

A framework for self-experimentation in personalized health

RECEIVED 11 May 2015

REVISED 6 August 2015

ACCEPTED 29 August 2015

PUBLISHED ONLINE FIRST 7 December 2015

Ravi Karkar¹, Jasmine Zia², Roger Vilardaga³, Sonali R Mishra⁴, James Fogarty¹, Sean A Munson⁵, Julie A Kientz⁵



OXFORD
UNIVERSITY PRESS

ABSTRACT

Objective To describe an interdisciplinary and methodological framework for applying single case study designs to self-experimentation in personalized health. The authors examine the framework's applicability to various health conditions and present an initial case study with irritable bowel syndrome (IBS).

Methods and Materials An in-depth literature review was performed to develop the framework and to identify absolute and desired health condition requirements for the application of this framework. The authors developed mobile application prototypes, storyboards, and process flows of the framework using IBS as the case study. The authors conducted three focus groups and an online survey using a human-centered design approach for assessing the framework's feasibility.

Results All 6 focus group participants had a positive view about our framework and volunteered to participate in future studies. Most stated they would trust the results because it was their own data being analyzed. They were most concerned about confounds, nonmeaningful measures, and erroneous assumptions on the timing of trigger effects. Survey respondents ($N=60$) were more likely to be adherent to an 8- vs 12-day study length even if it meant lower confidence results.

Discussion Implementation of the self-experimentation framework in a mobile application appears to be feasible for people with IBS. This framework can likely be applied to other health conditions. Considerations include the learning curve for teaching self-experimentation to non-experts and the challenges involved in operationalizing and customizing study designs.

Conclusion Using mobile technology to guide people through self-experimentation to investigate health questions is a feasible and promising approach to advancing personalized health.

Keywords: individualized medicine, self-experimentation, irritable bowel syndrome, human-centered design, technology

BACKGROUND AND SIGNIFICANCE

Understanding individual variation and treating individual needs is important in medicine and clinical science, an approach also known as *personalized medicine*.¹ Although this term historically emphasized genetics and pharmacology, it is now expanding to include other areas of health and disease,² recapturing the importance of personalized health.

Knowledge in clinical science primarily originates from the use of group designs, such as epidemiological surveys, longitudinal studies, and randomized controlled trials. Such methods provide a good understanding of the epidemiology, clinical course, and effects of specific treatments for certain medical conditions. Despite their merit, these traditional methods cannot address the question most relevant to any given individual: what is the likelihood that a treatment, whose average effect is well-documented, will have an effect on the symptoms of that specific individual? A novel methodological framework is needed for personalized health.

Single case designs (SCDs), sometimes referred to as *n-of-1* trials, can potentially advance personalized health.³ In this type of design, an individual serves as their own control, highlighting an individual's response to an intervention rather than a group's. Their contribution to more rapid, responsive, and relevant health research can be rather advantageous.⁴ Presently however, SCDs are not widely applied to personalize health nor has such personalization been scaled to larger groups of individuals.

Motivated by these gaps, we are pursuing an interdisciplinary research program to develop and test a methodological framework for SCDs in health self-experimentation. We view self-experimentation as a subset of self-tracking, but one that provides much-needed improvement in methodological rigor.⁵ This paper details components of this framework and presents a case study: a mobile application supporting self-experimentation for people with irritable bowel syndrome (IBS). We also discuss opportunities and challenges with this framework and identify areas of future work.

A FRAMEWORK FOR SELF-EXPERIMENTATION IN PERSONALIZED HEALTH

Individuals can currently access countless websites, applications, and sensing technologies intended to improve personal health (e.g., Fitbit, MyFitnessPal, Weight Watchers, RunKeeper). However, neither widespread, sustained adoption nor the promised health benefits of these technologies has been realized.⁶ Although there is broad interest in using technology to track health information, people often lack scientific rigor in their analyses.⁵ Analysis is also frequently done without consultation of healthcare providers. People seek answers to specific health questions, such as "does caffeine impact my sleep quality?" However, current tools support only data collection, not a systematic approach to answering such questions. For example, self-tracked data may suggest associations between sleep quality and caffeine, but it is impossible to determine if caffeine is causing poor sleep quality or if a

Correspondence to Julie A Kientz, Associate Professor, Human Centered Design & Engineering, University of Washington, Sieg 423C, Box 352315, Seattle, WA, 98195, USA Office Phone: 206-221-0614

© The Author 2015. Published by Oxford University Press on behalf of the American Medical Informatics Association. All rights reserved. For Permissions, please email: journals.permissions@oup.com For numbered affiliations see end of article.

person is consuming more caffeine because they are tired. Such uncertainty may prevent people from making lifestyle changes that can lead to improved health outcomes (e.g., eliminating caffeine).

We believe our framework can support everyday people in successfully applying self-experimentation to understand the cause of their symptoms and possibly take effective action. Although self-experiments may be more complex than simple self-tracking, it is our hope that this framework can reduce the burdens and challenges of tracking through targeted data collection while also providing more rigorous and concrete answers to specific health questions.

Technology for self-experimentation fits into a larger process of personalized health (Figure 1a). Traditional self-tracking methods and correlational approaches, such as food journals or fitness trackers, can be used to generate hypotheses (Step 1). Self-experimentation technology then robustly tests those hypotheses (Step 2). A person can then use findings to target the most appropriate health behavior change to address their needs (Step 3).

This paper focuses on Step 2 of this process (Figure 1b). Self-experiments begin with identifying hypotheses an individual wants to test (e.g., “does caffeine impact my sleep quality?”), then proceed with systematically testing hypotheses until results can support a person in making a decision about their health (e.g., “should I eliminate caffeine from my diet?”). This process includes defining *independent variables* (e.g., causes, triggers) and the *dependent variables* they may affect (e.g., symptoms, health outcomes). A person then conducts a multi-day self-experiment where they are randomly assigned to either apply the independent variable (e.g., drinking at least 100 mg of caffeine) or not (e.g., avoiding caffeine). Dependent variables are measured throughout the experiment (e.g., subjective sleep quality on a 5-point scale). Data is then analyzed and visualized using techniques

suitable for single-case study designs, yielding a confidence value for the self-experiment (e.g., “there is strong confidence ($P < .05$) that drinking caffeine reduces sleep quality by half”). A person can then review results to determine if they are compelling enough to make health behavior changes.

