perceived demands, and evening pain intensity were lower, and evening self-regulatory fatigue was higher. In participant 4, motivation to conserve resources was higher on days when perceived demands

TABLE 1 | Baseline descriptive information for each participant.

	Participant 1	Participant 2	Participant 3	Participant 4
Age	48	41	50	59
Gender	Female	Male	Female	Female
Pain condition(s)	Neck, shoulder, and lower back pain	Arthritis, trapped nerve in neck, and diabetic neuropathy	Persistent pain	Osteoarthritis and polymyalgia rheumatica
Comorbid condition(s)	–	Diabetes type 1, retinopathy, nephropathy, high blood pressure, and angina	Suspected spastic paraplegia	Post viral depression
Pain duration	2–5 years	10–20 years	10–20 years	1–2 years
Current Pain intensity	5	9	6	5
Average pain intensity	7	8	10	8
Physical functioning	27	18	10	13
Self-regulatory fatigue	49	67	43	68
Pain self-efficacy	33	22	10	19
Fear of movement	27	19	–	36
Anxiety	6	16	–	8
Depression	11	12	–	9
Goal	Enjoy activities more	Manage emotions when unexpected setbacks arise	Improved management and maintenance of relationships	Feeling more confidence in managing pain
PMP length	10 weeks	12 weeks	12 weeks	10 weeks

Scale ranges are as follows: current pain = 0–10; average pain intensity = 0–10; physical functioning = 8–40; self-regulatory fatigue = 18–90; pain self-efficacy = 0–60; fear of movement = 13–52; anxiety = 0–21; depression = 0–21.

and previous days' motivation to conserve resources were higher, and goal striving, and goal self-efficacy were lower.

Physical activity was higher for participant 1 on the weekends and on days when perceived demands from 2 days' previous were higher. Physical activity was higher for participant 2 on days when physical activity was higher the previous day. In participant 4, physical activity was higher on days when perceived demands were higher.

Sedentary time was higher for participant 1 on weekdays, when there was higher morning self-regulatory fatigue 2 days' previously, and when the previous days' perceived demands were lower. In participant 2, sedentary time was higher on days when motivation to conserve resources was higher, and goal striving and evening pain were lower. In participant 3, sedentary time was higher on days when sedentary time was higher the previous day and when motivation to conserve resources was higher. Sedentary time was higher for participant 4 on days when previous days' perceived demand and pervious days' goal striving were lower.

DISCUSSION

Main Findings

The purpose of this study was to examine the interindividual motivational dynamics involved in motivation to conserve resources and activity levels over time in people with chronic pain. In line with the GRASSP model, the associations between the outcomes and goal-related and self-regulatory variables were unique across individuals. Goal striving and perceived demand were most frequently associated with outcomes across participants. Goal striving was related to less motivation to conserve resources (participants 1, 3, and 4) and less sedentary time (participants 2 and 4). Perceived demands were associated with higher motivation to conserve resources and physical activity, and lower sedentary time in two participants (1 and 4). Perceived demands were also associated with less motivation to conserve resources in another participant (participant 3). Higher self-regulatory fatigue predicted higher motivation to conserve resources (participants 1 and 3) and sedentary time (participant 1). Evening pain intensity was related to motivation to conserve resources, but in opposing directions (participants 1 and 3), and also to higher sedentary time (participant 2). The direction of the relationship between motivation to conserve resources and sedentary time was in opposing directions for two participants (2 and 3). Finally, goal self-efficacy was negatively associated with motivation to conserve resources in one participant (participant 4).

Relationship to Past Research

The findings of this study are generally supportive of motivational accounts of activity patterns in people with chronic pain (Van Damme, 2014; Van Damme and Kindermans, 2015; Karoly, 2018) and previous research demonstrating that the context of a goal pursuit episode is associated with activity patterns (Karsdorp et al., 2010; Karsdorp and Vlaeyen, 2011; Schrooten

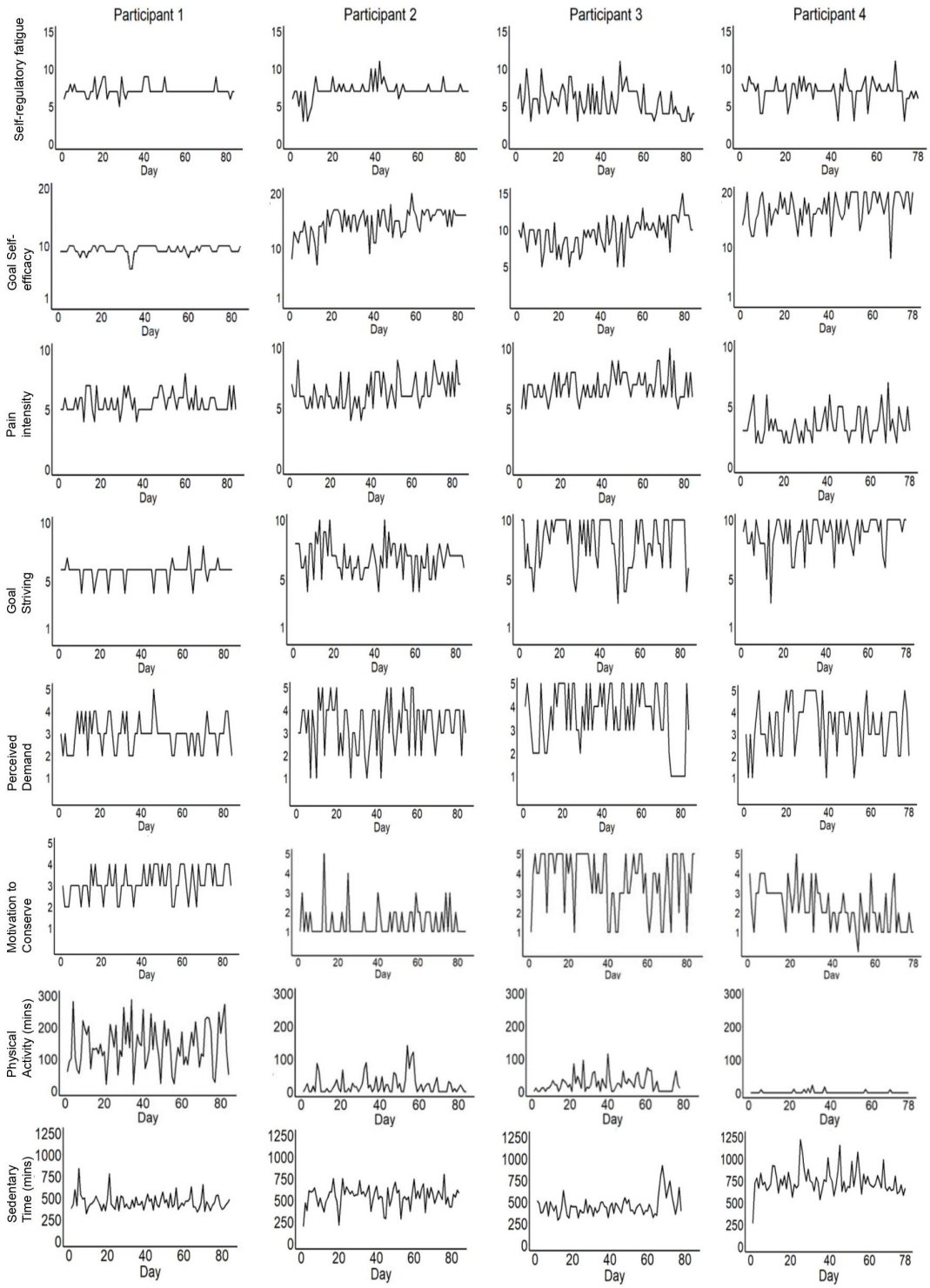


FIGURE 1 | Time plots for all variables over time for each participant.

et al., 2012; Van Damme et al., 2012; Pastor-Mira et al., 2019). The most consistent determinants of motivation to conserve resources and sedentary time in this study were goal striving and perceived demands. Perceived demands were also the only determinant of physical activity. The findings of this study are also partially in line with theory and past research asserting that

TABLE 2 | Descriptive statistics of daily assessment of all study variables.

	Participant 1	Participant 2	Participant 3	Participant 4
Variable	***M* (SD)**	***M* (SD)**	***M* (SD)**	***M* (SD)**
Motivation to conserve	3.1 (0.7)	1.4 (0.8)	3.7 (1.4)	2.3 (1.0)
Physical activity (mins)	141.1 (66.4)	20.7 (29.1)	24.1 (24.4)	1.4 (4.4)
Sedentary time (mins)	453.2 (89.6)	541.3 (114.9)	455.6 (113.0))	744.2 (143.4)
Pain	5.7 (0.9)	6.8 (1.2)	7.3 (1.1)	3.7 (1.3)
Self-regulatory fatigue	7.1 (0.7)	7.5 (1.2)	7.6 (1.7)	7.1 (1.5)
Goal self-efficacy	9.3 (0.8)	14.7 (2.4)	9.5 (2.1)	17.2 (2.7)
Goal striving	5.9 (0.7)	6.8 (1.3)	8.2 (2.0)	8.9 (1.4)
Perceived demand	3.0 (0.7)	3.3 (1.1)	3.6 (1.3)	3.5 (1.1)

The possible goal self-efficacy score ranged from 1 to 15 for participant 3 and 1 to 20 for participants 1, 2, and 4.

TABLE 3 | Multivariate associations between predictor variables and outcomes in all participants.

	Participant			
Predictors	**1**	**2**	**3**	**4**
	Motivation to conserve resources			
Week	0.06***			
Weekday				
SRF (morn)	0.16* (lag 1)			
Goal striving (morn)	−0.25** (lag 0)		−0.24*** (lag 0)	−0.24*** (lag 0)
Goal self-efficacy (morn)				−0.08* (lag 0)
SRF (even)			0.31*** (lag 0)	
Pain (even)	0.18** (lag 0)		−0.39* (lag 0)	
Perceived demand (even)	0.22* (lag 0)		−0.24* (lag 0)	0.37*** (lag 0)
MCR (even)				0.24** (lag 1)
	Physical activity			
Week				
Weekday	88.61***			
Physical activity		0.43*** (lag 1)		
SRF (morn)				
Goal striving (morn)				
Goal self-efficacy (morn)				
SRF (even)				
Pain (even)				
Perceived demand (even)	24.65** (lag 2)			1.38*** (lag 0)
MCR (even)				
	Sedentary Time			
Week				
Weekday	−57.55**			
Sedentary time			0.51*** (lag 1)	
SRF (morn)	26.16* (lag 2)			
Goal striving (morn)		−22.49** (lag 0)		−34.27*** (lag 1)
Goal self-efficacy (morn)				
SRF (even)				
Pain (even)		−37.16*** (lag 0)		
Perceived demand (even)	−40.35** (lag 1)			−35.63*** (lag 1)
MCR (even)		31.05* (lag 0)	−18.09*** (lag 0)	

*MCR, motivation to conserve resources; SRF, self-regulatory fatigue; morn, morning; even, evening. Lag 1 of morning variables refers to previous morning; lag 1 of evening variables refers to previous evening, etc. *p < 0.05, **p < 0.01, and ***p < 0.001.*

people with chronic pain experience self-regulatory fatigue which negatively impacts self-regulatory performance by increasing motivation to conserve resources (Solberg Nes et al., 2010, 2011; Eisenlohr-Moul et al., 2013; Vervoort and Trost, 2016; Rost et al., 2017). In turn, motivation to conserve resources was related to sedentary time. Taken together, these findings suggest that there

is a continuing evaluation of the costs and benefits of pursuing valued goals (Karoly, 2018; Van Damme et al., 2018).

However, no predictor variables were consistently associated with the outcomes across all participants. The direction of some observed relationships were contrary to expectations given previous research and theory (Karsdorp et al., 2010; Van Damme et al., 2012, 2018; Van Damme, 2014; Karoly, 2018). For example, while one participant reported higher motivation to conserve resources on days with higher pain and perceived demands (participant 1), the opposite associations were reported in another participant (participant 3). Further, there was a negative relationship between motivation to conserve resources and sedentary time in participant 3. The differences in the direction of relationships in this study are likely accounted for by whether physical activity levels were maintained despite pain, increased demands, and motivation to conserve resources (participant 1), or whether physical activity decreased due to motivation to conserve resources, meaning lower perceived demands and pain in the evening (participant 3). Meanwhile, goal self-efficacy was related to motivation to conserve resources in one participant, but it was generally not predictive of outcomes. This contrasts with past research demonstrating that self-efficacy predicts engagement in physical activity from groups-based studies (McAuley et al., 2011; Huffman et al., 2015).

Past evidence of the effect of self-regulatory fatigue and goal pursuit in people with chronic pain have often used experimental methods and retrospective self-report questionnaires and the average of group-aggregated data. The aggregation of group data can mask the direction of relationships within individuals (Johnston and Johnston, 2013; Yeo and Neal, 2013; McDonald et al., 2017a) and cannot account for the dynamic nature of self-regulatory processes. In addition, this study used objective measurement of physical activity and sedentary time with accelerometers as opposed to self-report measures. Self-report frequently results in biased estimation in people with chronic pain (Gosney et al., 2007; van Weering et al., 2011; Liu et al., 2014; Schaller et al., 2016), and for measurement at the individual level (Loney et al., 2011). The differences in the patterns of relationships observed in this study highlights the need to consider the individual goal-guided context (Karoly, 2018; Mun et al., 2019). The theory and methods used in this study have illustrated the heterogeneity in determinants of motivation to conserve resources and activity levels in people with chronic pain.

Strengths and Limitations

Unlike most past research, the methods used in this study accounted for the dynamic nature of self-regulatory processes. Another strength of this study was the use of wearable accelerometer devices in conjunction with digital daily diaries. Previous diary studies in people with chronic pain have tended to assess either physical activity or the pursuit of personal goals but there is a lack of integration of both types of data (Van Damme, 2014). Additionally, using smartphones enabled participants to complete their dairy immediately after receiving the text message with the link to the diary. Studies which use paper-and pencil diaries can suffer from poor adherence and falsification of data and it is difficult to ascertain reliably the time at which they were completed (Stone et al., 2003; Broderick et al., 2008). Within this study, adherence was very high (the participant with the lowest adherence completed 96% of diary occasions). Furthermore, the use of N-of-1 observational methods and dynamic regression modeling allowed for models to be estimated for individuals over time while accounting for time trends and autocorrelation, thus reducing potentially biased estimates (Vieira et al., 2017).

