Hypothesis Formation and Hypothesis Testing: Design Challenges in Self-Experimentation

Ravi Karkar
University of Washington
Seattle, WA 98195, USA
rkarkar@cs.washington.edu

Julie A. Kientz
University of Washington
Seattle, WA 98195, USA
jkientz@uw.edu

Jessica Schroeder
University of Washington
Seattle, WA 98195, USA
jesscs@cs.washington.edu

Sean A. Munson
University of Washington
Seattle, WA 98195, USA
smunson@uw.edu

James Fogarty
University of Washington
Seattle, WA 98195, USA
jfogarty@cs.washington.edu

Jasmine Zia
University of Washington
Seattle, WA 98195, USA
jasmine@uw.edu

Abstract

Personal informatics applications often fail to effectively support diagnostic self-tracking, wherein people seek to use data in answering specific questions about themselves. Diagnostic self-tracking is therefore difficult, tedious, and error-prone. To address these shortcomings, we designed a framework for self-experimentation in personalized health that helps guide the design of applications to support diagnostic self-tracking. We instantiated multiple aspects of the framework in our work assisting people with irritable bowel syndrome (IBS) in identifying their individualized food-based triggers. In this workshop paper, we describe our experiences from two studies to highlight key findings surrounding validity and reliability of such data. We then propose discussion topics for the workshop.

Introduction

In their examination of quantified selfers, Choe et al. describe diagnostic self-tracking, or self-tracking to answer a specific question about oneself [1]. One way to help people find the answers to such questions is by supporting them through a *self-experimentation* process.

We developed a framework for self-experimentation in personalized health [3], modeling self-experimentation as a three-step process: (1) formulating a hypothesis, (2) testing a hypothesis, and (3) interpreting a result. These steps can be repeated to test multiple hypotheses.

In our work, we focus on a case study of irritable bowel syndrome (IBS), a chronic illness in which people get episodic gastrointestinal symptoms despite normal laboratory test results. IBS symptoms are often caused or worsened by specific foods, but different foods can be problematic for different individuals. Identifying an individual's personal triggers is therefore difficult. The current standard of care for IBS involves the patients recording their food and symptoms in a journal, which they often bring to their healthcare providers for help identifying potential triggers. Unfortunately, providers currently lack both the time and the tools to help them make sense of the data, and the majority of IBS patients are dissatisfied by the feedback their providers give based on their data. Due to the uncertainty and highly individualized nature of the condition, IBS is an ideal domain for self-experimentation.

Hypothesis Formation
People who have chronic illnesses often want to determine what factors trigger their individual symptoms. However, determining a specific hypothesis to test is often difficult. For many chronic conditions, healthcare providers often supply overwhelming lists of possible symptom triggers. Attempting to perform a self-experiment on each possible trigger would be frustrating and time-consuming. People therefore need a method to narrow possible triggers to a subset that are likely triggers for that individual. Our work on hypothesis formation investigates tools to support people and their providers through the process of interpreting self-tracking data to determine what triggers they want to investigate in a self-experimentation process [4].

To investigate whether better tools can help people and their providers better identify likely personal triggers, we developed interactive, exploratory visualizations to help people and their providers explore data in a person's food and IBS symptom journal (e.g., Figure 1). We interviewed IBS patients and providers, both separately and together, about their interpretation of the visualizations. We found that collaboratively reviewing the visualizations helped patients and providers navigate food and symptom data and helped build mutual trust in their collaboration. Both patients and providers gave unprompted comments on incorporating the tools into the longer-term management of IBS: suggesting using the visualizations to form hypotheses, then following up the hypothesis formation step with an experimental step.

Hypothesis Testing
After a person has decided on a hypothesis to test, they need to design, conduct, and analyze the findings of a self-experiment. In the absence of tools to explicitly support hypothesis testing, people often conduct unsound and invalid experiments, reaching dubious conclusions [1].