Self-Experiment Study Designs and Analysis Methods

Comparing the effects of 2 or more interventions to groups of individuals, under controlled conditions, and the statistical analysis of the reliability of such effects has been referred as the “gold standard” for providing the best clinical evidence.⁷ These designs, also known as group randomized controlled trials, can provide a rational process to extrapolate the generalizability of certain interventions. However, this claim is based on the assumption that individuals are randomly selected from a population, a condition rarely met in clinical studies. Even when randomly sampling a population is possible, population-based estimates do not provide specific support for how individuals and their symptoms will respond to certain interventions.⁸ SCDs can possibly offer a solution,^{3,9} with Neuringer first proposing the idea of using SCDs for self-experimentation in 1981.¹⁰ Our research advances this vision in developing a framework for conducting self-experiments and examining the application of that framework with IBS.

Although SCDs are traditionally used in the early stages of an intervention's development to determine its feasibility and estimate its potential effects, our framework proposes using SCDs at the later stages of the research cycle continuum, after randomized controlled trials are conducted and general guidelines for successful treatments are proposed (Figure 2).¹¹ Using SCDs at the later stage can bridge the gap between average treatment effects for groups and specific treatment effects for individuals. SCDs can thus improve health interventions by

Figure 1: (a) Overall process of personalized health framework. (b) Expanded view of the self-experimentation process of our framework.

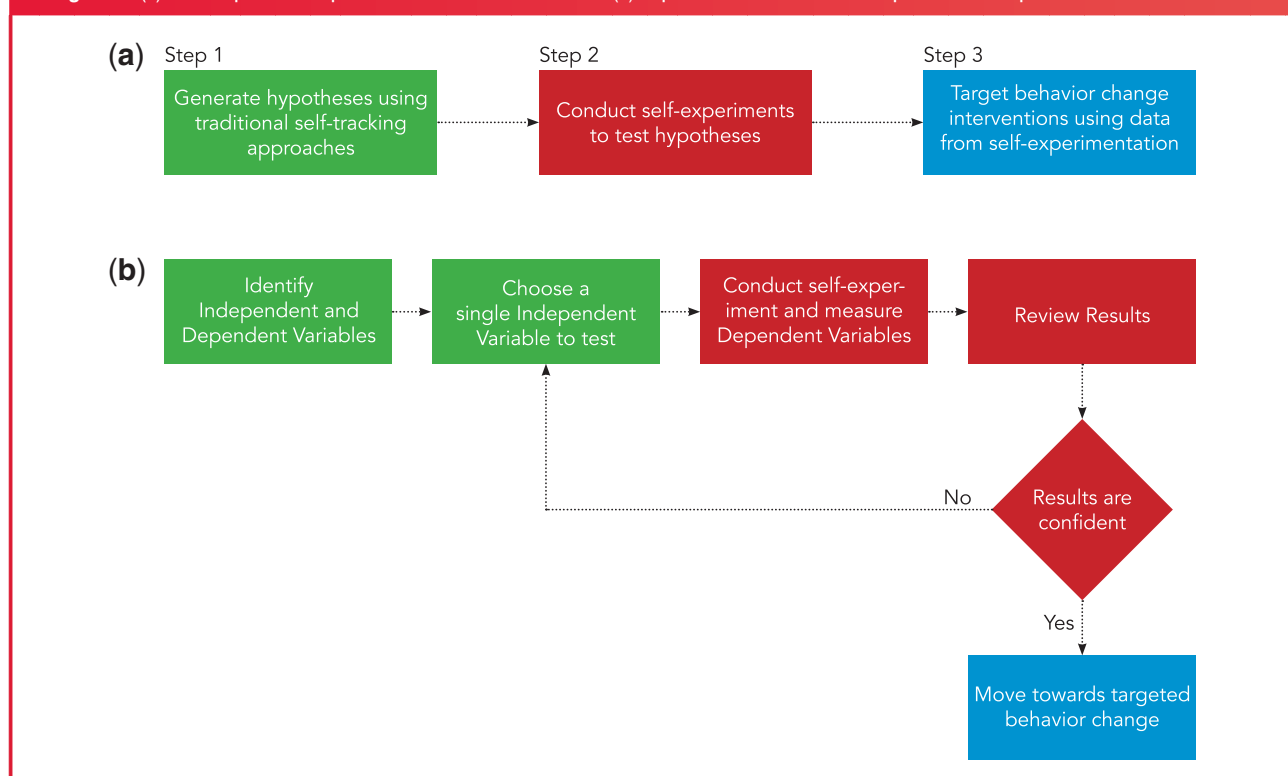
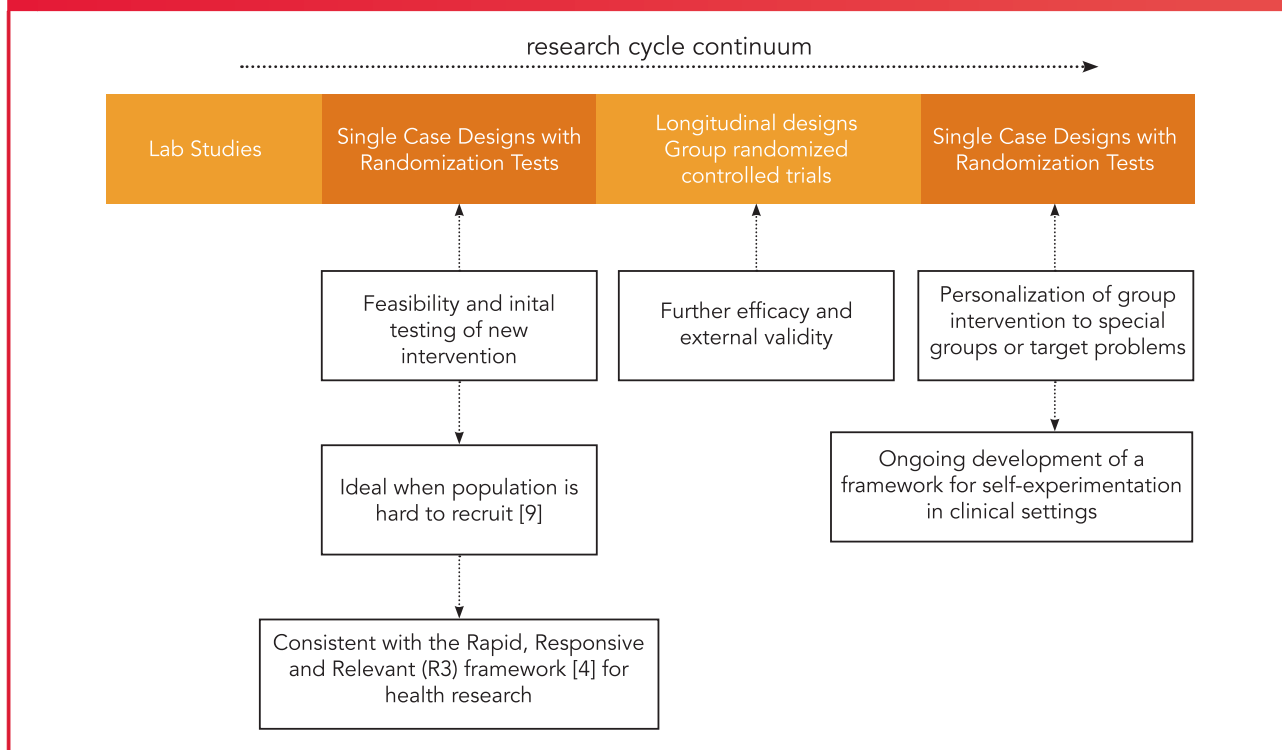