Some limitations of the study should be noted. First, the pattern of relationships between the predictors and outcomes are unique to the individual participants and so different patterns may be observed in the future. Additionally, this study examined a limited number of goal related predictors and other self-regulatory, cognitive or affective processes may predict the outcomes in this population. The study measured some self-report variables retrospectively (e.g., motivation to conserve resources and perceived demand). As the nature of the study involved repeated measures within individuals, as opposed to measurement of a group, the reliability and validity of the self-report items used in this study is unknown. Preliminary data on the three item self-regulatory fatigue measure indicated that construct validity was acceptable but internal consistency was not satisfactory due to the low number of items while attempting to preserve the measurement of each subscale (McMillan, 2019). However, this data was from a student sample, not a sample of people with chronic pain, which may have affected interpretation of the items (Bonetti et al., 2001). We acknowledge that the unknown validity of single item, self-report measures and the three item self-regulatory fatigue measure is a limitation of this study. That said, it is the case that the longitudinal nature of data collection required a balance between the number of items and the need to reduce participant burden and the potential amount of missing data and participant retention in the study.

Additionally, while the purpose of the study was to examine factors which may affect activity levels during goal pursuit, goals chosen by participants were emotion regulation goals, not physical activity goals, and progress toward goal achievement was not measured in this study. Measuring goal progress may have provided further useful information about self-regulatory mechanisms. Future research, which uses ambulatory methods to measure the variables "in the moment" may be useful and provide more reliable estimate of relationships, as opposed to using retrospective items (Bentley et al., 2019).

Implications for Methodology and Clinical Practice

Identifying the individual determinants of fluctuations in pain, motivation and self-regulatory processes will provide insight to people with chronic pain to enable them to manage to better their own condition and ultimately to pursue meaningful personal goals. Currently, psychological treatment programs evaluate whether the mean scores of psychosocial functioning indices have changed in the desired direction for groups of patients from pre to post intervention. For some patients, controlling

fluctuations and reducing variability in pain and motivation may have a more significant impact on quality of life than changes from baseline scores (Mun et al., 2019; Winger et al., 2019). The effect of fluctuations or variability in pain and motivation are rarely assessed within treatment programs. This study has demonstrated that N-of-1 observational studies using accelerometers and digital daily diaries, where a link is delivered by text message, can be implemented with patients engaged in a pain management program. Further, it has been argued that N-of-1 trial designs could become the "gold standard" for assessing treatment efficacy (Bradbury et al., 2020) and could also be used to examine changes in variability in pain and motivation from pre to post intervention (Mun et al., 2019; Winger et al., 2019).

Evidence that fluctuations in self-regulatory fatigue, self-efficacy, pain, goal striving and perceived demands have differential effects on motivation to conserve resources, physical activity and sedentary time suggests that people with chronic pain would benefit from individualized treatment plans targeting motivational processes that affect the pursuit of their valued goals. For example, Acceptance and Commitment Therapy focuses on acceptance and mindfulness strategies as well as commitment to values and behavior change. Acceptance and mindfulness may increase self-regulatory capacity (Azam et al., 2016) while commitment to values may affect the cost-benefit analysis in undertaking activities. For the participants in this study, increasing self-regulatory capacity and goal striving, and decreasing the perceived demands of activities may result in more effective goal-directedness (Karoly, 2018).

It has been suggested that individually tailored activity pacing which takes into account the psychosocial context of activity, such as motivation for engagement in activity, is needed (Murphy, 2015; Mun et al., 2019). Data-driven tailored interventions to facilitate physical activity have been conducted previously with action planning and control cognitions in people with osteoarthritis (O'Brien et al., 2016). Further, a tailored, data driven activity pacing intervention which used accelerometer data reduced fatigue interference in those with osteoarthritis (Murphy et al., 2010).

The use of some digital health technologies and software can be expensive (e.g., Ecological Momentary Assessment platforms where cost for use of software, data storage on remote servers and cost per signal can be high), making it less accessible for some researchers. This study used a low-cost and easily implemented method of sending automated text messages using a bulk SMS provider, and the text messages included a link to the online digital diary. There are also free applications providing automated SMS schedulers which can be accessed from the Google Play Store and the Apple Store.

Conclusion

This study demonstrated that the effect of self-regulatory fatigue, goal self-efficacy, goal striving and perceived demands on motivation to conserve resources, physical activity and sedentary time varied across participants. The observed relationships generally supported the GRASSP model which suggests that activity patterns in people with chronic pain can be accounted for by goal guided self-regulatory processes. This study illustrated that N-of-1 observational studies with digital health technologies can be conducted during pain management programs at low cost. The results from this study support the need for further research on within-individual variability of goal processes, the development of measures to support these research designs, and the development of individually tailored activity pacing interventions.

DATA AVAILABILITY STATEMENT

The datasets used within this study will be made available upon request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NHS South West-Central Bristol Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GM and DD contributed to the conception, design of the study, contributed to the manuscript revision, and read and approved the submitted version. GM recruited participants, collected data, conducted statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

GM was supported by a Ph.D. scholarship from the University of Strathclyde.

ACKNOWLEDGMENTS

We would like to thank Dr. Martin Dunbar for providing advice regarding study procedure and recruitment. We thank the pain management team for their help during the data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.516485/full#supplementary-material

REFERENCES

Aaron, L. A., Turner, J. A., Mancl, L., Brister, H., and Sawchuk, C. N. (2005). Electronic diary assessment of pain-related variables: is reactivity a problem? *J. Pain* 6, 107–115. doi: 10.1016/j.jpain.2004.11.003

Affleck, G., Tennen, H., Urrows, S., Higgins, P., Abeles, M., Hall, C., et al. (1998). Fibromyalgia and women's pursuit of personal goals: a daily process analysis. *Health Psychol.* 17, 40–47. doi: 10.1037/0278-6133.17.1.40

Andrews, N. E., Strong, J., and Meredith, P. J. (2012). Activity pacing, avoidance, endurance, and associations with patient functioning in chronic pain: a systematic review and meta-analysis. *Arch. Phys. Med. Rehabil.* 93, 2109-2121.e7. doi: 10.1016/j.apmr.2012.05.029

Azam, M. A., Katz, J., Mohabir, V., and Ritvo, P. (2016). Individuals with tension and migraine headaches exhibit increased heart rate variability during post-stress mindfulness meditation practice but a decrease during a post-stress control condition – A randomized, controlled experiment. *Int. J. Psychophysiol.* 110, 66–74. doi: 10.1016/j.ijpsycho.2016.10.011

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bentley, K. H., Kleiman, E. M., Elliott, G., Huffman, J. C., and Nock, M. K. (2019). Real-time monitoring technology in single-case experimental design research: opportunities and challenges. *Behav. Res. Ther.* 117, 87–96. doi: 10.1016/j.brat.2018.11.017

Bonetti, D., Johnston, M., Rodriguez-marin, J., Pastor, M., Martin-aragon, M., Doherty, E., et al. (2001). Dimensions of perceived control: a factor analysis of three measures and an examination of their relation to activity level and mood in a student and cross-cultural patient sample. *Psychol. Health* 16, 655–674. doi: 10.1080/08870440108405865

Bousema, E. J., Verbunt, J. A., Seelen, H. A. M., Vlaeyen, J. W. S., and Knottnerus, A. J. (2007). Disuse and physical deconditioning in the first year after the onset of back pain. *PAIN* 130, 279–286. doi: 10.1016/j.pain.2007.03.024

Bradbury, J., Avila, C., and Grace, S. (2020). Practice-based research in complementary medicine: could N-of-1 trials become the new gold standard? *Healthcare* 8:15. doi: 10.3390/healthcare8010015

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., and Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *PAIN* 139, 146–157. doi: 10.1016/j.pain.2008.03.024

Bushnell, M. C., Ceko, M., and Low, L. A. (2013). Cognitive and emotional control of pain and its disruption in chronic pain. *Nat. Rev. Neurosci.* 14, 502–511. doi: 10.1038/nrn3516

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J. Clin. Epidemiol.* 63, 1179–1194. doi: 10.1016/j.jclinepi.2010.04.011

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., et al. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337:a1655. doi: 10.1136/bmj.a1655

Crombez, G., Eccleston, C., Van Damme, S., Vlaeyen, J. W. S., and Karoly, P. (2012). Fear-avoidance model of chronic pain: the next generation. *Clin. J. Pain* 28, 475–483. doi: 10.1097/ajp.0b013e3182385392

Eccleston, C., and Crombez, G. (1999). Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol. Bull.* 125, 356–366. doi: 10.1037/0033-2909.125.3.356

Eisenlohr-Moul, T. A., Burris, J. L., and Evans, D. R. (2013). Pain acceptance, psychological functioning, and self-regulatory fatigue in temporomandibular disorder. *Health Psychol.* 32, 1236–1239. doi: 10.1037/a0030150

Eyler, A. A., Brownson, R. C., Bacak, S. J., and Housemann, R. A. (2003). The epidemiology of walking for physical activity in the United States. *Med. Sci. Sports Exerc.* 35, 1529–1536. doi: 10.1249/01.MSS.0000084622.39122.0C

Ferreira-Valente, M. A., Pais-Ribeiro, J. L., and Jensen, M. P. (2011). "Psychosocial factors as predictors of depression and anxiety in patients with chronic pain," in *Proceedings of the Conference, 7 Congress of the European Federation of Pain Chapters: Pain in Europe VII, EFIC*, Vol. 5, (Hamburg Germany: Conference Publication), 72. doi: 10.1016/s1754-3207(11)70244-9

Fordyce, W. E. (1976). *Behavioural Models for Chronic Pain and Illness*. St. Louis, MO: Mosby.

Francis, J. J., Eccles, M. P., Johnston, M., Walker, A. E., Grimshaw, J. M., Foy, R., et al. (2004). *Constructing Questionnaires Based On The Theory Of Planned Behaviour: A Manual For Health Services Researchers*. Woolloongabba QLD: Centre for Health Services Research.

Freedson, P. S., Melanson, E., and Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Med. Sci. Sports Exerc.* 30, 777–781. doi: 10.1097/00005768-199805000-00021

Gosney, J. L., Scott, J. A., Snook, E. M., and Motl, R. W. (2007). Physical activity and multiple sclerosis: validity of self-report and objective measures. *Fam. Community Health* 30, 144–150. doi: 10.1097/01.FCH.0000264411.20766.0c

Hasenbring, M. I., Hallner, D., Klasen, B., Streitlein-Bohme, I., Willburger, R., and Rusche, H. (2012). Pain-related avoidance versus endurance in primary care patients with subacute back pain: psychological characteristics and outcome at a 6-month follow-up. *PAIN* 153, 211–217. doi: 10.1016/j.pain.2011.10.019

Hasenbring, M. I., and Verbunt, J. A. (2010). Fear-avoidance and endurance-related responses to pain: new models of behavior and their consequences for clinical practice. *Clin. J. Pain* 26, 747–753. doi: 10.1097/AJP.0b013e3181e104f2

Hobbs, N., Dixon, D., Johnston, M., and Howie, K. (2013). Can the theory of planned behaviour predict the physical activity behaviour of individuals? *Psychol. Health* 28, 234–249. doi: 10.1080/08870446.2012.716838

Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *Am. Psychol.* 44, 513–524. doi: 10.1037/0003-066X.44.3.513

Honaker, J., King, G., and Blackwell, M. (2011). Amelia II: a program for missing data. *J. Stat. Softw.* 45, 1–47.

Huffman, K. M., Pieper, C. F., Hall, K. S., St Clair, E. W., and Kraus, W. E. (2015). Self-efficacy for exercise, more than disease-related factors, is associated with objectively assessed exercise time and sedentary behaviour in rheumatoid arthritis. *Scand. J. Rheumatol.* 44, 106–110. doi: 10.3109/03009742.2014.931456

Johnston, D. W., and Johnston, M. (2013). Useful theories should apply to individuals. *Br. J. Health Psychol.* 18, 469–473. doi: 10.1111/bjhp.12049

Karoly, P. (2010). "Goal systems and self-regulation," in *Handbook of Personality and Self-Regulation*, ed. R. H. Hoyle (Oxford: Wiley-Blackwell), 218–242. doi: 10.1002/9781444318111.ch10

Karoly, P. (2018). "A goal-centered, self-regulatory model of motivation and its relevance for advancing the study of chronic pain," in *Motivational Perspectives on Chronic Pain: Theory, Research, and Practice*, eds P. Karoly and G. Crombez (Oxford: Oxford University Press), 3–61.

Karoly, P., and Ruehlman, L. S. (2007). Psychosocial aspects of pain-related life task interference: an exploratory analysis in a general population sample. *Pain Med.* 8, 563–572. doi: 10.1111/j.1526-4637.2006.00230.x

Karsdorp, P. A., Nijst, S. E., Goossens, M. E. J. B., and Vlaeyen, J. W. S. (2010). The role of current mood and stop rules on physical task performance: an experimental investigation in patients with work-related upper extremity pain. *Eur. J. Pain* 14, 434–440. doi: 10.1016/j.ejpain.2009.07.003

Karsdorp, P. A., and Vlaeyen, J. W. S. (2011). Goals matter: both achievement and pain-avoidance goals are associated with pain severity and disability in patients with low back and upper extremity pain. *PAIN* 152, 1382–1390. doi: 10.1016/j.pain.2011.02.018

Kelly, L. A., McMillan, D. G., Anderson, A., Fippinger, M., Fillerup, G., and Rider, J. (2013). Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. *BMC Med. Phys.* 13:5. doi: 10.1186/1756-6649-13-5

Kindermans, H. P., Roelofs, J., Goossens, M. E., Huijnen, I. P. J., Verbunt, J. A., and Vlaeyen, J. W. (2011). Activity patterns in chronic pain: underlying dimensions and associations with disability and depressed mood. *J. Pain* 12, 1049–1058. doi: 10.1016/j.jpain.2011.04.009

Koenig, J., Loerbroks, A., Jarczok, M. N., Fischer, J. E., and Thayer, J. F. (2016). Chronic pain and heart rate variability in a cross-sectional occupational sample evidence for impaired vagal control. *Clin. J. Pain* 32, 218–225. doi: 10.1097/AJP.0000000000000242

Kwasnicka, D., and Naughton, F. (2020). N-of-1 methods: a practical guide to exploring trajectories of behaviour change and designing precision

behaviour change interventions. *Psychol. Sport Exerc.* 47:101570. doi: 10.1016/j.psychsport.2019.101570

Liu, X., Liu, Y., Li, L., Hu, Y., Wu, S., and Yao, S. (2014). Overgeneral autobiographical memory in patients with chronic pain. *Pain Med.* 15, 432–439. doi: 10.1111/pme.12355

Loney, T., Standage, M., Thompson, D., Sebire, S. J., and Cumming, S. (2011). Self-report vs. objectively assessed physical activity: which is right for public health? *J. Phys. Act. Health* 8, 62–70. doi: 10.1123/jpah.8.1.62

Marceglia, S., and Conti, C. (2017). A technology ecosystem for chronic pain: promises, challenges, and future research. *mHealth* 3, 6–6. doi: 10.21037/mhealth.2017.02.03

McAuley, E., Szabo, A., Gothe, N., and Olson, E. A. (2011). Self-Efficacy: implications for physical activity, function, and functional limitations in older adults. *Am. J. Lifestyle Med.* 5, 361–369. doi: 10.1177/1559827610392704

McDonald, S., Quinn, F., Vieira, R., O'Brien, N., White, M., Johnston, D. W., et al. (2017a). The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol. Rev.* 11, 307–323. doi: 10.1080/17437199.2017.1316672

McDonald, S., Vieira, R., Godfrey, A., O'Brien, N., White, M., and Sniehotta, F. F. (2017b). Changes in physical activity during the retirement transition: a series of novel n-of-1 natural experiments. *Int. J. Behav. Nutr. Phys. Act.* 14:167. doi: 10.1186/s12966-017-0623-7

McMillan, G. E. (2019). *An Examination of the Role of Self-Regulatory Fatigue in Adaption to Chronic Pain*. Unpublished Doctoral Dissertation, Unversity of Strathclyde, Glasgow.