We designed and examined TummyTrials [2], an app that scaffolds our self-experimentation framework to help people with IBS design, conduct, and analyze self-experiments with the goal of identifying their individualized food triggers (Figure 2). TummyTrials uses a wizard to support a person in configuring a self-experiment: choosing independent and dependent variables, setting daily reminders, and selecting a meal plan. In a field study with 15 IBS patients each completing a 12-day self-experiment, we found TummyTrials effectively supports self-experimentation. However, interviews with participants revealed a tension between scientific rigor and uncertainties of lived experience. Although they understood and appreciated the need for a certain level of rigor in the self-experiment, many participants expressed a desire for more customizability and flexibility in experimental design and execution.
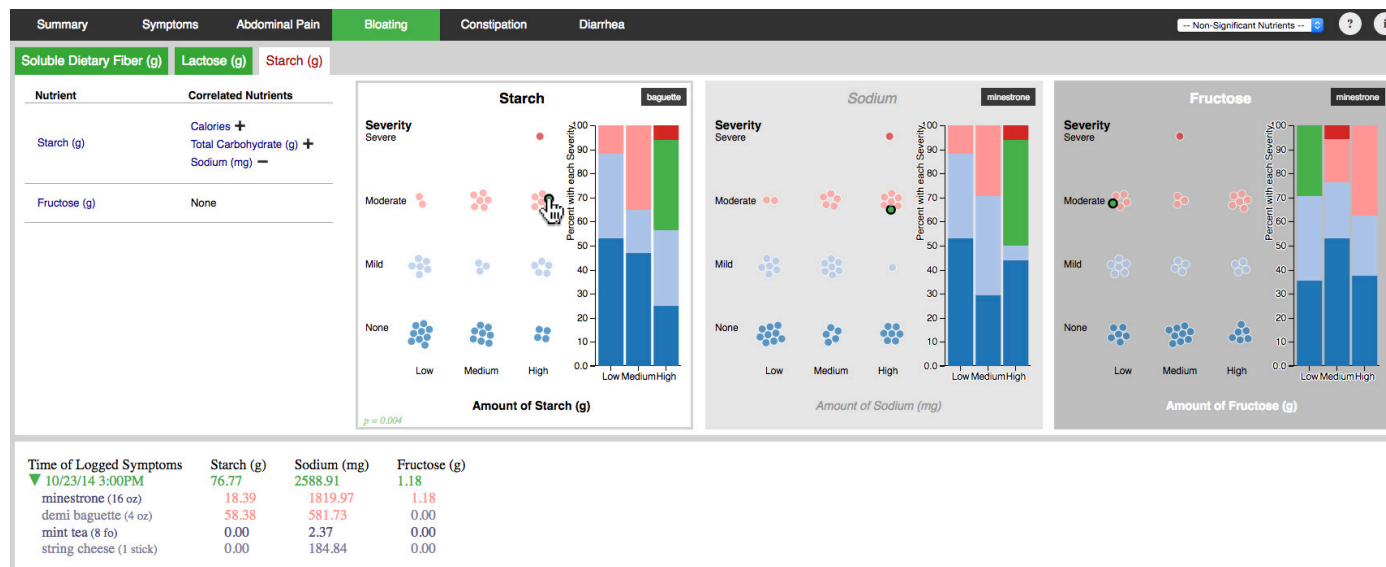
Figure 1: We created interactive, exploratory visualizations to help patients and their providers explore data from food and symptom journals. We then interviewed patients and providers to examine how the visualizations helped support collaboration and facilitate hypothesis formation.

Implications

Our work on hypothesis formation and hypothesis testing reveals opportunities and challenges in validity and reliability.

Respecting Different Forms of Expertise
A primary finding when investigating the visualizations developed for hypothesis formation was the need to respect the different expertise patients and healthcare providers bring to a collaboration. The knowledge of a health provider is often considered superior to that of a patient. However, the collaborative interviews we conducted, in which a patient and a provider collaboratively reviewed the patient's data, revealed the necessity of using both the patient's and the provider's knowledge to successfully interpret the data and form an appropriate hypothesis for testing. Throughout

the collaborative interviews, providers often remarked on how data interpretation was easier when the patient could provide context for the data. They appreciated learning about the patient's goals, their willingness to change, and their interpretations and opinions regarding the results. Because of this patient context, provider conclusions in the collaborative interviews were sometimes completely different from those made when considering data without the patient. Ultimately, provider recommendations depended as much on a patient's knowledge of their life as they did on the provider's knowledge of IBS.