Figure 2: Single case designs' role within the research cycle continuum.

providing more definitive, rigorous, and actionable guidance to individuals.

However, traditional SCDs have several limitations. First, according to some methodologists, their *internal validity* is questionable because decisions about the stability of a baseline (Phase A) before implementation of an intervention (Phase B) are not based on randomization.¹² Second, statistical inference with these methods has been challenging until most recently [see ^{13–16}, for most recent developments in quantitative analysis of SCDs], so visual analysis has been the primary method of evaluating data from single-case experiments.^{17,18} Third, SCDs are criticized for not providing *population-based estimates* for effect size of an intervention. Finally, SCDs in clinical settings will typically generate measurements during face-to-face contact with patients (e.g., weekly counseling sessions), producing very *limited number of observations*.

These limitations can be overcome by applying randomization tests to SCDs, where a random assignment is given to a population of occasions rather than a population of individuals.^{19,20} For example, in a traditional group design, each individual who belongs to a group of 100 people is assigned to treatment A or B. Instead, in a SCD with randomization tests, each measurement in a group of 100 measurement occasions is assigned to treatment A or B. After the data are obtained, a permutation procedure can be used to create all possible combinations of treatment exposures (A, B) and outcome measures and render a *P*-value indicating the probability of the null hypothesis (i.e., no differences between treatment A and B exposures). This procedure, originally envisioned by Fisher in the 1930s, ensures the same internal validity as group experiments and eliminates the statistical assumptions of parametric tests (i.e., normally distributed sample, homoscedasticity, independence of errors).²¹ The result is a statistical test that can be used in both individuals and groups of individuals. Its primary limitation remains that it does not provide *external validity*. This limitation is not applicable in our case, as we use it to personalize known

group-based or clinical guidelines to specific individuals. Lastly, advances in mobile technology now allow more frequent and ecologically valid measurements,²² reducing challenges regarding the need for face-to-face contact. Thus, SCDs with randomization tests overcome many limitations of traditional SCDs.

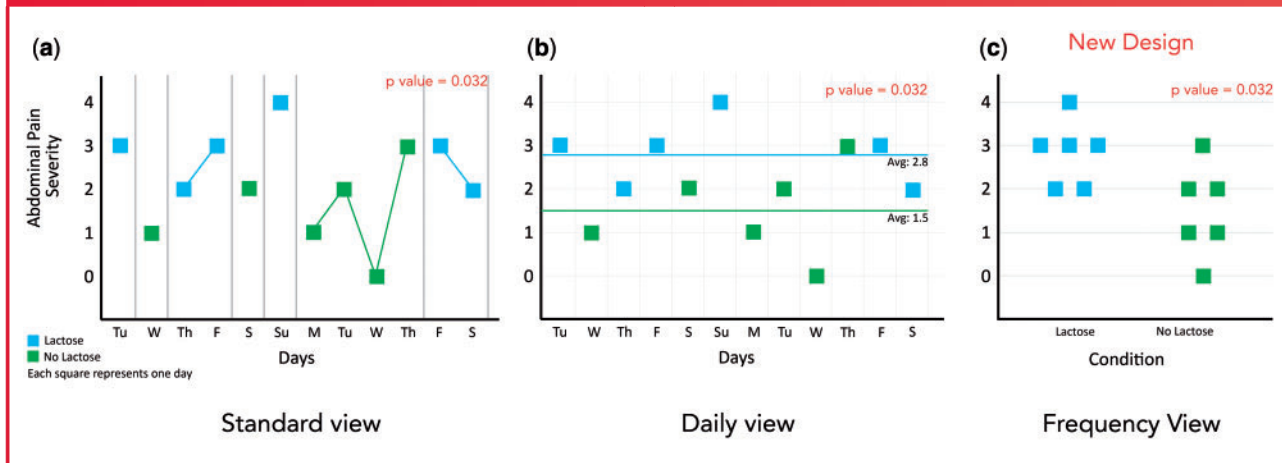
In the framework we propose, the choice of a specific SCD depends on: 1) the nature of *dependent variables*, 2) the lag effect of *independent variables*, 3) the statistical power of the design, and 4) effective human-centered design of the technology. For example, completely randomized SCDs, a form of alternation design,¹⁰ can be combined with randomization tests to provide the highest level of statistical power, but they require an independent variable with quick effects and the absence of carry over effects.²³ In the case study section, we describe the preparation of a study that required completely randomized SCDs, which is the best fit for IBS.

SCDs are often visually analyzed to look for trends and infer relationships, but a completely randomized SCD can make it challenging to apply visual analysis in identifying trends. Conditions (e.g., lactose or no lactose) can be distributed in any order, with no fixed phase lengths. It can therefore be difficult to find patterns across varying phase lengths. To overcome this we designed a new visualization, inspired from a violin plot,²⁴ that removes the temporal information and instead focuses on the distribution of the dependent variable (e.g., abdominal pain) across different conditions (Figure 3c).