Miller, R. P., Kori, S. H., and Todd, D. D. (1991). The Tampa Scale: a Measure of Kinisophobia. *Clin. J. Pain* 7:51. doi: 10.1097/00002508-199103000-00053

Mun, C. J., Suk, H. W., Davis, M. C., Karoly, P., Finan, P., Tennen, H., et al. (2019). Investigating intraindividual pain variability: methods, applications, issues, and directions. *PAIN* 160, 2415–2429. doi: 10.1097/j.pain.0000000000001626

Muraven, M., Shmueli, D., and Burkley, E. (2006). Conserving self-control strength. *J. Pers. Soc. Psychol.* 91, 524–537. doi: 10.1037/0022-3514.91.3.524

Murphy, S. L. (2015). Overactivity in daily life: a crucial step in understanding how to tailor treatments. *PAIN* 156, 1831–1832. doi: 10.1097/j.pain.0000000000000291

Murphy, S. L., Lyden, A. K., Smith, D. M., Dong, Q., and Koliba, J. F. (2010). Effects of a tailored activity pacing intervention on pain and fatigue for adults with osteoarthritis. *Am. J. Occup. Ther.* 64, 869–876. doi: 10.5014/ajot.2010.09198

Naughton, F., and Johnston, D. W. (2014). A starter kit for undertaking n-of-1 trials. *Eur. Health Psychol.* 16, 196–205.

Neal, A., Ballard, T., and Vancouver, J. B. (2017). Dynamic Self-Regulation and Multiple-Goal Pursuit. *Annu. Rev. Organ. Psychol. Organ. Behav.* 4, 401–423. doi: 10.1146/annurev-orgpsych-032516-113156

Nicholas, M. K. (1989). Self-efficacy and chronic pain. *Paper Presented at the Annual Conference of the British Psychological Society*, St. Andrews.

Nielson, W. R., Jensen, M. P., Karsdorp, P. A., and Vlaeyen, J. W. S. (2013). Activity pacing in chronic pain: concepts, evidence, and future directions. *Clin. J. Pain* 29, 461–468. doi: 10.1097/AJP.0b013e3182608561

Noar, S. M., Benac, C. N., and Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychol. Bull.* 133, 673–693. doi: 10.1037/0033-2909.133.4.673

O'Brien, N., Philpott-Morgan, S., and Dixon, D. (2016). Using impairment and cognitions to predict walking in osteoarthritis: a series of n-of-1 studies with an individually tailored, data-driven intervention. *Br. J. Health Psychol.* 21, 52–70. doi: 10.1111/bjhp.12153

Pastor-Mira, M.-A., López-Roig, S., Martínez-Zaragoza, F., León, E., Abad, E., Lledó, A., et al. (2019). Goal preferences, affect, activity patterns and health outcomes in women with fibromyalgia. *Front. Psychol.* 10:1912. doi: 10.3389/fpsyg.2019.01912

Philips, H. C. (1987). Avoidance behaviour and its role in sustaining chronic pain. *Behav. Res. Ther.* 25, 273–279. doi: 10.1016/0005-7967(87)90005-2

Pincus, T., Smeets, R. J. E. M., Simmonds, M. J., and Sullivan, M. J. L. (2010). The fear avoidance model disentangled: improving the clinical utility of the fear avoidance model. *Clin. J. Pain* 26, 739–746. doi: 10.1097/AJP.0b013e3181f15d45

Rost, S., Van Ryckeghem, D. M. L., Schulz, A., Crombez, G., and Vögele, C. (2017). Generalized hypervigilance in fibromyalgia: normal interoceptive accuracy, but reduced self-regulatory capacity. *J. Psychosom. Res.* 93, 48–54. doi: 10.1016/j.jpsychores.2016.12.003

Rubin, D. B. (1996). Multiple imputation after 18+ years. *J. Am. Stat. Assoc.* 91, 473–489. doi: 10.1080/01621459.1996.10476908

Schaller, A., Rudolf, K., Dejonghe, L., Grieben, C., and Froboese, I. (2016). Influencing factors on the overestimation of self-reported physical activity: a cross-sectional analysis of low back pain patients and healthy controls. *BioMed Res. Int.* 2016, 1–11. doi: 10.1155/2016/1497213

Schrooten, M. G., Van Damme, S., Crombez, G., Peters, M. I, Vogt, J., and Vlaeyen, J. W. S. (2012). Nonpain goal pursuit inhibits attentional bias to pain. *PAIN* 153, 1180–1186. doi: 10.1016/j.pain.2012.01.025

Smith, G., Williams, L., O'Donnell, C., and McKechnie, J. (2019). A series of n-of-1 studies examining the interrelationships between social cognitive theory constructs and physical activity behaviour within individuals. *Psychol. Health* 34, 255–270. doi: 10.1080/08870446.2018.1500576

Solberg Nes, L., Carlson, C. R., Crofford, L. J., de Leeuw, R., and Segerstrom, S. C. (2010). Self-regulatory deficits in fibromyalgia and temporomandibular disorders. *PAIN* 151, 37–44. doi: 10.1016/j.pain.2010.05.009

Solberg Nes, L., Carlson, C. R., Crofford, L. J., de Leeuw, R., and Segerstrom, S. C. (2011). Individual differences and self-regulatory fatigue: optimism, conscientiousness, and self-consciousness. *Personal. Individ. Differ.* 50, 475–480. doi: 10.1016/j.paid.2010.11.011

Solberg Nes, L., Ehlers, S. L., Whipple, M. O., and Vincent, A. (2013). Self-regulatory fatigue in chronic multisymptom illnesses: scale development, fatigue, and self-control. *J. Pain Res.* 6, 181–188. doi: 10.2147/JPR.S40014

Stone, A. A., and Broderick, J. E. (2007). Real-time data collection for pain: appraisal and current status. *Pain Med.* 8, S85–S93. doi: 10.1111/j.1526-4637.2007.00372.x

Stone, A. A., Broderick, J. E., Shiffman, S. S., and Schwartz, J. E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *PAIN* 107, 61–69. doi: 10.1016/j.pain.2003.09.020

Stone, A. A., Schwartz, J. E., Broderick, J. E., and Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers. Soc. Psychol. Bull.* 31, 1340–1346. doi: 10.1177/0146167205275615

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., and Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Control. Clin. Trials* 24, 182–199. doi: 10.1016/S0197-2456(02)00320-3

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., and Mcdowell, M. (2008). Physical activity in the united states measured by accelerometer. *Med. Sci. Sports Exerc.* 40, 181–188. doi: 10.1249/mss.0b013e31815a51b3

Van Damme, S. (2014). Avoidance and persistence: capacity or motivation?. *Clin. J. Pain* 31, 843–844. doi: 10.1097/AJP.0000000000000164

Van Damme, S., Becker, S., and Van der Linden, D. (2018). Tired of pain? Toward a better understanding of fatigue in chronic pain. *PAIN* 159, 7–10. doi: 10.1097/j.pain.0000000000001054

Van Damme, S., and Kindermans, H. (2015). A self-regulation perspective on avoidance and persistence behavior in chronic pain: new theories, new challenges? *Clin. J. Pain* 31, 115–122.

Van Damme, S., Van Ryckeghem, D. M., Wyffels, F., Van Hulle, L., and Crombez, G. (2012). No pain no gain? Pursuing a competing goal inhibits avoidance behavior. *PAIN* 153, 800–804. doi: 10.1016/j.pain.2011.12.015

van Weering, M., Vollenbroek-Hutten, M., and Hermens, H. (2011). The relationship between objectively and subjectively measured activity levels in people with chronic low back pain. *Clin. Rehabil.* 25, 256–263. doi: 10.1177/0269215510380828

Vervoort, T., and Trost, Z. (2016). The interpersonal function of pain: conserving multiple resources. *PAIN* 157, 773–774.

Vieira, R., McDonald, S., Araújo-Soares, V., Sniehotta, F. F., and Henderson, R. (2017). Dynamic modelling of n-of-1 data: powerful and flexible data analytics

applied to individualised studies. *Health Psychol. Rev.* 11, 222–234. doi: 10.1080/17437199.2017.1343680

Vlaeyen, J. W. S., and Crombez, G. (1999). Fear of movement/(re)injury, avoidance and pain disability in chronic low back pain patients. *Man. Ther.* 4, 187–195. doi: 10.1054/math.1999.0199

Vlaeyen, J. W. S., and Linton, S. J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *PAIN* 85, 317–332.

Vlaeyen, J. W. S., and Linton, S. J. (2012). Fear-avoidance model of chronic musculoskeletal pain: 12 years on. *PAIN* 153, 1144–1147. doi: 10.1016/j.pain.2011.12.009

Voodoo, S. M. S. (2020). *Voodoo SMS Bulk SMS Messaging*. Available online at: https://www.voodoosms.com/ (accessed July 14, 2020).

Williamson, A., and Hoggart, B. (2005). Pain: a review of three commonly used pain rating scales. *J. Clin. Nurs.* 14, 798–804. doi: 10.1111/j.1365-2702.2005.01121.x

Winger, J. G., Plumb Vilardaga, J. C., and Keefe, F. J. (2019). Indices of pain variability: a paradigm shift. *PAIN* 160, 2411–2412. doi: 10.1097/j.pain.0000000000001627

Yeo, G. B., and Neal, A. (2013). Revisiting the functional properties of self-efficacy:a dynamic perspective. *J. Manag.* 39, 1385–1396. doi: 10.1177/0149206313490027

Zigmond, A. S., and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* 67, 361–370.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RV declared a shared affiliation, with no collaboration, with one of the authors, DD, to the handling Editor at the time of review.

frontiers in Computer Science

REVIEW
published: 15 October 2020
doi: 10.3389/fcomp.2020.00031

Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and "Small" Data

Steven M. Schwartz[1]*, *Kevin Wildenhaus*[2], *Amy Bucher*[3] and *Brigid Byrd*[1]

[1] *IndividuALLytics, Inc., Ann Arbor, MI, United States,* [2] *Janssen, Brunswick, NJ, United States,* [3] *Mad*Pow, Boston, MA, United States*

The technology currently available for quantifying various biometric, behavioral, emotional, cognitive, and psychological aspects of daily life has become increasingly diverse, accurate, and accessible as a result of ongoing and continuous improvements. These burgeoning technologies can and will profoundly alter the way lifestyle, health, wellness, and chronic diseases are managed in the future. For those pursuing the potential of such digital technologies in the creation of a compelling and effective connected healthcare experience, a number of new concepts have surfaced. We have taken these concepts (many of which originate in engineering) and extended them so they can be incorporated into managing health risk and health conditions via a blended digital health experience. For example, the advent of mobile technology for health has given rise to concepts, such as ecological momentary assessment and ecological momentary intervention that assess the person's (digital twin) status and delivers interventions as needed, when needed—perhaps even preemptively. For such concepts to be fully realized, the experience design of mobile health (mHealth) program(s) (aka connected care) should and now can actually guide end users through a series of self-experiments directed by data-driven feedback from a version of their digital twin. As treatment development and testing move toward the *precision* of individual differences inherent in every person and every treatment response (or non-response), group data and more recent big data approaches for generating new knowledge offer limited help to end users (including practitioners) for helping individuals evaluate their own digital twin–generated data and change over time under different conditions. This is the renaissance of N-of-1 or individual science. N-of-1 evaluation creates the opportunity to evaluate each individual uniquely. The rigor and logic of N-of-1 designs have been well articulated and expanded upon for over a half century. For the clinician, this revitalized form of scientific and behavioral interaction evaluation can help validate or reject the impact a given treatment has for a given patient with increased efficiency and accuracy. Further, N-of-1 can incorporate biological (genomic), behavioral, psychological, and digital health data such that users themselves can begin to evaluate the relationships of their own treatment response patterns and the contingencies that impact them. Thus, emerges the self-scientist.

Keywords: digital health, digital twin, digital phenotype, self-scientist, mHealth, small data, N-of-1

OPEN ACCESS

Edited by:
Eric Hekler,
University of California, San Diego, United States

Reviewed by:
Steven Demets,
Quant ICT, Belgium
Kenneth Loh,
University of San Diego, United States

***Correspondence:**
Steven M. Schwartz
steve@individuallytics.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *29 November 2019*
Accepted: *06 July 2020*
Published: *15 October 2020*

Citation:
Schwartz SM, Wildenhaus K, Bucher A and Byrd B (2020) Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and "Small" Data. Front. Comput. Sci. 2:31. doi: 10.3389/fcomp.2020.00031

WHAT HAPPENED TO THE "QUANTIFIED SELF" MOVEMENT?