Tension between Scientific Rigor and Lived Experience
Our examination of TummyTrials also revealed challenges of applying clinical research methods in everyday life. For
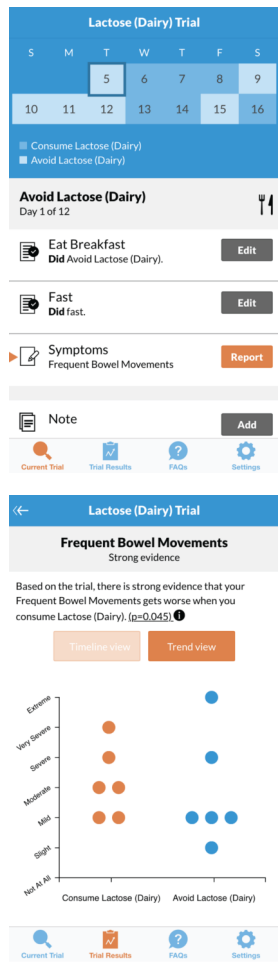
Figure 2: Top: From the TummyTrials app, a 'Current Trial' page used to report compliance and symptom levels and a 'Result' page with self-experiment results.

example, participants wanted more flexibility in experimental design, execution, and analysis (e.g., to accommodate spontaneity or unforeseen events). This tension motivates a need to design self-experimentation systems to balance rigor with the uncertainties of everyday life.

Discussion
Informed by our experiences in this work, we would like to propose several discussion topics for the workshop.

Designing for Validity and Reliability
in General-Purpose Self-Experimentation
We believe the self-experimentation framework is applicable across many domains, and our prior work explores absolute and desired requirements for applying the framework in a domain [3]. However, substantial expertise was required to design a self-experiment that maximized potential for a statistically significant result, minimized risk of confounds, identified appropriate measures, and chose potential hypotheses that were most likely to have an impact on health outcomes. Although a completely customizable platform for designing self-experimentation may be possible, such a platform will bring challenges in people conducting self-experiments that do not reach meaningful results. Incorporating knowledge of domain experts would help minimize this risk. For example, our development of TummyTrials included a gastroenterologist to ensure the supported self-experiments would help people with IBS. One approach might be for experts to design valid self-experiments for different questions, which people can choose as a starting point for adaptation to their needs.

Empowering Different Forms of Expertise
An individual is an expert on their own behaviors and habits. As we saw in our interviews with IBS patients and providers, successful collaboration relied on patients engaging a

provider on an equal footing, bringing in their experiences and data. By empowering people with tools and methods to explain their experiences, they can have more engaging and effective interactions with their providers.

Supporting Post-Experiment Actionability
Although validity is important, the primary goal of a self-experiment is knowledge to support action or behavior change based in an experimental result. When reviewing the results of their self-experiments, some TummyTrials participants seemed to display a confirmation bias. This bias might indicate a need to present a more comprehensive results section, rather than a summary of evidence or lack thereof. Results can also be presented as pathways to potential next steps, thus helping people in deciding how to act on the results of their self-experiments.

Conclusions
We look forward to participation as an opportunity to discuss self-experimentation, its potential benefits and challenges, and opportunities for improved tool support. The CHI 2017 workshop on Digital Health & Self-Experimentation is ideal for such a conversation.

Acknowledgements

REFERENCES

1. Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1143–1152. `http://doi.org/bbpd`

2. Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. 2017. 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 (CHI 2017)*. `http://dx.doi.org/10.1145/3025453.3025480`

3. Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R. Mishra, James Fogarty, Sean A. Munson, and Julie A. Kientz. 2015. A Framework for Self-Experimentation in Personalized Health. *Journal of the American Medical Informatics Association* (2015). `http://doi.org/bwtt`

4. Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean A. Munson, and Jasmine Zia. 2017. Supporting Patient-Provider Collaboration to Identify Individual Triggers using Food and Symptom Journals. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW 2017)*. `http://doi.org/bz28`