Requirements for Health Conditions to Be Eligible for Self-Experimentation

We believe self-experimentation can be applied broadly across many health conditions, though some are better suited than others. Analyzing literature across different fields of health, we identify *absolute* and *desired* requirements of health conditions appropriate and ideal for this self-experimentation for personalized health framework (Table 1).

Figure 3: Example visualizations for a completely randomized single-case design of 12 observations showing a statistical significant effect of lactose on an individual's abdominal pain. (a) Standard view. (b) Proposed visualization with daily view. (c) Proposed visualization with frequency view.



DESIGNING A MOBILE APPLICATION FOR SELF-EXPERIMENTATION IN IBS: A CASE STUDY

This section describes how we have applied our self-experimentation framework to IBS using dietary triggers as the independent variable.

Overview of IBS

IBS is a chronic functional disorder characterized by episodic abdominal pain associated with diarrhea and/or constipation despite normal blood tests, X-rays, and colonoscopies. It affects up to 20% of the US population and is one of the top 10 reasons people seek primary care.^{28,29} People with IBS report a lower quality of life and consume 50% more healthcare resources than non-IBS counterparts.^{30,31}

There are many potential triggers for IBS symptom flare-ups: certain foods, eating behaviors, stress, sleep disturbances, and menstruation. The most common trigger is food, and thus we choose it as the independent variable for our initial case study. Elimination diets (e.g., low-fat, low-carbohydrate, gluten-free) surpass traditional medications in their effectiveness in reducing IBS symptoms, but are difficult to comply with.^{32–35} Fortunately, complete elimination of all known possible trigger foods is not necessary for most people. An individual's response to specific foods is variable, with a given food triggering bowel symptoms in some people but not others.²⁵

Current methods for identifying individualized trigger foods generally include a 2-week complete elimination diet followed by serial re-introduction of main IBS trigger foods.^{36–38} People are asked to simultaneously journal their food and IBS symptoms. This process is flawed for multiple reasons. First, journals are typically handwritten with incomplete, disorganized, and unreliable data.^{39,40} Important information such as meal time, food ingredients, and symptom severity are often missing because journaling is complex and high burden.⁴¹ Second, this process is complex and lengthy. It can take more than 3 months to complete an elimination-reintroduction diet, with no guarantee of results. Finally, there is no validated methodology for determining an individual's trigger foods. Clinicians do not receive formal training on how to review journals, and such interpretations result in a high degree of interobserver variability.⁴² Not surprisingly, most people with IBS are dissatisfied with the journal feedback they receive from healthcare professionals.²⁵

Designing a Mobile Application for Self-Experimentation in IBS

IBS patients and their providers need more efficient and effective methods to determine individualized food triggers for symptom reduction and improved quality of life. Following a human-centered design approach, we created a mobile application prototype that can support self-experimentation for people with IBS. Although we believe self-experimentation can be applied more broadly across many conditions, working with a specific population and concrete variables helps ground our design toward a real solution.

To design the application, we assembled a team of researchers in medicine, behavioral psychology, computer science, and human-centered design to work with people suffering from IBS using an iterative human-centered design process.⁴³ Through multiple rounds of iteration, we generated process flows, storyboards, and prototypes of a mobile application.

Our initial application focuses on the set-up and deployment of self-experiments (Step 2 in Figure 1a). In our process, individuals generate a list of hypotheses that narrow which trigger foods to test (i.e., independent variables). The application will use relevant medical knowledge, the individual's personal experience, and/or the expertise of a medical provider to help generate testable hypotheses (storyboard in appendix). The individual then configures a self-experiment by setting their personalized symptoms (i.e., dependent variables). They also choose a start date, a study duration, and a time for daily reminders to enter symptoms.

When self-experimentation begins, the application shows the individual a schedule of which days to avoid and which days to consume the experimental trigger food, following a completely randomized SCD as previously described. The application shows sample meals for trigger and non-trigger food days (i.e., experimental and control). The individual is instructed to eat an otherwise consistent meal during each day, varying only the food component being tested. During each day of the experiment, the individual is also instructed to log peak symptom severity using a subjective scale at a defined time after eating the test meal (e.g., 4 hours). At the end of the experiment, findings are summarized and interpreted (Figure 4). The results include a *P* value on the likelihood the trigger food is causing IBS symptoms by chance. The individual may choose to re-run the experiment with a different possible trigger, or share the results with their medical provider for recommendations on how to avoid this trigger.

Preliminary Evaluation: Human-Centered Design Research and Findings

To learn about past experiences of identifying potential gastrointestinal food triggers and assess the feasibility of implementing our self-experimentation framework, we conducted focus groups and an online survey. We recruited participants from primary care and gastroenterology clinics associated with an academic center (University of Washington Medical Center, Seattle, WA, United States). Survey respondents were also recruited through social networks.

Focus Groups

We conducted three focus groups with people with IBS, totaling 6 participants (5 female, 1 male). To gather reactions to the self-experimentation process and how self-experimentation fits with participant priorities, focus groups included walkthroughs of the overall process using our flowchart (Figure 1b) and storyboard (see Appendix).^{44–46} To elicit reactions to our interface designs, focus groups included a click-through of a prototype of our mobile application (Figure 5).^{47,48} Questions focused on feasibility of the framework, understandability of the process, interface, and visualizations, and overall usability and usefulness of the application.

Survey

We complemented the focus groups with an online survey targeted toward people experiencing any type of gastrointestinal food intolerances, which commonly overlaps with the diagnosis of IBS (see supplemental documentation). Questions were internally tested for face validity and focused on specific opinions raised during focus groups, such as how to rate symptoms and trade-offs in design (e.g., shorter studies vs more confident results) and on factual information about food intolerances (e.g., “What food intolerance symptoms do you experience? You can select more than one option”). We received 60 responses to the survey (53 female, 7 male), with 75% aged between 25 and 45 and 90% having a B.S. degree or higher.

Prior experiences.

Most focus group participants described feeling “overwhelmed and frustrated” with the trial-and-error process they used to help determine possible IBS triggers. They instead wanted more guidance during this process. Most had tried elimination diets as recommended by health providers, friends, family, and/or online research. Participants were concerned about labeling certain foods as triggers if their symptoms were more delayed and/or subtle, and many were also concerned about findings being confounded by nonfood factors (e.g., mood, stress, food preferences).

Overall reactions to the framework.