The vision of a "quantified self" really began with Gary Wolf and Kevin Kelly (then editors at *Wired* magazine) in 2007 (Wolf, 2007). Its original intention was to promote the value of self-monitoring facilitated by emerging mobile (and other) connected digital technologies (mobile apps, wearables, wireless peripherals, etc.) for data acquisition and self-reflection. The data covered a range of overt and covert behaviors (steps, mood, diet, stress, medication adherence, etc.) and biomarkers (sleep, heart rate, weight, etc.) of which the end user's ultimate goal was to gain greater self-insights and share those insights with others. The movement itself centered largely on the activity of self-measurement (aka tracking or self-monitoring) as the primary functional component of the experience. Much has changed since Wolf and Kelly originally coined this term [see (Heyen, 2020)]. The technology currently available for quantifying various biometric, behavioral, emotional, cognitive, and psychosocial factors of daily life has become increasingly diverse, accurate, and accessible. Many believe these burgeoning technologies can and will profoundly alter the way lifestyle, health, wellness, and chronic disease are managed in the future although, as Heyen (2020) notes, the quantified-self phenomenon has had minimal impact on the collective scientific knowledge to date. For those pursuing the potential of such digital technologies, several concepts have surfaced and/or resurfaced that when conceptually and practically integrated may help facilitate the as-yet-unrealized potential of connected care.

The first of these concepts is that of a *digital twin*. A digital twin is a digital representation of a real-world entity or system that offers information on the functional status of that system. The *digital twin* has its origins as an engineering paradigm for predictive problem solving of dynamic systems with early applications at NASA (Marr, 2017; Tao and Qi, 2019). The concept has been extended into many manufacturing- and process-related contexts to map out potential system failures. Gartner named the digital twin concept one of the Top 10 Strategic Technology Trends for 2017, 2018, 2019, and 2020 (Cearley et al., 2020).

(Call out box: *The average consumer knows more about the operation of their car than their own bodies. Today's automobile is equipped with more than 50 sensors/minicomputers that continuously monitor the functioning (i.e., health) of the car. The driver is signaled when one of these key functions falls outside a specific set of parameters via simplified and displayed dashboards. Such information allows drivers to determine the current functional state of their vehicle and intervene to avoid more extensive and costly problems prior to catastrophe. When serious trouble arises, these integrated computers alert the dealership to diagnostic issues and even contact emergency services based on continuously collected data from the car).*

Imagine if people had access to a similar set of self-generated biobehavioral information via a dashboard connected to the increasingly sophisticated and diverse set of commercially available devices, biosensors, technologies, and related data representing their own operational health and lifestyle. These digital twins could be the by-product of a networked set of biosensors, wearables, peripherals, smart pill dispensers, smart inhalers, ingestible smart pills, implantable devices (e.g., implantable cardio defibrillators), smart injectors, smartphone applications, and/or smart speakers all connected to an intelligent home ecosystem. Data emanating from these varied sources and sensors would be rendered back to the person, reflecting everything from their ongoing blood pressure to degree of hydration. The rendered data, with supporting content, would drive personalized and actionable health choices and behavior change guidance to each person uniquely throughout the day based on their own configuration and biobehavioral readings.

The technology for creating a usable digital twin largely exists today for addressing wellness, prevention, and ongoing management of focused health conditions. **Figure 1** lays out a conceptual (albeit incomplete) digital health technology map as it might be applied to the range of monitoring possibilities by organ system based on commercially available digital technology. The map also attempts to display likely clinical goals, potential digital health tools, and the biometric and behavioral data gathered from them to be used for clinical purposes.

If the concept of a digital twin is currently conceivable with existing commercially ready digital health and therapeutics technologies, then data derived from such enhanced self-monitoring technology represents the individual's *digital phenotype* (Onnela and Rauch, 2016; Huckvale et al., 2019). As such, this digital phenotype is the sum of an individual's *ad libitum* behavior expressed through digital media (sensors, tools, devices, apps, and related software, such as machine learning or artificial intelligence, etc.) *in vivo* and *in situ*. Today, these digital phenotypes do not necessarily reflect an a priori attempt by the individual to make use of their digital information as reflected in their phenotype. However, the collected data, when organized, has the potential to typify an individual's behaviors, lifestyle, and related baseline biomarkers as they relate to targeted risks and health end points consistent with what is now being referred to as P4 (predictive, preventive, personalized, participatory; Flores et al., 2013; Sanger et al., 2016).

The link between the digital twin and the digital phenotype would likely be a set of algorithms patterned off the current scientific knowledge base. For example, findings from the ongoing Framingham Heart Study have been used to establish a 10-year coronary risk prediction algorithm (D'Agostino et al., 2008). If an individual has contributed the necessary input data of age, diabetes diagnosis, smoking status, treated and untreated systolic blood pressure, total cholesterol, HDL cholesterol, and/or body mass index (BMI) from any combination of sources, including consumer devices, clinical records, or self-report, then the system could provide real-time feedback that also ties to a risk score for cardiovascular health as well as related evidence-based insights for cardiovascular risk modification. Additional algorithms could provide similar scores for other biological functions. Collectively, the dashboard could demonstrate changing future risk based on real-time, present-time performance. This can also prompt individuals to

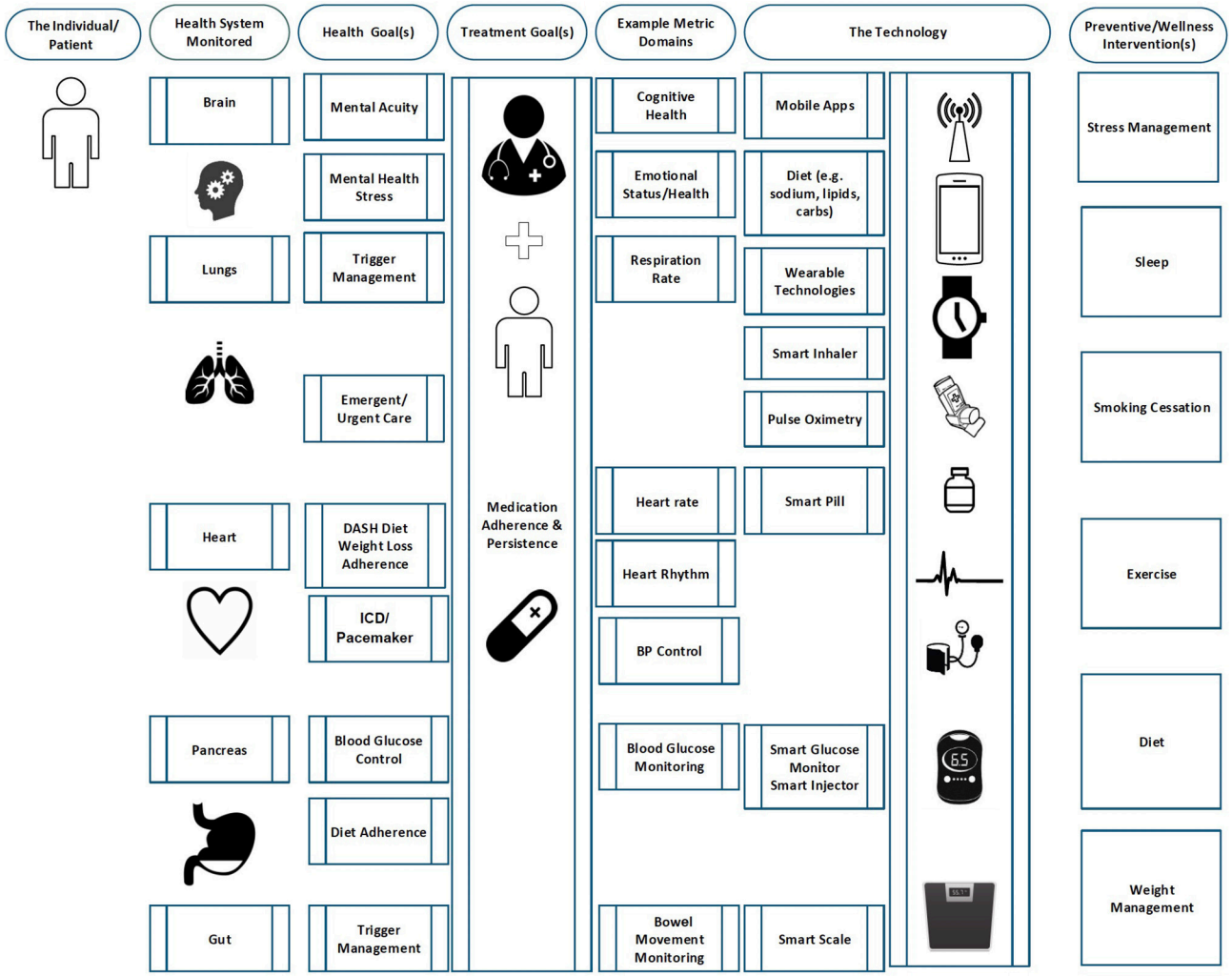


FIGURE 1 | This map describes one set of possibilities for configuring a digital twin with existing technologies.

connect other data sources if key variables are missing from certain algorithms.

One concern, which we address in more detail in the conclusion, is basing digital twin technology on data sets that capture patterns of bias. Because the technology, by nature, hones in on patterns, training the algorithm with flawed data can exacerbate and perpetuate those flaws. For example, a study examining an algorithm used by Optum to assign risk levels to patients systematically under-risked Black patients, likely due to the use of data that included racially based care disparities (Obermeyer et al., 2019). The resulting algorithm treated less care tendered to Black patients as indicative of lower need rather than less access.

Several considerations are important here. First is the integrity of the data itself as is true of all data operations. Second, any data included in the digital twin technology should be critically reviewed to identify and remediate issues that perpetuate historical bias. Third, and most importantly (and often confused with the second consideration), is the interpretive lens. Reliable and valid data do indicate biases in healthcare because they truly exist (and are now commonly and collectively referred to as the "social determinants of health"). What is deeply needed and will not be addressed by data or analytics alone is the strong interpretive lens, and that is about values (NEJM Catalyst, 2017). The history of intelligence testing and the eugenics movement is a sober reminder (Gould, 1981) of what happens when the interpretive lens is not considered.

COMMERCIALLY AVAILABLE DIGITAL HEALTH TECHNOLOGY

Digital health technology exists within and beyond the boundaries of the formal medical system. Although some healthcare sectors have enthusiastically pursued this technology, adoption by practitioners and patients has lagged. Recent current events prompted by the COVID-19 pandemic have altered

both the public and practitioner perception of the value of remote patient monitoring by means of digital technology. There are growing impressions that digitally enabled remote patient monitoring is no longer just an interesting innovation or "nice to have" but rather a clinical delivery infrastructure imperative.

Any given individual has likely generated an extensive digital footprint already in their lifetime from multiple data sources, of which most fall into one or more of these four broad and potentially overlapping categories:

- **Clinically generated data**. This is the full spectrum of data generated by a person's interactions with the formal healthcare system, including electronic medical records, lab test results, pharmacy data, and health insurance claims, etc.
- **Commercial real-world health data**. This information is data generated by programs focused on population health management that complement traditional healthcare, including wellness and disease management, to targeted populations intended to improve risk pools.
- **Consumer digital health device–generated data**. The increasingly sophisticated array of commercially available connected digital technologies are now available to leverage in the care and well-being of key patient populations and generate clinically relevant data from non-clinical sources.
- **Health-suggestive data**. Digital data is generated by people from a variety of non-health, non-clinical data that are not explicitly tied to health but do reflect other aspects of lifestyle and secondarily can provide additional insights into health (social determinants such as zip code, local weather, buying habits, etc.).

CREATING THE DIGITAL TWIN

To bring the digital twin concept to life, people must have access to an integrated set of tools, content, and services all existing within a single internally consistent live and digital experience that helps both patient and practitioner make data-based health choices. These data create the person's health data repository. There must also be a mechanism for people to access the resultant insights. In one model, people could create personalized accounts via a website or downloaded app. It is also possible that entities, such as health systems or regional or national governments, might create the digital twin system for enrolling their members or citizens.

Once an account is created for an individual, including unique identity markers, the user could permission various data sources to interface with the digital twin to avoid potential data misuse or abuse. Once data sources are connected to the system, the individual would then return to the account to view insights and feedback over time. The value of the system is 4-fold: 1. data capture, 2. communication, 3. intervention delivery, and 4. outcome evaluation. As with any digital technology, there is also the opportunity to design deliberate outreach to users. For example, users might receive a cellphone alert if their digital twin data indicates an acute health issue or, on the positive side, if their data indicates behavioral changes are leading to risk reduction.

To be explicit, the typical user of digital twin technology is not required to have the health or science expertise to form conclusions about behavioral responses based on their own data. In many circumstances, with better information, people themselves are in the best position to weigh the costs vs. the benefits of a given treatment. Therefore, a core component of the digital twin must be coaching or feedback to guide users through the "so what" of their data insights. By the nature of the basis of that personalized coaching and feedback, the digital twin harnesses the strengths of tailored interventions, which are consistently seen to produce more sustained changes than static or generic health education (Noar et al., 2007; Strecher et al., 2008).

CURRENT HEALTHCARE AND THE CLINICAL TRIAL

Clinical trials have been the primary mechanism for generating clinical knowledge and depend on measures of central tendency for assessing a treatment's benefits and side effects/risks relative to some comparator (i.e., true control or standard of care). This produces reliable and valid findings centered on group averages and variation around those averages. As the reigning gold standard in clinical research, the randomized control trial (RCT) has had great success demonstrating efficacy of treatment for most common conditions, syndromes, and diseases, yielding a portfolio of effective, evidence-based treatments that most helps the most people. This group lens and the supporting deductive inference provide a *forest view* of clinical outcomes relative to determining the greatest overall good. However, no matter how rigorously the group data is derived, using group statistics alone can never fully address the need to treat individuals uniquely.

SMALL DATA AND N-OF-1 INDIVIDUAL SCIENCE

Putting all the value, promise, and hype of *Big Data* aside, the digital twin for health has a small data requirement. It is becoming increasingly clear that RCT and group methods, although still quite valuable, are insufficient as treatments and testing move toward the precision of inherent individual differences, which are reflected in every person and every treatment response (positive or negative) (Gagne et al., 2014; Richter et al., 2015; Hilgers et al., 2016; McMenamin et al., 2018). The quantified self, with the "self" as the primary unit of analysis, was always intended for the individual to benefit from the added detail of self-observation. If self-quantification is to promote health and help manage chronic conditions, the data generated from a specifically constructed digital twin must be processed into a consumable and actionable form. Importantly, this information is embedded in every treated patient (if data were collected and analyzed properly) but remains largely latent. N-of-1 captures that value by rigorously evaluating each user, which provides *the tree level* of observation and evaluation.