During focus groups, participants expressed excitement by pointing out that this was a better approach to finding out their triggers than what they had been trying so far. In their view, part of their excitement was due to the fact that the proposed design was showing them their data and not just presenting a vague number. This in turn boosted their confidence in the framework. Participants stated they could trust the results because they were rigorous and based on their own data, rather than an average of other people's data. One participant mentioned liking that the application was “honest about its limitations and shortcomings.”

Participants appreciated that our process gave them the decisional authority on whether or not to eliminate certain foods after weighing the magnitude of the effect, the confidence level, and their food preferences. Participants preferred this over a simple recommendation

(e.g., “eliminate lactose”). The application acted like a “decision assistant” instead of a “decision maker.” Most also said the application would give them more “credibility” and empower them to talk with their medical providers because self-experiments were conducted in a controlled manner.

Meal choice.

Participants requested an app with more detailed sample meals for both trigger and nontrigger days, as they feared erroneously consuming trigger foods on a nontrigger day. Focus group participants were also concerned about the assumption that trigger foods result in symptom onset within 4 hours. The majority of survey respondents (80%) did however agree that their symptoms occurred within 4 hours after eating. Participants were also concerned that trigger foods could result in symptoms lasting longer than 24 hours. They were also wary of foods being categorized either too broadly or narrowly (e.g., testing “nuts” vs “almonds”). These concerns suggest additional design opportunities such as experimenting with different symptom onset periods or iteratively categorizing foods.

Seventy percent of participants in our survey selected breakfast as their preferred test meal. Focus group participants shared part of their rationale behind this preference: they thought breakfast would be the most feasible mealtime to eat the same type of food and drink for the duration of the study, unless the trigger food was an unconventional breakfast food (e.g., alcohol). Participant response to our framework's requirement that independent variables being controllable and actionable therefore influenced test meal preferences. Further exploration should be conducted to determine how this influence might affect the framework's efficacy.

Recording symptoms.

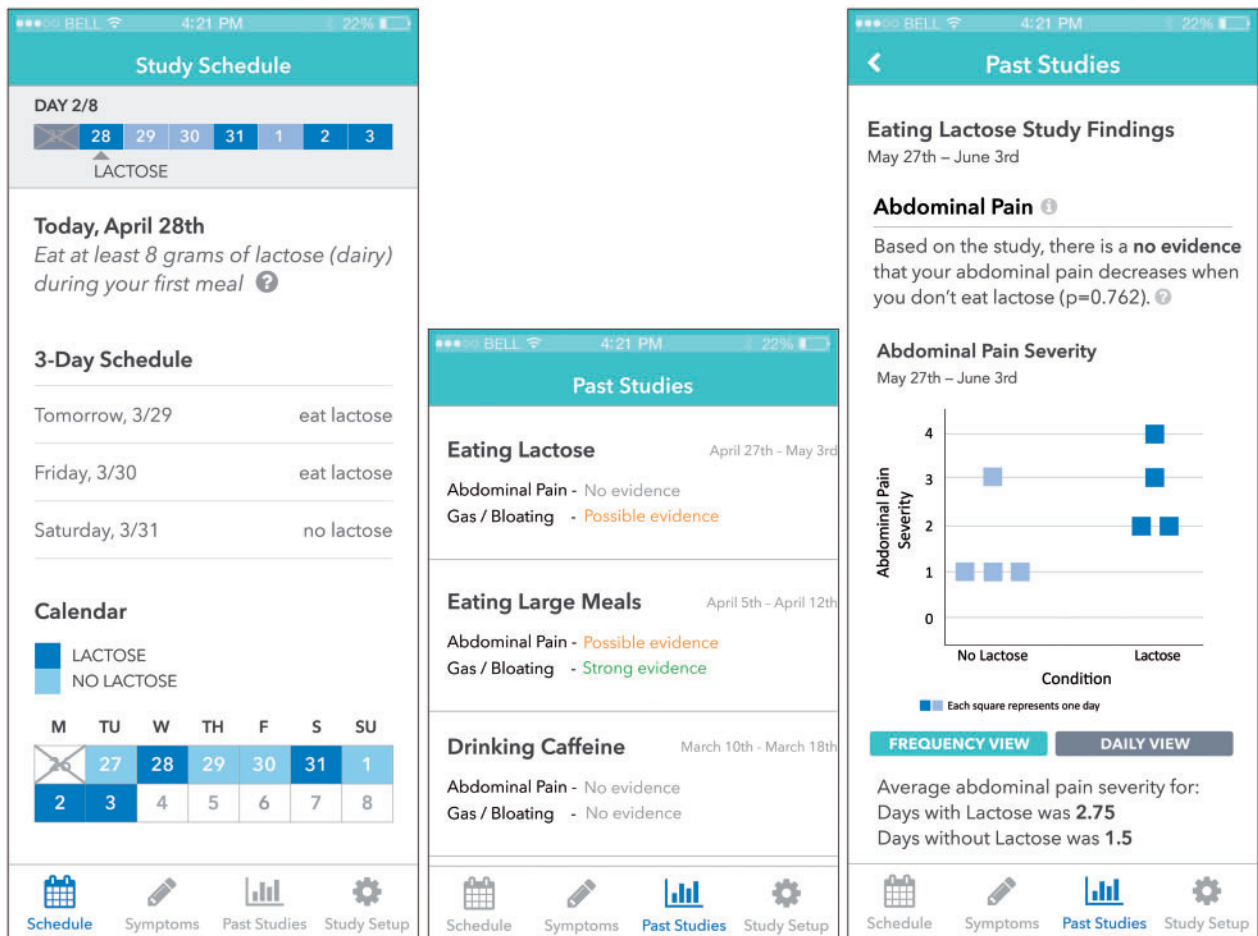
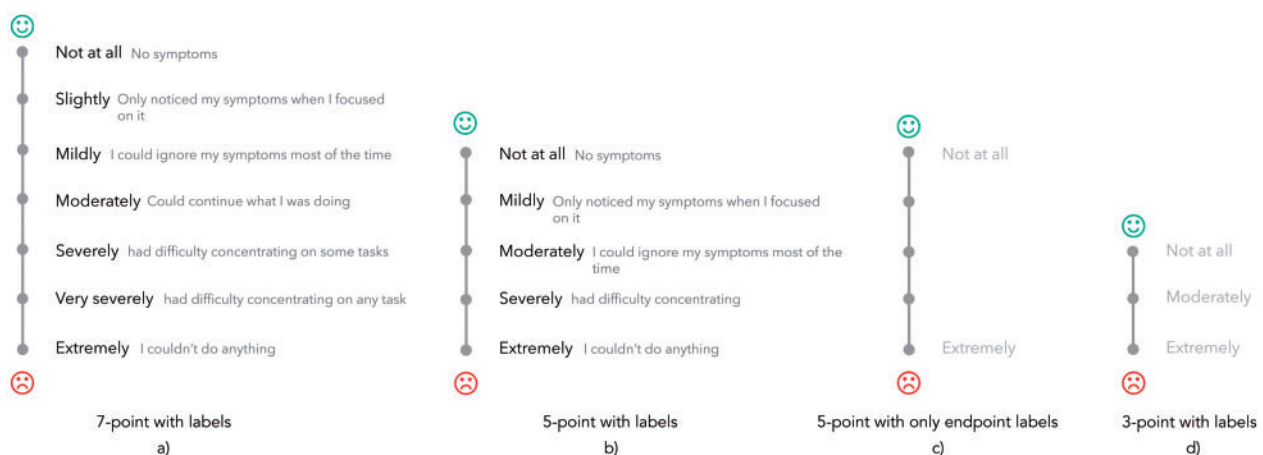
Focus group participants were concerned about inconsistent symptom ratings over time. The symptom scales allowed “too much room for interpretation.” They expressed the desire to customize symptom scales to be personally meaningful. We explored this issue further in the survey. We first asked the survey participants to generate their own custom scale with labels, then offered them a selection of four scales to choose from: 7-point with labels, 5-point with labels, 5-point with only endpoint labels, and 3-point with labels (Figure 5). Participants developed custom scales ranging from 4 to 10 points. Fifty-seven percent preferred one of our predefined scales over their custom scale whereas 25% preferred their custom scale. Of the 57%, 47% preferred scale 6b.

Duration of study and meal choice.

Although focus group participants felt they could adhere to self-experiments, most preferred shorter 8-day study, even when told that a longer 12-day study would likely result in higher confidence levels. In the survey, 83% said they would be “extremely likely” or “likely” to complete an 8-day study without giving up or missing days, compared to 67% for a 12-day study.

Understanding results.

All focus group participants preferred the frequency visualization over the standard and daily visualization designs (refer to Figure 3). However, participants also wanted access to the standard or daily visualizations. Participants felt the calendar associated with these could help them recall details of individual measures and events leading up to them. For all visualizations, participants valued seeing individual data points, as each represents a specific measure to which they

Figure 4: Screenshots of pages from our mobile app prototype. From left to right: study schedule, result summary of past studies, and results of a study.**Figure 5:** Selection of predefined scales shown during the online survey. From left to right: 7-point scale with labels, 5-point scale with labels, 5-point scale with only endpoint labels, and 3-point scale with labels.

could relate. Because data corresponded to each individual, and not to a group of other IBS patients, participants felt more trust that summary statistics described their situation. Participants with stronger backgrounds in statistics preferred interpreting the graphs over the summary statistics because the graphs allowed them to better identify and understand the impact of outliers. All participants expressed a desire for an easy way to share and communicate results with their health-care providers.

DISCUSSION AND FUTURE CHALLENGES

The results from our focus groups and survey both provide insight into the feasibility of our self-experimentation framework and the design of our mobile application. Despite this evidence of initial feasibility, our self-experimentation framework is still in its initial stages of development and must overcome a number of future hurdles. First, the feasibility and clinical efficacy of applications using our framework must still be evaluated across a larger sample of individuals with IBS and then other health conditions. We are currently finalizing the development of our mobile application for an initial feasibility study in people with IBS. If successful, we will generalize the framework to other health conditions and continue subsequent testing. Second, self-experimentation is more difficult to understand than simple journaling. In our focus groups, it was certainly more complicated to explain the self-experimentation process and rationale to participants than it would have been to simply tell them to record all of the food that they eat. However, after participants understood the process, they believed our framework would be much easier to adhere to than traditional methods, especially with the guidance of the mobile application. Third, a lack of control over confounds can bias and diminish the effects of self-experimentation. Unlike laboratory or more controlled clinical studies, our framework has an emphasis on “real-world” deployment thus may diminish its scientific rigor.

Despite these hurdles, self-experimentation has the potential to improve existing tracking methods across multiple health conditions and even non-health domains. Our self-experimentation framework can be applied to any condition or situation that meets the absolute requirements (Table 1). We can extend our framework to other health conditions with possible dietary triggers such as gastroesophageal reflux disease, migraines, and gout. Our self-experimentation framework can also be applied in mental health and addiction, where the specific skills or behavioral strategies that help individuals meet their goals varies by person. Insomnia is another possible application of our framework, wherein individuals identify behavioral or environmental variables impacting their sleep. Our framework can even be applied to nonhealth conditions such as increasing work productivity, testing whether a new habit saves energy, or evaluating if a skincare product acts as advertised.

Much work remains to be done in expanding to new domains. At this point, it is not clear how much customization is required for each new health condition. Additional challenges can arise in adapting to conditions where the desired requirements are not met (e.g., conditions where the duration of the independent variable's effect is not defined). We ultimately hope to support people and their healthcare providers with customizable experiments, but first need to explore how to support people in planning valid self-experiments and understanding trade-offs between study design choices. Our initial work on defining absolute and desired requirements (Table 1) and eligible independent variable types (Table 2) has helped establish the trajectory for this expansion, but more work is needed on how to translate these issues to everyday people.