A variety of authors from diverse disciplines have spoken about the value of N-of-1 research (Guyatt et al., 1986; Lillie et al., 2011; Barnett et al., 2012; Parker and Vannest, 2012; Dallery et al., 2013; Duan et al., 2013; Kravitz and Duan, 2014; Schork, 2015; Strathmann, 2015; Vohra et al., 2015; De Groot and Martin-Sanchez, 2017; Lobo et al., 2017; Mirza et al., 2017). The U.S. Department of Health's Agency for Healthcare Research and Quality has even published its own user guide to N-of-1 trials (AHRQ, 2014), and CONSORT has issued reporting guidelines (Vohra et al., 2015). N-of-1 evaluation creates the opportunity to evaluate each individual uniquely, which complements the existing evidence-based framework. The rigor and logic of these designs have been well articulated and expanded upon for over a half century (Mirza et al., 2017). By using more refined, time-ordered data to optimize individual-level understanding via N-of-1 analysis, more timely feedback can be provided to the patient (end user) that indicates what in their specific lives is influencing their behavior and biobehavioral health outcomes. The use of N-of-1 approaches is more true to life and clinical practice by providing individualized feedback to each patient and clinician about the quality and strength of their unique response to a given course of treatment. Further, as with more traditional approaches, N-of-1 can incorporate biological (genomic), behavioral, psychological, and digital health data such that users themselves can begin to evaluate the relationships of their own treatment response patterns and the contingencies that impact them in context. The approach can also evaluate and inform the combined treatments for comorbid condition management for which there are virtually no randomized clinical trials. For the clinician, this revitalized form of scientific and behavioral interaction evaluation methodology can help confirm or reject the impact of any given intervention for any given patient with increased efficiency and accuracy and greater insight to lifestyle: hence, the execution of precision medicine.

Because the classic clinical trial cannot fully answer all the relevant clinical questions or address the variants of an individual patient's treatment response, it must be coupled with other rigorous and valid clinical evaluation methods appropriate for the individual patient level of analysis. An N-of-1 perspective does not challenge the value of the RCT or Big Data, but rather complements it. This allows for personalization through a different lens and strategizes around time-ordered data within a single patient with multiple attributes. Further, through the use of the N-of-1 methodology, time-ordered data gets optimized by providing a new simple-to-interpret metric from the growing deluge of time-ordered data now coming from the advances in ecological momentary assessment and intervention coming from new and expanding measurement technologies (wearable devices, nanotechnology, pervasive wireless connectivity, Internet of Things, improved personal privacy and data protection technology, etc.) (Smyth and Stone, 2003; Kuntsche and Labhart, 2013; Runyan and Steinke, 2015; Spoelmann et al., 2016; Versluis et al., 2016; Dai and Bikdash, 2017).

Current standard of care for assessing treatment response to many chronic conditions takes place with limited frequency. Secondary clinical objectives (weight, diet, adherence to prescribed medicine, sleep, etc.) are almost never or only superficially addressed. Time and cost constraints do not allow for this type of care in our current healthcare delivery model. By bringing in scalable digital technology, the frequency of assessment can be increased while algorithms based on aggregate science offer evidence-based feedback. The architecture and metric strategy of any digital program that structures and optimizes time-ordered data with this technology can arrive at a user value level that is, by definition, personalized and without the classic paradox of requiring big data. One way to consider the digital twin is as a tool to pinpoint the specific individual response within the known variance of the aggregate responses, essentially locating the individual user within the distribution of the broader sample treatment responses.

DESIGN CONSIDERATIONS FOR THE DIGITAL TWIN

For the digital twin concept to produce its promised benefits, the technology must be designed to engage one or more end user groups in a meaningful way. The concept of N-of-1 must be operationalized in a way that a user can actually engage with and learn from it, including granting permission for data sources to be gathered and analyzed, viewing their personal data against the digital twin, and taking action based on insights from the comparison.

Any design of an actual N-of-1 product should take the needs of both provider and patient or caregiver user groups into consideration. Ideally, the group designing any interface will conduct original primary research with the intended users with the specific aim of informing the design. That said, prior experience with creating data-driven health interfaces for consumer use suggests several best practices that are likely to be relevant here. The design needs for a provider end user group vary somewhat from those for a patient and caregiver end user group due to the context in which they might access the system and the level of expertise they bring to interpret the data. As long as medically expert users are able to opt out of or skip instructional content that may lengthen their workflow without adding value, it is better to design the entire system to be appropriate for the patient than to take on a provider as the primary design target. Designers can then modify only with regards to the specific practical clinical care needs of the practitioner. Accordingly, the suggestions below focus on the patient and caregiver user audience.

We assume that an interface or dashboard will be created as the primary mechanism through which users interact with the N-of-1 system. This interface should accommodate the following.

Clear Data Visualization

The N-of-1 process draws from an enormous volume of data. Any user interface must offer a way of cutting through the noise to amplify the signal. Importantly, visual inspection of the data has been a critical part of the N-of-1 framework from its conception. Consider that, even if providers can discern meaningful trends in large volumes of data, they are unlikely

to have the time to do so. Therefore, N-of-1–enabled P4 medicine requires a data-processing step to optimize accuracy and efficiency without the practitioner losing autonomy to set a treatment plan with the patient. Patients and caregivers may have the time but likely lack the expertise and require a processing step that helps make the data more understandable and actionable (Fisher et al., 2003; Parker et al., 2011a,b). A presentation layer that prioritizes meaningful information and makes trends easily visible—perhaps through use of colors, icons, or other visual elements—is a prerequisite for adoption of the digital twin.

Plain Language

For patients and caregivers, an additional level of interpretation is likely to be needed for the digital twin concept to be useful. Although many patients do become experts in their own condition over time, the average health literacy and numeracy levels in the United States are quite low (see Kirsch, 1993; U.S. Department of Health Human Services., 2010). Particularly for people engaging with the digital twin prophylactically, plain language explanations of health phenomena and clear directives about action to take will increase the odds that they take positive action based on their data.

Access

Access refers to the ability of individuals to acquire and use technology and may be limited by either financial or geographical concerns. Additionally, in some countries, such as the United States, there is often a large cost burden on the patient for health care utilization that may affect whether and what data exists for a given individual. There is a role for organizations, such as health plans, governments, or employers, to subsidize or provide devices and internet access to facilitate people's use of the digital twin, but designers can also minimize the access burden wherever possible. One example is to offer options between data collected by often costly connected devices and web-based self-report; this will also make the digital twin more palatable to people who are not early adopters of technology and may not yet use connected devices. Another example is to design interfaces to require minimal data downloads so as to not max out limited cell phone plans or be unusable in areas with slow connectivity.

Accessibility

Accessibility best practices, such as those put forth in the Web Accessibility Initiative Worldwide Web Consortium (3Wc) (2019) should be observed in the design of the digital twin. The World Health Organization estimates that about 15% of the global population, or one billion people, have at least one disability World Health Organization (2020).

Although certainly demographic attributes can be important in personalizing an approach, we do not see pervasive demographically based needs. For example, research suggests that, despite stereotypes, older adults are increasingly likely to own and regularly use smartphones and computers (see Yoon et al., 2020). There is heterogeneity in technology usage among older adults as with all age groups with more educated and affluent people being more likely to use technology skillfully and regularly (Hargittai et al., 2019). However, as people age and experience normal physical and mental declines, they often benefit from the accommodations included in general accessibility standards, such as high color contrast between fonts and backgrounds and larger clickable areas on websites. Accessible design can, therefore, benefit multiple user groups. Importantly, it is always best practice to research the target users of any technology to understand and design for any limitations they are likely to experience.

Prioritization of Interventions

As the self-scientist identifies needs for intervention—for example, as data trends suggest an increased possibility of a health event without a change in behavior or medication—it will be important to offer a clear order of operations to follow (i.e., call to action and action steps). For health optimization situations, such as healthy self-scientists seeking to attain greater well-being, the order of steps to try may be entirely based upon their N-of-1 data and prioritize those activities most likely to produce the desired result. For more serious medical issues, the suggested steps could include the professional care team on the part of the patient and then focus on the provider as the audience for any other prioritized suggestions.

Ease of Adding and Removing Data Sources

The N-of-1 dashboard should make it simple for people to choose which self-generated data sources to include in their profile. Just as Mint (a financial budgeting program in North America) allows users to select financial accounts to connect with the service, the N-of-1 should allow a self-scientist to log into their consumer health apps, workplace wellness programs, and health risk assessments so that all relevant health data can be included. At the same time, it should be easy for any self-scientist to exclude a data source as they wish. Although theoretically optimal results come from including more data rather than less, there may be compelling reasons for a person to sever their relationship with a particular data source (e.g., privacy breaches or known errors in the data).

Ability to Add Context

A potential frustration with the N-of-1 approach is that some data may be better interpreted by both humans and algorithms with context. For example, a prolonged trend of low physical activity and weight gain would be interpreted differently if it happened during a stressful work period, a high-risk pregnancy, or without any precipitating life factors. Allowing users to specify contextual events that may have influenced their data would be helpful. In some cases, the context might have been provided by data that is not available for some reason. In the example above, a medical record would have revealed a pregnancy co-occurring with the activity cluster, but it may be that the electronic medical record (EMR) does not integrate with the N-of-1 system yet despite being technically possible and necessary for system-level optimization.

Integration Into Clinical Workflow

The more accessible the N-of-1 interface during the normal clinical workflow, the greater the chances of its wide adoption by providers. Within a clinical setting, providers' technology usage might be limited to an EMR on an intranet. The technological challenge in that case is facilitating access to the digital twin dashboard within the clinic, whether through EMR integration or offering new access to other systems. It is also worth pursuing policy-level inclusion of the N-of-1 approach in value-based reimbursement models. Providers are unlikely to dedicate significant time to a tool that does not contribute to their success metrics.

Logical Longitudinal Use of Phase Shifts

Because the individual and their time-series data are paramount in this framework, the experience is best designed around treatment phases (Pertschuk et al., 1978; Hayes et al., 1999; Dallery et al., 2013; Pham et al., 2016; Michie et al., 2017) and changes in those phases based on a new (novel) treatment circumstance. There exist a number of designs, such as ABAB reversal designs, that when strung together over the course of treatment can enhance the causal logic for associating a change in a monitored outcome with a specific treatment phase. Hayes et al. (1999) have described the clinical value of such approaches in detail.

PERSONA USE CASES

To bring clinical and consumer design realism to the discussion, the use of personas represents a common and useful methodology when designing a user-centered experience or program. By way of example, we wish to consider two personas (father and son) and their digital twins for what a risk modification health experience might look like today. Each has a connected care configuration based on their health profiles and life circumstances. In addition, we speculate on each persona with regards to what their health experience might look like in 5 years.

Raymond (father) and Josh (son): Current Connected Care Experiences

Call Out Box. Raymond (age 59).

- ☐ Firefighter Retired Secondary to Type II Diabetes Complications
- ☐ Type II Diabetes Treatment Metformin
- ☐ BMI 29.4
- ☐ Hypertension Treatment ACE inhibitor
- ☐ Spotty Medication Adherence
- ☐ Moderately Active
- ☐ Health Risk Assessment (HRA) Predicted Health
- ☐ Technology Suboptimizer[1].

[1]Person who uses some but not all and does not optimize the features of current digital connected technologies.

Background

Raymond lives with his wife Jeanne of 38 years. The couple have resided in the same suburban neighborhood their entire marriage. They raised their son Joshua (age 35) and daughter Casandra (age 24) in that home. Raymond retired 3 years ago following complications secondary to poorly controlled diabetes including retinopathy. Since that time, he has occupied himself with his garden, his grandkids, and general work around the home.

Health Status

Diagnosed with hypertension 10 years ago and type II diabetes 7 years ago, Raymond has struggled to keep his HbA1c and blood pressure levels in the range his doctor has recommended. Unfortunately, he has also struggled keeping both under control due to poor medication adherence and an unhealthy diet. As a result, Raymond has never been able to sustain a healthy blood pressure or HbA1c. Consultation with his internist indicated he may soon need to switch to injectable medications to better manage his blood glucose.

Psychographics

Raymond has always prided himself on his independence and being "a man's man." He finds it hard to ask for help and is much more comfortable caring for others than being cared for. He admits to some conflict with his wife over his health, diet, and activity level. He never liked the idea of being dependent on medication and would prefer to handle his health issues with diet and exercise. But he admits to dietary weaknesses, including a strong sweet tooth. His love of gardening and chasing around grandchildren are his primary sources of exercise, but he admits to very little physical activity in the last 4 months, which, in turn, raised his weight another 10 pounds. His history of habitual exercise as a firefighter and desire to retain his independence are potential strengths.

Technographics

Raymond can best be described as a technology "suboptimizer" (i.e., someone who uses current technology but not to its fullest potential). He uses a smartphone, laptop, and tablet, but does not get the most out of all the technology could provide. He uses apps on his phone for a limited number of practical activities, such as banking, scheduling, and checking the weather. He never uses the health apps that came preinstalled on the phone. He does use several online social sites.

RAYMOND'S CONNECTED CARE CONFIGURATION

Raymond and his physician have agreed to try several technological tools to help him take his medication as directed as the primary and most immediate goal. Raymond's doctor pointed out that Raymond needs to monitor his health, including his blood pressure and blood sugar, more regularly just as he once needed to monitor his firefighting equipment for proper

functioning. Based on a match between Raymond's needs and the available technology, they agree to the following tool set.

- A wearable device that can help monitor basic biobehavioral functions, such as sleep, activity (steps), and heart rate.
- A smart bottle that can message Raymond and select others when medication is to be taken and/or when a dose is missed.
- A smart scale that can assess weight as well as other secondary metrics, such as level of hydration, percent body fat, and body mass index.
- A mobile app that helps connect his support system, which includes his wife, children, and sister.
- Home smart glucometer that tracks blood glucose, aids in decision making, connects with the other technology (wearable and smart scale), and connects to support and practitioners.