Table 1: Table of *absolute* and *desired* requirements for health conditions to which our self-experimentation framework can be applied.*

Absolute Requirements	Case Study: Irritable Bowel Syndrome (IBS)
People must be uncertain about the effect of the independent variable on the dependent variable(s)	Most people with IBS are uncertain about their personalized food triggers ²⁵
Independent variables must be: A. Controllable and actionable B. Well-specified	People with IBS are able to consume or not consume certain foods, including their amounts, during specific times of the day
Dependent variables of condition being tested must: A. Recurrent episodes or flare-ups B. Follow the application of the independent variable within a defined period C. Be quantifiable and measurable	A. IBS symptoms (abdominal pain, bloating, diarrhea and constipation) vary in daily intensity and/or flare-up intermittently for set periods of time B. The time period between eating a trigger food and onset of symptoms typically occurs within a 3-hour window for most people with IBS ²⁶ C. IBS symptoms can be quantified and measured using self-reported scales
Independent and dependent variables must not result in any serious health risks (immediate and/or long-term)	Exposing people with IBS to potential triggers will not result in any life-threatening consequences and/or immediate or long-term serious health risks ²⁷
Desired Requirements	Case Study: Irritable Bowel Syndrome (IBS)
Independent variables should be: A. Low-burden for people B. Reliably applied in the same way C. Easy to apply frequently	A. Altering people's diet is relatively low-burden B and C. It can be done frequently and with minimal daily variation.
Dependent variables should be: A. Tolerable for people to endure B. Easy to measure frequently C. Measured consistently over time	A and B. Symptoms can be measured frequently and consistently over time with well-defined scales. C. Although unpleasant, enduring provoked symptoms should generally be tolerable for most people with IBS.
Duration of the independent variable's effect on dependent variables should be defined	The exact duration of trigger foods on IBS symptoms is unknown at this time but assumed to be no more than 3 days (unpublished focus group)

*For each requirement, we give an example from our IBS case study. We also identify *classes of independent variables* and assess their eligibility for our framework (Table 2).

Table 2: Classes of independent variables and their eligibility for the self-experimentation framework.

Independent Variable	Example Health-Related Question	Eligibility
Dietary	Does drinking coffee give me heartburn?	Eligible
Actionable Behavior	Does exercising in the morning after waking up give me more energy later in the day?	Eligible
Medications/Supplements		
As needed	Does this inhaler reduce my cough?	Eligible
Daily	Is my antidepressant improving my mood?	Ineligible
Environmental		
Controllable	Does elevating the head of my bed reduce my heartburn?	Eligible
Uncontrollable	Do I get headaches every time it rains?	Ineligible
Illness/pathogen	Is my mood worse when I'm coming down with the flu?	Ineligible
Combination of variables	Do I sleep better if I both exercise early in the day and stop working one hour before bedtime?	Eligible

CONCLUSION

This paper presents the background and conceptual foundation of a new framework for self-experimentation with person-generated health and wellness data. We described the methodological framework, its advantages over other proposed methods, the conditions in which this framework is applicable, a case study applying this framework to a specific health condition, and initial reactions among people with IBS and gastrointestinal food intolerances.

Understanding individual variation is an essential aspiration in clinical science and medicine. The framework proposed here combines advanced statistical methods and SCDs to assist individuals in their own process of self-experimentation. Developing such a framework requires an interdisciplinary team of researchers that combines knowledge in medicine, behavioral psychology, computer science, and human-centered design. If successful, our framework will provide for increased efficiency and rigor in person-generated health data and will help extend the effectiveness and reach of existing treatments to those in most need.

FUNDING

This work was funded by the University of Washington Innovation Award, the Intel Science & Technology Center on Pervasive Computing, Nokia Research, the National Science Foundation grant numbers OAI-1028195 and SCH-1344613, the Agency for Healthcare Research Quality grant number 1R21HS023654, and the National Institute on Drug Abuse grant number 1K99DA037276-01.

COMPETING INTERESTS

The authors have no competing interests to declare.

CONTRIBUTIONS

Conception and design of framework R.K., J.Z., R.V., S.R.M., J.F., S.A.M., J.K.; design and development of new visual analysis tools R.K., J.Z., R.V., J.F., S.A.M., J.K.; focus group study design, implementation, and analysis of data R.K., J.Z., R.V., J.F., S.A.M., J.K.; survey design, critical revision, interpretation, and analysis of results R.K., J.Z., R.V., J.F., S.A.M., J.K.; drafting and critical revisioning of manuscript R.K., J.Z., R.V., S.R.M., J.F., S.A.M., J.K.; design of the mobile application R.K., J.Z., R.V., J.F., S.A.M., J.K.

ACKNOWLEDGEMENTS

We acknowledge Jonathan Cook, Yen Truong, and Jessica Schroeder for their assistance in the research and our study participants. This work was reviewed by the University of Washington Human Subjects Division.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at <http://jamia.oxfordjournals.org/>.

REFERENCES

- Jorgensen JT. New era of personalized medicine: a 10-year anniversary. *Oncologist*. 2009;14(5):557–558.
- Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *Int J Environ Res Public Health*. 2009;6:492–525.
- Lillie EO1, Patay B, Diamant J, Isseil B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med*. 2011;8(2):161–173.
- Riley WT, Glasgow RE, Etheredge L, et al. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. *Clin Transl Med*. 2013;2(1):10.
- Choe EK, Lee NB, Lee B, et al. Understanding quantified-selfers' practices in collecting and exploring personal data. *Proc. the 32nd Annual ACM Conference on Human Factors in Computing Systems*. ACM 2014;1143–1152.
- Dunne E. The unrealized potential of mHealth. *J Mobile Technol Med News*. 2014. <http://www.journalmtm.com/2014/the-unrealized-potential-of-mhealth/>. Accessed May 1, 2015.
- Meldrum ML. A brief history of the randomized controlled trial: from Oranges and Lemons to the gold standard. *Hematol/Oncol Clin North Am*. 2000;14(4):745–760.
- Pedhazur EJ, Schmelkin LP. *Measurement, Design, and Analysis. An Integrated Approach*. New Jersey: Erlbaum; 1991.
- Rizvi SL, Nock MK. Single-case experimental designs for the evaluation of treatments for self-injurious and suicidal behaviors. *Suicide Life Threat Behav*. 2008;38(5):498–510.
- Neuringer A. Self-experimentation: a call for change. *Behaviorism*. 1981;9(1):79–94.
- Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. *J Med Int Res*. 2013;15(2):e22.
- Kratochwill TR, Hitchcock JH, Horner RH, et al. Single-case intervention research design standards. *Rem Spec Educ*. 2013;34:26–38.
- Parker RI, Vannest KJ, Davis JL. Effect size in single-case research: a review of nine nonoverlap techniques. *Behav Modif*. 2011;35:303–322.
- Shadish WR, Hedges LV, Pustejovsky JE. An SPSS Macro for a d-Statistic for Single-Case Designs. *Workshop presented at the 2013 Spring Conference of the Society for Research on Educational Effectiveness*. Washington, D.C., USA. 2013.
- Maggin DM, Swaminathan H, Rogers HJ, et al. A generalized least squares regression approach for computing effect sizes in singlecase research Application examples. *J School Psychol*. 2011;49:301–321.
- Moeyaert M, Ugille M, Ferron J, Beretvas S, Van Den Noortgate W. Three-level analysis of single-case experimental data: empirical validation. *J Exp Educ*. 2013;82:1–21.