Figure 2 displays Raymond's treatment plan laid out in a series of three consecutive treatment phases. Each phase represents a new unique set of treatment conditions. The A phase of Raymond's experience represents a baseline period of management as usual. In Raymond's case this includes a combination of historical data (past blood glucose values and other lab values, rough estimates of medication adherence, etc.) with 2 weeks of biobehavioral baseline (i.e., run in) data (blood sugars, medication adherence, activity level), average daily steps. The B phase represents the starting intervention that is focused on supporting medication adherence for both hypertension and type II diabetes by way of increased monitoring and intervention via smart bottle–driven reminders and increased coordination of family support via a support network app. There is also a secondary goal of increasing his average daily steps by 10% (from 5,800 to 6,380). The choices made for phase C are dependent on the phase B response data. They are presented in these personas as a "happy path" in which phase C builds on positive gains made in phase B. Importantly, the methodology would also allow for early detection of a non- or negative response to inform clinical decision making. Therefore, in our examples, the phase C has defined new goals for continued improvement in medication adherence and more programmatic increases in physical activity and diet (which now incorporate his wife Jeanne as the primary grocery shopper and cook). Each of these interventions is also tied to biometrics of blood pressure and blood sugar, the data of which is regularly collected and shared among patient, practitioner, and primary support (wife).

Call Out Box. Josh (age 35).

- ❑ Employed Community College Math Instructor
- ❑ BMI 27.6
- ❑ Prehypertension Considering Start of ACE Inhibitor
- ❑ Moderately Active
- ❑ HRA Health Prediction
- ❑ Technology Optimizer[2].

Background

Josh is Raymond's son and oldest child. He lives with his second wife, Adrian, of 3 years, and he has no children. The couple live in a condo not far from where Josh and Adrian work. Josh is employed as a math instructor at a community college, and Adrian works as a real estate broker. The couple maintain an active personal and professional life and have recently considered starting a family. They are particularly close with Raymond and Jeanne who live nearby. Josh decides to partner with his father on health to support him and spend more quality time together.

Health Status

Josh has always been athletic and still plays softball once a week during the season, but his regular exercise routine has become increasingly less regular and his expanding waistline reflects it. His doctor says that, unless he can get his blood pressure under control with diet, exercise, and weight loss, he will need to start medication. Having seen his father struggle with weight, high blood pressure, and diabetes, Josh is determined to avoid medication and get back to a healthier level of activity and eating. The prospect of fatherhood is an added motive.

Psychographics

Josh has always viewed himself as fit and athletic and, prior to marrying Adrian, was just that. It was a classic example of domestic comfort and contentment. The recent weight gain and increased blood pressure have him concerned but also motivated. He does not see himself as overweight, but his BMI says otherwise. His wife is very supportive of his plan to eat better and get more active and is eager to help him to develop a system that fits his busy and stressful schedule. Josh intends to begin with increased daily walking and biking at least once a week. Strong intrinsic motivation is a potential strength for Josh.

Technographics

Josh grew up with technology. He got his first X-Box at age 12 and his first cellphone when he obtained his driver's license. He uses a variety of technology, including a multitude of apps for managing personal affairs, work, and entertainment. He has recently purchased an updated fitness tracker and has used one off and on for the last several years.

The A phase of Josh's experience (see **Figure 3**) also has a baseline that parallels his father's. It combines data, such as blood pressure measures and historical data from his wearable, with the "run in" biobehavioral data (i.e., steps, sleep, BMI, etc.). Josh's B phase is designed to prevent and reverse his progression toward hypertension by way of lifestyle change alone. Therefore, his B phase is focused on weight loss by way of increased activity level and initial diet changes. He will increase his coordination of family support by better connecting with Adrian and Raymond as well as his mother and sister. Activity goals are to increase his average daily steps by 20% (from 8,900 to 10,680) and return to his old habit of biking a minimum of 1 but no more than 2 days per week. As with Raymond, the choices for phase C as presented here are a "happy path" in which phase C builds on positive gains made in phase B. Therefore, phase C has defined phase goals for continued improvement in physical activity and the addition of the DASH diet framework into his lifestyle (which

[2]Person who fully uses all the features of current digital connected technologies.

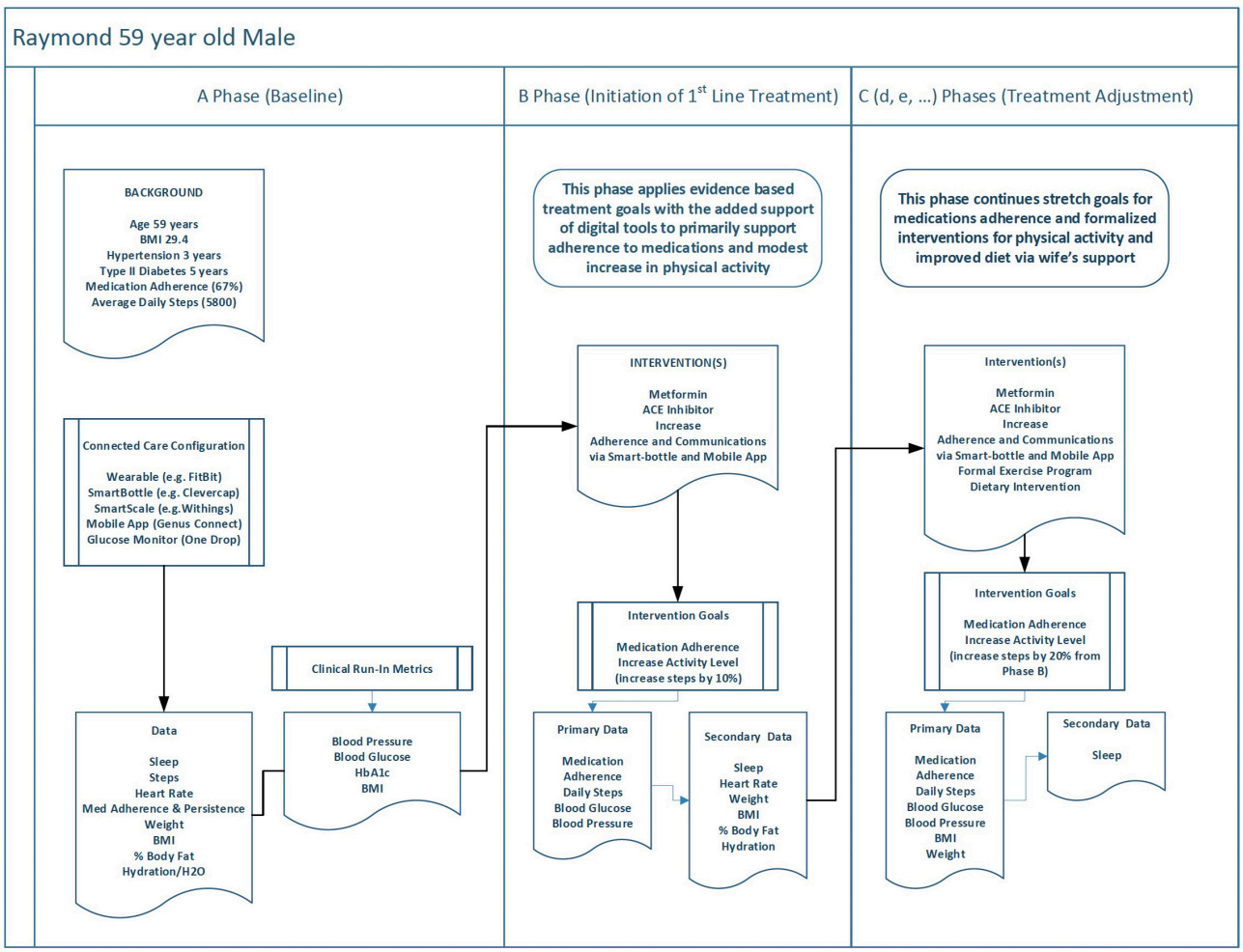


FIGURE 2 | This figure describes how Raymond's personal health data might be used with a digital twin across three phases of behavior change to achieve increasingly tighter outcome goals.

now incorporates his wife as the primary support for dietary change which she will share). Each of these interventions is also tied to primary biometrics of blood pressure and BMI.

RAYMOND AND JOSH—FAST FORWARD 5 YEARS

Figure 4 displays the health status, connected care configuration, and health goals for both Raymond and Josh as they might look 5 years into the future. Again, it is based on a digital twin model that helped both father and son achieve their baseline primary and secondary health goals and established healthier lifestyle habits (with their wives as participatory supporters).

RAYMOND HEALTH STATUS (AGE 64)

Raymond continues to manage both hypertension and type II diabetes and has been well controlled on both for 4+ years. His retinopathy has slowed, but visual impairment has been a challenge. The basics for what he must do to manage his health are the same (i.e., stay physically active, watch what he eats, and take his medications as prescribed) but how he does it (the tools he uses) has changed over time. Now, due to visual impairment, he brings his exercise indoors and uses an exercise bike and an elliptical machine, both connected to specific live and on-demand training programs. Data collected by the machines are also now available as part of his digital twin with full details (METS, etc.) for each bout of exercise.

JOSH HEALTH STATUS (AGE 40)

Josh also continues to manage his health. He has, thus far, avoided hypertension but developed dyslipidemia a year and a half ago. Josh now takes a statin and has been well controlled. Josh's goal is to stay physically active and watch what he eats.

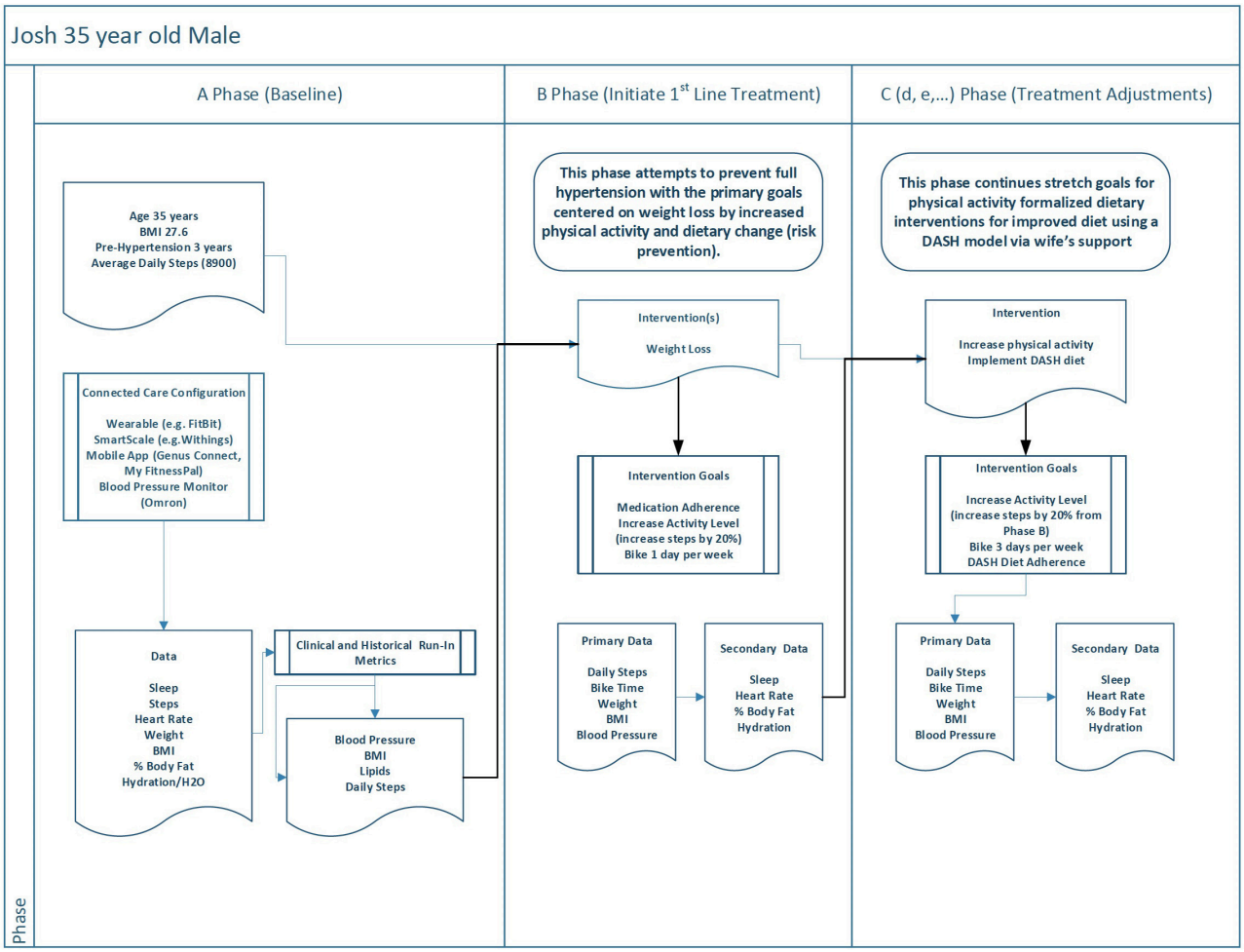


FIGURE 3 | This figure offers a contrast to **Figure 2** by showing how a similar phased digital twin process might unfold for a relatively healthier individual in order to accomplish health goals.

PERSONA SUMMARY

Connected care in this model included selected use of currently commercially available digital health and therapeutic technologies that will likely be more enhanced 5 years in the future. In both the current and future configurations, the assumption is that interconnectivity and visualization of individual patient data back to the patient, practitioner, and caregiver network can be accomplished and integrated from a technical and practical perspective. The data in these examples is integrated with more traditionally collected healthcare data (i.e., labs, utilization data, etc.) and evaluated using established N-of-1 methods and presented back with design considerations that make the personal data accurate (valid), easy to understand quickly, and with clear relevance for clinical decision making.

THE EXPERT PATIENT PROBLEM

It is now generally recognized that regardless of the health condition(s) being managed, individuals are a rich source of experiential information about the signs and symptoms of disease, common and unique responses to interventions, and successes and failures for self-managing health, all within the context of everyday living. These *expert patients* (Tattersall, 2002; Cordier, 2014) and their primary caregivers are the ultimate source of information for patient-centered processes and outcomes as they are shaped by each individual's experience of illness, social circumstances, attitudes to risk, values, preferences, and problem solving.

The challenge is in how to best facilitate and leverage that collective experience for the benefit of the whole health community, including practitioners. If the original intent of the quantified self movement was to gain greater self-insights and share those insights with others, then what better way to leverage them at the personal and community level than to use data and digital phenotypes? This aggregated N-of-1–level data can then facilitate and accelerate (e.g., crowdsource) the intelligence that is latent within the collective expert patient community (Levy, 2005; Eysenbach, 2008; Buecheler et al., 2010; McAfee, 2010; Li et al., 2012; Ranard et al., 2013; Crequit et al., 2018).