17. Kahng S, Chung K, Gutshall K, et al. Consistent visual analyses of intrasubject data. *J Appl Behav Anal.* 2010;43(1):35–45.
18. Hayes SC. Single case experimental design and empirical clinical practice. *J Consult Clin Psych.* 1981;49(2):193–211.
19. Edgington ES, Onghena P. *Randomization Tests.* Boca Raton: Chapman & Hall; 2007.
20. 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(1):51–64.
21. Fisher RA. *The Design of Experiments.* Oxford, England: Oliver & Boyd; 1935.
22. Vilardaga R, Bricker JB, McDonnell MG. The promise of mobile technologies and single case designs for the study of individuals in their natural environment. *J Contextual Behav Sci.* 2014;3(2):148–153.
23. Onghena P, Edgington ES. Customization of pain treatments: single-case design and analysis. *Clin J Pain.* 2005;21(1):56–68.
24. Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. *Am Stat.* 1998;52(2):181–184.
25. Jamieson AE, Fletcher PC, Schneider MA. Seeking control through the determination of diet: a qualitative investigation of women with irritable bowel syndrome and inflammatory bowel disease. *Clin Nurse Specialist.* 2007;21(3):152–160.
26. Eswaran S, Tack J, Chey WD. Food: the forgotten factor in irritable bowel syndrome. *Gastroenterol Clin North Am* 2011;40(1):141–162.
27. Canavan C, West J, Card, T. The epidemiology of irritable bowel syndrome. *Clin Epidemiol.* 2014;6:71–80.
28. Elsenbruch S. Abdominal pain in Irritable Bowel Syndrome: a review of putative psychological, neural and neuroimmune mechanisms. *Brain, Behav Immun.* 2011;25(3):386–394.
29. Lovell RM, Ford AC. Effect of gender on prevalence of irritable bowel syndrome in the community: systemic review and meta-analysis. *Am J Gastroenterol.* 2012;107:991–1000.
30. Mitra D, Davis KL, Baran RW. All-cause healthcare charges among managed care patients with constipation and comorbid irritable bowel syndrome. *Postgrad Med.* 2011;123(3):122–132.
31. Ladabaum U, Boyd E, Zhao WK, et al. Diagnosis, comorbidities, and management of irritable bowel syndrome in patients in a large health maintenance organization. *Clin Gastro Hepatol.* 2012;10(1):37–45.
32. Staudacher HM, Whelan K, Irving PM, et al. Comparison of symptom response following advice for a diet low in fermentable carbohydrates (FODMAPs) versus standard dietary advice in patients with irritable bowel syndrome. *J Hum Nutr Dietetics.* 2011;24(5):487–495.
33. Biesiekierski JR, Newnham ED, Irving PM, et al. Gluten causes gastrointestinal symptoms in subjects without celiac disease: a double-blind randomized placebo-controlled trial. *Am J Gastro.* 2011;106(3):508–514.
34. Halmos EP, Power VA, Shepherd SJ, et al. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. *Gastroenterology.* 2014;146(1):67–75.
35. Biesiekierski JR, Peters SL, Newnham ED, et al. No effects of gluten in patients with self-reported non-celiac gluten sensitivity after dietary reduction of fermentable, poorly absorbed, short-chain carbohydrates. *Gastroenterology.* 2013;145(2):320–328.
36. Barney P, Weisman P, Jarrett M, et al. *Master Your IBS: An 8-Week Plan Proven to Control the Symptoms of Irritable Bowel Syndrome.* Bethesda: American Gastroenterological Association; 2010.
37. Catsos P. *IBS Free at Last! Second Edition: Change Your Carbs, Change Your Life with the Fodmap Elimination Diet.* Portland: Pond Cove Press; 2012.
38. Shepherd S, Gibson P. *The Complete Low-FODMAP Diet: A Revolutionary Plan for Managing IBS and Other Digestive Disorders.* New York City: The Experiment; 2013.
39. Jhaveri M, Lee E. Performance of electronic diaries in diabetes clinical trials measured through overall satisfaction of site coordinators. *J Diabetes Sci Tech.* 2007;1(4):522–530.
40. Heinonen R, Luoto R, Lindfors P, et al. Usability and feasibility of mobile phone diaries in an experimental physical exercise study. *Telemed e-Health.* 2012;18(2):115–119.
41. Cordeiro F, Epstein DA, Thomaz E, Bales E, Jagannathan AK, Abowd GD, Fogarty J. Barriers and Negative Nudges: Exploring Challenges in Food Journaling. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 1159–1162.
42. Kueper T, Martinelli D, Konetzki W, et al. Identification of problem foods using food and symptom diaries. *Surgery.* 1995;112:415–420.
43. Maguire M. Methods to support human-centred design. *Int J Hum Comput Stud.* 2001;55(4):587–634.
44. Truong KN, Hayes GR, Abowd GD. Storyboarding: an empirical determination of best practices and effective guidelines. In: *Proceedings of the 6th conference on Designing Interactive systems*, ACM, 2006, pp. 12–21.
45. Haesen M, Meskens J, Luyten K, Coninx K. Draw me a storyboard: incorporating principles & techniques of comics. In: *Proceedings of the 24th BCS Interaction Specialist Group Conference*, British Computer Society, 2010, pp. 133–142.
46. Dow S, Saponas TS, Li Y, Landay JA. External representations in ubiquitous computing design and the implications for design tools. In: *Proceedings of the 6th conference on Designing Interactive systems*, ACM, 2006, pp. 241–250.
47. Beyer H, Holtzblatt K. Contextual design. *Interactions.* 1999;6 (1):32–42.
48. Hartson R, Pyla PS. *The UX Book: Process and guidelines for ensuring a quality user experience.* Waltham, Massachusetts: Elsevier; 2012.

AUTHOR AFFILIATIONS

¹Computer Science & Engineering, University of Washington, Seattle, WA, USA

²Division of Gastroenterology, Medicine University of Washington, Seattle, WA, USA

³Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

⁴Information School, University of Washington, Seattle, WA, USA

⁵Human Centered Design & Engineering, University of Washington, Seattle, WA, USA