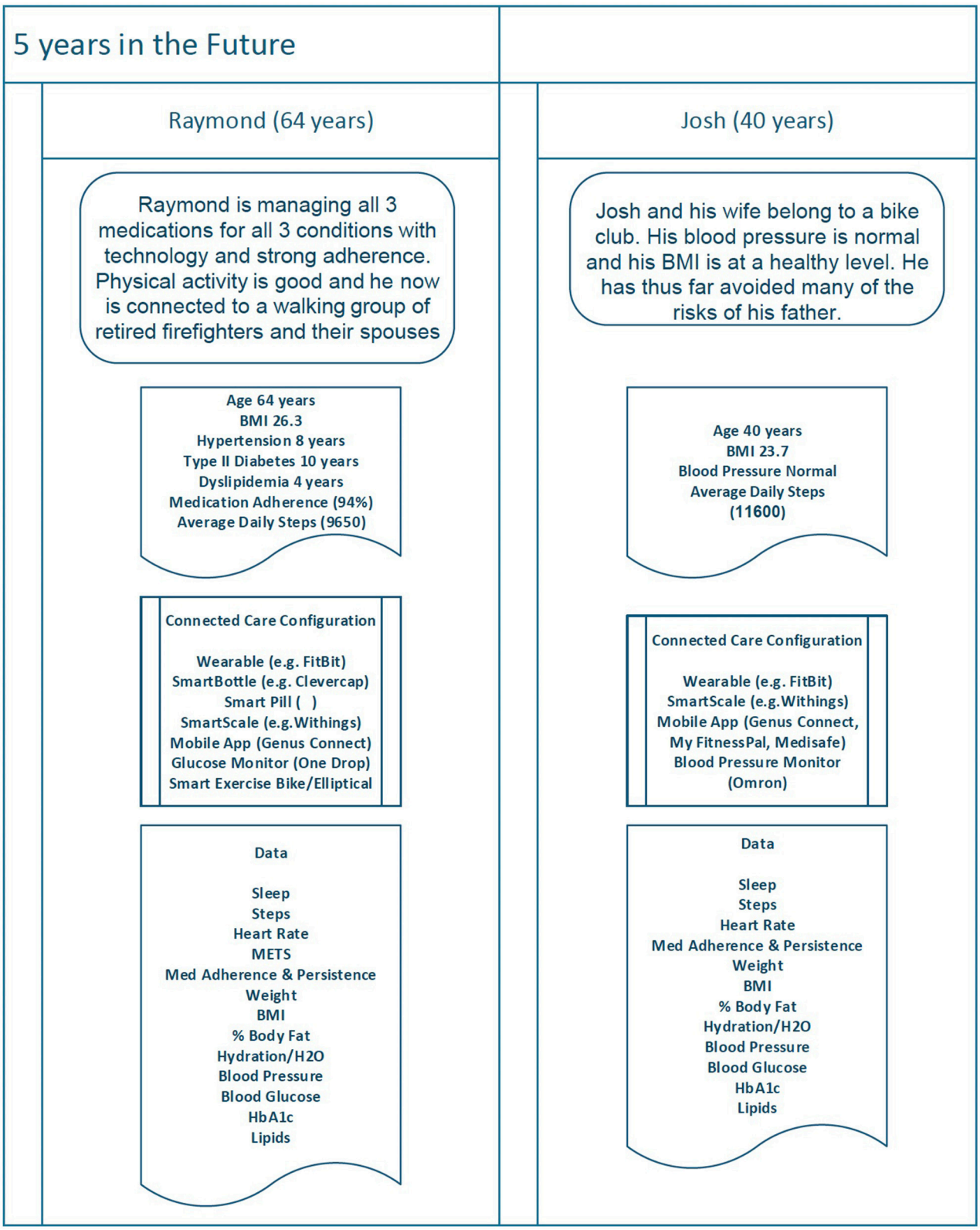


FIGURE 4 | As technology improves and both Raymond and Josh transition to maintenance of goals, they continue to monitor their data using a variety of connected devices and make behavioral adjustments in order to achieve desired outcomes.

ACCELERATED PATIENT INSIGHTS

Digital technology today provides the mechanism to harness patients' (expert or not) latent collective wisdom with the potential to provide value and self-insights for the collective patient-consumer experience (e.g., patientslikeme, a patient-focused social network for crowdsourcing treatment options and clinical trials). Research supports the notion that large groups, when properly facilitated and applied to a topic, including health problems, will self-organize in ways that are well-suited to developing behavioral patterns that, with proper analysis can generate a greater number of novel and valid insights and alternative conclusions.

Although Von Hippel (1986) lead user concept suggests that expert patients may push the edges of the digital twin technology to discover issues and opportunities to be incorporated into the design, less expert patients also contribute to learning. Indeed, many people using digital twin technologies will, especially at the outset, have low levels of health knowledge and may not be able to extract meaningful insights without coaching or support. Fortunately, the individual group members do not necessarily need subject matter expertise for the aggregated group to display this form of *emergent* or *collective intelligence* (Roskams and Popovic, 2016; Khatib et al., 2019). The digital twin technology can capture and sort the data provided by such users to provide expert researchers with information that can jump-start their research as well as provide visibility to expert patients to help provide that coaching and support. Such connected care can realize the quantified-self goals of greater self-insights (and very likely improve health literacy as well).

THE SELF-SCIENTIST

One of the historical shortcomings of the quantified-self movement was an overemphasis on *tracking* (i.e., self-monitoring) functionality within the technology (which still exists in many apps today) as the primary action. Practically speaking, health data and behavioral tracking generally suffered from two fundamental challenges in converting such data into insights and actions.

1. Nearly all methods for tracking were tedious and burdensome to the end user (e.g., patient or practitioner), making compliance to data collection problematic.
2. Perhaps more importantly, it has been repeatedly demonstrated that tracking in isolation does not produce sustained behavior change. Tracking must be coupled with analysis-driven intervention and feedback. This latter component is largely based on the overall experience design, functionality, and visualization.

What is known today from behavioral science research is that self-monitoring is a necessary but not sufficient condition for sustained engagement and subsequent behavior change (Karkar et al., 2016; Wicks, 2018). Insights and behavior change will not simply blossom out of self-quantification. The experience design and generated data must facilitate valid and practical insights with clear, well-timed, and motivating calls to action through the integration of an evidence-based approach.

Digital health and therapeutic technologies are increasingly solving the first problem by limiting or eliminating the end users' input requirements, thereby refining the digital twin and phenotypes. Consequently, the streaming data from these technologies can now become a rich source of time series data (aka repeated measures, trends) such that single case and small sample research designs can now be integrated into the analytical space to drive the ability of the end-user to draw valid conclusions about themselves and the contextual factors that influence their health behaviors. When the overall user experience is well designed around small case design and clinical reasoning, only then can technology-enabled experience go beyond tracking and actually enable end users to become more rigorous *"naturalistic observers"* of themselves…*self-scientists* (Parker and Vannest, 2012; Strathmann, 2015; Karkar et al., 2016; De Groot and Martin-Sanchez, 2017; Wicks, 2018).

The advent of mobile technology for health has also given rise to concepts, such as ecological momentary assessment and ecological momentary intervention. By combining and integrating data from the assessed person's (digital twin) status, technology can be used to deliver interventions as needed, when needed (even preemptively). For such concepts to be fully realized, the experience design of mobile health (mHealth) program(s) (connected care) should actually guide the end user through a series of self-experiments directed by data-driven feedback from a version of their digital twin and provide ongoing feedback about the experiments' outcomes.

PULLING IT ALL TOGETHER FOR THE GREATER GOOD (CROWDSOURCING N-OF-1)

The implications of the digital twin for healthcare at the individual and collective levels are exciting from both an outcome and an experience perspective. A clear benefit to successful implementation of the digital twin is the more rapid identification of treatment needs to ultimately improve health outcomes and patient quality of life with ancillary benefits across the ecosystem.

The digital twin also offers the opportunity to create an empirically validated understanding of how medical treatments (e.g., medications, surgical procedures) might be enhanced or supported by behavior change interventions, such as formally or informally crafted lifestyle management programs. These programs have yielded a mixed bag of outcomes without a strong shared understanding of how they might be developed and implemented most effectively. For example, it is only in 2019 that a group of experts put forth their consensus statement on basic standards for digital mental health applications (Torous et al., 2019). The digital twin concept would enable more rapid capture of usage of such apps and related outcomes to leapfrog our current understanding of

when and how to leverage them and how to identify best-in-class options.

Anticipating that digital twin technology will follow a standard diffusion of innovation curve (Rogers, 2003), we urge teams working on developing the digital twin to consider the lead user concept first advanced by Von Hippel (1986). Lead users are very early adopters of new technologies who, as the first users, are also the first people to encounter problems and unmet needs. By including lead users as part of a beta test group in which feedback is actively sought, designers will be better prepared with a useful and usable product when a less savvy segment of the population begins to set up digital twins. Research has shown that leveraging lead users as part of a product development process accelerates both volume and variety of projects completed (Ho-Dac, 2020).

A consumer-facing outcome of digital twin could be an algorithm-driven recommendation engine. The combination of individual and aggregate data at the heart of digital twin mimics the model used by companies, such as Amazon and Netflix to surface suggestions for purchases or entertainment. Such an engine could support shared decision-making tools, improve the quality of information within digital health apps, and reduce the frequency of patient panic when a search engine suggests dire but unlikely health outcomes.

Aggregation of N-of-1 replications (replications being those self-experiments that have serially been proven to work for N of x number of people) based on a set of common contextual attributes allows for a bottom-up approach to developing new knowledge (i.e., inductive reasoning) that combines and leverages the concept of the expert patient. Development of recommendation engines based on accumulating replications with ongoing evaluation of group based significance grows knowledge from direct experience accelerated by N-of-1 (Levy, 2005; Eysenbach, 2008; Buecheler et al., 2010; McAfee, 2010; Li et al., 2012; Ranard et al., 2013; Crequit et al., 2018).

ETHICAL AND POLICY CONSIDERATIONS

As we consider a future in which each individual can cross-reference their digital twin against their own personalized health data, it is important to note a few potential ethical issues with the intention of preempting them with patient protections. It is feasible that entities with access to patient data may use insights from the digital twin concept in ways that are against patients' best interests. For example, employers could use digital twin insights to deny offers to employees who are likely to incur high healthcare costs in the next few years[3]. Or insurance plans may raise premiums on people whose health data suggests an impending negative event, ironically making care more difficult to access at just the time when it is most critical. These would be misuses of the digital twin concept. We suggest three principles to safeguard against them.

[3]Indeed, some employers already prohibit employees from using tobacco and require a biometric test before confirming employment offers.

Patients Own Their Data

Patients are the source of their data. When the patient is the primary entity associated with the data, it enables the digital twin concept to work. When data is scattered amongst tens or hundreds of databases under the control of organizations that are not talking to each other, the burden of compiling an individual's digital twin data set becomes enormous. Maintaining patient data ownership across the life span helps to solve for that (Dorey et al., 2018). It also creates an opportunity for new economic models that permit patients to monetize the use of their data if they choose to share it with third-party entities for research and development or other purposes. It offers patients a mechanism to protect the privacy of their data against misuse by other entities by maintaining ownership and granting or withholding access to specific data in a granular fashion.

Patients Must Provide Explicit Informed Consent for the Use of Their Data

There are many legitimate reasons why third parties may want to use patient-generated data for research or investigation. It's also clear that there may be a public health benefit to this type of data usage that we would not want to disrupt. Informing patients in plain language about the potential ways in which their data may be used—and offering them regular opportunities to review that information and change their consent—provides them the agency to participate or not, similar to the code of ethics used in human subjects research. As patterns emerge in what people are willing to consent to and not, it also may be possible to determine whether and when it's appropriate to compensate patients for use of their data.

Advocacy Efforts Should Enshrine Patient Data Ownership and Access Into Law

Research shows that although physicians and researchers are sympathetic to the notion that patients should have a say in the use of their data, they are also skeptical that patients can make appropriate choices based on the potential benefits of research to society (Dorey et al., 2018). It cannot be assumed that physicians and researchers will honor patient data ownership in practice whether because they truly believe the positive impact of their work is sufficiently large to override patient rights or because they are able to convince patients to consent when they otherwise might not through their position authority. The potential for data misuse or abuse speaks to the need for advocacy for strong patient protection laws and a role for policy advocacy from subject matter experts and organizations (e.g., HIMSS, Xcertia, etc.). The forms this advocacy might take will differ depending on the form of government and existing data protections in place in a given region as will any resultant policy (see the GDPR in the European Union as an example).

In the United States, some of the specific advocacy targets to enable the digital twin concept to work include establishing a national standard data format, which will enable interoperability and easier sharing of data across platforms. The Standard Health Record (http://standardhealthrecord.org/)

and the Fast Healthcare Interoperability Resources (FHIR) are two examples of U.S.-based projects to accomplish a standard format.

There is also a call to make patient data open source, meaning it is available to be used for free. With appropriate privacy protections in place (e.g., deidentification of individual data in public data sets, securing individual informed consent for data inclusion in the data set), having clinically generated health data made open source would facilitate integration into a digital twin system.

Product Designers Must Critically Evaluate Data Sources

The success of the digital twin relies on building a data model against which any individual's data can be compared. Unfortunately, many existing data sets incorporate racial (see Obermeyer et al., 2019), gender (see Criado Perez, 2019), or other bias. Using those data sets without correction can perpetuate and intensify the bias, which would result in the digital twin making suboptimal or inappropriate recommendations for anyone not fitting the "right" demographic profile. Designers must do a deliberate and critical review of candidate data to identify potential sources of bias, such as the following:

- Historical inequities in care
- Studies that exclude women, non-white people, or members of other groups
- Biased or motivated data-collection methods
- Understand potential interpretive biases and the values that shape access to resources, technology, and information

Any identified biased data should be excluded from the digital twin, adjusted and corrected, or specified to apply only to the relevant demographic groups. Designers should also specifically question any assumptions in using one piece of data as a proxy for another (as was the case in the (Obermeyer et al., 2019), study in which prior access to care was used as a proxy for need, obscuring a history of discrimination that limited access among Black patients).

FUTURE WORK

In many ways the future is here. The technology to create the beginnings of a digital twin to support the care and management of individuals with multiple chronic conditions exists today. The next challenge is for adventurous companies and institutions with quality technologies and subject matter expertise to start testing such systems (albeit in limited form to start) in real environments to address the following problems:

1. Data integration, quality, and security across platforms and systems.
2. Iterative development of the connected care experience that includes practitioners and caregivers.
3. Iterative development of the data visualization feedback to each stakeholder.
4. Broaden the N-of-1 library of analytical approaches designed for specific clinical data circumstances.
5. Application of the continuously improving design to high value patient populations to evaluate impact on engagement, clinical outcomes, utilization of services and costs.
6. Develop a mechanism to compile N-of-1 learnings to detect aggregate trends and insights and bring them into the shared scientific knowledge.
7. Identify, within regional or national subject matter expert coalitions, the appropriate political advocacy targets and organize lobbying efforts toward making health data accessible and usable for tools such as the digital twin.

Assuming the ability to build out success for the first four short-term goals above, the more distal goals would be to experiment with the aggregation and dissemination approaches back to key audiences to better foster the value of N-of-1 insights; continuously improve the experience and design of the digital twin; and create replications for the relevant communities in ways that leverage and reinforce collective intelligence of the group.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Barnett, S. D., Heinemann, A. W., Libin, A., Houts, A. C., Gassaway, J., Sen-Gupta, S., et al. (2012). Small N designs for rehabilitation research. *J. Rehabil. Res. Dev.* 49, 175–186. doi: 10.1682/JRRD.2010.12.0242

Buecheler, T., Sieg, J. H., Füchslin, R. M., and Pfeifer, R. (2010). Crowdsourcing: open innovation and collective intelligence in the scientific method: a research agenda and operational framework. *Proc. Alife XII Conf. Odense Denmark* 2010.

Cearley, D., Burke, B., Smith, D., Jones, N., Chandrasekaran, A., and Lu, C. K. (2020). *Gartner Top 10 Strategic Trends for 2020*. Available online at: https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/

Cordier, J. F. (2014). The expert patient: towards a novel definition. *Euro. Respir J.* 44, 853–857. doi: 10.1183/09031936.00027414

Crequit, P., Mansouri, G., Benchoufi, M., Vivot, A., and Ravaud, P. (2018). Mapping of crowdsourcing in health: systematic review. *J. Med. Internet Res.* 5:e187. doi: 10.2196/preprints.9330

Criado Perez, C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men*. New York, NY: Abrams.

D'Agostino, R. B. Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., et al. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 117, 743–753. doi: 10.1161/CIRCULATIONAHA.107.699579

Dai, X., and Bikdash, M. (2017). Trend analysis of fragmented time series for mHealth apps: Hypothesis testing based on adaptive spline filtering method with importance weighting. *IEEE Access.* 5, 27767–27776.

Dallery, J., Cassidy, R. N., and Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *JMIR* 15:e22. doi: 10.2196/jmir.2227

De Groot, M., and Martin-Sanchez, F. (2017). Single-subject (N-of-1) research design, data processing and personal science. *Methods Med.* 56, 416–418. doi: 10.3414/ME17-03-0001

Dorey, C. M., Baumann, H., and Biller-Adorno, N. (2018). Patient data and patient rights: Swiss healthcare stakeholders' ethical awareness regarding large patient data sets - a qualitative study. *BMC Medical Ethics* 19:20. doi: 10.1186/s12910-018-0261-x

Duan, N., Kravitz, R. L., and Schmid, C. H. (2013). Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *J. Clin. Epidemiol.* 66, s21–s28. doi: 10.1016/j.jclinepi.2013.04.006

Eysenbach, G. (2008). Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J. Med. Internet Res.* 10:e22. doi: 10.2196/jmir.1030

Fisher, W. W., Kelley, M. E., and Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *J. Appl. Behav. Analysis* 36, 387–406. doi: 10.1901/jaba.2003.36-387

Flores, M., Glusman, G., Brogaard, K., Price, P. D., and Hood, L. (2013). P4 medicine how systems medicine will transform the healthcare sector and society. *Personal. Med.* 10, 565–576. doi: 10.2217/pme.13.57

Gagne, J., Thompson, L., O'Keefe, K., and Kesselheim, A. S. (2014). Innovative research methods for studying treatments for rare disease: methodological review: *BMJ* 349:g6802. doi: 10.1136/bmj.g6802

Gould, S. J. (1981). *The Mismeasure of Man.* Ney York, NY: Norton & Company.

Guyatt, E., Sackett, D., Taylor, D. W., Chong, J., Roberts, R., Pugsley, S., et al. (1986). Determining optimal therapy – randomized trials in individual patients. *NEJM* 314, 889–892. doi: 10.1056/NEJM198604033141406

Hargittai, E., Piper, A. M., and Morris, M. R. (2019). From internet access to internet skills: digital inequality among older adults. *Univ. Access Inform. Soc.* 18, 881–890. doi: 10.1007/s10209-018-0617-5

Hayes, S. C., Barlow, D. H., and Nelson-Gray, R. (1999). *The Scientist Practitioner: Research and Accountability in the Age of Managed Care.* Needham Heights, MA: Allyn and Bacon.

Heyen, N. B. (2020). From self-tracking to self-expertise: the production of self-related knowledge by doing personal science. *Public Unders. Sci.* 29, 124–183. doi: 10.1177/0963662519888757

Hilgers, R. D., Konig, F., Molenberghs, G., and Senn, S. (2016). Design and analysis of clinical trials for small rare disease populations. *J. Rare Disease Res. Treatm.* 1, 53–60. doi: 10.29245/2572-9411/2016/3.1054

Ho-Dac, N. N. (2020). The value of online user generated content in product development. *J. Business Res.* 112, 136–146. doi: 10.1016/j.jbusres.2020.02.030

Huckvale, K., Venkatesh, S., and Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality and safety. *Digital Med.* 2:88. doi: 10.1038/s41746-019-0166-1

Karkar, R., Zia, J., Vilardaga, R., Mishra, S. R., Fogarty, J., Munson, S. A., et al. (2016). A framework for self-experimentation in personalized health. *J. Amer Info Assoc.* 23, 440–448. doi: 10.1093/jamia/ocv150

Khatib, F., Desfosses, A., Players, F., Koepnick, B., Flatten, J., Popović, Z., et al. (2019). Building de novo cryo-electron microscopy structures collaboratively with citizen scientists. *PLoS Biol.* 17:e3000472. doi: 10.1371/journal.pbio.3000472

Kirsch, I. S. (1993). Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey. US Government Printing Office, Superintendent of Documents. Washington, DC 20402 (Stock No. 065-000-00588-3).

Kravitz, R. I., and Duan, N. (eds.) (2014). The DEcIDIE Methods Center N-of-1 Guidance Panel. (2014. Design and implementation of N-of-1 Trials: A User's Guide. AHRQ Pub No 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality. Available online at: www.effectivehealthcare.ahrq/gov/N-of-1Trials.cfm

Kuntsche, E., and Labhart, F. (2013). Using personal cell phones for ecological momentary assessment: an overview of current developments. *Eur. Psychol.* 18, 3–11. doi: 10.1027/1016-9040/a000127

Levy, R. (2005). Collective intelligence, a civilization: towards a method of positive interpretation. *Int. J. Polit. Cult. Soc.* 18, 189–198. doi: 10.1007/s10767-006-9003-z

Li, B., Lee-Urban, S., Scott Appling, D., and Riedl, M. O. (2012). Crowdsourcing narrative intelligence. *Adv. Cog. Systems* 1, 1–18.

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., Schork, N. J., et al. (2011). The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Person. Med.* 8, 161–173. doi: 10.2217/pme.11.7

Lobo, M. A., Moeyaert, M., Cunha, A. B., and Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *J. Neurol. Phys. Ther.* 41:187–197. doi: 10.1097/NPT.0000000000000187

Marr, B. (2017). *What is Digital Twin Technology and Why is it so Important. Forbes,* Available online at: https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important/#7c3690342e2a

McAfee, A. (2010). *Why Crowdsourcing Works.* Available online at: http://forbes.com/2010/08/23/science-proteins-foldit-technology-breakthrough-crowdsourcing.html

McMenamin, M., Berglind, A., and Wason, J. M. S. (2018). Improving the analysis of composite endpoints in rare disease trials. *Orphanet J. Rare Diseases* 13:81. doi: 10.1186/s13023-018-0819-1

Michie, S., Yardley, L., West, R., Patrick, K., and Greaves, F. (2017). Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *JMIR* 19:e232. doi: 10.2196/jmir.7126

Mirza, R. D., Punja, S., Vohra, S., and Guyatt, G. (2017). The history and development of the N-of-1 trials. *J. R. Soc. Med.* 110, 330–340. doi: 10.1177/0141076817721131

NEJM Catalyst (2017). Social Determinants of Health (SDOH). *NEJM Catalyst.*

Noar, S. M., Benac, C. N., and Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychol. Bull.* 133, 673–693. doi: 10.1037/0033-2909.133.4.673

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 447–453. doi: 10.1126/science.aax2342

Onnela, J. P., and Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 41, 1691–1696. doi: 10.1038/npp.2016.7

Parker, R. I., and Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *J. Behav. Educ.* 21, 254–265. doi: 10.1007/s10864-012-9153-1

Parker, R. I., Vannest, K. J., and Davis, J. L. (2011a). Effect side in single-case research: a review of nine nonoverlap techniques. *Behav. Modif.* 4, 1–20. doi: 10.1177/0145445511399147

Parker, R. I., Vannest, K. J., Davis, J. L., and Sauber, S. B. (2011b). Combining Nonoverlap and trend for single case research: Tau-U. *Behav. Therapy* 42, 284–299. doi: 10.1016/j.beth.2010.08.006

Pertschuk, M. J., Edwards, N., and Pomerleau, O. F. (1978). A multiple-baseline approach to behavioral intervention in anorexia nervosa. *Behav. Therapy* 9, 368–376. doi: 10.1016/S0005-7894(78)80080-X

Pham, Q., Wiljer, D., and Cafazzo J. A., (2016). Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. *JMIR* 4:e107. doi: 10.2196/mhealth.5720

Ranard, B., Ha, Y., Meisel, Z., Asch, D., Hill, S., Becker, L., et al. (2013). Crowdsourcing – Harnessing the masses to advance health and medicine, a systematic review. *J. General Internal Med.* 29, 187–203. doi: 10.1007/s11606-013-2536-8

Richter, T., Nestler-Parr, S., Babela, R., Khan, Z. M., Tesoro, T., Molsen E., et al. (2015). Rare disease terminology and definitions—a systematic global review: report of the ISPOR rare disease special interest group. *Value Health*. 18, 906–914. doi: 10.1016/j.jval.2015.05.008

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edn.* New York, NY: Simon & Schuster.

Roskams, J., and Popovic, Z. (2016). Power to the people: addressing big data challenges in neuroscience by creating a new cadre of citizen neuroscientists. *Neuron* 92:45. doi: 10.1016/j.neuron.2016.10.045

Runyan, J. D., and Steinke, E. G. (2015). Virtues, ecological momentary assessment/intervention and smartphone technology. *Front. Psychol.* 6:481. doi: 10.3389/fpsyg.2015.00481

Sanger, M., McNeil, A., Puska, P., Auffray, C., Price, N. D., Hood, L., et al. (2016). The P4 health spectrum–A perspective, preventive, personalized, and participatory continuum for promoting healthspan. *Progress Cardiov. Diseases* 59, 506–521. doi: 10.1016/j.pcad.2016.08.002

Schork, N. J. (2015). Time for one-person trials. *Nature* 520, 609–611. doi: 10.1038/520609a

Smyth, J. M., and Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *J. Happiness Stud.* 4, 35–52. doi: 10.1023/A:1023657221954

Spoelmann, W. A., Bonten, T. N., de Waal, M. W. M., Drenthen, T., Smeele, I. J. M., Nielen, M. M. J., et al. (2016). Effect of an evidence-based website on healthcare usage: an interrupted time-series study. *BMJ Open* 6:e013166. doi: 10.1136/bmjopen-2016-013166

Strathmann, F. G. (2015). N-of-1 clinical trials: removing the hay to find the needle. *Clin. Chem.* 61:5928. doi: 10.1373/clinchem.2015.245928

Strecher, V. J., McClure, J., Alexander, G., Chakraborty, B., Nair, V., Konkel, J., et al. (2008). Web-based smoking cessation program: results of a randomized trial. *Am. J. Preventive Med.* 34, 373–381. doi: 10.1016/j.amepre.2007.12.024

Tao, F., and Qi, Q. (2019). Make more digital twins: virtual models boost smart manufacturing by simulating decisions and optimization, from design to operations. *Nature* 573, 274–277. doi: 10.1038/d41586-019-02849-1

Tattersall, R. (2002). The expert patient: a new approach to chronic disease management for the twenty-first century. *Clin. Med.* 2, 227–229. doi: 10.7861/clinmedicine.2-3-227

Torous, J., Andersson, G., Bertagnoli, A., Christensen, H., Cuijpers, P., Firth, J., et al. (2019). Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry* 18, 97–98. doi: 10.1002/wps.20592

U.S. Department of Health and Human Services. (2010). *Office of Disease Prevention and Health Promotion. National Action Plan to Improve Health Literacy*. Washington, DC. Available online at: https://health.gov/communication/hlactionplan/pdf/Health_Literacy_Action_Plan.pdf

Versluis, A., Verkuil, B., Spinhoven, P., van der Ploeg, M. M., Brosschot, J. F., et al. (2016). Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. *JMIR* 18:e152. doi: 10.2196/jmir.5642

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., et al. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ*. 350:h1738. doi: 10.1136/bmj.h1738

Von Hippel, E. (1986). Lead users: a source of novel product concepts. *Managem. Sci.* 32, 791–805. doi: 10.1287/mnsc.32.7.791

Wicks, P. (2018). Patient, study thyself. *BMC Med.* 16:2. doi: 10.1186/s12916-018-1216-2

Wolf, G. (2007). *Quantified Self*. Available online at: https://www.webcitation.org/66TEHdz4d

World Health Organization (2020). *Disability Inclusion*. Available online at: https://www.worldbank.org/en/topic/disability

Worldwide Web Consortium (3Wc) (2019). Web Content Accessibility Guidelines (WCAG) 2.1. Available online at: https://www.w3.org/TR/WCAG21/

Yoon, H., Jang, Y., Kim, S., Speasmaker, A., and Nam, I. (2020). Trends in internet use among older adults in the United States, 2011-2016. *J. Appl. Gerontol.* doi: 10.1177/0733464820908427. [Epub ahead of print].

Conflict of Interest: SS was employed by IndividuALLytics and Diplomat. KW was employed by Janssen. AB was employed by Mad*Pow. BB was employed by IndividuALLytics.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership