

©Copyright 2020

Jessica Schroeder

Goal-Directed Self-Tracking in the Management of Chronic Health Conditions

Jessica Schroeder

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

James Fogarty, Chair

Sean A. Munson, Chair

Wanda Pratt

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Goal-Directed Self-Tracking
in the Management of Chronic Health Conditions

Jessica Schroeder

Co-Chairs of the Supervisory Committee:

Professor James Fogarty
Computer Science & Engineering

Professor Sean A. Munson
Human Centered Design & Engineering

Health technologies are increasingly prevalent and important to support the management of chronic health conditions. Such health technologies often rely on self-tracking, or the practice of collecting and reflecting on personal data, to help people better understand their symptoms and learn how their habits and behaviors affect those symptoms. Many people with chronic conditions and their health providers believe that self-tracking offers a potential for a more complete and accurate understanding of an individual's personal health. However, people and their health providers often struggle to collect, interpret, and act on self-tracked data, leaving this potential largely unmet.

My dissertation research examines how health technologies can better support the goals an individual may bring to their chronic condition management, focusing on three distinct health contexts. I first summarize my work in mental health and irritable bowel syndrome, in which I investigated how systems could support common condition management goals that people and health providers pursue within these domains. My research in these areas revealed nuanced and personalized goals that individuals often developed throughout the management of their condition. I then describe my development of *goal-directed self-tracking*, a novel method that aims to explicitly elicit, represent, and support the goals people and their health providers may have for their personalized health management. I explored goal-directed self-tracking in the context of migraine, synthesizing

findings from my prior work and formative studies to explore how tools could help people and their health providers: 1) express their tracking goals; 2) collect *exactly* and *only* the data they need to achieve those goals; and 3) interpret the resulting data appropriately for those goals. Finally, I discuss next steps and future challenges for goal-directed self-tracking, including opportunities for additional research in supporting holistic health management, patient-provider collaboration with patient-generated data, and artificial intelligence to further personalize health technologies. My dissertation therefore contributes new understanding, methods, and tools to support people and their health providers in *expressing* and *pursuing* their *multiple, distinct, and evolving* goals.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	iv
Chapter 1: Introduction	1
1.1 Thesis Statement	3
1.2 Thesis Overview	3
Chapter 2: Background and Related Work	5
2.1 Health Contexts	5
2.2 Related Work in Personal Informatics and Health Technology	9
Chapter 3: Supporting Mental Health Management through a Mobile Web App	16
3.1 Providing Holistic Support for Dialectical Behavioral Therapy	17
3.2 Investigating and Predicting Skill-Level Effectiveness	39
3.3 A Goal-Directed Perspective on Supporting Mental Health	62
3.4 Summary	63
Chapter 4: Supporting IBS Management through Personalized Trigger Identification	64
4.1 Identifying Correlations between Nutrients and Symptoms	65
4.2 Supporting Interpretation and Patient-Provider Collaboration	67
4.3 Investigating Holistic Support for IBS Trigger Identification	90
4.4 A Goal-Directed Perspective on Supporting IBS Management	95
4.5 Summary	96

Chapter 5:	Supporting Migraine Management through Goal-Directed Self-Tracking . . .	98
5.1	Identifying Goals and Needs in Migraine Tracking	99
5.2	Investigating Goal-Directed Self-Tracking	122
5.3	Summary	155
Chapter 6:	Discussion and Conclusion	156
6.1	Next Steps in Goal-Directed Self-Tracking	157
6.2	Future Challenges for Goal-Directed Self-Tracking	166
6.3	Conclusions	172
Bibliography		173

LIST OF FIGURES

Figure Number	Page
3.1 Pocket Skills app design.	18
3.2 Validated scale results throughout the Pocket Skills feasibility study.	25
3.3 Theoretical model of how participants felt Pocket Skills supported DBT.	29
3.4 Skill use throughout the Pocket Skills feasibility study.	46
3.5 Average improvement after using Pocket Skills skills, by module.	48
3.6 Average Pocket Skills skill effectiveness for different subgroups of participants. . .	51
3.7 Accuracy of machine learning model predictions for Pocket Skills skill effectiveness.	57
4.1 IBS visualizations: bubble and bar chart.	70
4.2 IBS visualizations: parallel coordinates plot.	71
4.3 IBS visualizations: trend identification for each visualization type	72
4.4 IBS visualizations: tutorial.	73
4.5 IBS visualizations: symptoms page for each visualization type.	74
4.6 Visualizations of Bayesian predictive distributions of future IBS symptoms.	94
5.1 Example of migraine tracking evolution over time.	110
5.2 Goal-Directed Self-Tracking for migraine management: paper prototype.	130
6.1 Goal-Directed Self-Tracking for migraine management: configuration designs. . .	159
6.2 Goal-Directed Self-Tracking for migraine management: post-configuration designs.	162

LIST OF TABLES

Table Number	Page
3.1 Pocket Skills feasibility study: demographic information.	21
3.2 Pocket Skills feasibility study: usage statistics.	24
3.3 Pocket Skills modules, skills, and subskills examined in our quantitative analyses.	42
3.4 Pocket Skills feasibility study: groupings and distributions for quantitative analysis.	43
3.5 Pairwise differences in post-ratings between Pocket Skills skills.	49
3.6 Features used for training predictive models of Pocket Skills skill improvement. . .	55
3.7 Model performances for Pocket Skills skill effectiveness predictions.	56
4.1 IBS visualization feasibility study: demographic information.	75
5.1 Migraine formative study: demographic information of participants with migraine.	100
5.2 Migraine formative study: demographic information of provider participants. . . .	101
5.3 Migraine tracking routines by goal described by formative study participants. . . .	123
5.4 Supported migraine tracking goals in the paper prototype.	124
5.5 Examples of migraine-related data tracked by formative study participants.	128
5.6 Migraine paper prototype study: demographic information of provider participants.	133

ACKNOWLEDGMENTS

I would not have been able to complete my PhD without the legion of people who provided endless support, consistently going above and beyond to help me throughout the past six years.

I first want to thank my advisors, James Fogarty and Sean Munson. I did not realize how important the advisor/advisee relationship was when I joined the program, so I feel extremely lucky to have stumbled upon the perfect advising situation for me. Much of my growth over the past six years can be directly traced back to their thoughtful mentoring. They were always happy to listen to even the most half-baked ideas, help me develop those ideas when they were good, and take the time to explain why when they were not. Equally importantly for me was how much I like them just as people, which helped me enjoy meetings I might otherwise have dreaded and be productive even when feeling unmotivated. I would not have completed my PhD without such fantastic advisors.

I also want to thank my other mentors throughout the pursuit of my PhD. From my thesis committee, Wanda Pratt helped mentor me not just with respect to my dissertation, but also throughout my job hunt, and Jennifer Mankoff was the reason I pursued my PhD in the first place: she mentored me for a summer internship while I was an undergrad and encouraged and contributed to my application. I feel very lucky to have had the chance to work with such a positive, supportive, and meaningful committee. Mary Czerwinski mentored me during a summer internship at Microsoft Research and continues to provide mentorship and support even now, as I start to look at the next steps in my career. Not every woman in a STEM field is lucky enough to have a fantastic female mentor in the industry they want to pursue. Tim Althoff taught only one of my classes, but he still took the time to subsequently mentor me on one of my publications and help with my job hunt, ultimately connecting me with my new employer. Finally, without my medical collaborators Jasmine Zia,

Chelsey Wilks, Melissa Schorn, and Natalia Murinova, I would not have been able to complete my work. Beyond providing necessary medical expertise, they were all patient, kind, and supportive as they helped me define and iterate on ideas and better understand the health contexts I examined.

I was extremely fortunate to be in such a collaborative environment, so I also want to thank the other collaborators, coauthors, and friends I had the chance to work with throughout my PhD. Some of my closest friendships were born from a machine learning study group my first quarter in the program; Jane Hoffswell, Alex Fiannaca, and Brandon Haynes helped me get through that class and navigate countless challenges since, including class projects, publications, and my general sanity and wellbeing. Stacey Haynes and Maggie Thompson were not even in my program, but they still spent time listening to practice talks and editing document drafts. Arvind Satyanarayan helped me increase my confidence to contribute to the HCI community, both within and beyond UW. I was also lucky to work closely with Elena Agapie, Christina Chung, Daniel Epstein, Liwei Jiang, Ravi Karkar, Matthew Kay, Sonali Mishra, Laura Pina, Alex Okeson, Annie Ross, and Jina Suh; working with them not only made the research itself better, it also made doing the work much more enjoyable. More broadly, I want to thank my labmates, the personal informatics group, the human-computer interaction group within CSE, and the broader DUB and CSE communities for all of the support, guidance, and perspectives that helped me become a better researcher.

Finally, I also want to thank my friends and family outside of the program. My parents, Kathy and Scott Schroeder, always prioritized and encouraged my education; anything they could do to set me up for success, they enthusiastically did. My grandmothers, Ann Schroeder and Anne Kleist, also provided constant encouragement and support. My undergraduate provider, Professor Tzu-Yi Chen, encouraged me to major in computer science and apply to PhD programs. Finally, thanks to my brother Andy Schroeder and my friends Joanne Reid, Sammy Chao, and Caitlyn Hynes for helping me stay as sane and happy as possible throughout all the ups and downs. I would not have been able to pursue and complete my PhD without these many sources and types of support.

I was also extremely fortunate for my many sources of funding. My dissertation research was supported in part by a University of Washington Innovation Research Award; the Marilyn Fries Endowed Regental Fellowship; the ARCS Foundation, with sponsorship from Lesa Sroufe, Matt Barnes, Kathy Petrait, and Richard Petrait; the Allen Institute for Artificial Intelligence (AI2); Supportiv; Microsoft Research; Nokia Research; an Adobe Data Science Research Award; the Intel Science and Technology Center for Pervasive Computing; the Agency for Healthcare Research Quality, under award 1R21HS023654; the National Institute on Drug Abuse, under award 1K99DA037276-01; the National Center For Advancing Translational Sciences of the National Institutes of Health, under award UL1 TR002319; the National Institutes of Health, under awards P50 MH115837 and R01 LM012810; and the National Science Foundation, under awards DGE-1256082, IIS-1553167, IIS-1813675, IIS-1901386 and SCH-1344613. The content is solely the responsibility of the author and does not necessarily represent the official views of the funders.

Chapter 1

INTRODUCTION

Chronic health conditions are a leading cause of sickness, disability, and death worldwide. 60% of US adults have at least one chronic health condition today, and that number is expected to continue to grow [30]. This trend is unsustainable for our health system [247], necessitating new approaches to better support chronic condition management. Health technology is likely to play an increasingly vital role in these new approaches due to its potential to help people understand and manage their conditions (e.g., [12, 103]). Technologies designed or appropriated to support chronic condition management often include self-tracking, or the practice of collecting and reflecting on personal data (e.g., [34, 40, 70, 159]). Self-tracking can help people monitor health-related data not just in the clinic, but also at home, enabling ongoing self-management of their condition (e.g., [34, 70, 155]). Many patients and providers believe self-tracked data has the potential to capture a more complete understanding of patient health [41, 42, 251]. Providers therefore often recommend their patients self-track data related to their conditions [1, 7, 41, 75, 232].

Unfortunately, the tools people use to manage their chronic conditions often leave the goals they have for their condition management unstated and lacking explicit support. For example, people with chronic conditions often self-track towards the goal of better understanding how their habits and behaviors affect their symptoms [7, 152]. However, current self-tracking tools tend to emphasize the data itself without providing such insights (e.g., in charts and timelines that primarily show the magnitude of data) [64]. Providers also struggle with their goals for patient-tracked data, as they often lack time to analyze the data or feel unequipped to interpret it [41, 251].

Due to these disconnects between the goals people and their providers have for their chronic condition management and the designs of the tools they use to support that management, people often:

1. find self-tracking overwhelming and without purpose [63]
(i.e., *struggle to define goals for tracking*),
2. find their self-tracking data does not support their questions [39]
(i.e., *struggle to align data with goals*),
3. find their tools fail to changes in their tracking over time [66]
(i.e., *struggle to adapt tracking to evolving goals*),
4. find collaboration with providers difficult [42]
(i.e., *struggle with multiple goals of patients and providers*)

Prior research has investigated how to better support *common* goals that individuals with specific chronic conditions pursue, often with encouragement from their health providers. For example, previous work in diabetes management examined how to help people understand relationships between the food they consume and their subsequent blood glucose levels, which can help them develop habits to better manage their diabetes [153]. Discovering and supporting these common condition-specific goals is an important step in helping people manage those conditions. However, not only do people with different conditions often have different goals, but different individuals *with the same chronic condition* can also have different goals. These goals often depend on the individual's personal habits, behaviors, experiences, and approaches related to their chronic condition management. For example, a study with overweight and obese women found the need for tools to support their specific, varied health goals (e.g., more energy, better stamina, consistent exercise routines, myriad definitions of healthier eating) rather than emphasizing weight loss [16]. To fully support the management of an individual's chronic condition, tools must allow people to explicitly define and pursue their nuanced and personalized condition management goals.

1.1 THESIS STATEMENT

My dissertation research has therefore focused on investigating the following:

Tools that support people and providers in expressing and pursuing their multiple, distinct, and evolving goals can help them overcome barriers to personalized management of complex, chronic health conditions.

I have examined supporting chronic condition management in three health contexts: 1) mental health disorders, in which I investigated support for dialectical behavioral therapy; 2) irritable bowel syndrome, in which I investigated support for personalized trigger identification; and 3) migraine, in which I developed *goal-directed self-tracking*, a novel method to help people and health providers express and pursue their goals. Each of these contexts has its own unique challenges, but each is characterized by complex, chronic symptoms that require personalized interventions to address. My research has therefore focused on designing, developing, and evaluating novel approaches to better support these personalized interventions. Together, these projects contribute new understanding, methods, and tools for eliciting and supporting the multiple, distinct, and evolving goals people and their health providers may have for their chronic condition management.

1.2 THESIS OVERVIEW

In this document, I report my findings from investigations in the three health contexts I examined. As all of my dissertation research has been collaborative, I focus on describing the overall contributions of the work, emphasizing my personal contributions where appropriate.

In Chapter 2, I first provide necessary background for each specific health context, motivating my research in those contexts and highlighting both the similarities between them and the unique challenges each introduces. I then describe how the literature has characterized opportunities and pitfalls in self-tracking, with a focus on self-tracking for health management. I finally discuss related work in patient-provider collaboration around patient-generated data.

Next, I describe my research in mental health (Chapter 3) and irritable bowel syndrome (Chapter 4). My work in each of these contexts investigated how to support *common* goals (i.e., goals that individuals within the context often pursue). Although the methods and tools I describe often helped people achieve these common goals, my evaluations revealed nuanced and personalized goals that individuals and health providers often developed throughout the management of their condition. Within each chapter, I therefore first describe the original research I completed in that context. I then summarize and reflect on how the context could benefit from a goal-directed perspective.

Chapter 5 describes my work in migraine, in which I developed *goal-directed self-tracking* as a novel method to help people with migraine and their health providers express and pursue their personalized goals. I first present a formative study I conducted to characterize needs and opportunities in migraine. I then describe how my findings informed my iterative development of *goal-directed self-tracking* to explicitly elicit, represent, and support the goals people have for their migraine-related tracking. I discuss my evaluation of my ideas via a paper prototype study that investigated how a goal-directed self-tracking system could elicit and support migraine-related goals.

Finally, in Chapter 6, I reflect on future challenges and opportunities I see for goal-directed self-tracking, including immediate next steps and higher-level future challenges in personalized health.

Chapter 2

BACKGROUND AND RELATED WORK

In this chapter, I first provide background information on the three health contexts I examined to motivate my work in those contexts (Section 2.1). I then summarize related work in personal informatics, health-related HCI research, and patient-provider collaboration (Section 2.2).

2.1 HEALTH CONTEXTS

In my dissertation, I investigated supporting the management of three complex, chronic health conditions: mental health disorders (e.g., borderline personality disorder, major depressive disorder), irritable bowel syndrome (IBS), and migraine. Each of these contexts is characterized by personalized symptoms and “triggers” (i.e., different people with the same condition will experience a different set of symptoms and have a different set of factors that contribute to those symptoms). Each context therefore requires the development of personalized management strategies to try to avoid symptoms. Here, I provide background information on these contexts, describing needs and opportunities for novel methods and tools that help people manage health challenges related to each context.

2.1.1 Mental Health Disorders

Mental health disorders (e.g., mood disorders, personality disorders) are characterized by disturbed cognition, emotional regulation, or behavior, and are a leading cause of disability and death worldwide [9]. Approximately 18% of US adults suffer from a mental illness in a given year [182], and suicide is the 10th leading cause of death in the US [74]. Despite the prevalence and seriousness of mental illness, engagement in treatment remains relatively low, often due to barriers such as stigma, perceived ineffectiveness or lack of need, and inability to access or afford treatment [167].

Mental health disorders are often considered fundamentally different than health conditions that affect other parts of the body. However, they share a number of characteristics with other complex, chronic health conditions. Symptoms—which manifest as negative moods, behaviors, or coping mechanisms—can be thought of as being “triggered” by particular social and physical environments [85]. Mental health disorder management often requires avoiding triggers when feasible (e.g., eliminating negative relationships, avoiding upsetting media). However, many triggers may be unavoidable (e.g., upsetting life events, difficult interactions with others). Successful management therefore requires developing strategies to defend against symptoms, even when exposed to triggers (e.g., positive coping mechanisms people can employ instead of negative ones in upsetting situations).

Evidence-based psychotherapy is effective for many mental health disorders, with different types of disorders often requiring different interventions [129,248]. A common psychotherapy for people with high clinical complexity (i.e., those at risk for suicide who meet criteria for multiple mental health disorders) is dialectical behavioral therapy (DBT). DBT is designed to support people in developing concrete coping skills to help them solve problems, maintain positive relationships, and navigate negative events and emotions, regardless of their particular mental health disorders [140,141]. These DBT skills are traditionally separated into modules, each of which supports the development of specific types of skills. For example, the *Mindfulness* module contains skills dedicated to teaching people to accept the moment without judgment (e.g., by observing their breathing, by describing their thoughts). The *Emotion Regulation* module is designed to help people understand, manage, and adjust their emotional states. The *Distress Tolerance* module gives people specific healthy alternatives to unhealthy behaviors (e.g., instead of self-harming, people can take a cold shower or gently snap a rubber band on their wrist). By working through the modules and learning specific skills that can replace unhealthy behavior, people can start to apply these positive coping mechanisms in their lives when they need them, helping them better manage their mental illnesses [177]. Traditionally, people learn DBT skills in individual and group therapy sessions, and practice them using worksheets [141]. They also complete daily *diary cards*, with which they self-track the skills they practice and their positive and negative emotions and behaviors [141].

2.1.2 Irritable Bowel Syndrome

Irritable bowel syndrome (IBS) is characterized by chronic gastrointestinal symptoms (e.g., abdominal pain, constipation, diarrhea) despite normal test results. IBS affects up to 20% of the US population and is one of the top ten reasons people seek primary care [62, 147]. People with IBS report a lower quality of life and consume 50% more healthcare resources than those who do not have IBS [128, 163].

IBS symptoms can be triggered by a variety of factors (e.g., menstruation, stress, diet), but up to 70% of people with IBS that report certain foods can cause or worsen their symptoms [81]. As different foods and nutrients are problematic for different people, population-level recommendations are often inappropriate [168, 223, 260]. To try to avoid symptoms, individuals with IBS currently must choose between eliminating all nutrients thought to correlate with IBS symptoms (e.g., FODMAPs, gluten, caffeine [81, 168]) and trying to identify particular nutrients that cause their individual symptoms. Dramatic elimination diets can be highly burdensome and can even lead to malnutrition [81, 137] so many people with IBS want to identify their personal triggers.

Unfortunately, identifying which nutrients correlate with an individual's IBS symptoms is a difficult and unreliable process. The American Gastroenterological Association advises providers to have patients keep food and symptom journals to attempt to identify correlations between the foods they eat and the symptoms they experience [7]. However, despite their effort to identify such correlations, providers often lack the time and training necessary to do so [41]. An analysis of the recommendations given by 8 experienced providers after reviewing 17 paper food and symptom journals found poor reliability in their recommendations: recommendations for the same patient were inconsistent across providers, and individual providers showed biases toward identifying particular triggers regardless of the patient's data [262]. Perhaps because of these issues, the majority of patients with IBS are dissatisfied with the feedback their providers give based on their journals [102].

2.1.3 Migraine

Migraine is a prevalent and debilitating condition that can severely affect people's quality of life and productivity [225]. An estimated 14% of the population experiences migraine, making it the third most prevalent disease in the world [227]. Symptoms can be debilitating, often including pain, sensitivity to light, and impaired cognition. Migraine is therefore a leading cause of loss of workplace productivity [225] and is the seventh-highest cause of years lost due to disability [227]. As with IBS, people with migraine often struggle to manage their symptoms [10].

Although migraine symptoms can also be “triggered” by different personalized factors [5, 119], potential migraine contributors span a wide range of possible domains, including diet (e.g., alcohol, caffeine), environment (e.g., certain lighting or weather), menstruation, sleep patterns, and mood and stress. In addition, multiple factors often must accumulate before precipitating symptoms [119, 250]. Identifying and avoiding single “triggers” can therefore be helpful to some extent, but successful management often requires identifying the full range of factors that may contribute to symptoms and avoiding exposure to too many of those factors. As with mental health disorders, some migraine contributors may be unavoidable (e.g., changes in weather), so migraine management also requires developing strategies to defend against symptoms despite exposure to potential contributors (e.g., sleep, exercise, and dietary changes that reduce the likelihood of symptoms).

People with migraine often attempt to identify factors that cause or worsen their symptoms [152]. Because these factors vary across individuals, health providers often encourage self-tracking as part of attempts to identify personal triggers [222]. Both patients and providers believe self-tracked migraine data can support communication and collaboration [15]. However, self-tracking in migraine can be even more difficult than self-tracking in other health conditions, due to the wide range of potential contributing factors and the necessity of those factors to accumulate. Approaches that focus on identifying single triggers, rather than considering multiple potential contributors, can therefore be misleading (e.g., by identifying spurious correlations due to confounds). As with IBS, these challenges surrounding data interpretation can frustrate people with migraine and health providers, and patients often decide to discontinue clinical consultations [143, 144].

2.2 RELATED WORK IN PERSONAL INFORMATICS AND HEALTH TECHNOLOGY

In this section, I summarize relevant related work in personal informatics and health technology. This section does not summarize all work related to health, technology, and personal data; for example, a rich body of work has investigated mobile health applications and technologies (e.g., [122, 185, 241]) and technological support of behavior changes for health (e.g., [48, 121, 255]). Here, I focus on summarizing related work that most directly informed and relates to my research.

2.2.1 Models and Barriers Personal Informatics

Self-tracking has become increasingly popular as technology has advanced to reduce associated barriers and burdens. Although self-tracking tools can help people understand their habits and behaviors, many self-trackers struggle to find value in their data [133]. Li et al. developed a five-stage model of personal informatics and characterized barriers people encounter in each stage [133]:

1. *Preparation*: identifying motivation to track and determining what and how to track.
2. *Collection*: gathering data over time
3. *Integration*: transforming and combining collected data to enable reflection
4. *Reflection*: synthesizing findings to understand behavior and habits
5. *Action*: deciding what to do based on findings

This model highlights how challenges *cascade* into later stages; for example, failure to *collect* appropriate data undermines *reflection* and *action* based on that data. Even “expert” self-trackers (i.e., technology-fluent people that have the skills and motivation to build custom tools) encounter pitfalls; an examination of members of the Quantified Self community found that expert self-trackers often encounter insurmountable barriers, including: 1) burdens of tracking too many things, 2) failure to track necessary triggers and context, and 3) lack of rigor in experimental design and analysis [39].

The lived informatics model [66] extends Li et al.’s model, surfacing that burdens associated with tracking often lead to lapses and highlighting the fact that realities of everyday tracking are not well supported in current tools. The model includes three stages: 1) *deciding and selecting*, in which people choose *why* and *how* to track; 2) *tracking and acting*, in which people engage in the ongoing

process of *collecting, integrating, and reflecting on* data; and 3) *lapsing and resuming*, in which people forget, intentionally skip, or suspend tracking. My research in goal-directed self-tracking examines how all stages of self-tracking can be aligned to an individual's goals.

Setting personalized and realistic self-tracking goals and showing progress towards those goals can help encourage healthy behaviors (e.g., [47,49,173]). However, current tools often fail to support the personalized and evolving goals individuals want to pursue [16,87,131]. Tools can sometimes even encourage behaviors that are *contrary* to a person's goals. For example, people tracking their meals using commercial food journaling tools reported that the relative ease of journaling foods with barcodes encourages them to eat highly processed foods, contrary to their healthy eating goals [51]. Setting unrealistic goals can also be detrimental; for example, people with obesity who have unreasonable expectations for their weight loss are more likely to discontinue treatment [52]. People can also encounter pitfalls when their tools do not acknowledge their personalized goals. For example, some menstrual tracking apps assume a pregnancy-related goal, which can be frustrating or painful for people not pursuing that goal [65]. Similarly, people with chronic fatigue syndrome often use fitness trackers to ensure they do not expend too much energy, but those trackers often encourage people to be more active, resulting in feelings of frustration and shame [55]. The tracker goal evolution model characterizes relationships between people's *qualitative* goals (e.g., losing weight, having a healthy lifestyle) and their *quantitative* goals (e.g., meeting a certain step count, eating a specific diet) and describes challenges people face when attempting to defining appropriate *quantitative* goals based on their *qualitative* goals [179]. These barriers and negative nudges highlight the need to explicitly support an individual's goals and help them navigate analogous challenges in setting appropriate *tracking* goals based on their *condition management* goals.

2.2.2 Health and Diagnostic Self-Tracking

Self-tracking health-related information is particularly common; 69% of U.S. adults report tracking a health factor, with 14% using technology to do so [76]. A challenging, prevalent, and important form of health-related tracking is *diagnostic self-tracking*, in which a person tracks to answer questions they have about their health. Challenges in effective diagnostic self-tracking

motivate recent research in support for self-experimentation (e.g., [18, 53, 54, 132]), in which an individual varies one or more factors in a controlled manner with the intent of making causal inferences about the effect of those factors [138, 205]. The framework for self-experimentation in personalized health [110] provides guidelines for such self-experimentation and considers properties of health conditions appropriate for self-experimentation, including behaviors to be experimentally manipulated (i.e., independent variables) and outcomes (i.e., dependent variables).

Another way to support diagnostic self-tracking is to appropriately visualize data to answer the specific questions people have. Prior research has examined transforming self-tracking data into actionable information with visualizations (e.g., [59, 64]). Tailored visualizations can also help people reflect on their goals and define future goals [134]. Visualizations of health-related self-tracked data can also support more objective identification of trends in that data [204]. Visualizations can be particularly useful when combined with textual information to improve interpretability and help health providers draw conclusions more rapidly [235].

Self-tracking research has often examined supporting self-tracking of specific health-related factors (e.g., diet [19, 50, 109, 154], menstruation [65], physical activity [49, 139], sleep [115, 201], stress [169]). Current tools often isolate data in factor-specific silos [66, 68], impeding people from forming a holistic picture of their health. Prior research has investigated importing data from multiple tools (e.g., [96, 135]), which enables analysis but cannot support earlier self-tracking stages.

To allow self-tracking of *any* data, regardless of the domain, my work builds upon recent research in customizable approaches to self-tracking. Such an approach can allow people to configure exactly *what*, *when*, and *how* they track. For example, semi-automated tracking combines automated tracking (e.g., through devices and sensors) with manual tracking (e.g., of data that cannot be automatically tracked, of data that brings positive awareness when manually tracked), preserving tracking benefits while lowering associated burdens [38]. Kim et al.’s OmniTrack builds on the concept of semi-automated tracking to support a form-based approach, similar to Google Forms and other popular tools for configuring data entry [120]. OmniTrack extends the basic form-based approach by providing an ability to link automatically-tracked data to the configured fields (e.g., linking sleep data from a Fitbit or other wearable to a field in the authored form). OmniTrack also allows configuring

shortcuts for lightweight data entry, as well as interactively specifying notification-based reminders to track. Although these approaches can help people track *exactly* and *only* their desired data, they do not provide guidance or support in identifying, expressing, or pursuing goals for data. *Goal-directed self-tracking* therefore builds upon demonstrated approaches to customized self-tracking, focusing on new questions in effective support for goal-based tracking and patient-provider collaboration.

2.2.3 Health Technology for the Management of Chronic Conditions

Prior research has examined personal informatics practices and tools in the management of myriad chronic conditions, including asthma (e.g., [103]), diabetes (e.g., [44, 70, 111, 113, 154]), Parkinson's (e.g., [244]), and hypertension (e.g., [34, 78, 84]). These conditions have fairly well-understood relationships between symptoms and contributors, and technology generally focuses on helping people understand and manage those relationships. For example, research in multiple sclerosis has examined how to help people maintain control of their lives despite debilitating and unpredictable symptoms [12, 258]. Previous work in migraine management emphasized a need for tools designed to better support chronic conditions with unpredictable, intermittent, and poorly-understood symptoms [190]. Tools supporting personalized self-tracking for enigmatic diseases can help support the development individual and population-level knowledge of those diseases [160].

Recent work has also explored technological support for people with mental health disorders, often including self-tracking to help people monitor their moods and behaviors and visualize their progress. For example, systems designed to support cognitive behavioral therapy (CBT) can increase engagement [17] and reduce depressive symptoms [24, 123, 195]. Systems can also help people practice positive coping skills; for example, a suite of skills-based apps has been shown to reduce depression and anxiety [166], and a number of systems provide constantly-available skills to help people during times of distress (e.g., [28, 206, 207]). In general, for a mental health system to be effective, it must be engaging enough to encourage people to use it [17, 60, 61, 158, 164]. One possible way to encourage engagement is through conversational agents, which can increase engagement in mental health systems (e.g., [22]) and behavior change interventions (e.g., [104, 145]); reduce depressive symptoms [72]; and encourage adherence to antipsychotic medications (e.g., [23]).

Recent work in the mental health space has focused on investigating how technology can provide more personalized interventions (e.g., by recommending activities to manage stress [210] or prevent negative moods [94] based on an individual's past sleep, diet, and activity data; by recommending interventions based on an individual's personal characteristics and context [187]). A large body of work has examined how machine learning techniques can support detection, diagnosis, and treatment of myriad mental health conditions (e.g., [221]). The concept of suggesting specific interventions when people need them, known as “just in time interventions” [174], has been pursued in myriad health contexts (e.g., promoting physical activity [89], stress management [101], weight management [226]), with recent interest in applying them to positive coping skill use [107].

Although self-tracking can sometimes help people manage chronic conditions, some people find it a negative experience that can emphasize challenges and health-related difficulties [3, 12] or encourage unhealthy behavior [118]. These negativities can be exacerbated because chronic conditions often require long-term management, but intensive self-tracking is generally highly burdensome and unsustainable (e.g., [39, 51, 63]). Even semi-automated tracking can be burdensome when it fails to accommodate a person's evolving tracking routines and goals [31]. In addition, many tools include insufficient support for customization, which can result in people: 1) feeling forced to track data they do *not* want to track (i.e., increasing burden for data irrelevant to their goals), and 2) feeling unable to track data they *do* want to track (i.e., preventing tracking of data necessary for their goals) [214]. Tools also generally fail to support lapsing and resumption [46, 63, 66], making the prospect of tracking to manage a chronic condition particularly daunting.

My dissertation research investigates supporting the long-term management of complex, chronic conditions, examining the distinct, multiple, and evolving goals people and their health providers bring to self-tracking to manage a condition over time. By focusing on the specific goals an individual self-tracker has at any given time, I aim to minimize the burdens they face, helping them track exactly and only what they need to track to achieve their goals. My research draws upon insights from previously-studied chronic conditions and investigates new approaches and perspectives required for the specific health contexts my dissertation examines.

2.2.4 Patient-Provider Collaboration

People who try to use self-tracked data to support their chronic condition management often face a number of challenges when attempting to interpret that data [133]. One approach to overcoming these barriers is to seek expert help. Many people turn to providers for help making sense of their data [39, 41, 42, 57, 75, 259], both because interpretation is often difficult [39, 133] and because tracking is often initiated at a provider's suggestion [1, 7, 41, 75, 232]. Patients and providers often rely on data to inform diagnoses and treatment, traditionally obtained through verbal summary (e.g., patients recalling experiences with symptoms) or infrequent clinical collection (e.g., clinical tests). However, patients and providers are increasingly interested in augmenting this data with patient-generated data (e.g., mobile journals, wearable sensor data) [57, 252]. Reviewing patient-generated data can help providers understand and contextualize patient health [232], which can lead to improved diagnoses, treatments, and health outcomes [41, 186, 228]. Data can also help ground patient-provider interaction and foster a collaboration, with providers contributing medical expertise while patients contribute complementary knowledge of personal data and behaviors [14, 215]. Collaborative review can help people identify trends, hypothesize about symptom contributors, generate action plans, and identify new information needs [198, 215].

Despite potential and enthusiasm for patient-generated data, patients and providers often struggle to get value from such data. Providers generally lack time to review data or feel unequipped to interpret it, especially because data is often incomplete, disorganized, and unreliable [41, 251, 252]. Current tools are usually not designed to support data sharing or collaboration in a manner compatible with provider workflows [41, 75]. When tools fail to allow patients to collect the data providers want, providers sometimes recommend paper-based tracking [41, 42]. Paper provides flexibility, but interpretation generally requires manually sifting through pages attempting to identify patterns. No validated method exists for providers to interpret journal data, so the process is generally difficult and error-prone [92, 105]. Patients and their providers are often dismayed by the lack of tools to help identify trends in food and symptom journals [27, 42], and many patients who share their data with health providers report disappointment in provider engagement with that data [75].

Prior work has investigated enabling patient-provider collaboration by encouraging patient engagement in the decision-making process (e.g., [4, 161, 232]) and by supporting long-term care management [130, 197, 237]. Systems designed to support collaboration over patient-generated data often help providers contextualize the data necessary for informed diagnoses [232] and facilitate patient-provider communication [1, 154, 191, 239], which can often improve patient health outcomes [186, 228]. However, patient-provider collaboration around self-tracking data can break down when patients and providers have different goals or expectations for the data or the consultation [42, 199, 202]. Chung et al. investigated goals patients and providers bring to collaborative review of patient-generated data and discussed the value of modifying food tracking protocols according to an individual's goals [42]. They also suggested that collaboratively designing a tracking routine could help patients and providers align and fulfill goals. Luo et al. had dietitians construct patient profiles and design custom tracking routines for those profiles, similarly finding a need to help patients and providers collaboratively design tracking routines that best fit individuals and their goals [149]. Other recent work includes a pre-visit note through which patients can summarize their tracking and communicate goals for a clinic appointment [43]. West et al. developed a 6-stage workflow model for the use of patient-generated data in clinical settings, identifying barriers in each stage: *aligning objectives*, *evaluating data quality*, *judging data utility*, *rearranging the data*, *interpreting the data*, and *acting on the data* [253]. The workflow focuses on the use of data within a single clinic appointment, and therefore assumes the data have already been collected. Aligning goals and expectations for the data *before* the collection stage of self-tracking and supporting collaboratively-defined collection, integration, interpretation, and action based on those goals could help avoid barriers and support collaboration across these stages.

My dissertation research has investigated helping people with chronic conditions and their providers align their goals and expectations for their data by enabling them to explicitly state and pursue their goals. By doing so, I hope to help people and their health providers: understand each other's goals; ensure they are both on the same page with regards to their goals; collect the data they need with respect those goals; review appropriate analyses and visualizations of their data to support their goals; and decide whether and how to act on their data and change their goals moving forward.

Chapter 3

SUPPORTING MENTAL HEALTH MANAGEMENT THROUGH A MOBILE WEB APP

Evidence-based psychotherapy, such as Dialectical Behavioral Therapy (DBT), can often help people manage mental health conditions [140,248]. However, an estimated 1 in 5 adults in outpatient psychotherapy quits treatment prematurely [183]. Technology-delivered mental health treatments have the potential to reduce barriers and increase engagement in mental health treatments by mitigating issues surrounding stigma [80, 82, 148] and time and money [73]. Unfortunately, few mental health apps that are available today support evidence-based psychotherapy [97].

I therefore aimed to help people engage with DBT and develop positive coping skills via Pocket Skills, a multi-media mobile web app designed to provide holistic support of DBT, including goal setting, educational components, skill practice, and self-tracking of positive and negative moods and behavior. I joined the Pocket Skills project as an intern at Microsoft after some initial formative work and development had already taken place, working on an interdisciplinary team of HCI experts, clinical psychologists, and mobile app developers and designers who focused on translating content from the DBT skills training manual and workbooks [141] into the mobile app. I then led a field study to assess the feasibility of Pocket Skills. We found Pocket Skills helped support common DBT goals, as well as goals individual participants had for their own DBT. We also developed a model based on our qualitative findings of how Pocket Skills supported DBT. I present our paper on this development and evaluation of Pocket Skills [219] in Section 3.1.

In addition to investigating how Pocket Skills could help support DBT, we also wanted to examine whether *particular* skills delivered by the app were more or less effective for different groups of people. We therefore performed a quantitative analysis of our feasibility study data to develop design implications for translating evidence-based psychotherapies into technology-delivered interventions. I present our paper on this quantitative analysis [218] in Section 3.2.

Finally, Pocket Skills was implemented before I developed and investigated *goal-directed self-tracking* for migraine management. However, the concept of explicitly eliciting, representing, and supporting the goals an individual brings to their chronic condition management is also highly relevant to the mental health domain. I therefore reflect on how a goal-directed perspective could further support personalized mental health management in Section 3.3. I conclude by summarizing how my work in the mental health domain supports my thesis statement in Section 3.4.

The research described in this chapter was done in collaboration with Jina Suh, Chelsey Wilks, Ann Paradiso, Kael Rowan, Arturo Toledo, Tim Althoff, Mary Czerwinski, James Fogarty, Marsha M. Linehan, Gloria Mark, and Sean A. Munson.

3.1 PROVIDING HOLISTIC SUPPORT FOR DIALECTICAL BEHAVIORAL THERAPY

In this section, I first describe the design and development of Pocket Skills. I then present and discuss our findings from the feasibility study we conducted to evaluate the app [219].

3.1.1 Design of Pocket Skills

Pocket Skills consists of a conversational interface in which a conversational agent (*eMarsha*) walks people through the content via a text-like interface (Figures 3.1c and 3.1d). The Pocket Skills content is based on Dr. Marsha Linehan’s DBT Skills training manual and workbooks [141]. Pocket Skills includes modules for *Mindfulness*, *Emotion Regulation*, and *Distress Tolerance*, and adds a module for *Addiction Skills* to help people with addictive behaviors overcome their addictions.

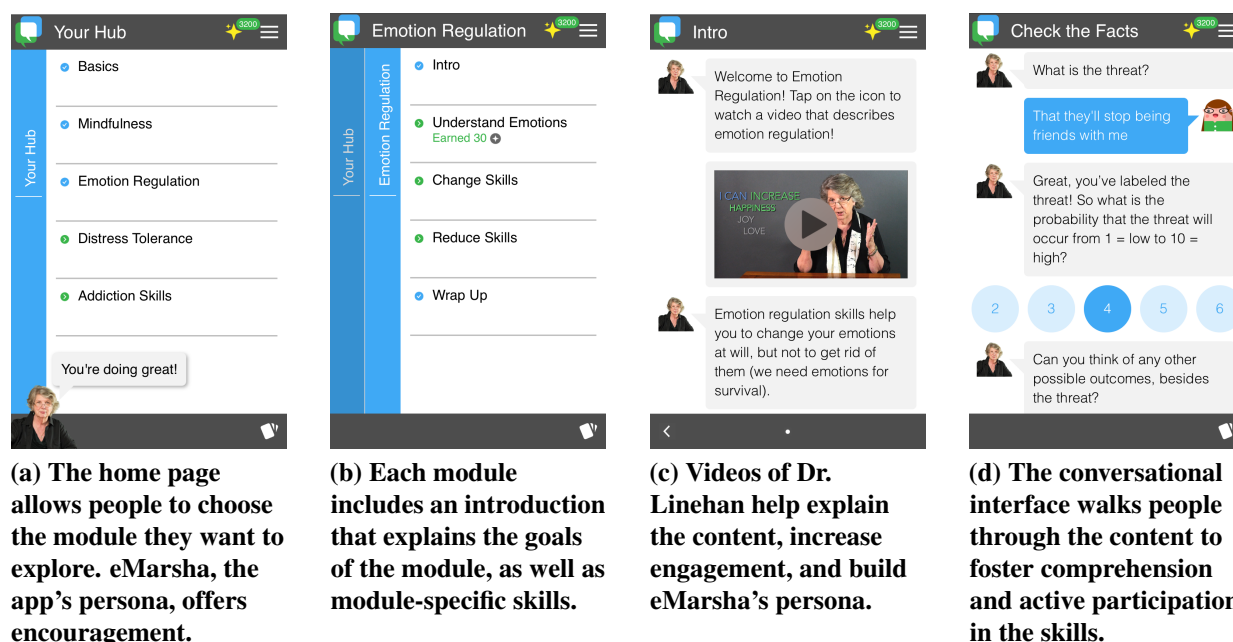


Figure 3.1: Pocket Skills allows people to learn and practice skills on the go via a conversational UI.

Prior to the 4-week feasibility study I discuss next (and before my involvement in the project), the Pocket Skills team conducted a pilot study in which they presented early iterations of the app to 13 people who were enrolled in DBT to assess usability and inform iterative design.

Pocket Skills is web-based to allow access across platforms, but designed for mobile devices to increase availability of the content. As cell phone ownership among mental health patients approximately equals cell phone ownership in the general population [32], designing for mobile screens also provides the potential to reach the majority of mental health patients. At the time of the study, Pocket Skills was hosted via Microsoft Windows Azure, as was the database storing app content and app usage data. Pocket Skills is accessed via an activation code and secure sign in.

App Flow

The Pocket Skills home page, or “Hub”, displays the supported DBT modules: *Mindfulness*, *Emotion Regulation*, *Distress Tolerance*, and *Addiction Skills* (Figure 3.1a). Before accessing the modules, people must first complete the “Basics” section, which walks people through the format of Pocket Skills and helps them set their overall DBT goals. Possible goals include increasing

positive emotions or behaviors (e.g., exercise, sleep, coping skills, productivity, positive friendships); decreasing negative emotions or behaviors (e.g., nonsuicidal self-injury, procrastination, addiction); and changing one's environment (e.g., by getting a new job, by managing friendships).

Each module contains an introduction to the module and the module-specific skills (e.g., Figure 3.1b). The module introductions begin with a high-level module description, often including videos describing the objectives of the module and explanations of what that module can help with (e.g., Figure 3.1c). After describing the module, the introductions walk people through setting module-specific goals. For example, someone using the *Emotion Regulation* module may have the goal of decreasing the frequency of unwanted emotions. After completing the introduction, people can learn about and practice the module-specific skills (e.g., Figure 3.1d). Some skills contain subskills, or different methods or options for practicing the skill (e.g., the *Mindfulness* skill of *Observing* includes subskills for observing *breathing*, *sounds*, *visuals*, and *everyday life*). Skills cannot be accessed until the associated module's introduction has been completed.

In addition to the modules, Pocket Skills also includes a *diary card* where people can record their emotions and behaviors. For familiarity reasons, this diary card is modeled on the paper diary cards often used in DBT to encourage mindfulness and track progress over time [140]. The data people track with the diary card are customized based on the goals they select in the “Basics” section.

People earn *points* on Pocket Skills when they practice a skill, fill out a diary card, or complete a module. Originally, points were included to encourage continued engagement with the app, and did not have any other function. However, participants in the pilot study suggested the addition of extra content people could access using the points. A virtual store therefore contains guided meditations narrated by Dr. Linehan that people can purchase with the points they earn.

Pocket Skills also includes daily text messages to remind people to consistently engage with the app. Text messages are sent in the morning to encourage participants to think about DBT and apply their DBT skills throughout the day. The texts link to Pocket Skills so people can access the app directly from the message. As repetitive messages can quality bore recipients, diversifying health-related messages is important to increase activity completion [124]. Pocket Skills messages are therefore selected from an extensive list to ensure variability between messages.

Conversational Interface Design

Pocket Skills uses a conversational interface to help people feel more engaged with the app: by having people actively participate in conversations, rather than passively reading information, the user experience resembles therapy more closely than traditional DBT skill practice worksheets. People can choose their avatar, but the conversational agent is always an image of Dr. Linehan, for a number of reasons. For one, as she narrated the instructional videos, using her as the agent provided a sense of continuity and cohesiveness: rather than being a collection of instructional media aimed to help people learn about and implement DBT skills, the app feels like a single entity supporting the persona through the development of their skills. In addition, since Dr. Linehan developed DBT, people pursuing DBT tend to be familiar with her. Using her image could therefore foster trust in the content.

Although the majority of the pilot study participants liked the conversational interface, many wanted a faster way to access the skills once they had learned about them. A *Skill Practice* section was therefore added to enable quick access to the full list of supported DBT skills. With this section, people can use a particular skill in the moment they need it, without the delay of the introductory material. Having the full list of skills can also help people remember every skill option.

3.1.2 Feasibility Study Methods

To evaluate Pocket Skills, we conducted a 4-week field study with 73 participants.

Recruitment

Participants were recruited via a DBT listserve. All participants needed to be enrolled in psychotherapy for the duration of the study and agree to go to their primary therapist for any questions or concerns they have about their health or activities in which they participated for the study. Participants also needed to own a smartphone, have a US phone number, and be willing to receive daily text messages. Participation required consenting to the study, completing an intake survey, and logging into the app within a week of receiving login instructions. 84 people initially enrolled in the study; however, 11 participants dropped out over the course of the study. These

Gender	65 female, 7 male, 1 Genderqueer/androgynous
Age	18-63($\bar{X}$ =37.3)
Education	High school degree (7), Some college (16), 2-year degree (5), 4-year degree (21), Professional degree (2), Master's degree (14), Doctorate (4)
Depression at Intake (Via PHQ-9)	Minimal or None (6), Mild (17), Moderate (21), Moderately Severe (16), Severe (13)
Anxiety at Intake (Via OASIS)	No anxiety disorder (12), Anxiety disorder (61)
Diagnosed Disorders	Depression (38), General Anxiety Disorder (35), Borderline Personality Disorder (29), PTSD (20), Bipolar Disorder (10); 17 disorders with <10 people each

Table 3.1: Demographic and condition information of the Pocket Skills feasibility study participants.

11 participants did not explain their decision to drop out. I therefore report on data from 73 participants (Table 3.1). The vast majority of participants were female, perhaps because women are more likely to be diagnosed with borderline personality disorder (BPD), the disorder for which DBT was initially developed (although prevalence is thought to be approximately equal between genders) [211]. Women are also more likely to seek therapy, both for BPD [83] and in general [246].

Procedures

Participants first completed an intake survey, which consisted of demographic questions and three clinical scales: the OASIS, which assesses anxiety [181]; the PHQ-9, which assesses clinical depression [127]; and the DBT Ways of Coping Checklist (“DBT WOCC”), which assesses frequency of coping via DBT skills use, general dysfunctional strategies, and blaming others [178]. Participants were then emailed an activation code, as well as instructions on how to log into Pocket Skills.

Throughout the study, participants received text messages every morning to encourage continued engagement. Participants were randomized into two messaging groups. One group received semi-personalized messages, which reminded them of their goals, their progress in the app, and the skills they had done (e.g., “One of your mindfulness goals is to reduce pain, tension, and

stress! Keep practicing mindfulness skills!”). I describe these messages as “semi-personalized” because we did not use machine learning to further tailor the messages to the participant, as prior work has examined (e.g., [104, 188]). The other group received non-personalized messages, which described app features in general (e.g., “the ‘Practice Skills’ section gets you to the skills ASAP!”). Prior research suggests that text messages are often better received when the content is personalized [171, 196], and personalized messages have been shown to reduce drinking [172] and support smoking cessation [234]. We therefore hypothesized that people who received semi-personalized messages would use the app more and show more improvement.

Every week, participants completed a short survey containing the OASIS, the PHQ-9, a 5-point Likert scale question about their satisfaction with Pocket Skills, and a text box for any feedback they had. The weekly survey did not include the DBT WOCC, as that scale is designed to be administered monthly. Due to the feedback we received in weeks one and two requesting more content, additional *Distress Tolerance*, *Emotion Regulation*, and *Mindfulness* skills were added two weeks into the study. In addition, *Emotion Regulation* skills were initially locked until participants went through the entire *Mindfulness* module, but was unlocked two weeks into the study due to the feedback we received.

At the end of the study, participants completed an exit survey that contained the OASIS, the PHQ-9, the DBT WOCC, and the Mental Health Self-Efficacy Scale [45], as well as optional questions about their mental health history. The survey also contained questions about the usability and usefulness of Pocket Skills, including the System Usability Scale [212], which evaluates the usability of the system, and primarily open-ended questions about their experience with Pocket Skills (e.g., their favorite and least favorite aspects of Pocket Skills overall, their favorite and least favorite modules, whether and how Pocket Skills supported their therapy).

Participants were compensated with a \$50 American Express gift card for each completed survey (excluding the intake survey), for a maximum total of \$200. Compensation did not depend on app use.

For the qualitative analysis, two authors (myself and Gloria Mark) participated in an open coding exercise of approximately 20% of the 1015 total participant responses to the 15 open-ended exit survey questions to identify emergent themes. I then coded the remaining data. Themes included learning, practicing, implementing, results, self-efficacy, engagement with conversational interface, and availability and context. I then applied grounded theory analysis for model development [233].

Limitations

Our recruitment method introduced some limitations. As we used a DBT listserve, all participants were at least somewhat familiar with DBT, and all participants but one were either currently enrolled in DBT or had been enrolled in the past. We therefore cannot say whether people without any DBT experience would have found Pocket Skills as beneficial as the participants.

Another limitation was the fact that all participants were concurrently enrolled in psychotherapy, due to legal and ethical considerations. Although our team included a clinical psychologist, we wanted participants to have a therapist to whom they could reach out in a crisis. As we did not have the resources to offer such support, participants needed to have a personal therapist. However, since all participants were enrolled in psychotherapy, we do not know whether Pocket Skills could support someone who is *not* in therapy. In addition, while we observed significant improvements on all clinical measures of interest and participants described positive results after using Pocket Skills, we did not have a control group that did not receive the intervention, nor did we have groups that only had access to subsets of the app features. I therefore cannot comment on the clinical efficacy of Pocket Skills versus normal therapy, nor the effectiveness of specific app features, given the lack of distinct comparison groups. Such an evaluation of clinical efficacy is an opportunity for future research. Such a study progression is consistent with best practices in health-related research in HCI [121].

Finally, a limitation to our methods is the fact that we did not measure self-efficacy in the intake survey. We therefore could not quantify any change in self-efficacy over the course of the study. Prior work has described improved self-efficacy after using systems for mental health management (e.g., [29]), and some participants reported increased self-efficacy in the open-ended questions. Self-efficacy should be studied more thoroughly in this context in the future.

Feature	<i>All Participants</i>		<i>Semi-Personalized Messages</i>		<i>Non-Personalized Messages</i>	
	Average	Average Per Day	Average	Average Per Day	Average	Average Per Day
Days Used	13.95	–	14.21	–	13.7	–
Time Used	2:18:08	0:04:36	2:30:37	0:05:01	2:06:17	0:04:12
Diary Cards Saved	7.74	0.26	7.87	0.26	7.63	0.25
Skills Practiced	21.37	0.71	23.68	0.79	19.18	0.64
Videos Watched	8.74	0.29	9.16	0.31	8.35	0.28

Table 3.2: Usage statistics for Pocket Skills features over the 28-day feasibility study. On average, the 73 participants logged into Pocket Skills at least once on almost half of the study days.

3.1.3 Results

In this section, I first present our quantitative results on Pocket Skills feasibility and usability. I then describe our qualitative results that explain how participants felt Pocket Skills supported DBT.

Pocket Skills Feasibility and Usability

Participants generally found Pocket Skills very usable. They reported a mean of 81.82, median of 85, and standard deviation of 16.01 out of 100 on the System Usability Scale (SUS). The suggested threshold for acceptable usability for the SUS is 68 [212]. Of the 73 participants, 21 were “satisfied” with Pocket Skills overall, and 37 were “extremely satisfied”. 71 participants said they plan to keep using Pocket Skills now that the study is over. Participants used Pocket Skills for almost half the study days, on average, with total time averaging more than 2 hours (Table 3.2).

Participants reported decreased depression, anxiety, and dysfunctional coping, as well as increased DBT skills use, after the study (Figure 3.2). To analyze our scale data, we used the statistical methods our clinical psychologist (Chelsey Wilks) had been trained to apply in such a study, with the hopes that using what the psychology community considers a rigorous analysis would encourage proliferation of the work beyond the HCI community. We used hierarchical linear modeling (HLM; e.g., [26]) with restricted maximum likelihood estimation (REML) to analyze the data. Compared to other analytical methods, HML is more flexible, as it treats time as a continuous variable to allow for the variability of

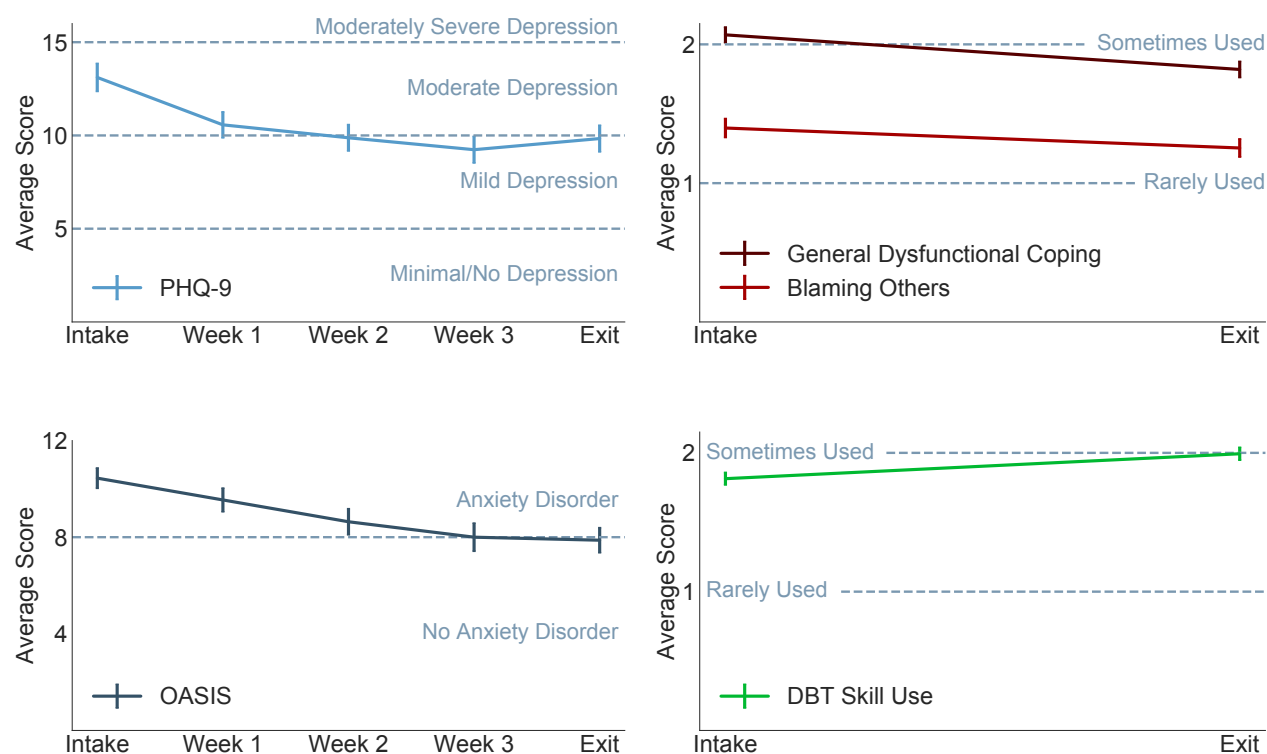


Figure 3.2: Scale results throughout the Pocket Skills feasibility study. On average, anxiety (OASIS), depressive symptoms (PHQ-9), general dysfunctional coping (WCCL subscale), and blaming others (WCCL subscale) decreased, while DBT skill use (WCCL subscale) increased.

the actual time assessment for each participant. HML also compensates for incomplete data across time, increasing power. We built separate unconditional growth models for the PHQ-9 and OASIS from baseline to exit. Parameters were systematically added to the model, and deviance statistics were analytically compared [245]. For both scales, the effects of time were estimated as linear and quadratic. The time variable was centered at the mid-point (the 2-week time point) to reduce collinearity between the linear and quadratic components. For both variables, the intercept and linear components were set to random. Analyses were completed using SPSS.

We fitted the growth models with intercept, slope, and quadratic effects capturing variation at baseline, linear change over time, and subsequent flattening over time. For the PHQ-9, both the linear component ($B = -0.79$, $p < 0.001$) and quadratic component ($B = 0.45$, $p < 0.001$) were significant, indicating participant PHQ-9 scores decreased by 0.79 each week, but that the rate of change slowed

by 0.45. For the OASIS, a similar pattern emerged where the linear ($B = -0.66$, $p < 0.001$) and quadratic ($B = 0.13$, $p < 0.05$) components were significant. These results illustrate that participants significantly improved on both the OASIS and PHQ-9 between the intake survey and week 1 of use, and continued improving (though at a slower rate) throughout the study (see Figure 3.2). For the DBT Ways of Coping Checklist (which was evaluated only at the intake and exit surveys), participants also saw an average decrease in dysfunctional coping ($B = -0.06$, $p < 0.001$) and blaming others ($B = -0.04$, $p < 0.05$), as well as an increase in their use of DBT skills ($B = 0.04$, $p < 0.001$), from baseline to post-study.

Once we had built the unconditional growth models, we evaluated whether message type and app use affected the slopes for each scale. The following group-level predictors were systematically evaluated: condition (i.e., semi-personalized or non-personalized messages); Borderline Personality Disorder (BPD) diagnosis; and total skills practiced. Each predictor was evaluated as a main effect and as an interaction term with each other on the linear and quadratic components. For the sake of parsimony, ill-fitting parameters were removed [245]. For the PHQ-9, the main effect of BPD was significant ($B = 4.82$, $p < 0.001$), suggesting that at week 2, individuals diagnosed with BPD had PHQ-9 scores that were nearly five points higher, on average, than those without a BPD diagnosis. The main effects of condition and total number of skill practice sessions were not significant; however, the total skills practiced was significant at the slope ($B = -0.15$, $p < 0.01$), indicating that individuals who practiced more skills had faster improvements on the PHQ-9. The three-way interaction of condition, total skills practiced, and time was significant for both the PHQ-9 ($B = -0.22$, $p < 0.01$) and the OASIS ($B = -0.11$, $p < 0.05$), indicating that individuals randomized to the semi-personalized messaging condition practiced more skills and reported faster improvements.

To further distinguish the *meaningfulness* of Pocket Skills on a clinical population, we then calculated reliable change (i.e., positive change better than statistical chance) and clinically significant change (i.e., reliable change above a clinically meaningful threshold) using criteria suggested by Jacobson and Truax [100]. The reliable change index (RCI) was calculated for each participant and outcome using the following formula: $RC = (\text{score at baseline} - \text{post-intervention score}) / SDIFF$, where $SDIFF$ is the standard error of difference based on baseline standard deviation and Cronbach's α . Clinically significant change

(CSC) was operationalized as reaching a level of functioning, reflecting in outcome scores, that was closer to the mean score of the non-clinical population than that of the clinical population. Participants showing both reliable change ($RCI > 1.96$) and clinically significant improvement were classified as “recovered”; those showing reliable change only were classified as “improved”; and those showing neither were classified as “unchanged/deteriorated”. RC indices, CSC cut-offs, sources of reliability, and norming data were as follows: PHQ-9 ($RC = \pm 6.61$, $CSC \geq 2.31$ [127]); OASIS ($RC = \pm 3.26$, $CSC \leq 13.44$ [180]); DBT Skills Use ($RC = \pm 0.47$, $CSC \leq 2.62$ [178]).

Overall, 41.7% of participants were considered “recovered” on the OASIS, while 8.3% improved. For the PHQ-9, 22.2% of participants were recovered, while 4.2% improved. Finally, 6.9% of participants were recovered on DBT skills use, while 12.5% improved. More individuals randomized to the semi-personalized message condition were recovered than those in the non-personalized condition for the OASIS (45.9% vs 37.1%) and the PHQ-9 (27.0% vs 17.1%), but more participants in the non-personalized condition were recovered on DBT skills use (8.6% vs 5.4%).

The mental health self-efficacy score was highly negatively correlated with general dysfunctional coping (a subscale of the Ways of Coping Checklist), as rated in the final survey: higher mental health self-efficacy correlated with lower general dysfunctional coping (i.e., *less* frequent use of these negative coping mechanisms) at the end of the study ($r = -0.658$, $p < 0.001$).

How Participants Felt Pocket Skills Supported Their DBT

In addition to examining participant improvements based on the validated scales we used, we also wanted to know whether and how participants felt that Pocket Skills supported their DBT. When we asked whether participants felt that Pocket Skills helped them *set* constructive DBT goals, we got an interesting range of responses. Some participants reported strongly feeling that Pocket Skills helped them set goals; *P71* explained that “*I had to choose what I wanted to work on and the constant reminders of my goals was really helpful*”. However, others felt that Pocket Skills had not helped them set DBT goals at all. Although participants could choose from common goals (both overall DBT goals and module-specific goals), the app did not support setting personalized goals. Some participants seemed to not even consider selecting predefined goals as setting DBT goals at all; *P36*

said that “*I didn’t realize it was meant to set a goal. Maybe a prompt to set that goal at the start of each section and a prompt to review the goals and monitor them would be helpful*”, and 4 others expressed similar sentiments. P54 felt that “*it might’ve been more helpful to create my own instead of choosing from a list*”, and P3 explained that she felt DBT only somewhat helped her set constructive goals because she “*didn’t feel like [she] could write down my goals*”. Allowing people to express their *personalized* DBT goals may have helped them feel that Pocket Skills supported their goal setting.

Although participants expressed mixed feelings about whether Pocket Skills helped them *set* their DBT goals, most seemed to feel that Pocket Skills *supported* their goals. They reported a mean of 7.4 (median 8) when asked how well Pocket Skills supported their DBT goals on a scale from 1 (“not at all well”) to 10 (“extremely well”). When asked how the past 4 weeks would have been different without Pocket Skills, P51 said that she “*would not have been so focused on my specific goals*”.

Finally, from the analysis of the qualitative exit survey data, we developed a model of how participants described Pocket Skills supporting their DBT (Figure 3.3). The model illustrates that participants felt Pocket Skills increased their *engagement* with DBT, which enabled them to *learn* the DBT principles and skills. This *learning* helped them *implement* the DBT skills in their daily lives when they needed them. They could then see how using those skills can lead to *positive results*, increasing their *self-efficacy*. I discuss these themes and provide representative examples. I include counts of participants that expressed specific perspectives; however, as the majority of the questions were open-ended, participants were not explicitly asked about these themes. I therefore note that more participants than indicated in the counts may have agreed with a given perspective, if asked.

Pocket Skills Increases Engagement with DBT

Prior research has emphasized the need for mental health technology to be engaging, so people will use the technology and benefit from the intervention [17, 60, 61, 158, 164]. Engagement is particularly important for DBT, as development and adoption of DBT skills is critical to effective treatment [177]. We discovered that Pocket Skills engaged participants not only with the app itself (see Table 3.2), but also with DBT more generally, due to a number of factors. One major aspect participants found helpful was the conversational interface. Thirty-one participants reported

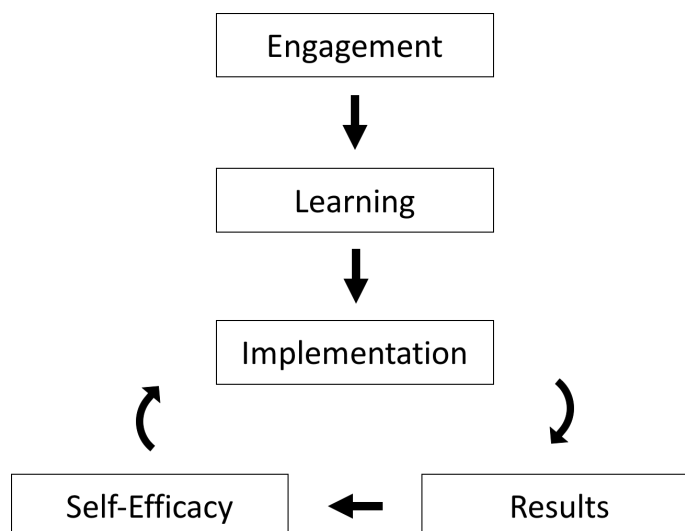


Figure 3.3: Our model illustrates how participants felt Pocket Skills supported their therapy.

enjoying the interactive interface and finding it engaging, especially compared to traditional methods of learning and practicing DBT skills such as using worksheets [141]. For example, *P1* thought that “*having the interactive app was vastly superior to just having a book*”. Participants described particularly liking the conversational interface, which helped them more actively engage with the material. *P62* explained that what she liked most about Pocked Skills was that “*the Q&A / conversational tone [helped] me think about how to help myself*”. By actively guiding participants through their DBT skills, the app helped them engage with the skills they were practicing.

One of the reasons participants reported enjoying Pocket Skills was the simulated therapy with Dr. Linehan, the creator of DBT and a well-respected figure in the DBT community. Because Pocket Skills was based on Dr. Linehan’s material, participants found it trustworthy. Twenty-nine participants discussed appreciating the fact that Pocket Skills used “Marsha” to teach them the material and walk them through the skills. *P37* commented that she “*knew [Pocket Skills] was a legit product, not someone jumping on the mindfulness wagon, because of Dr. Linehan*”. One of the top barriers to using apps for people with depression is perceived intervention efficacy: people who do not believe that the intervention is correct are not likely to engage with an app [230]. Fostering this trust in the material is therefore important to encourage engagement.

Perhaps due to their familiarity with Dr. Linehan, participants described personifying *eMarsha*, which helped them feel supported by the app despite lack of human involvement. Some participants described *eMarsha* as though they were actually interacting with a human coach. For example, *P4* explained why she found Pocket Skills helpful: “[*With Pocket Skills*] I could have someone help me through what I was supposed to be doing instead of me just winging it by myself and getting lost in my head”. *P22* similarly felt supported by *eMarsha*, stating that “‘*Marsha’s*’ presence gave me the support that I needed to step back ... and look at my own feelings and how my actions and behavior were contributing to those feelings”. *P11* particularly liked the daily text messages because “It really helped to know someone was there”. This sense of bonding with a conversational agent was also reported for Woebot, an agent designed to support CBT [72]. However, *eMarsha* is entirely scripted with some branching, whereas Woebot includes natural language processing so “he” can intelligently respond to free text. Despite our more rudimentary approach to a conversational agent, participants still personified *eMarsha*, which helped them stay engaged with Pocket Skills.

Although many participants liked *eMarsha*, 8 mentioned being unable to bond with the agent, which kept them from being as engaged. These participants generally felt the agent was too generic and impersonal. For example, what *P40* disliked most about Pocket Skills was that “*the Marsha canned responses felt stupid*”. For some, the lack of human involvement was viewed as an asset; *P5* explained: “*I am introvert so I love it. I don’t have to sit in a room of people so I can learn DBT. My first round of class was so uncomfortable because of that that I quit*”. However, the fact that some people thought *eMarsha* felt so “virtual” (*P30*) could negatively impact engagement for others. By incorporating machine learning and natural language processing into Pocket Skills, we could make *eMarsha* more personalized and responsive, and therefore more engaging.

Another reason people reported feeling engaged with Pocket Skills was the fact that it “complimented” (*P41*) their in-person therapy. Thirty-nine participants felt that Pocket Skills worked best as a supplement to traditional DBT, rather than a replacement. *P63* explained that she found Pocket Skills “*a really excellent supplement. I wished that others in my program had the chance to use it, also. I think it is a really helpful review/refreshers*”. The fact that Pocket Skills generally had a “consistent message” (*P8*) with their in-person therapy may have also helped with engagement.

Participants also felt Pocket Skills helped them stay engaged with DBT by consistently reminding them to interact with the app, and therefore to engage with DBT in general. Fifty-nine participants described Pocket Skills helping remind them of and reinforcing their DBT skills. One of the ways Pocket Skills helped people keep DBT consistently on their minds was through the text messages. *P4* explained: *“I lose time and will forget to practice so the daily reminders were very helpful”*. What *P40* liked most about Pocket Skills was *“just getting a text message each day”*. Perhaps contributing to the fact that the semi-personalized messages did not make as much difference to usage as we expected, some participants explained that the content of the messages was moot. *P63* said, *“the texts were helpful as reminders. I did not pay close attention to their actual content too much”*. Because people often tire of too-frequent messages [157], we thought participants might dislike daily messages. However, when asked how frequently they would like to receive messages, 21 participants wanted *more* frequent texts, and 28 liked getting them every day. No one reported wanting to receive texts less frequently than once per week. Participants reported appreciating the reminders because it helped them stay engaged with Pocket Skills and their DBT.

In addition to reminding people to think about DBT through the text messages, participants reported that just having Pocket Skills installed on their phones helped reinforce the principles and skills that DBT teaches. *P32* explained: *“The app was a concrete reminder that I have a wealth of skills - which I already know quite well - to help me get through everything that has been going on, without crashing. It gave me a quick refresher on the details of skills I hadn’t thought much about for awhile”*. Similarly, *P31* said: *“the app really sharpens my focus, reminds me of all the skills there are to draw on, and reminds me to practice!”*. *P18* felt that Pocket Skills helped her engage with DBT even when she did not use the app: *“Sometimes just having the app in my phone even without opening it was a good reminder to practice my skills a couple of times at work this week”*. Simply having Pocket Skills available helped participants stay more engaged with DBT.

Engagement with Pocket Skills Enables Learning DBT Skills

Engagement with Pocket Skills, and DBT in general, seemed to help participants learn the principles and skills associated with DBT. Sixty-six participants felt that Pocket Skills helped them learn useful skills. One of the reasons people reported being able to learn with Pocket Skills was its clear explanations, which 35 participants mentioned appreciating. *P51* explained, “*I like it the best of the other apps I have tried, because it actually teaches the skills rather than just reminding you of them or tracking things*”. By including explanations, Pocket Skills ensured that the content was broadly accessible, even people who had struggled with, forgotten, or not previously learned the material.

Some participants also felt Pocket Skills helped them learn by including “*actual realistic scenarios and situations*” (*P54*). Ten participants mentioned that the examples that Pocket Skills included facilitated learning the material. *P24* explained, “*The Mindfulness skills and explanation of it and examples in each section really helped to clarify the concept of mindfulness itself, which I struggle with*”. Through concrete examples, participants felt they could better understand the content.

Although Pocket Skills helped many participants learn, 15 felt their learning was hindered due to their inability to ask questions. For example, *P7* felt she needed someone to help her through the material: “*I needed more guidance on what units to practice, on how to work my way through the app. So, it needs an online coach (a live person, not just the eMartha avatar), or it needs to be used under the direction of a real-world therapist*”. *P36* summarized that one of the biggest weaknesses of Pocket Skills is that it “*doesn’t allow you to discuss anything with a real human*”. Prior work has investigated allowing people to message their therapists about technology designed to support mental health, or even through it (e.g., [17, 60]). Pocket Skills may benefit from such a feature.

In addition to gaining knowledge about DBT, 18 participants felt Pocket Skills helped them gain insight about themselves. For example, *P19* felt that Pocket Skills helped her learn which skills were helpful to her personally: “*In high emotional state it is good to know these are effective for me*”. Pocket Skills also helped people learn about their emotions, which is a major goal of the *Emotion Regulation Module*. *P54* said, “*[Pocket Skills] helped me be able to recognize my emotions better*”.

This self-knowledge often helped people use more positive coping skills; *P25* explained: “*knowing exactly what I’m feeling helps me be able to intercede appropriately*”. The fact that Pocket Skills helped people learn more about themselves enabled them to better relate to the DBT material.

Although the self-knowledge people obtained through Pocket Skills was generally positive and constructive, some participants learned aspects about themselves that they found upsetting. For example, *P15* mentioned that his least favorite module was the *Addiction Skills*. When asked why, he explained: “*Mostly because I learned that I engage in addictive behavior*”. Such a realization is difficult, but may ultimately be positive if it helps the participant overcome the addiction. However, other participants learned information that was less constructive. For example, *P50* said, “*community reinforcement [an Addiction skill] reminded me that I am alone and have no support*”. Given these comments, future iterations of Pocket Skills may need to take the possibility of negative self-knowledge into account and either suggest resources for anyone struggling with the information they learn about themselves or encourage them to contact their therapist.

In addition to learning DBT material, participants felt Pocket Skills gave them the means to practice DBT skills. Sixty participants felt they practiced skills more with Pocket Skills than they would have without it. For example, *P17* said, “*I’ve been in therapy for years, but I’m not great about practicing my therapeutic homework. Pocket Skills helped me practice my skills in between therapy appointments*”. Similarly, *P26* found Pocket Skills motivated her to recommit to her DBT skills: “*Pocket Skills got me motivated to brush up on the various skills in each [module] and to stay in practice with them*”. Practicing DBT skills is an essential step to fully understanding them (and to implementing them, as I discuss in the next section) [177]. The fact that participants felt Pocket Skills encouraged them to practice is therefore a key piece to helping them internalize the material.

Learning Enables Implementation

Participants reported that Pocket Skills helped them not only learn and practice skills, but also implement those skills in their lives when they needed them, instead of choosing a negative coping mechanism. As discussed above, we found significant increases in skill use between the intake and exit surveys. Fifty-four participants described Pocket Skills helping them apply DBT skills in their

daily lives. For example, *P53* explained: “*I don’t just learn the skills but Pocket Skills helps remind me what the skills is and how to use it step by step to apply it on a daily basis so it becomes a part of me*”. Even participants who had already started implementing DBT skills felt that Pocket Skills helped them: “*[Pocket Skills helped me] increase my depth and consistency of applying skills on a daily basis from prior to using the app*” (*P31*). With help from Pocket Skills, participants were able to better bridge the gap between simply practicing their DBT skills and actually applying those skills.

One reason participants described feeling Pocket Skills helped them implement DBT skills was the fact that it contained a variety of skills from which they could choose. Seventeen participants said the clear options Pocket Skills provided helped them make good choices about what skills to implement when. For example, one of the aspects *P14* liked most about Pocket Skills was “*learning different ideas to self soothe*”. The ability to see and access such a wide range of options was particularly helpful when people were experiencing emotional distress. *P5* mentioned: “*[the ‘Skill Practice’ section was] a place I could go to find a skill that might help me when I was feeling badly or having an emotional issue and I couldn’t think about what skill was the best. I could be reminded by seeing its name*”. By providing options of positive coping mechanisms when participants were struggling, Pocket Skills helped them learn how to apply the skills when they most needed them.

The availability of Pocket Skills was also key in enabling participants to implement DBT skills. Pocket Skills was available whenever participants had access to their phone and the Internet, which helped keep the skills prominently in their mind. Forty-one participants mentioned the importance of this availability, with *P58* describing it as the link between learning and implementing DBT skills:

Having something to explain [the skills] and refer back to in the moment anytime I want or need was an invaluable resource ... The app allows you to practice in the moment, every day, any skill and have quick access right at your fingertips, since you always have a phone (but don’t always have a worksheet say). So it decreases the barriers and burden of practicing which is great.

The constant availability of Pocket Skills also helped enable participants to use DBT skills *in context*, at times of emotional intensity or distress. What P28 liked most about Pocket Skills was “*the security of having the app readily available to help [her] if a crisis arose*”. Similarly, 23 participants mentioned the importance of being able to access their skills in context, when they most needed them. P46 explained: “*It’s useful with the app to be able to scroll through and pick out quickly what skill I need to practice. Without it, it can be hard (especially in distress) to remember what my options are*”. By allowing people to access their skills in the right context (e.g., when they were in distress), Pocket Skills helped them learn how to implement those skills in their daily lives.

Implementation Enables Results and Self-Efficacy

Participants felt implementing DBT skills in context helped them see the positive results of using those skills. In addition to the significant results we found via the scales, 32 participants explained that they saw concrete results in their lives. For example, P2 felt Pocket Skills helped improve her relationships: “*I think there were a number of times that I would not have been as successful in interactions with my friends and family [had I not had Pocket Skills these past 4 weeks]. Just thinking about DBT on a daily basis helped keep my skills in the forefront!*”. The self-tracking of positive and negative moods and behaviors via the diary cards was often particularly helpful in seeing results. What P17 liked most about Pocket Skills was “*seeing my progress when the app would tell me my ratings are improving*”. P27 similarly saw a difference in her distress when she used Pocket Skills to apply her DBT skills: “*When I actively used Distress Tolerance skills I was able to reduce my SUDS [(Subjective Units of Distress Scale)] score*”. Pocket Skills also helped some participants change their behaviors. For example, P47 said if she had not had Pocket Skills, she “*Likely would have used SIB [(Self-injurious behavior)] More frequently*”. Not only did the scales show reduced depressive and anxiety symptoms, but the qualitative results also illustrate that participants felt Pocket Skills helped them see significant improvement in their lives.

Participants also described feeling an increased ability to change after using Pocket Skills. Prior research suggests self-efficacy is important for positive symptom outcomes when using a mental health app for depression management [45, 71], and as discussed above, our own results similarly

showed a significant negative correlation between self-efficacy and general dysfunctional coping. Sixteen participants mentioned using Pocket Skills increased their self-efficacy. *P1* explained Pocket Skills helped her realize she was capable of using DBT skills: “*They helped me break down something that seemed overwhelming into manageable bits I could do on my own*”. Similarly, *P46* said:

[Pocket Skills gave me] increased independence. I used it during moments of distress when I might otherwise have called someone for help. Phoning friends/family/doctors often leaves me feeling embarrassed after I’ve calmed down. It is rewarding to feel like I can make it through intense moments with a little more independence, and Pocket Skills was helpful with that.

Increasing self-efficacy can in turn enable future skill implementation. *P22* explained, “*I think that I would feel less confident in my ability to use DBT skills now if it weren’t for Pocket Skills*”. Because she has more confidence in her ability to implement DBT skills, *P22* felt more likely to continue to do so in the future. Over time, I hope that an app such as Pocket Skills could reinforce the cycle of implementing skills, seeing concrete results, and increasing self-efficacy to manage mental illness.

3.1.4 Discussion and Future Work

Here, I discuss design implications our feasibility study revealed for technologies supporting mental health, as well as opportunities to better support people and their therapists in the future.

Model Implications: Need for Availability and Engagement

Our model reveals the importance of enabling people to access skills in context to help bridge the gap between *practicing* DBT skills and *implementing* those skills in their daily lives. The importance of *availability* suggests technologies designed to support mental health should have mobile components, especially if the mental health intervention includes skill implementation.

Our model also reiterates the importance of *engagement* for mental health technology described by prior work (e.g., [17, 60, 61, 158, 164]). One of the prominent ways participants found Pocket Skills engaging was the conversational agent. Participants appreciated the appearance of actively

conversing to learn material and complete skills, finding the interface more engaging and thought provoking than passively completing worksheets. Future work could incorporate machine learning to make *eMarsha* even more engaging. For example, “she” could suggest skills based on what has worked well in the past, or even based on what works well for other people with the same disorders.

We were surprised to see so little difference between semi-personalized and non-personalized messages, as we expected semi-personalized messages to increase engagement. Participants described paying little attention to the content of the messages, simply treating them as a reminder. In this context, message content may be immaterial. However, the fact that participants in the semi-personalized messages condition practiced more skills and had faster rates of change in the OASIS and PHQ-9 suggests personalization may be more significant over a longer time period. Future work should investigate whether more personalized content could better support engagement with the app, and could assess the effects of different messaging types on longer-term engagement.

Visualizing History and Improvement

Participants discussed appreciating the ability to see the difference Pocket Skills made in their behaviors and emotional regulation. However, although Pocket Skills collects data on skills practice sessions and diary cards, it does not currently fully support people in reflecting on and learning from that data. People can access past diary cards, which some participants mentioned finding useful; what P8 liked most about Pocket Skills was “*Watching the change in [her] diary cards*”. However, Pocket Skills does not currently visualize conversational history with *eMarsha* or any data around app use over time. Prior work has discussed the importance of visualizations of self-assessment data in a system designed for people with bipolar disorder [17]. By including ways for people to visualize their behaviors and emotions over time, they could better understand the results of engaging with DBT. Allowing people to access histories of their conversations with *eMarsha* could allow them to view the evolution of their skill use and remember which skills have been helpful for them in the past.

Considering Therapist Needs

Pocket Skills was designed by an interdisciplinary team that included clinical psychologists. However, Pocket Skills focuses on supporting individuals enrolled in therapy, and does not currently consider or support therapist needs. Health provider needs are often overlooked in tools for mental health, but providers generally have a strong interest in having such tools [220]. A number of features could help Pocket Skills better support therapists. For example, four participants mentioned wishing they could print or email their diary cards or skill practice history so they could show them to their therapists. By designing ways to share such data, Pocket Skills could help people and their therapists collaboratively review their progress and make decisions regarding future treatment.

Pocket Skills could also help people better manage their mental health conditions by supporting collaborative use. The current design assumes that a single person pursuing DBT is using the app, but three participants mentioned using Pocket Skills with their therapists. Furthermore, a therapist to one of the participants reached out to us via email praising Pocket Skills, mentioning that it was “[v]ery helpful in stimulating great discussions today about mindfulness” in her group DBT session. By including designs for use alone, with a therapist, and in a group setting, Pocket Skills could better support the myriad ways people engage with DBT. Future work could investigate exactly how people and therapists would like to use a tool such as Pocket Skills, both individually and collaboratively.

Supporting People Without Access to Therapy

Due to ethical considerations, our study only included people enrolled in psychotherapy. However, many people who would benefit from psychotherapy cannot access it [125,249], due to factors such as cost [243] and stigma [150,254]. Ultimately, I would like an app such as Pocket Skills to not only support people who are in therapy, but also to provide help to those who cannot access it. As Pocket Skills is designed to holistically support DBT, including educational materials explaining its principles and skills, it may include sufficient information to help some people who are unfamiliar with DBT. However, participants felt Pocket Skills still might be confusing or uninteresting to someone without prior DBT experience. P73 explained that “it is important to have a base knowledge

of the stuff so an individual knows how to appropriately engage". Pocket Skills currently assumes people understand enough about DBT to know they want to pursue it. To support people with no DBT experience, it may need to include information about why DBBT was developed and who and how it helps. As people who do not know what DBT is will not be familiar with Dr. Linehan, it may also need to work harder to develop the same level of trust in *eMarsha* that participants reported.

3.1.5 Study Conclusion

We created Pocket Skills, a mobile web app designed to support Dialectical Behavioral Therapy (DBT). Pocket Skills guides people through DBT education and skills practice via *eMarsha*, a conversational agent modeled on Marsha Linehan, the developer of DBT. We conducted a 4-week field study of 73 participants to test the feasibility of using Pocket Skills to support DBT. After the study, participants showed significant improvement in depression, anxiety, and DBT skills use. In addition, we developed a model of how Pocket Skills supported DBT based on our qualitative analysis of open-ended questions in the exit survey. Participants reported that Pocket Skills helped them *engage* both in the app and in DBT in general. This engagement helped them *learn* the principles and *practice* the skills in context, helping them *implement* those skills in their everyday lives. Participants were therefore able to see the *concrete results* of implementing their DBT skills and improve their *self-efficacy*. I discussed the necessity of technology being available and engaging to support mental health treatment, as well as the importance of visualizing history and improvement, considering therapist needs, and supporting people without access to in-person therapy. This work therefore motivates future study, design, and development of tools to support mental health.

3.2 INVESTIGATING AND PREDICTING SKILL-LEVEL EFFECTIVENESS

Our Pocket Skills feasibility study provided a concrete example of translating an evidence-based treatment into a mobile app. Because we collected usage data throughout the study, it also provided a unique opportunity to investigate whether specific skills were more or less effective in fostering mindfulness or reducing emotional distress and intensity, which could inform designs that help

ensure people receive the best possible support. Comparing the effectiveness of different skills for different subgroups of people (e.g., those with particular disorders, demographics, or other characteristics) could also reveal individuals that may need additional support. Such analyses of the effectiveness of *particular skills* in *real-world contexts* have generally been infeasible in traditional DBT practice and research. Working closely with Jina Suh, I therefore conducted a quantitative analysis of our survey and app usage data to identify differences in the effectiveness of different skills, both overall and for different subgroups of people (e.g., people with specific diagnosed disorders, people of different ages). We investigated the following questions:

RQ1: When did study participants use the skills?

RQ2: Were particular skills more effective (e.g., in reducing distress, in fostering mindfulness)?

RQ3: Were skills more effective for different subgroups of people (e.g., those with certain disorders)?

RQ4: Did skill-level effectiveness influence study-wide depression, anxiety, or skill use improvement?

RQ5: Can we *predict* a skill's effectiveness, given participant and skill characteristics?

3.2.1 Dataset

To examine our research questions, we reanalyzed data from the Pocket Skills feasibility study. The data required additional processing for our analyses, and have associated limitations.

Additional Data Processing

As discussed in Section 3.1.2, we collected data on participant characteristics via surveys throughout the Pocket Skills feasibility study, including: 1) demographic information (e.g., gender, age); 2) their anxiety, depression, and coping skill use (measured with the OASIS [181], PHQ-9 [127], and DBT Ways of Coping Checklist [178]); and 3) other characteristics (e.g., what disorders they have been diagnosed with, whether they take medication, what modules they preferred). We also collected app

usage data throughout the study, including app navigation and participant inputs for the skills. Many skills included Likert-scale ratings of how a participant felt before and/or after completing the skill, which I refer to in this chapter as *pre- and post-ratings*. For example, *Mindfulness* skills often asked people to rate how mindful they felt after completing the skill; *Emotion Regulation* skills often asked people to rate their emotional intensity before and after completing the skill; and *Distress Tolerance* skills asked people to rate their level of distress before and/or after completing the skill.

To process the Pocket Skills study data, we first reviewed the app content to examine the skills themselves (see Table 3.3). We identified 11 skills that had pre- and post-ratings and 26 skills that had only post-ratings: 18 distinct *Mindfulness* skills (all with post-ratings only); 5 distinct *Emotion Regulation* skills (all with post- and pre-ratings); and 14 distinct *Distress Tolerance* skills (9 with post-ratings only and 5 with post- and pre-ratings). We excluded from our analyses any skill that did not include any Likert-scale ratings. No participant completed the *Emotion Regulation* skill *Cope Ahead*; all other skills were practiced at least once by at least one participant. No *Addiction* skills had Likert-scale ratings, so we excluded all skills from that module. Table 3.3 shows our final list of *modules*, *skills*, and *subskills*. Four participants did not complete any of the included skills over the course of the study, and were therefore excluded from the dataset.

After we identified the skills to include, we extracted each distinct use of a specific skill from the usage log data and derived metadata (e.g., the ratings, the first use of a skill, the total order of practiced skills). Each interaction with the app (e.g., enter/exit skill screen, rate mindfulness) was logged with a local timestamp from each participant's device, from which we derived temporal metadata (e.g., skill rating time of day, day of week). Because participants were able to access the app even after the study was completed, we filtered the data to only include skills practiced between the study intake (July 19th, 2017) and the last submitted exit survey (August 23rd, 2017). We then standardized the ratings: some skills had Likert-scales from 1-10, some from 1-5, and some binary (yes/no). *Mindfulness* scores were better if higher (i.e., a higher score indicated more mindfulness), but lower scores were better for all other modules (e.g., lower *Distress Tolerance* scores indicated lower distress). We therefore shifted everything to a 5-point scale, and reversed *Mindfulness* ratings so lower ratings were better (i.e., indicated higher mindfulness). Finally, we

Module	Skill	Subskill
Mindfulness	Observe ^b	Breathing Sounds Visuals Everyday Life
	Describe ^b	Visuals Thoughts Expressions Everyday Life
	Participate ^b	Counting Jumping Laugh Club Saying Words Walking Everyday Life
	NonJudgementally ^b	Observe Judgments Rephrase Judgmental Statements
	One Mind ^b	—
Emotion Regulation	Check the Facts ^a	—
	/Brief CtF	—
	Opposite Action ^a	—
	/Brief OA	—
	Problem Solve ^a	—
Distress Tolerance	/Brief PS	—
	Distract ^b	Comparisons Emotions Pushing Away Sensations Thoughts
	Self-Soothe ^a	Hearing Smell Taste Touch Vision
	TIP ^b	Intense Exercise Paced Breathing Progressive Muscle Relaxation Temperature

^aSkills with both pre- and post-ratings (with which we could calculate skill improvement).

^bSkills with post-ratings only.

Table 3.3: Pocket Skills modules, skills, and subskills examined in our quantitative analyses.

Gender	62 female, 6 male, 1 Genderqueer/androgynous
Age	18-63 ($\bar{X}$ =37.3)
Age Buckets	<25 (12), 25-35 (17), 35-45 (21), >45 (19)
Education	High school (7), Some college (16), 2-year degree (5), 4-year degree (21), Professional degree (2), Master's degree (14), Doctorate (4)
Education Buckets	<4 year degree (28), 4 year degree (21), >4 year degree (20)
Depression at Intake (Via PHQ-9)	Minimal or None (6), Mild (17), Moderate (18), Moderately Severe (16), Severe (12)
Anxiety at Intake (Via OASIS)	No anxiety disorder (12), Anxiety disorder (57)
On Related Medication	12 no, 54 yes, 3 declined to answer
Types of Disorders	Mood (42), Anxiety (43), Personality (27), Neurodevelopmental (8), Eating (3)
Close Family Member Buckets	0 (13), 1-2 (24), 3-4 (20), >4 (12)

Table 3.4: Pocket Skills participant categories based on survey responses, as well as buckets developed for the participant groups and subgroups for more even distribution.

computed the difference between pre- and post-ratings to characterize improvement before and after completing a skill, referred throughout this dissertation as *skill improvement*. Because lower numbers indicate better ratings, negative numbers indicate more improvement.

We grouped participants within categories for analysis based on their survey responses (see Table 3.4 for categories and distributions). Categories included education; age; gender; intake survey scores on the PHQ-9 and OASIS; the number of family members living close (within a 50-mile radius); whether they were on any mental health related medication; and any mental health disorders they were diagnosed with. For age, close family, and education, we defined buckets to better balance the groups. Because participants had been diagnosed with a wide range of disorders, we grouped disorders by category based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) [8]. Categories included neurodevelopmental disorders, mood disorders, anxiety

disorders, personality disorders, eating disorders, and no diagnosed disorder. Each participant was then categorized as having or not having each disorder type, based on their reported disorders. Diagnosed disorders were all self-reported; the original study did not include clinical diagnostic interviews.

Additional Data Limitations

Our dataset has a several associated limitations for our analyses. First, we found evidence of incomplete data logging: one datapoint had a post-skill rating without a pre-skill rating, which should not have been possible given the app design. We excluded that datapoint, but its existence shows that our skill practice data is not complete. I also cannot say how many skills people practiced or applied in their daily lives without using the app. Participants reported that having the app reminded them to use skills in the moment, sometimes without the app; although the app included self-tracking of moods and behaviors, it did not allow users to log skills practiced outside of the app. Third, for many skills, participants may not have actually completed the skill at the time it was suggested. For example, some *Distress Tolerance* skills suggest activities such as taking a walk or a cold shower; the app did not support indication of whether or when participants actually completed those activities. Future tools designed to support DBT could include ways to report these aspects of DBT skill practice, which would allow further research and development of intelligent support for skill practice. Fourth, I can only comment on the usefulness of the skills as they were translated in the app, which may be different than the usefulness of skills more traditionally taught and practiced.

Finally, the clinical psychologists on the team (Chelsey Wilks) ordered the modules in the app based on traditional DBT practices. Although this ordering best preserved the traditional experience of practicing DBT, it may have encouraged more use of *Mindfulness* skills (which happened to be the ones with only post-ratings). Additionally, at the beginning of the study, participants could not access the *Emotion Regulation* module until they had gone through the entire *Mindfulness* module. Based on participant feedback, all modules were unlocked after the second week; however, this initial limitation encouraged more use of *Mindfulness* skills, which lacked the pre-ratings needed to observe skill improvements. The *Distress Tolerance: Self-Soothe* skills were also added halfway

through the study, limiting the amount of time participants were able to use them. Future studies investigating relative skill effectiveness should consider designs that collect pre- and post-ratings for every skill, and should examine and account for possible ordering effects.

3.2.2 Methods and Results

In this section, I describe our methods and results from the statistical analyses we conducted and machine learning (ML) models we created to investigate our research questions.

First, we addressed RQ1–RQ4 by examining the data along different individual dimensions through stratification and statistical tests. We first investigated overall skill use throughout the study as well as skill effectiveness (e.g., how much participant emotional intensity or distress improved after using the skill, how mindful participants felt after using the skill) both in general and between different subgroups of participants. We then examined how these participant and skill-level characteristics correlated with mental health improvements over the course of the study (i.e., improvements in anxiety, depression, and skill use). We used Python to clean and process the data and calculate descriptive statistics, then used R for our statistical analyses.

Finally, we addressed RQ5 by consolidating our findings and explorations in machine learning models that jointly use the variables and leverage interactions between variables to predict skill improvement given participant and skill characteristics and demonstrate the incremental value of additional information. We used scikit-learn, a Python ML package [192], to develop these models.

RQ1: Skill Use

We first examined overall module and skill use throughout the feasibility study.

Methods

To investigate overall skill use, we visualized the data and computed basic sums and averages.

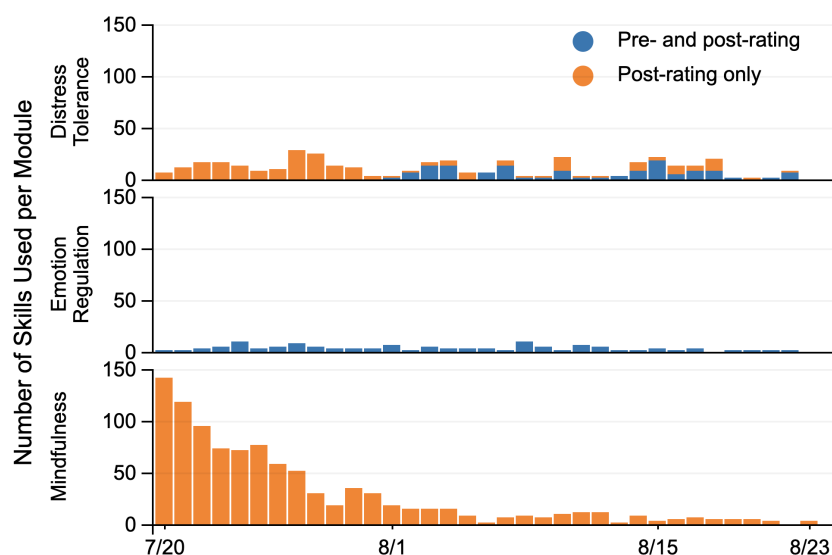


Figure 3.4: Distribution of the total number of DBT skills practiced with Pocket Skills over the course of the feasibility study, showing skills with pre- and post-ratings and skills with only post-ratings.

Results

We found 1483 total skills practiced by 69 participants: 974 *Mindfulness* skills (all with post-only ratings); 127 *Emotion Regulation* skills (all with pre- and post-ratings); and 382 *Distress Tolerance* skills (131 with pre- and post-ratings and 258 with post-ratings only). *Mindfulness* skills were most often used at the beginning of the study, with *Emotion Regulation* and *Distress Tolerance* skills more evenly distributed throughout (see Figure 3.4). Between July and August, *Mindfulness* skill use went from 78% of total usage to 37%, while *Distress Tolerance* went from 16% to 47%, perhaps because *Distress Tolerance: Self-Soothe* skills were added on August 1st. In the same time period, *Emotion Regulation* skill usage increased from 5% to 16%, perhaps because the *Emotion Regulation* module was universally unlocked (i.e., no longer depended on the completion of the *Mindfulness* module) on the same date. Usage patterns may also be partially due to an ordering effect (i.e., because the *Mindfulness* module was presented first in the app). Total usage of the skills dropped by more than half between July and August, indicating that skills were used less overall as the study progressed.

Each participant completed an average of 21.5 skills throughout the study (min=1, max=93, stdev=20.57). On average, participants completed 14.1 *Mindfulness* skills (min=0, max=70, stdev=15.6), 1.84 *Emotion Regulation* skills (min=0, max=10, stdev=2.15), and 5.54 *Distress Tolerance* skills (min=0, max=42, stdev=7.00). Of the 69 participants, 51 practiced at least one skill more than once. Of the skills practiced with pre- and post-ratings, 213 (82.6%) had pre-ratings of ≥ 3 (i.e., indicated high levels of emotional intensity or distress before completing the skill).

RQ2: Overall Skill Effectiveness

We next compared overall skill effectiveness for the different modules and skills across all participants.

Methods

After standardizing skill ratings, our data included skills with only pre-ratings, skills with only post-ratings, and skills with pre- and post-ratings. To investigate improvement trends, we examined skills with both pre- and post-ratings; to investigate post-rating trends, we examined skills with only post-ratings. For overall skill use, we examined skills at both the *module* and *subskill* granularities (see Table 3.3). We used t-tests to analyze effectiveness by module: one to examine differences in skill improvement between the *Emotion Regulation* and *Distress Tolerance* skills (the only modules with skills that had pre- and post-ratings) and another to examine differences in post-ratings between *Distress Tolerance* and *Mindfulness* (the only modules with skills that had only post-ratings).

For our *subskill*-level analyses, we used one-way analysis of variance (ANOVA) tests to examine differences in 1) skill improvement and 2) post-ratings by *subskill*. When we found significant results, we then investigated pairwise differences, employing Tukey's HSD procedure to correct for the increased risk of Type I error due to unplanned comparisons.

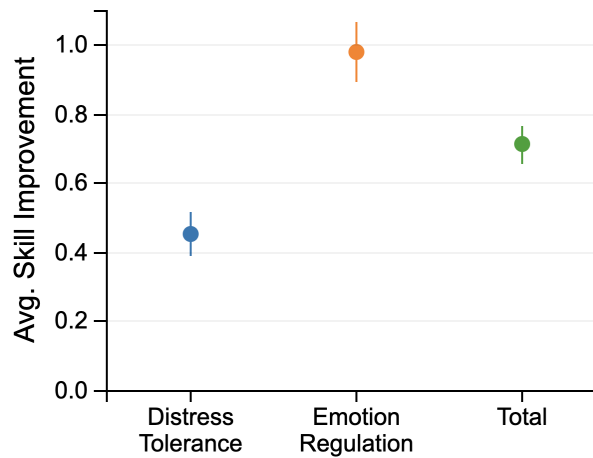


Figure 3.5: Average skill improvement (on the 5-point scales) using Pocket Skills, for all skill uses (purple), *Emotion Regulation* skills (blue), and *Distress Tolerance* skills (red), with standard error bars. People improved more using *Emotion Regulation* skills than *Distress Tolerance* skills.

Results

For skills with pre- and post-ratings, we found a significant effect of *module* on skill improvement ($t(233.16) = 4.956, p < 0.001$). *Emotion Regulation* skills correlated with more than half a point more improvement on our 5-point scale ($\bar{X} = -0.98$) than *Distress Tolerance* skills ($\bar{X} = -0.45$). Figure 3.5 illustrates the overall average skill improvements across different modules. We also found a significant effect of the specific *subskills* ($F(9, 248) = 3.901, p < 0.001$). Our Tukey HSD test accounting for unplanned comparisons revealed *Emotion Regulation* skills, particularly *Problem Solve*, generally correlated with significantly greater improvement than *Distress Tolerance* skills of *Self-Soothe*. Specifically, the *Emotion Regulation* skill of *Problem Solve* resulted in more improvement than the *Distress Tolerance* skills of *Self-Soothe via Hearing* (by 0.64 points; $t = -3.252, p < 0.05$), *Self-Soothe via Touch* (by 0.90 points; $t = -4.388, p < 0.01$), and *Self-Soothe via Vision* (by 0.63 points; $t = -3.361, p < 0.05$). The *Emotion Regulation* skill *Check the Facts* also resulted in more improvement than the *Distress Tolerance* skill *Self-Soothe via Touch* (by 0.79 points; $t = -4.108, p < 0.01$).

For skills with only post-ratings, we found no significant difference between *modules*. However, we did find a significant effect of the *subskills* ($F(25, 1199) = 4.461, p < 0.001$). Our Tukey HSD test accounting for unplanned comparisons revealed 14 significant pairwise differences (see Table 3.5).

Finding	Estimate	p Value
DT Distract via Pushing Away <i>vs</i> DT Distract via Comparisons	-1.20	0.0227
DT Distract via Sensations <i>vs</i> DT Distract via Comparisons	-1.43	0.0171
M Participate via Saying Words <i>vs</i> DT Distract via Comparisons	-1.04	0.0104
M Non-Judgementally Observe Judgments <i>vs</i> DT Distract via Pushing Away	0.992	0.0307
M Non-Judgementally Rephrase Judgmental Statements <i>vs</i> DT Distract via Pushing Away	1.25	< 0.0001
M Non-Judgementally Observe Judgments <i>vs</i> DT Distract via Sensations	1.23	0.0302
M Non-Judgementally Rephrase Judgmental Statements <i>vs</i> Distract via Sensations	1.49	0.000351
M Non-Judgementally Rephrase Judgmental Statements <i>vs</i> Distract via Thoughts	0.901	0.0472
M Participate via Saying Words <i>vs</i> M Non-Judgementally Observe Judgments	-0.836	0.00468
M Focus On One Thing at a Time <i>vs</i> M Non-Judgementally Rephrase Judgmental Statements	-0.710	0.00307
M Describe Visuals <i>vs</i> M Non-Judgementally Rephrase Judgmental Statements	-0.749	0.00234
M Observe Breathing <i>vs</i> M Non-Judgementally Rephrase Judgmental Statements	-0.819	< 0.0001
M Observe Sounds <i>vs</i> M Non-Judgementally Rephrase Judgmental Statements	-0.828	< 0.0001
M Participate via Saying Words <i>vs</i> M Non-Judgementally Rephrase Judgmental Statements	-1.10	< 0.0001

Table 3.5: Pairwise differences in post-ratings between Pocket Skills skills. Positive deltas indicate the first skill results in *worse* post-ratings than the second skill. M=*Mindfulness*, ER=*Emotion Regulation*, DT=*Distress Tolerance*.

Generally, *Mindfulness* skills of *Observing*, *Describing*, and *Participating* correlated with better post-ratings than *Mindfulness* skills of *Non-Judgementality*. *Distress Tolerance* skills of *Distracting* also correlated with better post-ratings than *Mindfulness* skills of *Non-Judgementality*.

RQ3: Skill Effectiveness Across Subgroups

After we had determined which skills led to the most improvement in general, we examined whether different subgroups of people reported different skill effectiveness.

Methods

We first separated participants into the groups described in Section 3.2.1 and Table 3.4. We ran t-tests on categories with two groups (i.e., whether or not the participant completing the skill: took medications to manage their mental health; had anxiety disorder at the intake survey based on their OASIS results; had a mood, anxiety, eating, personality, neurodevelopmental, or no disorder diagnosed) to investigate differences between subgroups for each type of skill effectiveness (i.e., skill improvement, post-ratings). For each type of skill effectiveness, we then used the Benjamini-Hochberg procedure [20] to correct for multiple comparisons. Our results report these adjusted p-values.

For groups with three or more subgroups (i.e., initial PHQ-9 class; gender; family, age, and education buckets), we ran one-way ANOVA tests to investigate subgroup differences in each type of skill effectiveness measure. When we found significant results, we again investigated pairwise differences using Tukey's HSD procedure to account for unplanned comparisons.

Results

Figure 3.6 illustrates differences in average skill improvements and post-ratings across subgroups.

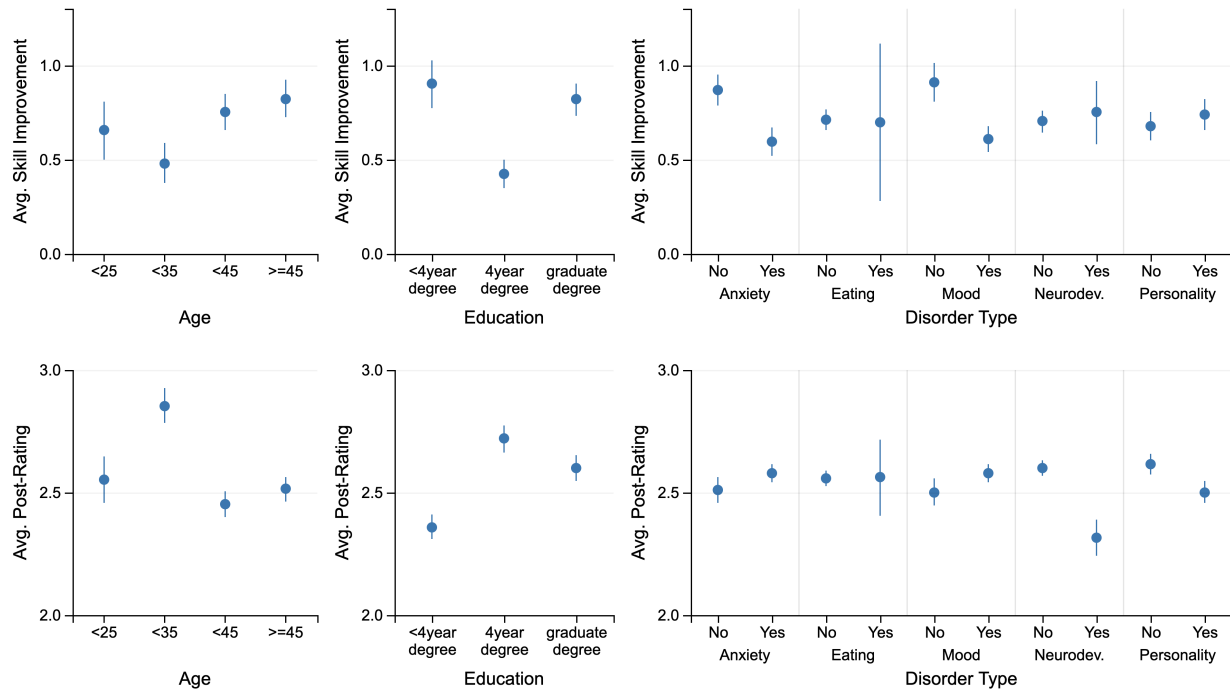


Figure 3.6: Average skill improvement (top) and post-ratings (bottom) using Pocket Skills across age subgroups (left), education subgroups (middle), and disorder types (right), with standard error bars. Higher improvement indicates more improvement (i.e., is better), while lower post-ratings indicate more positive ratings (i.e., is better). Some subgroups varied more, on average, than others.

First, for skill improvement, we found that *Education* had a significant effect ($F(2, 255)=7.068$, $p<0.01$), with having <4 year degree correlating with 0.48 points more improvement than having a 4-year degree ($t=3.404$, $p<0.01$) and having a graduate degree correlating with 0.4 points more improvement than having a 4-year degree ($t=-3.081$, $p<0.01$). We found no significant differences in skill improvements between subgroups in any of the other categories.

For skill post-ratings, we found a significant effect of *personality disorder* ($t(1220.7)=2.533$, $p<0.05$): people who have a *personality disorder* rated skills slightly more positively ($\bar{X}=2.40$) than those who do not ($\bar{X}=2.56$). We similarly found a significant effect of *neurodevelopmental disorder* ($t(252.68)=2.580$, $p<0.05$), with people who have a neurodevelopmental disorder rating skills slightly more positively ($\bar{X}=2.28$) than those who do not ($\bar{X}=2.52$). We also found a significant difference in post-rating given *medication use* ($t(214.39)=3.1043$, $p<0.05$), with people who took medication rating skills slightly more positively ($\bar{X}=2.43$) than people who did not take medication

($\bar{X}$ =2.75). *Education* also significantly correlated with post-rating ($F(2, 1222)=6.27, p<0.01$), where a <4 year degree correlated with slightly better post-ratings than having a 4-year degree (by 0.28 points; $t=3.384, p<0.01$) or a graduate degree (by 0.20 points; $t=2.561, p<0.05$). *Number of family members close* also had a significant effect ($F(3, 1221)=3.36, p<0.05$), with having no family close correlating with less positive ratings than having 1-2 (by 0.38 points; $t=-2.90, p<0.05$) or 3-4 (by 0.42 points; $t=-3.10, p<0.05$). Finally, we found a significant effect of *age* ($F(3, 1221)=4.19, p<0.01$); being 25-35 years old correlated with more negative ratings than either being between 35-45 years old (by 0.34 points; $t=-3.502, p<0.01$) or being greater than 45 years old (by 0.25 points; $t=-2.651, p<0.05$). We found no significant differences in post-ratings between any other subgroups.

RQ4: Skill vs. Validated Scale Improvement

We next investigated whether participant and skill characteristics correlated with improvements in the clinically validated scales that participants completed for anxiety (OASIS [181]), depression (PHQ-9 [127]), and skill use (DBT Ways of Coping Checklist [178]) throughout the study.

Methods

We first calculated score differences between the intake and exit surveys for the PHQ-9, OASIS, and DBT Ways of Coping Checklist scales. We then extracted additional participant characteristics, including: their favorite and least favorite module; how much they felt Pocket Skills helped their goals and skill use; and whether they felt they practiced more skills with Pocket Skills than they would have practiced without it. We also examined participant skill usage patterns, including the total skills they practiced in each module; whether they repeated any skills; average, best, and worse skill improvement and post-rating; and the proportion of unique skills practiced (i.e., an individual's number of unique skills over the number of total skills practiced).

We performed mixed model analyses of variance for each scale, treating the characteristics described above, medication use, and demographics (i.e., age, education, family member buckets) as fixed effects and the *specific disorder types* (i.e., whether they reported any disorders within each category) as random effects to account for heterogeneity within the overarching disorder types. We again investigated pairwise differences using Tukey's HSD procedure to adjust for repeated testing.

Results

For depression improvement, we found significant main effects of *age bucket* ($F(3, 9.3811)=4.5852$, $p<0.05$); *family bucket* ($F(3, 9.6203)=4.3922$, $p<0.05$); *education bucket* ($F(2, 9.5581)=5.0718$, $p<0.05$); *favorite module* ($F(3, 9.1457)=4.5373$, $p<0.05$); *best skill improvement* ($F(1, 9.3973)=5.3843$, $p<0.05$); and *best skill post-rating* ($F(1, 9.6632)=8.5388$, $p<0.05$). Participants with larger best skill improvements and more positive best post-skill ratings tended to improve more on their PHQ-9 score. Pairwise analyses revealed that being 35 or younger generally correlated with more improvement than being older than 35. Specifically, being under 25 ($z=3.414$, $p<0.01$) and 25-35 ($z=3.167$, $p<0.01$) years old correlates with more improvement than 35-45 years old. Being under 25 ($z=2.698$, $p<0.05$) and 25-35 ($z=2.869$, $p<0.05$) years old also correlates with more improvement than being greater than 45 years old. Our examination of *family bucket* revealed that having zero family members living close correlates with greater improvement than having 1 or 2 ($z=3.618$, $p<0.01$), 3 or 4 ($z=3.021$, $p<0.05$), or more than 4 ($z=-2.781$, $p<0.05$) close family members. Having a <4 year degree correlated with more improvement than having a graduate degree ($z=-3.161$, $p<0.01$). Finally, preferring the Addiction module correlated in more improvement than preferring the Distress Tolerance module ($z=3.638$, $p<0.01$), the Emotion Regulation module ($z=3.391$, $p<0.01$), or the Mindfulness module ($z=3.564$, $p<0.01$).

For anxiety improvement, we found a significant main effect of *age bucket*, with participants who were older than 35 again improving less than those who were 25-35 ($z=2.939$, $p<0.05$) or older than 45 ($z=3.278$, $p<0.01$). The skill use model yielded no significant results.

Consistent with our other analyses, these models use the buckets defined in Section 3.2.1 for the variables of *age*, *number of family members close*, and *education*. If these variables are instead treated as continuous, the resulting models yield no significant effects. This discrepancy may be due to a nonlinear relationship to those variables or due to our relatively limited dataset.

RQ5: Predictability of Skill Effectiveness

Finally, we examined the feasibility of predicting skill effectiveness for particular participants and skills. Given participant and skill use characteristics and each participant’s historical skill usage and rating data, we built four machine learning classifiers to predict whether a specific skill would result in a skill improvement for that participant (i.e., lower emotional intensity or distress after skill use).

Methods

To predict whether a skill would lead to a skill-level improvement for a participant, we used the subset of data (N=258) that included pre- and post-ratings to train binary classification models with positive prediction corresponding to skill improvement. We used participant and skill use characteristics as features (see Table 3.6), including contextual information (e.g., time of day, day of week, emotional intensity or distress prior to skill use); usage patterns (e.g., consecutive use of a skill within 15 minutes); historical skill improvement (e.g., cumulative average of skill improvements, skill improvement from the last skill use); and preferences (e.g., favorite module). Categorical features were one-hot encoded to transform each category into a binary representation necessary for model construction.

We trained binary classifiers using four different learning algorithms (i.e., decision tree, gradient boosting, random forest, logistic regression). We trained and evaluated each model using leave-one-participant-out cross validation. We used the average model accuracy and AUC on the hold-out participants as our metric to prevent overfitting and tune hyper-parameters of the learning algorithm (e.g., depth of the tree, max feature count, number of estimators, minimum samples for splitting nodes and leaf nodes, regularization strength, penalty). Because results from both test accuracy and test AUC were qualitatively similar, we discuss only test accuracy.

Feature	Description	Type
Skill ID	Unique skill identifier	Categorical
Day of week	Day of week for skill use	Categorical
Time of day	Time of day for skill use grouped by morning (6-12), afternoon (12-18), evening (18-24), and night (0-6).	Categorical
Consecutive use of any skill	Boolean indicating whether another skill is used within 15 minutes prior to the current skill use	Categorical
Consecutive use of the same skill	Boolean indicating whether the same skill is used within 15 minutes prior to the current skill use	Categorical
Cumulative average skill improvement	Cumulative average of skill improvements since the study intake	Numerical
Last improvement	Skill improvement of the last used skill (0 for the first skill use)	Numerical
Pre-rating	Pre-rating of emotional intensity or distress	Numerical

Table 3.6: Skill use features used for training predictive models of Pocket Skills skill improvement.

Finally, we performed a feature ablation study by training independent classifiers using all but one feature to understand the informativeness of each feature. We focused on the features or variables that we found to have significant effect on skill improvement in our statistical analyses (see Table 3.7).

Results

Given all features, the decision tree classifier yielded the best test accuracy (72.1%). All classifiers performed better than the base rate of the majority class (57.4%, or 148 out of 258 skill uses, were reported to have improved ratings). Our feature ablation study revealed that *skill ID* was the most impactful feature for the decision tree, gradient boosting, and logistic regression classifiers, leading to drops of 17.1%, 13.4%, and 4.9% in accuracy when removed. For the random forest classifier, *mood disorder* was the most impactful feature, leading to a drop of 1.1% in accuracy when removed.

We further found that different classifiers perform better or worse on individual skills. For example, Figure 3.7 shows that the accuracy ranges from 20% for a random forest classifier to 80% for gradient boosting and decision tree classifiers for the *Emotion Regulation: Brief Check*

Feature set	DT	LR	GB	RF
All	0.721 / 0.628	0.643 / 0.665	0.705 / 0.699	0.659 / 0.632
All but skill	0.62 / 0.457	0.636 / 0.616	0.64 / 0.565	0.663 / 0.648
All but pre-rating	0.721 / 0.628	0.663 / 0.69	0.698 / 0.708	0.667 / 0.657
All but mood disorder	0.721 / 0.628	0.655 / 0.685	0.709 / 0.703	0.64 / 0.621
All but anxiety disorder	0.721 / 0.628	0.64 / 0.666	0.698 / 0.707	0.671 / 0.609
All but education	0.721 / 0.628	0.643 / 0.671	0.671 / 0.703	0.698 / 0.677
All but favorite module	0.721 / 0.628	0.651 / 0.664	0.713 / 0.703	0.655 / 0.672
All but least favorite module	0.721 / 0.628	0.647 / 0.659	0.694 / 0.705	0.647 / 0.662

Table 3.7: Predictive model performances (denoted by “test accuracy / test AUC”) across four different learning algorithms—decision tree (DT), logistic regression (LR), gradient boosting (GB), and random forest (RF)—and various feature sets.

the Facts skill. Similarly, accuracy varied dramatically within specific classifiers for different skills. For example, for the logistic regression classifier, the accuracy for *Distress Tolerance* skills (54.2%) is consistently lower than the accuracy for *Emotion Regulation* skills (68.8%).

3.2.3 Discussion

In this follow-up study, we analyzed data from a month-long field study of Pocket Skills, a conversational mobile web app designed to provide holistic support for dialectical behavioral therapy (DBT). Based on our findings, I present design implications for future translations of evidence-based therapies into mobile apps, including the importance of designing for environmental context, emotional context, and personal context. I also discuss the promise for future work in personalized and context-aware recommendations for skill-based mental health interventions. These design implications and opportunities can inform future designs for the increasingly-prevalent mobile app interventions aimed to help people better manage their mental health.

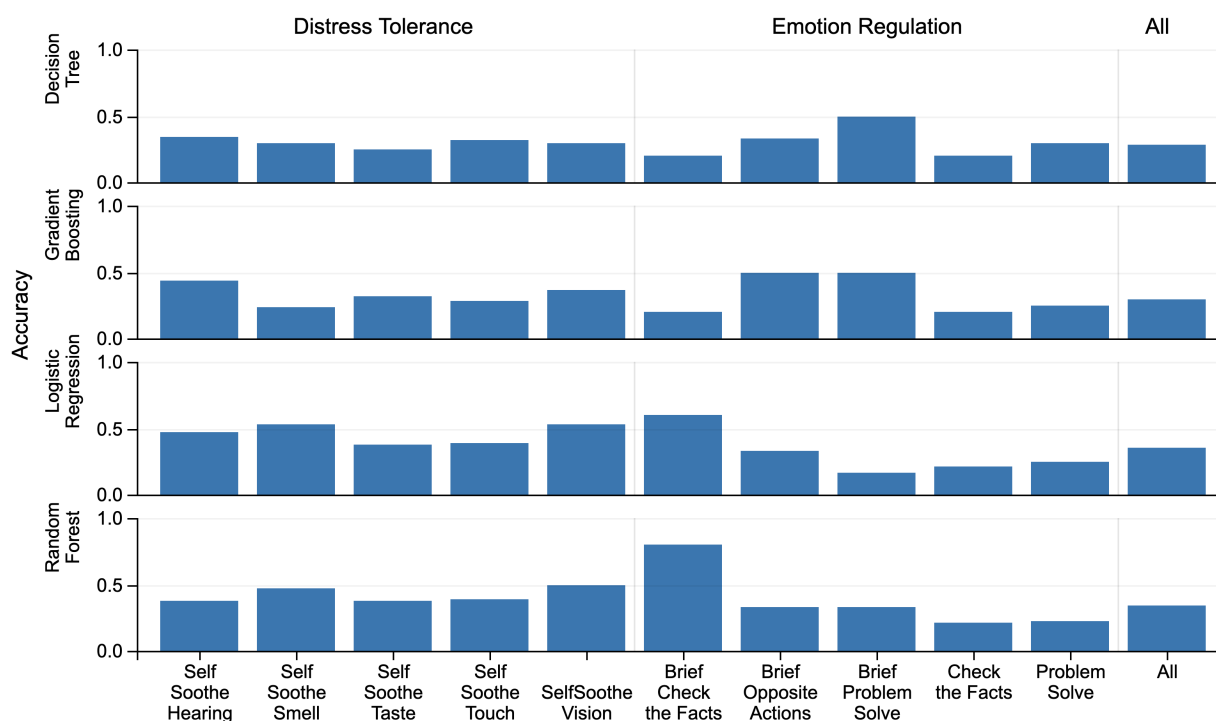


Figure 3.7: Accuracy of model predictions across individual DBT skills, with accuracy of a single model across all skills displayed on the right. The varying accuracy of different models for different skills, together with an ablation study highlighting the importance of *skill ID*, reveals a need for skill-specific models that can account for different context relevant to each skill.

Data-Driven Design Implications

Our findings highlight a range of contextual information that can influence skill effectiveness for an individual using a mobile app to identify and implement a positive coping skill.

Designing for Environmental Context

Overall, we found that participants consistently reported specific skills to be more effective than others. In Section 3.2.2, we described how participants generally improved more using *Emotion Regulation* skills and less using *Distress Tolerance: Self-Soothe* skills. This difference may be due to the immediate feasibility of the activities suggested in the respective skills. *Emotion Regulation* skills generally walk people through considering and solving the problem they have, and can therefore be

completed in most contexts. In contrast, the included *Distress Tolerance: Self-Soothe* skills suggest specific methods to distract people from distress that may be difficult to complete at certain times (e.g., eating something spicy, taking a cold shower). These activities are helpful when one has the ability to complete them, and are therefore useful in certain environmental contexts. However, they often cannot be immediately performed (e.g., when the person using the app is not at home). The post-ratings we collected therefore may not always reflect how people felt after completing the activity, as they may have been unable to do so when the post-ratings were solicited. Skills that suggest activities that may be infeasible in certain contexts should therefore be designed differently than those that can be done anywhere at any time, perhaps by asking the person if they can immediately complete the skill and enabling them to explicitly schedule the skill for a future time if not.

Designing for Emotional Context

In addition to enabling people to indicate later uses of a skill, apps designed to support skill-based interventions should recognize emotional circumstances in which it may be inappropriate to suggest skills that cannot be immediately completed. As I reported in Section 3.2.2, the majority of people using skills with pre- and post-ratings reported high distress or emotional intensity in the pre-rating of the skill: they were using the app *in the moment*, rather than practicing the skill so they could use it during *future* times of distress. Although all skills were directly adapted from the DBT Skills Training and Workbooks [219], Pocket Skills is a constantly-available resource and may therefore be used much differently than a traditional DBT skills worksheet. Instead of directly translating content, designers of tools to support skill-based interventions must consider how to best adapt skill-related materials to consider these different use cases. For example, a tool could differentiate between people practicing skills versus people who are currently in distress and attempting to use a skill *in the moment*. When a person indicates they are in distress, such a tool could focus on guiding them towards skills that are most likely to be feasible and useful in their current context. A tool could even *sense* emotional state (e.g., using sensing techniques such as those described in [108, 170]) and send push notifications during times of distress to directly suggest an appropriate skill. Such just-in-time interventions have been shown to support health behavior change (e.g., in promoting physical activity [89], in stress

management [101], in weight management [226], in smoking cessation [175]). Our results both support emerging interest in applying such techniques to positive coping skill use (e.g., [107]) and differentiate the additional need to support skill development when people are *not* distressed. For example, a tool could then focus on supporting discovery of new skills, so people can continue to expand their positive coping skill “toolbox” and identify the practices that work best for them.

Designing for Personal Context

In addition to the trends we found in overall skill use, we also found that individual characteristics were often correlated with different levels of effectiveness, both in terms of individual skills and overall improvement throughout the study. As I discussed, the type of disorder an individual had sometimes correlated with different levels of skill effectiveness, as did their education level, age, medication use, and the number of physically close family members. Many of those characteristics also correlated with differences in depressive symptom improvement throughout the study. Our investigations additionally revealed that people who had larger best improvements and post-ratings in individual skills tended to improve more in depressive symptoms, so helping people find skills that are effective for them could also help them improve their mental health overall.

Our preliminary results also indicate that individual preferences may influence effectiveness: as I discussed, participants who preferred the *Addiction Skills* module reported higher improvements in depression after the study than those who preferred other modules. However, I cannot confirm whether those skills lead to better skill-level improvements because the module did not have skill ratings.

Given these differences between subgroups of people, future designs should account for individual characteristics and preferences in intervention activities. Future studies should also investigate how to better understand, acknowledge, and counteract any disparities or detrimental effects that could result from an individual’s characteristics and preferences.

Opportunities for Intelligent Support

Researchers have investigated how technology could use predictive models to identify and support people with mental health conditions (see Section 2.2). Machine learning has also been used to match stress relief interventions to particular individuals and contexts (e.g., [187,210]). Mohr et al. recently examined a recommender system that identified skills an individual was particularly likely to use, finding that such recommendations resulted in improved depressive symptoms [165]. Preliminary evidence therefore suggests that incorporating such models into technology-delivered mental health interventions could better support people in engaging with psychotherapy.

People using Pocket Skills must currently discover which skills are most helpful to them on their own and must remember to use those skills during times of distress. However, our preliminary modeling results indicate promise for predicting whether a skill would yield improvement based on participant and skill use characteristics, indicating a potential to intelligently identify effective skills. Future studies with more participants, as well as skills designed to consistently elicit appropriate pre- and post-ratings, could generate data for more advanced modeling and prediction of skill effectiveness.

Predictive models would enable apps to generate skill recommendations, better supporting people in discovering and using effective skills. Our analyses suggested that some skills may be more effective than others, and our feature ablation study similarly revealed that the skill itself was consistently an important feature for predicting skill improvement. As I described, these differences in effectiveness may be due to the translation of the skill into app content rather than the skill itself. However, if certain skills do tend to be more effective than others, an app could suggest that people generally focus on those skills. Our results suggested that without such recommendations, people tend to go through the modules and skills in the order the app presents. Even with advanced predictive modeling techniques, people who are new to an app would lack necessary data for personalized recommendations. This common “cold-start problem” of recommender systems [213] can be alleviated through general recommendations based on someone’s demographics or other static features. Expert advice could

also inform recommendations based on specific conditions and characteristics (e.g., by working with a psychologist to develop general recommendations). Such recommendations would enable even new users to discover skills that are more likely to help them in a given moment.

In this study, we explored a one-size-fits-all predictive model, in which a single model is used to generate all predictions. We found that models trained on our current set of features can predict individual improvements more successfully (73.1%) compared to the base rate of the majority class (57.4%). In addition to demonstrating this potential for modeling approaches, we also revealed opportunities for future work to improve predictions. For example, we found that certain learning algorithms performed better at predicting skill improvements for different individual skills. Future approaches could include an ensemble of models to further improve skill recommendations, leveraging: 1) different learning algorithms, to account for any differences in model performance for different skill or participant characteristics, and 2) different models, to account for any differences in data needs for individual skills (e.g., a model for *Distress Tolerance: Self-Soothe* skill could leverage location data).

Our modeling approach was also limited by a lack of richer contextual information. Given more complete data, a predictive model could allow personalized recommendations based on the contexts described above (e.g., in-the-moment environment and emotions, preferences and goals, personal characteristics). For example, during times of distress, a recommender system could examine the individual's preferences, context, and historical app use to suggest skills that are most likely to result in improvements for that person. The same system could help people diversify their skill portfolio by recommending new skills that were helpful to similar individuals. Recommendations could also depend on the individual's environmental context, suggesting different skills when a person is at home, in transit, or on the bus. Such recommendations could further support people in identifying coping skills that work for them and implementing those skills in their lives when they need them.

3.2.4 Study Conclusion

Mobile mental health interventions are becoming increasingly ubiquitous, prompting a need for improved understanding to help designers more consistently base these interventions on evidence-based principles. We analyzed data from a month-long field study of Pocket Skills, a mobile web app

designed to provide holistic support for dialectical behavioral therapy (DBT) to help people develop positive coping skills. We identified several factors that contribute to skill effectiveness, including the skill itself and participant characteristics and preferences. We also developed machine learning models to predict skill-based improvements. Based on our findings, we presented design implications for translating evidence-based psychotherapies into application content, including the need to consider different environmental, emotional, and personal contexts. Finally, we discussed opportunities to use machine learning techniques for better mental health support by producing personalized and context-aware skill recommendations based on an individual's personal characteristics, preferences, past application use, and in-the-moment emotions and environmental factors.

3.3 A GOAL-DIRECTED PERSPECTIVE ON SUPPORTING MENTAL HEALTH

Pocket Skills was designed to support common goals that people often have for their mental health management, by helping them learn and implement DBT skills. However, because of the wide variety of *personal* goals people often bring to DBT, people using Pocket Skills would likely benefit from a more goal-directed perspective. Individuals who pursue dialectical behavioral therapy (DBT) have a wide range of personal experiences. They may have different mental health disorders, and may experience different symptoms even within the same disorder (e.g., negative moods, behaviors, or coping mechanisms) [140]. Pocket Skills does not currently provide explicit support for these different goals people may bring to their therapy. The app does ask people to categorize their goals for their own reflection, but the modules and skills themselves were taken directly from traditional DBT and not tailored or suggested based on the individual using the app. Just providing that level of support did sometimes help participants achieve personalized goals; as I discussed in Section 3.1.3, participants reported Pocket Skills helped them learn and implement skills to achieve the goals they had, such as more positive relationships, reduced distress, and behavioral changes. However, as participants themselves often conveyed in the exit survey, explicitly supporting those goals could enable further improvement. With such explicit support for an individual's goals, Pocket Skills could provide more personalized help, including tailored suggestions for skills during

times of distress (congruent with the personalized and contextually-appropriate recommendations proposed in Section 3.2.3); customized diary cards to allow tracking towards their personal goals; and visualizations of progress to help them see the results of engaging with their therapy.

3.4 SUMMARY

We designed, developed, and evaluated Pocket Skills, a conversational mobile web app designed to provide holistic support of DBT, from initial goal setting to skill learning and practice. After a 4-week field study with 73 participants, we found significant improvements in anxiety, depression, and dysfunctional coping mechanisms, as well as an increase in DBT skills use. We also contributed a model describing how participants felt Pocket Skills supported DBT, finding that Pocket Skills helped people *engage* with DBT and *learn* specific skills to apply given their personal needs and experiences. They then were able to *implement* those skills in their everyday lives, see *concrete results*, and increase their *self-efficacy* to improve their mental health. In our subsequent quantitative analysis of the survey and usage data from the feasibility study, we developed data-driven design implications for translating evidence-based psychotherapies into technology-delivered interventions and identified opportunities to provide personalized and contextually-appropriate skill recommendations.

My work in the mental health domain therefore supported people in *expressing* and *pursuing* common DBT goals, which resulted in positive outcomes both from the perspective of common goals (i.e., the improved depression, anxiety, and skill use seen in the scale results) and from the perspective of the personalized goals that individuals pursuing DBT may have (i.e., the improved relationships, moods, and behaviors that participants described seeing in their lives). I also investigated how tailored skill recommendations could further contribute to the *personalized management* of their mental health conditions. My work further revealed an opportunity to better support the wide range of goals people pursuing DBT may have by allowing them to explicitly communicate and pursue their personal goals.

Chapter 4

SUPPORTING IBS MANAGEMENT THROUGH PERSONALIZED TRIGGER IDENTIFICATION

People with irritable bowel syndrome (IBS) and their health providers often share a goal of trigger identification. The American Gastroenterological Association advises providers to have patients keep food and symptom journals to attempt to identify personal triggers [7], and when people with IBS were asked what they want to know about IBS, the most common response was a desire to learn what foods they should avoid [88]. However, as no validated method exists to interpret food and symptom journals, the goal of identifying triggers remains largely unsupported.

I therefore aimed to support people with IBS and health providers in their goal of identifying personalized dietary triggers. Working with a multidisciplinary team of health and technology researchers, I first developed analyses to identify correlations between the food nutrients an individual consumes and their subsequent IBS symptoms using data from previously-collected food and symptom journals [261,263]. I discuss this data processing and analysis process in Section 4.1.

After identifying correlations between nutrients and IBS symptoms, we wanted to communicate those correlations back to the individual with IBS and their health provider. We therefore designed, developed, and evaluated a system to help people explore relationships between the foods they eat, the nutrients those foods contain, and the symptoms that were correlated with those nutrients based on their food and symptom journals to support *hypothesis formation* of what foods in their diet may trigger their symptoms [215]. Our system included two visualizations designed to help people and their health providers identify and understand those relationships and communicate and collaborate with each other. I discuss our development and evaluation of this system in Section 4.2.

The analyses and visualizations I developed supported interpretation of traditional food and symptom journals. However, holistic IBS management requires additional support in collecting, interpreting, and acting on personal data. I therefore contributed to the design, development, and evaluation of a system to support *hypothesis testing* of potential IBS triggers via self-experimentation [109]. I also characterized nuanced and personalized goals surrounding IBS trigger identification that people and providers identified while using our tools, and designed analyses and representations to better address those goals [216]. Finally, I contributed to the design, development, and evaluation of a photo-based food and symptom journaling system, which aimed to lower burden and better support interpretation [43]. I discuss these additional studies in Section 4.3.

Finally, I again investigated supporting personalized IBS management before I developed and investigated *goal-directed self-tracking*. However, the concept of explicitly eliciting, representing, and supporting an individual’s goals is also highly relevant to IBS. I therefore reflect on how this goal-directed perspective could further support personalized IBS management in Section 4.4. I conclude by summarizing how my work in IBS supports my thesis statement in Section 4.5.

The work described in this chapter was done in collaboration with Chia-Fang Chung, Jane Hoffswell, Ravi Karkar, Matthew Kay, James Fogarty, Julie Kientz, Sean A. Munson, and Jasmine Zia.

4.1 IDENTIFYING CORRELATIONS BETWEEN NUTRIENTS AND SYMPTOMS

A main barrier to identifying IBS triggers is the difficulties associated with interpreting food and symptom journals. As part of feasibility and usability studies of paper [261] and app-based [263] journals, we developed a method to identify correlations between nutrients and subsequent symptoms.

4.1.1 Food and Symptom Data Collection

For both studies, data collection was designed to provide detailed data capture while limiting burden. Consistent with established dietary protocol [203, 238], participants journaled for 3 sets of 3-day “on” periods, separated by 3-day “off” periods (i.e., 9 days of food and symptom journals collected over a

15-day period). Researchers, trained by a registered dietitian, provided detailed instructions on how to record food. Participants were asked to record all food and drink with the corresponding time. At each meal entry, and again at bedtime, participants reported their peak symptoms since the prior entry.

4.1.2 Analysis of Food Data to Identify Constituent Nutrients

After we collected each participant's food and symptom journal, trained research dietitians analyzed the food data using the Nutrition Data System for Research (NDSR) [36], decomposing each food in a journal into its constituent nutrients. Wherever possible, foods were first broken into their component foods (e.g., "cheeseburger" would become "beef patty", "bun", "cheese", toppings, and condiments). As prescribed by the NDSR's "Data Entry Rules" [35], standardized assumptions were made for missing information or unfamiliar food. This decomposition allowed each journal entry to be mapped to 19 nutrients the gastroenterologist on the team (Jasmine Zia) identified as potentially correlated with IBS symptoms (e.g., sugars, caffeine, lactose).

4.1.3 Linear Regressions to Identify Correlations Between Nutrients and Symptoms

For each participant's journal, for each symptom they experienced, we performed regression analyses to identify correlations between food and symptoms. Informed by prior IBS patient reports on the timing of symptoms following foods they believe triggered those systems [69, 194], independent variables were defined as the sum of a nutrient in all foods reported within the 4-hour window preceding a symptom report. Dependent variables were defined as the corresponding symptom reports. Symptom entries without meal entries in this 4-hour window were excluded.

Our regressions assume a high degree of independence between predictors. However, nutrients can have high collinearity, due to both natural co-occurrences (e.g., foods with higher fat tend to have higher total calories) and personal habits (e.g., some people always drink their coffee with milk). Participant-specific feature selection was therefore performed prior to regression analysis.

For any pair of nutrients that had a high pairwise correlation in an individual's data (> 0.75), we removed the nutrient with the highest average correlation to all other nutrients (i.e., the most redundant nutrient). We then performed the regressions with the remaining nutrients.

4.2 SUPPORTING INTERPRETATION AND PATIENT-PROVIDER COLLABORATION

Once we had a method to identify potential triggers from food and symptom data, we wanted to be able to communicate that information back to patients and their health providers. However, we did not just want to give them a list of those nutrients, for two main reasons. For one, determining exactly what foods are *causing* an individual's IBS symptoms is impossible with the collected food and symptom data (e.g., due to concerns for correlations, confounds, and analysis limitations). In addition, we wanted to help people understand not just what nutrients were correlated with their IBS symptoms, but also what foods in their diet contained those nutrients, to help them gain a personalized understanding of when they might be exposed to those nutrients. We therefore wanted people to be able to review and reflect on their food and symptom data with the additional support of the information gained from the analyses. Working closely with Jane Hoffswell, I developed interactive, exploratory visualizations to support people with IBS and their health providers in the interpretation of their food and symptom data, both individually and collaboratively [215].

4.2.1 System Design

We designed two interactive, exploratory visualizations, created to help people with IBS and their providers interpret what nutrients correlate with an individual's IBS symptoms and what foods in the individual's diet contain those nutrients. Together, the visualizations explore tradeoffs in complexity and expressiveness. The *bubble and bar chart visualization* emphasizes the identification of high-level trends, sacrificing some detail in favor of easier interpretation. This visualization is more likely to be familiar, as its component visualizations are often covered in K-12 curricula [79]. In contrast, the *parallel coordinates visualization* emphasizes expressiveness, sacrificing some simplicity to expose greater detail. This visualization can display correlations among many

variables (e.g., between symptoms and nutrients, among multiple nutrients) [99], but is less likely to be familiar. Both designs visualize the same data, as analyzed via the methods described in Section 4.1. Both visualizations are presented in a web environment created using the D3 package [25]. We informally presented early iterations to two patients with IBS to confirm usability and inform iterative design prior to the patient and provider interviews I present in Section 4.2.2.

Design Goals and Tradeoffs

We identified four objectives motivating our designs for both visualizations. We considered these goals relative to the different expertise of patients and providers (e.g., patients have more familiarity with the data and the lived experience of IBS, providers have more clinical expertise).

Flexibility:

Journalized food and symptom data is highly particular to each individual patient. To personalize support for trend identification, a visualization must be able to present patient-generated data for vastly different diets and corresponding nutrient ranges, as well as different symptom experiences.

Simplicity:

To facilitate patient-provider collaboration within a clinic appointment, visualizations should allow both people with IBS and health providers to easily view and understand the data, illuminating important trends between the foods, nutrients, and symptoms. Early design iterations attempted to simultaneously visualize all nutrients and symptoms, but doing so violated this goal of simplicity. Our final designs instead present separate visualizations for each of the patient's symptom, allowing them to investigate a single symptom at a time and focus on whichever symptoms are most important to them. A common primary goal is to identify nutrients that *improve* or *worsen* symptoms, so visualizations also need to support simple identification of such trends in an individual's data.

Exploration:

Paired with simplicity is the need to support thorough exploration of the data. For example, the person with IBS or their provider may have a pre-existing belief that a particular nutrient affects the person's symptoms. In addition, given potential limitations in the data and the correlation analyses (e.g., the possibility of overfitting, potential confounds), we wanted to avoid implying greater confidence than is warranted by framing the analyses as *scaffolding* exploration, not replacing it. All nutrients and symptoms should therefore be available in the visualization, even if analyses did not identify them as significantly correlated. We also defined nutrients as *significant* to a symptom when they had a p value of $p < 0.1$, considering other nutrients *non-significant*. This threshold encourages consideration of more potential triggers, rather than only extremely confident results. Our visualizations also encourage exploration of *correlated nutrients* (i.e., those removed from analysis due to their high correlation with another nutrient). When nutrients are highly correlated, more data is needed to determine which is most relevant to a symptom, so encouraging exploration allows patients and providers to collaboratively apply their understanding and expertise.

Actionability:

Finally, visualizations should foreground information that helps people with IBS and their health providers collaboratively develop treatment plans based on that individual's data. Because strong correlations are most likely to be actionable, we default to visualizing only the *significant* nutrients for each symptom. Developing actionable plans also requires patients and providers to determine which foods in the patient's diet actually contain the identified nutrients. Our visualizations therefore emphasize relationships between the visualized nutrients and the foods the patient reported in their food and symptom journal (i.e., rather than providing generic examples of foods high in a nutrient).

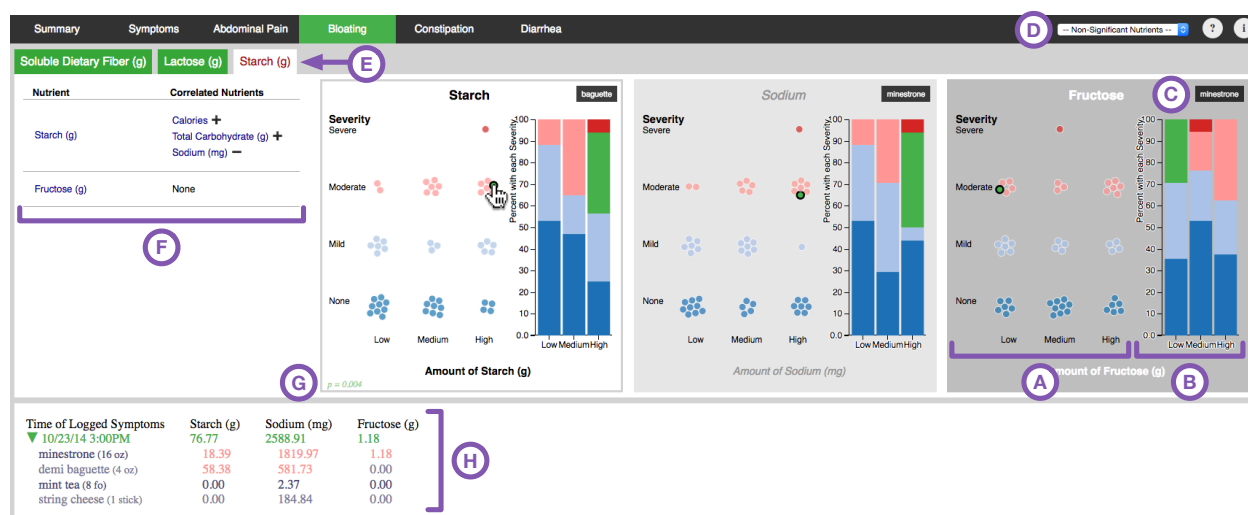


Figure 4.1: The bubble and bar chart visualizations consist of (A) a bubble chart and (B) a bar chart. (C) Hovering over a point shows the food in that meal highest in each nutrient. (D) Non-significant nutrients (e.g., fructose) can be added from a dropdown. (E) Each significant nutrient for the selected symptom is shown in a separate tab. (F) Nutrients correlated to the visualized nutrients (e.g., sodium) are shown in a table, from which they can be added to the visualization. (G) The p value appears when interacting with the plot. (H) Selecting points displays the food and nutrient information.

Data Visualizations

The *bubble and bar chart visualization* (Figure 4.1) displays the relationship between a selected symptom and a selected nutrient from a patient's data, enabling easy exploration of high-level trends. Each meal from the patient's journal is displayed as a point in a faceted bubble chart (Figure 4.1A.). Meals are bucketed by symptom severity (*None*, *Mild*, *Moderate*, *Severe*) and the amount of the nutrient in the 4-hour window preceding that symptom (*Low*, *Medium*, *High*, corresponding to thirds in the distribution of the nutrient in that individual patient's dietary data). A stacked bar chart aggregates the information in the bubble chart to further facilitate identification of trends (Figure 4.1B). The visualization incorporates multiple bubble and bar chart units to additionally enable comparison across different nutrients (e.g., correlated nutrients can be added from Figure 4.1F's table, non-significant nutrients can be added from Figure 4.1D's dropdown).

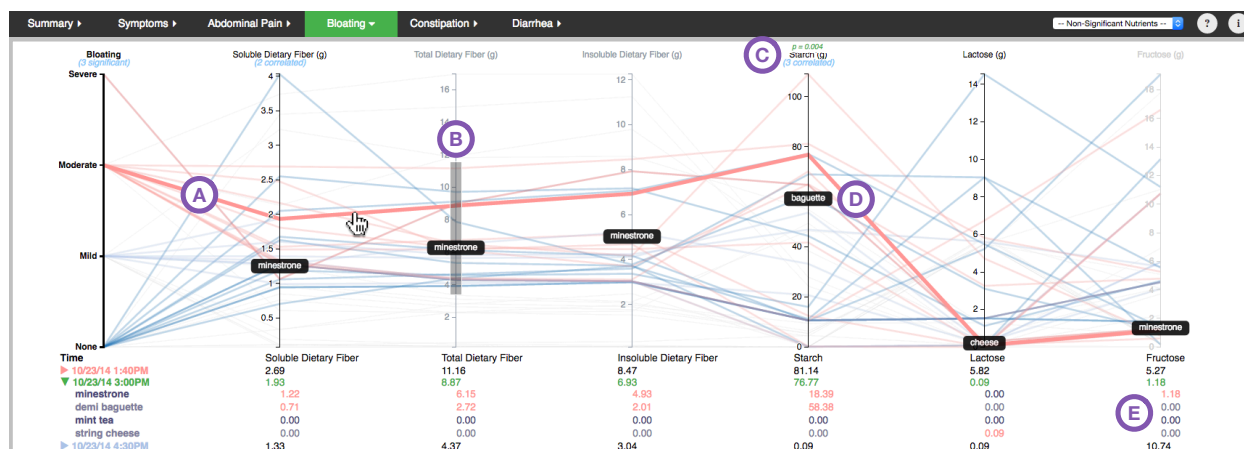


Figure 4.2: The parallel coordinates visualization. (A) A line corresponds to a single meal in the patient’s journal. **(B)** The user can specify a filter on a variable to only view meals within that range. **(C)** The axis title shows additional information above (the p value on hover) and below (the number of correlated nutrients). Clicking the number of correlated nutrients expands the visualization to include those nutrients. **(D)** Hovering over a line fades the other lines out and shows annotations of the food in that meal highest in each nutrient, at the appropriate position on the axis for that food and nutrient. **(E)** The foods and nutrient values for the hovered line are expanded in the food table below the visualization. The other meals correspond to ones within the current filter.

The *parallel coordinates visualization* (Figure 4.2) emphasizes relationships *between* nutrients and includes more detailed information (e.g., explicitly showing nutrient ranges in the patient’s journal). The visualization defaults to showing a symptom and each of its *significant* nutrients as axes in the parallel coordinates plot. Each line through the axes corresponds to a single datapoint, showing the symptom severity and the amount of each nutrient consumed in the preceding 4-hour window (Figure 4.2A). Axes can be rearranged by dragging an axis title to facilitate comparison of nutrients. Dragging vertically along an axis creates a filter that grays all excluded lines (Figure 4.2B). Correlated nutrients can be expanded from the label below the nutrient name (Figure 4.2C), and non-significant nutrients can be added from the menubar dropdown.

For both visualizations, hovering over a meal shows which food in that meal contained the highest amount of the given nutrient (Figure 4.1C, Figure 4.2D), using food names that were manually simplified to provide a more succinct label. For both visualizations, hovering over a datapoint shows the food that contained the highest amount of the nutrient (Figure 4.1C, Figure 4.2D).

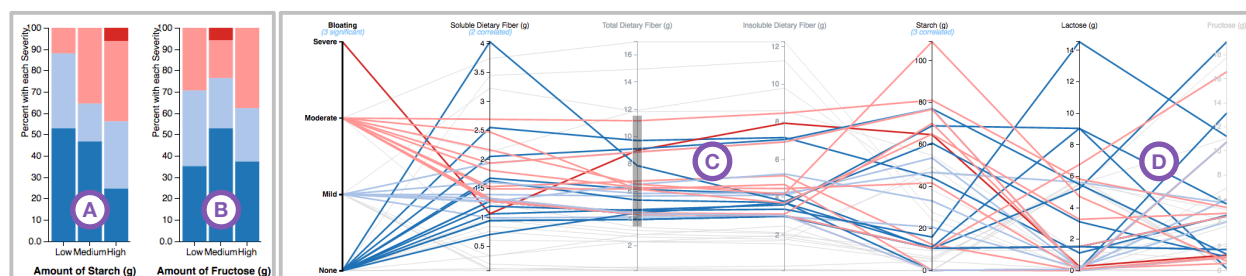


Figure 4.3: Trend identification with the visualizations. (A) A nutrient with worsening symptoms in the bar chart: the size of the red bars increases as the nutrient amount increases. (B) A non-significant nutrient does not show a strong trend: the bars are fairly equal for each symptom severity. (C) Correlated variables show mostly parallel lines between the axes in the parallel coordinates visualization. (D) Nutrients without a clear relationship are more jumbled.

Selecting a datapoint displays a food table, which includes food names and nutrient amounts for each visible nutrient (Figure 4.1H, Figure 4.2E). Foods with a nutrient amount in the top 75% of foods the patient consumed are highlighted in the color of the corresponding symptom severity.

Both representations aim to support interpretation by making correlations visually salient. In the *bubble and bar chart visualization*, correlation is indicated by the relative area of the stacked bars for each nutrient amount. For example, Figure 4.3A shows that as the amount of starch in this patient's diet increases, so does the severity of their symptoms (i.e., the portion of red in the bar grows from left to right). In contrast, Figure 4.3B shows that the amount of fructose does not seem to correlate with symptoms (i.e., no clear increasing or decreasing trend can be seen in the bars). For the *parallel coordinates visualization*, Figure 4.3C shows a strong positive correlation between soluble dietary fiber, total dietary fiber, and insoluble dietary fiber in an individual patient's data (i.e., straight horizontal lines between these nutrients). A strong negative correlation would be indicated by a tight grouping of line crossings (i.e., an "X" shape). Figure 4.3D shows no clear correlation between lactose and fructose (i.e., no clear trend in slopes between the nutrients).

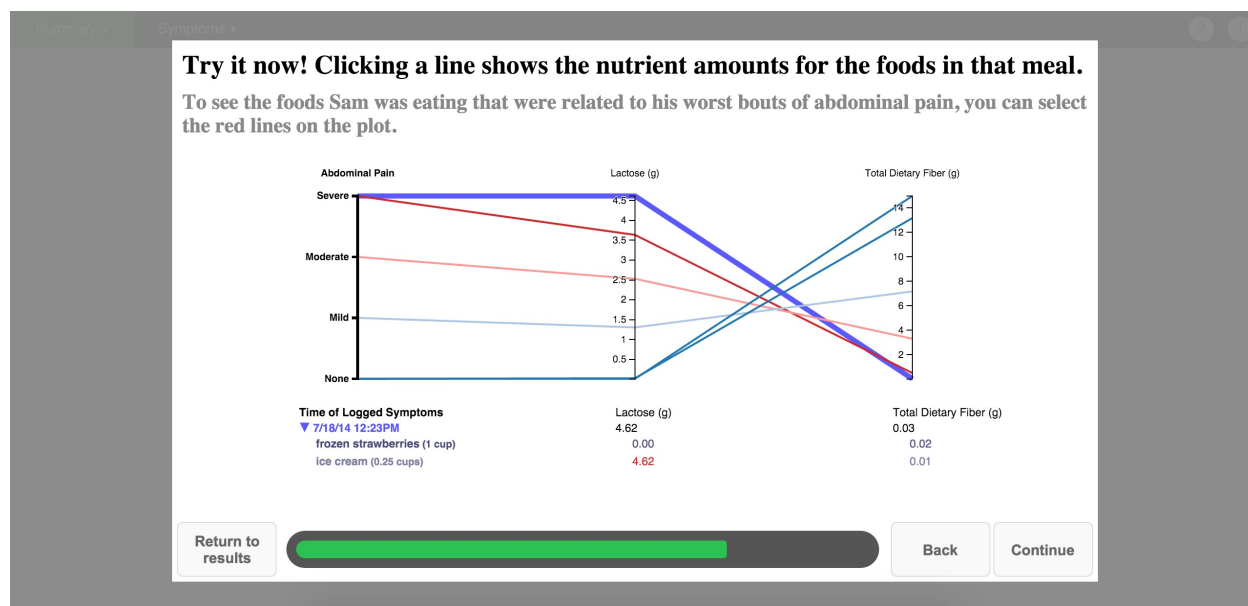


Figure 4.4: The tutorials aim to help people understand how to interpret the visualizations.

Interactive Experience

Upon opening either visualization, an interactive tutorial explains how to view, interpret, and interact with that visualization (Figure 4.4). Our goal for these tutorials was to familiarize patients and providers with the interface and its depiction of data, to increase their comfort interpreting the visualizations. After the tutorial, a summary page presents actionable entry points into the visualization. The summary page shows a table of the patient's symptoms, the significant nutrients to those symptoms, foods in the particular patient's diet that are high in those nutrients, and any nutrients that were correlated to the significant nutrients. Nutrients with positive effect sizes (i.e., for which symptoms tended to be more severe when more of the nutrient was consumed) are described as *worsening*, and those with negative effect sizes (i.e., for which symptoms tended to be less severe when more of the nutrient was consumed) are described as *improving*. Selecting a symptom, nutrient, or correlated nutrient navigates to that view in the visualization. Each symptom is also accessible from a separate tab in the top menubar, and defaults to showing nutrients determined

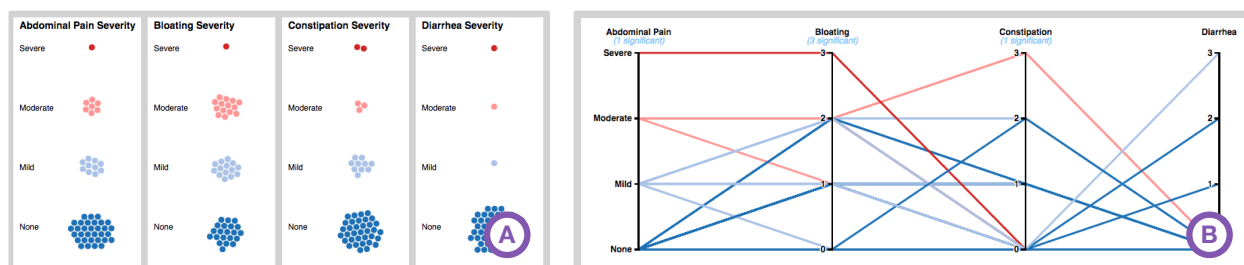


Figure 4.5: Both visualizations include a symptoms page, which summarizes reported symptoms. (A) The number and severity of symptoms is easy to see in the bubble and bar chart visualization. (B) Co-occurrences of symptom severity are easier to see in the parallel coordinates visualization.

to be significant for that symptom. If no nutrient was found to be significantly correlated, an initially empty tab is shown with a short explanation of why nothing is visible. A symptom summary tab also shows an overview of a patient’s reported symptoms without relating them to nutrients (Figure 4.5).

During our early informal iterations, we found that *patients* often found *p* values unfamiliar and somewhat intimidating. However, *providers* considered *p* values essential to a more complete understanding of the information. We therefore introduced two modes in the visualizations. In patient mode, *p* values are hidden from the summary table and correlated nutrients are described as “related”. In provider mode, *p* values are included in the summary table, the term “correlated” is used as appropriate, and the summary table emphasizes that example foods are taken from the individual patient’s diet (rather than generic examples of foods high in that nutrient). When interacting with the visualizations, *p* values are shown regardless of mode (Figure 4.1G, Figure 4.2C). The mode was set programmatically prior to each session, and cannot be changed in the interface.

4.2.2 Feasibility Study Methods

As we intended the visualizations we developed to be collaborative tools, we conducted semi-structured interviews with people with IBS and providers, both independently and together, to evaluate them.

<i>Number</i>	<i>Provider</i>			<i>Patient</i>			<i>Analysis Results</i>	
	Gender	Type	Years Experience	Gender	Age	Years Symptoms	Symptoms ^a	Nutrients ^b
1	F	MD	17	M	35	18	1	1
2	M	MD	5	F	43	3	1	2
3	F	NP	1	F	34	3	3	5
4	M	MD	20	M	25	2	0	0
5	M	MD	5	F	22	15	2	3
6	M	MD	6	F	34	10	1	2
7	F	Dietitian	16	F	30	9	2	4
8	F	MD	5	F	33	5	1	2
9	F	MD	2	F	46	33	3	7
10	F	MD	6	F	32	17	4	8

^aThe number of symptoms with at least one nutrient identified as significantly correlated in our analyses.

^bThe total number of significantly correlated nutrients identified in our analyses (across all symptoms).

Table 4.1: Demographic information for people with IBS and providers in the visualization study.

Recruitment and Procedures

We recruited 10 providers with experience working with patients with IBS (4 male) and 10 patients with IBS (2 male, age *mean* = 33, *min* = 22, *max* = 46) (Table 4.1). I note this patient gender distribution reflects that of IBS diagnoses [33]. Patients were compensated with a \$25 Amazon gift card, analysis and interpretation of their data, and a free consultation with a provider experienced in working with patients with IBS. Providers were compensated with a \$50 Amazon gift card.

Because visualizations of personal data are most meaningful to the person who provided that data, we recruited patients from participants in the prior paper food and symptom journaling study [261]. The study was completed about a year before our analyses and visualizations were developed, so the journals were not as recent as we would have liked. We nevertheless believed that personalizing the study with a participant’s own data would be the best representation of the system’s intended use, as patients are most likely to have detailed questions or hypotheses surrounding their own food and symptom data. At the time of the prior study, all patient participants met the Rome III criteria for

IBS [146] and had experienced IBS symptoms within the previous month. Patient participants had not met the providers prior to the study; recruiting an already-known provider for each patient was infeasible, and we wanted to be consistent in having all patient-provider pairings be new collaborations.

Interviews included two phases: individual and collaborative. In the individual phase, one researcher interviewed the patient while another interviewed the provider. This phase lasted 40 minutes, with approximately 20 minutes dedicated to each visualization and follow-up questions. The order visualizations were shown was held constant within a patient-provider pairing and counter-balanced across pairings. Participants completed the tutorial and then explored the visualization of the patient's data. They were encouraged to talk aloud during their data exploration and ask any questions they had. Researchers then asked about participant opinions of the visualizations, whether and when they would use them, and what conclusions they would draw from them. In the collaborative phase, the person with IBS and the provider were brought together to explore and interpret the visualizations collaboratively, simulating a clinic appointment. Participants were given their choice of which visualization(s) to use, and were able to switch between them at any time. We observed their collaborative interaction and interpretation, then interviewed them about collaborating with the visualizations and the contrast between individual versus collaborative interpretation.

We audio-recorded and transcribed all interviews. We then conducted an affinity diagram analysis to identify emergent themes. Two researchers (myself and Chia-Fang Chung) transformed interview transcripts into approximately 800 affinity notes and iteratively organized these notes into 75 categories. A third researcher (Jane Hoffswell) verified and discussed the emergent themes with the analysis team. We then identified several key themes regarding patient-provider collaboration and use or non-use of the tool to support collaboration. When quoting participants from these interviews, I will refer to people with IBS as *PX* and providers as *RX*. I specify identifiers due to our goal of monitoring and conveying the evolution of individual perspectives, as well as differing perspectives of paired patients and providers. However, due to the open-ended nature of our protocol, not every participant was explicitly prompted about every topic. Additional participants beyond those indicated may have agreed with any discussed perspectives if asked.

Limitations

We structured the interviews to be exploratory, with limited guidance, for multiple reasons. We wanted to observe how providers would use the tools to investigate their hypotheses and questions based on their years of experience treating patients with IBS. Similarly, we wanted patients to explore questions about their symptoms, foods, and nutrients according to their personal interests, rather than because we had artificially instructed them to complete a particular task. Our goal was to understand how patients with IBS and providers would individually and collaboratively use the visualizations to inform the management of the patient's IBS, based on the patient's experience and the provider's expertise. Directing participants to answer pre-formulated questions might have better compared how each visualization addressed those particular questions, but it would have been less representative of our real-world use-case. However, our method has limitations, as we cannot evaluate how the visualization techniques influenced the quality of the data interpretation. We were also unable to quantify differences between visualization strategies. We instead focused on the overall experience of using data visualizations for collaborative interpretation.

We also had a couple of inconsistencies during the interviews themselves. *P9* was uncomfortable with the prospect of a collaborative interview with a provider, expressing unease about “*bothering somebody to come in*”. We therefore only interviewed *P9* and *R9* separately (i.e., that interview had no collaborative phase). *P4* had no significant nutrients in their data. To help *R4* understand what the visualizations would look like with significant findings, we therefore showed *R4* visualizations for both *P4* and *P3*. For privacy reasons, patient participants were only shown their own data.

4.2.3 Results

Participants generally felt that the visualizations would help both people with IBS and providers better interpret food and symptom data. The interviews also surfaced requirements for successful patient-provider collaboration. These requirements span from design necessities for systems to support patient-provider collaboration to issues of trust and feeling trusted in the collaboration.

Trust in Patient-Provider Collaboration

A recurring theme was the concept of trust in patient-provider collaboration to interpret patient data. Providers needed to trust their patients could correctly interpret the visualizations. Both expressed concern about whether the other trusted them. Although visualizations helped build trust in some cases, they also introduced new considerations concerning trust during patient-provider collaboration.

Provider Trust in Patient Data Interpretation Abilities

An important question for many providers was whether they could trust patients to correctly interpret the data. Five providers (*R2, R6, R7, R8, R9*) worried the visualizations were too complicated and would confuse patients. This concern affected how they expected to use the visualizations and whether they wanted the patient to be able to access the visualizations before the clinic appointment.

During individual interviews, *R7* and *R8* explicitly mentioned they would likely focus on the summary page during a clinic visit, only showing patients the data visualizations if they had been able to identify a clear trend or an example they wanted to emphasize to the patient. Four providers (*R2, R6, R7, R8*) indicated being wary of giving a patient the ability to review the visualizations before the appointment, although three of the four (*R6, R7, R8*) indicated they still thought the patient should have access if they wanted. *R7* believed that the patient should have access, despite her worries, only because the patient should “*have the right*” to see their own data. Only *R2* said that he did not want a patient to be able to review the visualizations beforehand, predicting the patient would find the visualizations too complicated and would fixate on irrelevant things.

Patient Demonstration of Data Interpretation Abilities

The collaborative interviews ultimately assuaged many of the concerns providers initially had about patient data interpretation abilities. As I described, *R2* originally did not want patients to have independent access to the visualizations. However, he changed his mind after the collaborative interview. He appreciated that familiarity with the visualizations empowered the patient to set her own agenda, bringing things she considered important to the provider’s attention. The patient’s

familiarity with the data also allowed her to feel comfortable explaining different aspects of the data. The provider therefore felt more capable of making recommendations, because the patient could contextualize the data. By combining the provider's expert knowledge with the patient's personal knowledge, they were able to have a deeper conversation about possible problems and solutions, instead of the provider just *"coming in and being like, here's what I found"* (R2).

The other three initially skeptical providers (R6, R7, R8) also seemed more confident in the abilities of the patients after the collaborative interview. The patients they were paired with all said they would want to be able to review the visualizations before a clinic appointment, and the providers agreed the exposure would be helpful. R8 commented that having both collaborators be familiar with the data would help *"focus [their] energies during the visit"*. She explained:

I don't like to deliver news to patients that is surprising if I don't have to. If [the patient] saw this and [was] like, "Fiber and caffeine? That's not what I expected", [the patient] would be processing that ... as opposed to focusing on ... how I'm interpreting the data.

Need to be Trusted by the Collaborator

Patients and providers both commented on the need for their collaborators to trust what they said. Patients needed providers to understand their experiences, and providers needed patients to understand why they made their recommendations. Six providers (R2, R4, R6, R7, R9, R10) and five patients (P1, P5, P6, P7, P8) felt that being able to go through an objective view of the patient's data, rather than talking abstractly about the patient's experiences, helped them feel more capable of successfully communicating their views. P1 and P8 felt the tools helped them convey their symptoms in a way that was clear and easy to understand. P1 remarked that such information is *"kind of hard to put into words, but [the visualization] kind of quantifies it"*. Four providers (R2, R4, R6, R9) mentioned the importance of showing their patients the data to explain their recommendations. R4 explained:

It really helps them understand how what they eat affects their symptoms, which is the most important thing. If they don't understand that then they can't change it, right? I think it's a really objective way to show them "you've been telling me that every time you have pizza your belly hurts, look, this is it, believe it, stop".

However, the visualizations also caused some unease for *R8* and *R10*, who mentioned feeling “embarrassed” or “anxious” about giving the impression that they lacked necessary knowledge to explain the visualizations. Although they found collaborative review of the data helpful, they also doubted their abilities to correctly interpret the data or describe what foods contain certain nutrients. They therefore needed the tool to have enough information for them to quickly and accurately answer questions that patients might have. Although the visualizations helped assuage some concerns regarding trust, systems designed to support patient-provider collaboration must also be careful not to introduce new trust-related concerns in the collaboration.

Confidence in Data, Analyses, and Visualizations

Some patients and providers expressed uncertainty regarding the data quality, the correlation analyses, or the graphical depictions of the data, which affected their ability to interpret and act on the data.

Possible Data Confounds

The visualizations helped patients and providers raise questions regarding the quality of data collection and analysis with respect to their original expectations. Four providers (*R3*, *R4*, *R5*, *R8*) and four patients (*P2*, *P3*, *P5*, *P8*) described needing to know more about the context of the data in order to trust the displayed trends, citing possible confounds including emotional and physical health, hydration, exercise, and miscellaneous abnormalities (e.g., Thanksgiving Day, which was one of the days on which some participants journals, often involves abnormal dietary patterns). Similarly, two providers (*R2*, *R7*) and three patients (*P6*, *P8*, *P10*) expressed uncertainty because they lacked confidence in the patient's judgment of symptom severity at the time the data was collected. *P8* said:

[A]m I PMSing? If I am, maybe everything is terrible. I'm like, "Pain is terrible. Bloating terrible. Diarrhea? Didn't really happen, but it's terrible anyways".

In addition, five patients (*P1, P3, P4, P6, P8*) mentioned feeling their personal experience was not accurately reflected. As discussed in our correlation analyses, we defined potential triggers in terms of the preceding 4-hour window and discarded any symptom entries lacking corresponding meal entries. Some patients felt this strategy did not accurately reflect their experience with their symptoms and potential triggers, and others recalled symptoms that were not visualized. The time elapsed between data collection and our study also may have impacted patient recall and perspective on their data.

Personal versus Population-Level Information

Many participants wanted to verify the regression results were biologically plausible. Three patients (*P5, P9, P10*) and two providers (*R7, R10*) mentioned wanting to see corresponding population-level data, because it would help them determine whether the individual data and results were ‘normal’. *P10* said she needed to do more research about the significant nutrients from her analysis to learn whether they often cause IBS symptoms before deciding whether she trusted the results.

Participants were sometimes confused because the analysis was based on the nutrient range observed in that patient’s data, not some population-level range. For example, *P7* was perplexed her results indicated improvement when she consumed meals with higher fat. She knew from experience she consistently feels worse when she consumes high-fat meals. In fact, *P7* had that knowledge at the time of data collection and was already avoiding high-fat meals. Her analysis therefore did not suggest she experienced reduced symptoms when she consumed high-fat meals relative to the broader population, but rather meals that were higher in *her personal range* of fat content, which was low compared to the population average. Including typical diet information or population-level context may improve patient and provider understanding and confidence in personalized analyses.

Finally, participants also surfaced a need for the individual’s goals for the data to be understood for actionable interpretation. For example, during the individual interview, *R8* focused on identifying foods the patient could eliminate to reduce her symptoms. People with IBS often have the goal

of identifying foods to eliminate to avoid symptoms, so the assumption was not unreasonable. However, during the collaborative interview, the provider learned that the person with IBS already had restricted her diet to reduce her symptoms, and now wanted to focus on *diversifying* her diet without worsening her symptoms. Had the visualizations included that information, the provider would have been able to focus her time on addressing her patient's primary concern.

Visualization Complexity

Our visualizations illustrate different tradeoffs in simplicity versus expressiveness. The bubble and bar chart visualization shows high-level trends, while the parallel coordinates provide more detail. We thought some clear preference might emerge, but found diverse reactions to the alternatives. Some participants felt both were simple and straightforward, some expressed a preference, and others felt both were overwhelmingly complex. Participant confusion often affected confidence and interpretation. Two patients (*P2*, *P5*) and one provider (*R2*) indicated their ability to interpret the visualization directly informed the trust they had in the presented symptom analyses. *R2* explained:

I feel more comfortable with the interpretation than I did before where I really don't think, even toward the end and answering all your questions, that I actually understand [the parallel coordinates visualization]. Whereas I feel like I can understand [the bubble and bar chart visualization] and make my own conclusions here.

We also observed the opposite: *more* trust in a visualization participants found confusing. For example, *R6* started with the parallel coordinates visualization. Although he claimed to find it straightforward, he drew most of his conclusions from the summary page. When asked, he stated he had confidence in the results. However, upon switching to the bubble and bar chart visualization, he expressed less confidence, pointing out that the analysis “*relies on how accurate that data is that the person is inputting*”. *P8* similarly stated she had confidence in the bubble and bar chart visualization, but upon switching to the parallel coordinates remarked “*this is a really clear visualization to me of the range of how much fiber did I eat. I ate a really wide range of fiber from my ‘no symptoms’*”, thus revealing that she had not comprehended that range with the prior

visualization. We wanted the visualizations to provoke such critical thinking, as statistical analyses are limited (e.g., due to possible overfitting, due to correlated nutrients, due to other potential confounds). Results therefore require interpretation to determine whether and how they apply for a particular patient. Confusion regarding the visualizations often limited such critical thinking, leading participants to rely on the summary page and deemphasize the backing data.

Use Within and Beyond a Clinic Appointment

Although we intended the visualizations to be used individually and collaboratively in and surrounding a single clinic visit, participants identified opportunities for longer-term use.

Use During the Clinic Appointment

Patients and providers both wanted to use the visualizations collaboratively in clinic appointments. Providers valued learning more about the patient, and patients liked having an expert talk to them about the results. Eight providers (*R1, R2, R3, R4, R5, R7, R8, R10*) mentioned that talking to the patient helped their interpretation of the visualizations. They felt learning more about the patient helped them give effective recommendations: they were interested in patient goals, perception of their symptoms, opinion on their results, thoughts about their current diet, and willingness to change their diet. Five patients (*P5, P6, P7, P8, P10*) mentioned valuing having an expert's interpretation of their results, to ensure their own interpretations were correct (in some cases wanting interpretation guidance from an expert) and to verify the results made sense to a provider with experience with IBS.

Use Before the Clinic Appointment

Participants generally wanted to be able to access and review the visualizations before a clinic appointment. Eight providers (*R1, R2, R3, R4, R5, R6, R7, R8*) and six patients (*P1, P2, P4, P6, P7, P8*) wanted prior access to the visualizations to help prepare for the appointment. *P8* also expressed the need to be able to come to terms with any surprising or bad news:

If something came up [in the visualization] that I was unhappy about ... for me to be able to be unhappy about it on my own, and then come to the visit with, “Okay, what are we going to do?” [would be helpful]. Either I think this is bogus and we should try it again, or I’ve come to terms with the fact that we need to do this.

However, four providers (*R3, R6, R8, R10*) doubted whether they would have the time to review the visualizations between clinic appointments. Designs of future tools must therefore emphasize the ability to quickly interpret the visualizations and understand the main results.

Longer-Term Use Over Multiple Clinic Appointments

Eight providers (*R1, R4, R5, R6, R7, R8, R9, R10*) and six patients (*P1, P4, P5, P6, P8, P10*) expressed a desire for the visualizations to enable longer-term, iterative use. IBS is a chronic condition, and participants wanted to be able to view changes in a patient’s diet and symptoms over time. *R8* also mentioned wanting to be able to take notes about patients and review those notes before subsequent appointments. Although the analyses and visualizations were designed for a single food and symptom journal dataset, participants wanted to extend the system to aid the ongoing management of their IBS.

Use as a Tool for Hypothesis Formation

Six providers (*R4, R5, R6, R7, R8, R10*) and one patient (*P8*) expressed a desire to use the visualizations to form hypotheses about what nutrients might be impacting a patient’s symptoms. They then wanted to follow up this hypothesis formation with an experimental step, so they could test whether limiting the nutrient improved the patient’s symptoms. *R4* remarked:

It’d be interesting if you actually sort of did an experiment and you pick something that’s really high in one of these things ... and see what happens to your symptoms.

Integrating exploratory analyses for hypothesis formation with hypothesis testing through self-experimentation is one potential approach to providing such rigorous, actionable results [110].

4.2.4 Discussion

Patient-generated data is transformed into different types of boundary negotiating artifacts during patient-provider collaboration [42]. Building on the lens of this theory, we considered our visualizations as inclusion artifacts that summarize data to facilitate patient-provider discussion. Computer-supported analyses and visualization reduce the time and effort needed for a provider to perform on-the-spot mental analysis, thus allowing more interaction with the patient. Visualizations also help patients communicate their experiences clearly and concisely, making them feel more comfortable and involved in the clinic appointment. Our designs focus on exploratory, interactive visualization and interpretation, providing the opportunity for patients and providers to negotiate which potential triggers to consider in a treatment decision. Here, I discuss considerations for tools to support patient-provider collaboration in the interpretation of patient-generated data.

Encouraging Provider Trust in Patient Abilities

Many providers were initially skeptical of whether patients would be able to correctly interpret the data visualizations. Although this concern was often assuaged during the collaborative interviews, more dramatic differences between patient and provider modes might help mitigate their initial concerns. Another approach to increasing provider confidence in patient abilities may be to support different levels of exploration with different levels of complexity. For example, we designed: 1) the summary page as an overview of the main takeaways; 2) the bubble and bar chart visualization as a simpler view of high-level trends; and 3) the parallel coordinates visualization as a more detailed view of the patient's nutrients and symptoms. By supporting different levels of complexity, patients who are comfortable exploring complex visualizations can do so, while patients who are confused or intimidated can explore the simpler representations to assess their data before a clinic appointment. When the patient reviews the data with their provider, the provider could determine the patient's level of understanding by inquiring about the complexity level they reviewed at home and what interpretations they formed during this review. The provider could then explain the data at an appropriate level.

Providers often doubt a patient’s ability to correctly collect or interpret data [251], and patients often have difficulty understanding common visualizations of health data [176]. However, our study found that many patients were able to navigate a complex visualization when aided by a short interactive tutorial. Fully developed tutorials could support patient interpretation and help address provider concerns. Such an approach could also support provider-directed tracking (i.e., when a patient initiates tracking at the request of a provider, a common practice in IBS management). Educational materials and protocols could help ensure patients develop the knowledge necessary to both collect and interpret data, while simultaneously helping to mitigate provider concerns.

Providing Clear and Comprehensive Visualizations

Both patients and providers were concerned that their collaborator would not trust or understand what they were trying to communicate. Patients were concerned providers would not understand their symptoms, and providers were worried patients would not believe recommendations the providers could not illustrate or support. The visualizations helped assuage many of these concerns by providing clear and objective views of the patient’s experiences, rather than relying on abstract conversations about them. As I discussed with *R4*’s explanation of the importance of concrete examples, the visualizations helped foster a feeling of collaboration by providing something to examine together. Tools designed to help promote patient-provider collaboration should emphasize such explorations.

However, the visualizations also introduced some new potential concerns. Providers did not want the patient to see them as incompetent, and worried they might not be able to quickly and easily navigate and interpret visualizations in the presence of a patient. To ensure a provider is comfortable, systems should be simple to understand and interpret, with educational information immediately accessible (e.g., in case a patient asks a question that a provider cannot answer). In addition, systems should support patients and providers reviewing data individually before an appointment, so that both are confident about the data and agenda during collaborative interpretation.

Systems should also surface information about data processing. Prior work in algorithm transparency has suggested making system logic explicit to increase engagement [67, 224, 236]. We included details of data analysis as a help page in the interface, and the tutorial alerted participants

to its existence. However, few participants attempted to read this information. Participants may not have seen a need to consult this resource during the short interview study, but patients were sometimes confused about why the displayed data differed from their recollection. The fact that participants did not access the information may therefore illustrate a need to better integrate it into the visualization itself, rather than presenting it as a separate page. In addition to supporting understanding, better surfacing analysis details and assumptions can support patient and provider evaluation of whether those details and assumptions are appropriate for an individual patient's case.

Supporting Use Inside and Outside the Clinic

Patients and providers wanted to access the visualizations both before and during clinic appointments, and also expressed a desire for long-term use. Providers were concerned the visualizations may take too long to interpret, which would hinder their ability to review data before a clinic appointment. A key requirement for supporting provider use before an appointment is therefore ensuring that interpretation is easy enough for providers to quickly review and get a sense of the data between clinic appointments. Another approach is to help the provider prepare for the goals and questions the patient is bringing to an appointment. Prior research has found that patient-generated data prepares patients for questions and conversations in face-to-face and remote provider appointments [1,4,154]. If a system could support patient annotation before or between appointments, it might help providers understand a patient's concerns, questions, and goals, thereby allowing the provider to efficiently prepare [41]. Future work should further explore opportunities and challenges with longitudinal data.

We designed our visualizations to support independent use by a patient or provider as well as synchronous, co-located use by a patient-provider pair. We did not design or evaluate our visualizations for use in asynchronous or remote synchronous collaboration. Additional features may help support effective asynchronous collaboration, such as support for common ground and awareness suggested by Heer and Agrawala [91] or digital traces suggested by Trainer et al. [240]. Future research should develop and evaluate such additional enhancements.

Providing Understandable and Actionable Information

The visualizations sometimes lacked necessary context for accurate interpretation. For example, five patients (*P5*, *P7*, *P8*, *P9*, *P10*) and one provider (*R10*) mentioned not knowing exactly what certain nutrients were. This confusion was exacerbated when participants forgot that the data was based entirely on the patient's diet, rather than population-level information. *R10* was confused because the visualization indicated that coffee had a large amount of soluble fiber, which *R10* knew was false. However, *P10* consumed little dietary fiber. Coffee was *High* in soluble fiber only relative to that patient's diet, not relative to any absolute notion of a large amount at the population-level.

Similarly, I reported that *P7* was confused about the role fat played in her symptoms. If she understood that she tended to feel better when she ate meals that were *relatively* high in fat, that information might help her liberalize her diet (e.g., trying meals with slightly higher fat to see whether she continues to avoid symptoms). Even if *P7* did not want to experiment, better conveying the context for these analyses would prevent confusion over an indication that she tended to feel better with higher-fat meals. Providing the context necessary for interpretation is important for all data visualizations, and is particularly important for designing systems that are intended to provide actionable information.

Actionable insights are a common goal of patient-provider collaboration, so supporting the translation of results into action is essential. However, with data as complex as that considered here, designers should resist the temptation to jump from summaries and recommendations to automated plan generation. Three providers (*R6*, *R7*, *R10*) and two patients (*P5*, *P8*) wanted the system to tell them exactly what foods the patient should avoid. *R10* expressed a desire for the system “*to be smarter than a human and come up with its own recommendations ... which it could be, because it's a computer*”. Unfortunately, determining exactly what foods are causing an individual's IBS symptoms is beyond the scope of what can be done with food and symptom journals of the type considered in this work (e.g., due to concerns for correlations and confounds). We did not want the presentation of an analysis to imply greater confidence than is warranted, but instead focused on using analyses to scaffold exploratory visualizations. Patients and providers can then collaborate in applying their knowledge and expertise to interpret the data.

One approach to actionability in the face of uncertain results is to support hypothesis testing with self-experimentation [110]. IBS management is a long-term process that often involves iterative hypothesis formation and testing, as well as changes in the habits and goals of the patients [156]. Our visualizations provide support for identifying potential relationships between nutrients the patient eats and their subsequent symptoms. Utilizing self-experimentation for hypothesis testing could then provide actionable next steps for patients to determine whether dietary changes based on these relationships will help with the long-term management of their symptoms.

4.2.5 Study Conclusion

Building upon prior research in supporting patient-provider collaboration, we developed two visualizations to support patients with IBS and their providers in collaboratively identifying individual symptom triggers. We designed the visualizations to be both actionable and exploratory, allowing patients and providers to collaboratively apply their knowledge and expertise in interpreting the visualizations to reflect, negotiate, and make treatment decisions. We found that collaborative review helps both patients and providers better understand patient-generated food and symptom data, supporting mutual trust in their relationship. We also revealed a need to use such tools beyond a single clinic appointment to support the long-term management of the patient's illness.

Designers of such systems should help healthcare providers understand patient abilities for data interpretation and should support patients and providers through tutorials and other educational support. Tools should also be *flexible* to a range of patient data and experiences. To help patients and providers build confidence in *exploring* the data together, designers should strive to provide *simple* and understandable views of the data, as well as the process of data collection and analysis. Providing more context surrounding patient-generated data can help patients and providers form better interpretations and decisions. Finally, as symptom management is a long-term process, patients and providers need *actionable* support to help them test and adjust decisions throughout long-term care.

4.3 INVESTIGATING HOLISTIC SUPPORT FOR IBS TRIGGER IDENTIFICATION

With our interactive, exploratory visualizations, I aimed to help patients and providers use their data to form hypotheses about potential IBS triggers. In subsequent work, I also contributed to investigations in supporting: 1) hypothesis testing through scaffolded self-experimentation [109]; 2) interpretation of self-experimentation data based on the questions an individual may have [216]; and 3) lower-burden collection of food and symptom data [43]. Here, I briefly review this additional work.

4.3.1 Supporting Hypothesis Testing through Scaffolded Self-Experimentation

Although hypothesis formation about potential IBS triggers is a useful first step, many people with IBS and their health providers would rather not needlessly eliminate foods from their diet. They therefore often wish to test their hypotheses, so they can understand whether the identified nutrients actually *cause* their IBS symptoms. We therefore developed TummyTrials, a mobile application designed to scaffold low-burden self-experimentation to investigate whether a hypothesized IBS trigger causes symptoms [109]. This research was led by Ravi Karkar.

System Design

With TummyTrials, a person with IBS first identifies a single nutrient they want to examine, as well as the symptoms they want to measure. TummyTrials then randomly assigns an equal number of days throughout the experiment for the person to consume or avoid the trigger (thus configuring a fully randomized alternating treatment experimental design). Every morning throughout the study, the person will have the same meal for breakfast, the only change being whether they consumed or avoided the nutrient as directed by the app. For example, someone examining lactose might have cereal with cow's milk on the days they are directed to consume the nutrient, and the same cereal with lactose free milk on days they must avoid it. They then record the symptoms they experience in the next three hours on a scale from 0 to 6, avoiding all other food in that time period. At the end of the 12-day experiment, the app provides a visualization and summary of the results.

Feasibility Study

We evaluated TummyTrials via a feasibility study in which 15 people with IBS completed personalized 12-day experiments using the app. In general, participants found the app to be usable, and were able to successfully complete the 12-day self-experiment. However, the study also revealed tensions between the desire for scientific validity and the realities of the lived experience of self-experimentation.

4.3.2 Discovering Personalized Goals Surrounding IBS Trigger Identification

In the feasibility studies for both the interactive, exploratory visualizations and TummyTrials, we observed a trend of participants expressing a desire for more nuanced and personalized information than whether or not a particular nutrient is a trigger for them. Working closely with Matthew Kay and Ravi Karkar, I therefore led an investigation of the types of questions people want to answer via self-experimentation by 1) reanalyzing our feasibility study data and 2) conducting a survey to determine what questions individuals want to answer about their health and wellness. We then provided design recommendations for analyses and data representations that could help people answer and interpret such questions, finding that Bayesian methods may provide analyses that can better answer those questions than frequentist analyses (e.g., null-hypothesis significance testing, frequentist estimation) [216]. Here, I focus on the reanalysis of our interview data, as it illustrates the nuanced goals people have around identifying their personal IBS triggers.

Questions that Past Interview Participants Wanted to Answer

People with IBS and their health providers often express a desire to know *whether* certain foods trigger their symptoms. However, when using our tools, they often identified a need for more nuanced and detailed understandings of the relationship between potential triggers and symptoms (e.g., because the patient did not want to completely give up the identified nutrient, because the provider did not want to suggest elimination if the effect of the trigger was sufficiently

small). I refer to provider participants in the visualization study as *RX*, participants with IBS from the visualization study as *VX*, and participants from the TummyTrials study as *TX*. We identified two main types of questions participants in our feasibility studies wanted to answer.

By how much do my symptoms change when I consume the nutrient?

When informed that a nutrient was correlated with a change in symptoms, people often wanted to *quantify* that change. Having a detailed understanding of how a nutrient affects symptoms could help people perform cost-benefit analyses to decide whether to avoid that nutrient. Providers were also interested in quantifying differences between consuming and avoiding a nutrient because it influenced their recommendations. If the difference in symptoms between consuming and avoiding the nutrient was small, they might not advise the patient to eliminate the trigger, despite its significant *p-value*. Regardless of any *statistical* significance, nutrients with low magnitude and high variance of the effect had limited *clinical* significance.

“When you say improving, how much improvement?” (V9)

“Is being more awake worth potentially having stomachache? Which matters more to me at this particular moment?” (T8)

“I bet it’s correlated, but clinically, it doesn’t make any difference. That’s what my interpretation is” (R4)

“Even if the data is saying it’s significant, if it’s not helpful in the real-world then what good is it?” (R6)

How much of the nutrient is associated with increased or decreased symptoms?

People also wanted to understand how much of the trigger they could eat before their symptoms were affected. This information could help people understand how drastically their behaviors needed to change to avoid future IBS symptoms (e.g., *eliminating* versus *limiting* their caffeine consumption). Providers also wanted to understand how much of the nutrient was associated with changed symptoms because they did not want to advise patients to eat more or less of a nutrient than was practical and healthy. Simply indicating that more of a nutrient correlated with improved symptom was therefore not nuanced enough for providers to give clinical advice they were comfortable giving.

“if there is a threshold ... I would just stay beneath the threshold and not be the weird person who has to drink decaf” (T8)

“maybe having zero fiber is not as good as having five grams of fiber per meal. Maybe ‘target five grams’ is different than ‘have more’” (V8)

“ideally, we would ... quantify it a little bit more to say ‘moderate caffeine associated with no symptoms’ ... at higher levels of caffeine, [the patient is] just going to have diarrhea” (R8).

Using Bayesian Statistics to Answer the Kinds of Questions People Have

Once we had characterized different questions self-experimenters might want to ask, we investigated different statistical methods people might consider when analyzing self-experimentation data to answer those questions. For example, using Bayesian modeling, we can create predictive distributions of *future* symptoms (see Figure 4.6). With this information, people could predict their symptoms if they were to consume or avoid the trigger. They could therefore perform the kinds of cost-benefit analyses participants often desired. Our paper describes other possible representations of Bayesian methods and compares them to the more common, but often less appropriate, frequentist analyses [216].

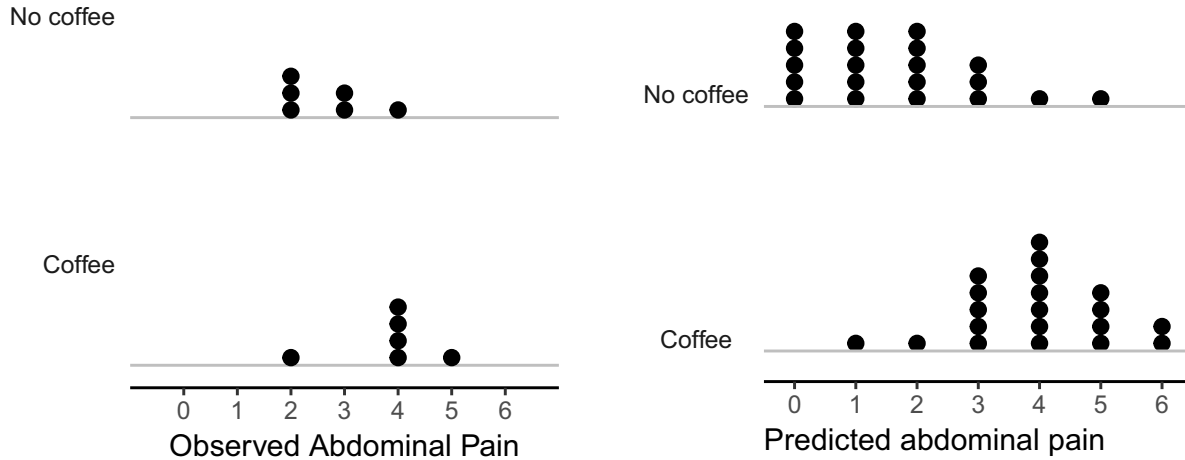


Figure 4.6: Simulated data from a self-experiment investigating whether coffee causes abdominal pain (left), and predicted *future* abdominal pain if the self-experimenter avoided or consumed coffee (right). Each plot shows 20 approximately equally likely predicted outcomes, illustrating that abdominal pain at level 4 or more is predicted to occur 2 times out of 20 when coffee is avoided (or 10% of the time) and 13 times out of 20 when coffee is consumed (or 65% of the time).

4.3.3 Lowering Burden of Food Journaling through Lightweight Food Diaries

Although health providers often recommend people with IBS journal their foods and symptoms, traditional food diaries are often highly burdensome [51]. Photo-based food diaries can not only be less burdensome than traditional journaling methods, but can also support reflection on eating behavior and context (e.g., people can determine whether they ate the logged meal at home or at a restaurant based on the information included in the photo) [40, 50]. We therefore developed Foodprint, a system designed to help people: 1) collect data about what they eat via photos augmented with lightweight manual tracking, and 2) interpret that data, both individually and collaboratively with a health provider [43]. This work was led by Chia-Fang Chung.

System Design

Foodprint consists of three main components: 1) a mobile app that supports collection via photos as well as lightweight manual tracking (e.g., added sugar, IBS symptom severity since the previous meal); 2) a web app that includes both photo-based visualizations and an adapted version of the bubble and bar chart I presented in Section 4.2; and 3) a pre-visit note with which patients could communicate expectations, questions, and goals to health experts before a clinic visit. To produce the nutrient-level information needed for the bubble and bar chart visualization, trained dietitians translated the photos into the meal's constituent nutrients via a similar process to that described in Section 4.1.

Feasibility Study

We evaluated Foodprint via two feasibility studies: one with 17 people with healthy eating goals and 7 health experts that examined how the system could support healthy eating, and another with 16 IBS patients and 8 health experts that examined how it could support trigger identification. We found that such a system could: 1) help people collect data relevant to their health goals; 2) help providers focus on the individual's eating context, experiences, and goals during collaborative review of the data; and 3) help patients and providers collaboratively develop individualized, actionable plans.

4.4 A GOAL-DIRECTED PERSPECTIVE ON SUPPORTING IBS MANAGEMENT

We designed our methods and tools to support the common goal of IBS trigger identification, based on collaborations and studies with medical experts that identified the goal as important and unsupported. We were also mindful that people might want to answer specific questions with their data. For example, as I described in Section 4.2.1, we wanted the visualizations to help people explore any pre-existing beliefs they may have rather than limiting their explorations. However, we did not explicitly elicit or support the personalized goals individuals may have for their data, which caused some pitfalls. As I discussed in Section 4.3.2, people often develop more nuanced and personalized goals than simply identifying a nutrient as a trigger. For example, I included a quote from T8, who found that caffeine was likely a trigger for her symptoms. She wanted to know whether

she could still have *some* caffeine without causing symptoms, illustrating a personalized goal based on her desired habits and management approaches. Although some people may be willing to completely eliminate a nutrient that causes IBS symptoms, this participant did not want to stop drinking coffee. Her personal goal was therefore to discover how much she could have without causing symptoms. Answering this question could help her develop a personalized management strategy she would be comfortable pursuing—which simply identifying caffeine as a potential trigger may fail to do.

We also found pitfalls when a patient’s and provider’s goals did not align. For example, in Section 4.2.3, I described how *R8* spent the individual interview looking through the data to identify foods to eliminate, only to find when they met in the collaborative interview that the patient actually wanted to focus on diversifying her diet. If the visualizations had included that information, the provider would have been able to focus her time on addressing her patient’s primary concern. This finding therefore revealed a need to better communicate not only the data itself, but also the reasons people have for sharing that data, when attempting to support patient-provider collaboration using self-tracked data. Based on this information, we included a pre-visit note in Foodprint so patients could indicate their goals for an appointment (see Section 4.3.3). This note helped ensure the provider understood those goals when reviewing their data, increasing the value of the data and the appointment.

Both of these issues illustrate opportunities for tools to explicitly acknowledge and support the multiple, distinct, and personalized goals that people may bring to their IBS management.

4.5 SUMMARY

We designed, developed, and evaluated systems to support *hypothesis formation* and *hypothesis testing* in the identification of personalized IBS triggers, a common goal for people with IBS and their health providers. Our findings suggest potential for interactive, exploratory data visualizations to support data interpretation and patient-provider collaboration to identify *potential* IBS triggers, helping people better understand food and symptom journal data that is generally difficult to interpret and act on. We also found that scaffolding self-experimentation can support people in using scientifically-rigorous methods to investigate whether those potential triggers *cause* IBS symptoms.

However, as people interacted with our tools, they often developed more personalized and nuanced goals than simply *identifying* symptom triggers. Based on these findings, we proposed the use of Bayesian analyses to help people interpret their data appropriately for those personalized goals.

I therefore supported people with IBS and their health providers in *pursuing* the common goal of trigger identification, helping them *overcome barriers to the personalized management* of their condition (i.e., barriers related to collecting and interpreting burdensome food and symptom data without validated methods and scaffolding). I also found a need to better support people in expressing and pursuing their *multiple, distinct, and evolving* goals for their IBS management.

Chapter 5

SUPPORTING MIGRAINE MANAGEMENT THROUGH GOAL-DIRECTED SELF-TRACKING

Although my work in mental health and IBS was able to support common goals in those health contexts, it also uncovered the need to support the *expression* and *pursuit* of each individual's personal goals over the course of their long-term condition management. These findings informed the design and development *goal-directed self-tracking*, a novel method of personalized self-tracking that explicitly elicits and supports the specific goals of the individual and their health provider throughout all stages of the self-tracking process. With goal-directed self-tracking, I aim to: 1) support people and their health providers in developing actionable goals; 2) ensure people are tracking *exactly* and *only* what they need to be tracking and when to achieve those goals; and 3) support people and their health providers in appropriately interpreting their data given their goals.

I examined goal-directed self-tracking in the context of migraine, a complex, chronic condition characterized by debilitating symptoms that can be induced by a wide range of personalized contributing factors. Working with an interdisciplinary team including technology experts and health providers from the UW headache clinic, I first completed formative work to better understand self-tracking by people with migraine. We identified distinct categories of goals people often pursue when tracking migraine-related data. I discuss our paper on this formative study [214] in Section 5.1.

Informed by our formative study findings, we iteratively developed *goal-directed self-tracking* to scaffold the process of deciding what, when, and how to track towards a specific goal and to analyze and visualize the resulting data to support that goal. We created a paper prototype for a goal-directed self-tracking system for migraine management and evaluated the prototype in interviews with 14 people with migraine and 5 health providers, finding potential for the method to elicit and support migraine tracking goals. I present our paper on this development [217] in Section 5.2.

Finally, I summarize how my work in migraine supports my thesis statement in Section 5.3.

The work described in this chapter was done in collaboration with Chia-Fang Chung, Daniel A. Epstein, Ravi Karkar, Adele Parsons, James Fogarty, Sean A. Munson, and Natalia Murinova.

5.1 IDENTIFYING GOALS AND NEEDS IN MIGRAINE TRACKING

To inform my research in methods and tools to represent and support a person's migraine tracking goals, I first conducted preliminary research investigating the goals people and providers bring to migraine management, as well as challenges they face in addressing those goals [214].

5.1.1 Formative Study Methods

To examine how people and providers self-track to manage migraine, we conducted a survey examining experiences with migraine and then conducted follow-up interviews with a subset of survey participants (see Table 5.1). We also interviewed 6 providers for their perspective on working with patients with migraine (see Table 5.2). Participants skew female, which matches migraine prevalence in the United States: women are three times as likely to experience migraines than men [200]. I refer to survey participants as SX, interview participants as IX, and provider participants as PX.

Survey of 279 People with Migraine

We first surveyed 279 people about how they understand, track, and manage their migraines. The survey asked primarily open-ended questions about experiences with migraine, tracking related to migraine, and consulting with health providers about migraine. We focused on understanding what people found particularly helpful and unhelpful about tracking, health consultations, and sharing their data. Participants were recruited through university mailing lists, Facebook posts, and posts to two Reddit subreddits: one dedicated to migraine and another to surveys. Participants were entered into drawings for a single \$100 and three \$25 Amazon gift cards.

Survey Participants (279 People)	
Gender	222 Female, 51 Male, 2 FtM transgender, 1 genderqueer, 1 genderqueer man (2 no answer)
Age	18–96, $\bar{x} = 33.8$ (7 no answer)
Years with Migraine	1–65, $\bar{x} = 17$
Past Apps (106 people)	Migraine Buddy (71); 40 apps with <10 people each (14 could not remember which they had tried)
Current Apps (56 People)	Migraine Buddy (30); 23 apps with <10 people each
Other Resources Tried (196 People)	paper journals (83), online articles about migraine (70), online forums such as reddit (44), custom spreadsheets (30), calendars (24), books about migraine (14); 17 methods with <10 people each
Interview Participants (14 People)	
Gender	11 female, 1 male, 1 genderqueer man
Age	20–61, $\bar{x} = 32.15$
Years with Migraine	1–30, $\bar{x} = 15.8$
Past Apps (8 people)	Migraine Buddy (5), Migraine Coach (2), Fitbit (1), Clue (1), Pacifica (1), Curelator (1), iHeadache (1), PTracker (1), Grid Diary (1) (1 could not remember which they had tried)
Current Apps (3 People)	Migraine Buddy (1), Migraine Coach (1), Clue (1), Pacifica (1)
Other Resources Tried (13 People)	paper journals (7), custom spreadsheets (5), online articles about migraine (4), online forums such as Reddit (4), calendars (3), books about migraine (1), tracking help from family (1), discussions with other people with migraine (1)

Table 5.1: Summary information for formative study participants with migraine.

To analyze the survey data, three researchers (myself, Ravi Karkar, and Daniel Epstein) conducted an affinity diagram analysis of open-ended survey responses. I then created a preliminary code book based on emergent themes. We coded three different categories of participant experiences: 1) using apps to understand, track, and manage migraines; 2) consulting with health providers about migraines; and 3) sharing data with providers during consultations. We discussed and iterated on

Participant	Area of Practice	Provider Type
P1	Family Medicine	Medical Doctor, Master of Public Health
P2	Family Medicine	Medical Doctor
P3	Headache Clinic	Nurse Practitioner
P4	Primary Care	Nurse Practitioner
P5	Headache Clinic	Nurse Practitioner
P6	Primary Care	Medical Doctor

Table 5.2: Summary information for formative study provider participants.

codes to develop the final code book, which consisted of 40 codes. I then coded 100% of the data, while the other two researchers each coded 10% of the data. We then calculated inter-rater reliability, defining “agreement” as use of the same code within a given question response. As preliminary code agreement varied ($\kappa = 0.16 - 1$, $\bar{x} = 0.67$ for the 28 codes used more than 10 times, with 0.8 or higher for 11 codes), the researchers discussed and resolved any disagreements. Most disagreements resulted from misunderstandings about the code book and in infrequently-occurring codes. The first author then took a second pass through the data based on the discussion.

Interviews of 13 Survey Participants

After analyzing survey data, we identified themes to explore in more detail. We wanted to learn more about people who commented on successes or failures in tracking. For example, we reached out to individuals who said having data had helped their health consultations, as well as those who had struggled to interpret and act upon data. We were particularly interested in people who had struggled with apps and turned to alternative methods. We emailed 23 survey participants who had agreed to be contacted about follow-up interviews; 13 responded to our request.

Semi-structured interviews were conducted via Skype or phone call. Interviews covered: 1) the participant's general experience with migraine; 2) how the participant collected data; 3) what they liked and disliked about their method; and 4) how and with whom they shared data (e.g., family, friends, coworkers, providers). We encouraged participants to discuss their experiences, thoughts, and behaviors. Interviews were approximately 30 minutes, compensated with a \$20 Amazon gift card.

A third-party service transcribed the audio recordings of the interviews. Two researchers (myself and Chia-Fang Chung) then open coded 2 transcripts. I open coded the remaining 11 transcripts, then affinity diagrammed the resulting codes and created a code book of 70 codes from emergent themes. Finally, three researchers (myself, Daniel Epstein, and Adele Parsons) coded the data using the code book. Inter-rater reliability is not often calculated on semi-structured interview data because different people can apply the same code to different parts of a conversation [6]. We therefore did not calculate reliability between researchers.

Interviews of 6 Health Providers

Finally, after learning about the expectations and experiences people with migraine had when bringing their data to providers, we interviewed 6 health providers from a range of specialties about their experiences with patients with migraine. We conducted 5 interviews in person and 1 by phone. Interviews were again semi-structured, covering three main topics: 1) general experiences working with patients with migraine; 2) experiences with patients who tracked information about their migraine, focusing on whether and how the provider interacted with patient data; and 3) reactions to specific examples of migraine tracking methods, collected from our interviews of people with migraine. Providers were recruited via email introductions from our medical collaborators and compensated with a \$30 Amazon gift card. A third-party service again transcribed audio recordings. Resulting transcripts were analyzed via a similar process of open-coding, affinity diagramming of codes to create a code book, and coding all data using the resulting code book.

Limitations

Many of the people we surveyed and interviewed were associated with professional and social networks of the researchers. Participants skew more highly-educated than the general population: 38% of survey participants have an advanced degree, and an additional 32% a bachelor's degree. Recruitment, particularly via the subreddit for migraine, may have biased our sample to people who seek an online community for migraine resources. Prior work suggests people who get migraines less frequently or less severely are less likely to dedicate as much time to tracking and management [190], so they may also not engage in such a community. Finally, 84% of survey participants were from the United States, as were all 6 provider participants. Tracking habits and care strategies may differ in other cultures or countries, so care must be taken when attempting to generalize our findings.

5.1.2 Results

Of 279 survey participants, 106 had used an app to try to understand, track, or manage their migraines, and 56 still did so. Participants primarily reported using apps for tracking migraines (n=80), determining triggers (n=59), and tracking medication use (n=13). Most participants had used resources or tools other than apps to try to understand, track, or manage their migraines (n=196), instead of or in addition to apps. The most common use of other technological and paper resources was tracking, but participants also described online research and community support (see Table 5.1).

Many survey participants found apps useful for migraine-related tracking. Those who had used apps often felt their app was easy to use (53/106) and helped them identify trends and triggers (42/106). However, many participants felt the tools they used failed to fully support their goals. Here, I discuss four challenges we found current technologies fail to fully support: 1) variety in personal tracking goals; 2) evolution of tracking goals over time; 3) variety in personal experiences with migraine; and 4) support for data sharing and collaboration with health providers.

Although I report counts of participants who expressed perspectives, the open-ended nature of the questions means not every participant was explicitly prompted about each perspective. As a result, additional participants beyond those in each count may have agreed with any given perspective.

Tracking for a Variety of Goals

Informed by and building upon prior research on self-tracking goals (e.g., [39,65,117,133,133,208]), we find four primary categories of goals people bring to migraine-related self-tracking: 1) answering questions about migraines, 2) predicting and preventing migraines, 3) monitoring and managing migraines over time, and 4) enabling motivation and social recognition. Participants often had simultaneous and overlapping goals, but explicitly distinguishing them better facilitates consideration of the design implications of each goal for self-tracking data.

Goal 1: Answering Questions About Migraines

Many people who track information about their migraines do so with a goal of *diagnostic self-tracking*, or self-tracking to answer a specific question about a person's health [208]. Common questions included what behaviors or other factors contribute to an individual's symptoms and whether a new treatment or preventative medication seems to improve symptoms. For example, of 106 survey participants who had tried an app, 65 had used it to attempt to identify factors that contributed to their migraines. Some apps provide graphs or analyses to help identify factors that correlate with symptoms, and 16 survey participants mentioned feeling that apps helped them interpret their data. However, another 16 survey participants explicitly discussed finding data hard to interpret. When asked during the interviews, 10 of 13 participants described similar challenges (I1, I4, I5, I6, I8, I9, I10, I11, I12, I13). I02 described her frustrations attempting to identify and understand the factors that contributed to her migraines: *"I thought ALL drinking was a trigger for me. Turns out it was the corn syrup in the tonic water and other mixers I was using. This was so hard to figure out and none of the current apps can help you with this"*.

In addition to failing to support investigations of sufficiently nuanced trends in migraine contributing factors, individual apps generally cannot support tracking the wide range of factors potentially relevant to an individual's migraine. Participants therefore often tracked data across multiple apps. To identify trends, participants then needed to integrate data across apps themselves. I10 explained she *"found it stressful ... there's no place in any of those apps for a whole health*

picture of, like, OK, I was nauseated today. I had a migraine today". S66 lamented that no single app could help him track and understand all of his data: *"I wish a program like a general health app (e.g., Lose-It) would track everything and do correlations."* S258 summarized that apps made it *"[t]oo difficult to put all the pieces together"*. When people collect data in separate apps, the data typically remains siloed, creating additional barriers and burdens in data interpretation.

Participants often used data to compare symptoms before and after making a change. If people feel symptoms improve after making a change, they may be more likely to commit to that change, especially if the change requires modifying behavior or incurs side effects. If symptoms do not improve after the change, people may return to the prior behavior or ask their provider for another solution. Data-driven "before and after" comparisons can therefore help people get a sense of how such a change affects their symptoms, helping them perform these cost-benefit analyses. 7 of 13 interview participants mentioned that tracking had helped them identify treatment-related trends (I1, I2, I6, I9, I10, I11, I12), with I11 explaining, *"you basically turn your life into a control of an experiment and try to change one thing at a time and see what works"*. All 6 providers also wanted their patients to track as part of enabling comparisons to determine effectiveness and whether to continue a treatment.

However, despite generally considering data useful for answering such questions, providers were often wary of patients focusing too much on a single potential trigger. 4 of 6 providers (P1, P3, P4, P5) expressed concern that patients might see false patterns, leading them to make changes that could be unnecessary or inappropriate. When discussing participants who track foods to try to identify dietary triggers, P1 explained that the practice was not only often futile, but also potentially dangerous:

There can be a lot of overlap between dysfunctional eating and tracking, and so sometimes [tracking foods] reinforces dysfunctional eating like, "Oh, I got a headache after I ate bread, so now eliminate gluten from my diet", when it's just a way to kind of normalize dysfunctional eating.

Goal 2: Predicting and Preventing Migraines

Distinct from tracking to determine what factors increase an individual's likelihood of symptoms (i.e., Goal 1), many people track with a goal of predicting and preventing migraines. Of 106 survey participants who had tried an app, 15 had used it to try to predict their migraines. Because migraines often occur due to an accumulation of multiple factors [119,250], people may not need to entirely eliminate factors that can contribute to a migraine. In addition, some factors that may contribute are outside of a person's control. When people have identified their potential migraine contributors, they therefore often attempt to be cognizant of exposure to factors that affect the likelihood of symptoms. 4 interview participants (*I1, I3, I6, I7*) described trying to predict likely migraines by paying attention to what they are exposed to, so they can compensate by making more of an effort to avoid other contributing factors in an attempt to prevent a migraine. For example, *I6* described using the MigraineX app to get alerts when the barometric pressure was going to drop. When she got those alerts, she said:

I do what I can to make sure that I'm really well hydrated ... that I get my best sleep ... that everything else in my body is as healthy as possible so that hopefully I can ride through that weather front without any other triggers.

Even when a migraine could not be prevented, participants appreciated predictions because they provided an opportunity to prepare. For example, *S257* often got migraines at specific times in her menstrual cycle, so she tracked her cycle to predict symptoms: *"The app can be validating: when I feel a migraine coming on I can see that it might be roughly the usual time in my [menstrual] cycle to get one. Instead of fighting it, I can just try to adjust my schedule and get through it."* *I7* additionally explained how predicting a migraine allowed her to prepare: *"Even if it's just throwing extra rescue meds in my purse. It's an incredibly helpful thing instead of sitting there and floundering while you're on the train ride on the way home like 'oh my God, I wish had something to take right now'"*.

Tracking for prediction and prevention is often complicated by people struggling to identify the factors that affect their symptoms, especially because some important factors are surprising to many people with migraine. For example, 4 of 6 providers described overuse of acute medications

as the top mistake among patients. Many people with migraine take acute medications (e.g., 224 of 279 survey respondents reported trying over-the-counter medication to stop a migraine, 201 of 279 reported trying a prescription acute medication). However, medication overuse can actually cause more frequent migraines, creating a cycle in which a person gets more migraines, which prompts them to take more medication, which leads to still more migraines [142]. Providers emphasized a need for education about trigger accumulation and medication overuse, as tracking other factors is likely futile absent this foundational knowledge. Many participants wanted technology to include such information: 5 interview participants (*I3, I4, I6, I11, I13*) thought apps should provide recommendations and make suggestions about what they might track or changes they might try.

Goal 3: Monitoring and Managing Migraines Over Time

Although many participants tracked to learn about their migraines (i.e., Goal 1) or predict migraines (i.e., Goal 2), others tracked to monitor and manage their condition. Tracking migraine symptoms was the most common reason survey participants reported using an app, with 80 of 106 people who had used an app reporting symptom tracking as among their goals. In contrast to prior goals, monitoring and managing migraines over time often focuses on symptom-related data, and therefore cannot help identify correlations with factors that might be related to symptoms. However, 10 interview participants discussed how such symptom tracking helped them remember and monitor their experiences (*I2, I3, I4, I5, I6, I7, I9, I10, I11, I13*). For example, *I2* talked about how tracking helped her monitor her symptoms, ensure she was not getting worse, and notice any need for changes:

[my provider] saw the January and February and she's like, "Oh, that's a reasonable number of headaches", and then she saw March and April and was like, "That's too many headaches" ... And she was like, "Okay if you're still getting six or seven headaches a month in a couple months, we're gonna talk about doing something else".

Many providers encouraged patients to track migraine-related data, and providers often use such data to support diagnoses and treatment decisions. All 6 providers had recommended patients track migraine symptoms under some circumstances. Of the 47 survey participants who had shared data

with a provider, 10 felt that having the data directly influenced their diagnosis or treatment, and all 6 providers said the same. To support such monitoring and management, patients and providers need data about migraine frequency and severity, as well as information about pursued treatments.

Goal 4: Enabling Motivation and Social Recognition

Although providers often recommend patients track to gain information, tracking can also help motivate people to make changes to improve their symptoms. *P3* and *P5* work at a headache clinic, where they recommend behavior changes that can help reduce migraine frequency and severity (e.g., increasing exercise, modifying diet, relieving stress). Patient-tracked symptom data helps providers show patients that their behavior changes actually improve their migraines. *P3* explained:

[a patient] might think, “I’m doing so horrible, I’m doing so horrible”, and then you look at it, and you’re like “you actually cut back your severity by 50%” ... you kind of shift your perspective as you go on, and the new horrible becomes just the same horrible, even though it’s improved. So, it’s helpful to show that improvement more than anything.

The “before and after” comparison *P3* describes has similarities to interpreting what factors impact symptoms (i.e., Goal 1) and to monitoring migraines (i.e., Goal 3), but also differs in its purpose. For example, a person may make multiple small behavior changes, and then need to examine a larger, cumulative comparison to motivate themselves that all of these changes together are justified because they result in an overall improvement. Although data collection may be similar, a person seeking motivation to continue the overall process can therefore need different representations of data than when investigating a specific trigger or monitoring symptoms.

In addition to using data to motivate by demonstrating improvement, *P5* used it for accountability:

We want to know whether they are really following the plan, the recommendations, because sometimes some patients say, “Oh, I’m not getting better, I’m doing everything.” Well, what is everything? ... besides giving me information for diagnosis, treatment, and plan management, I think when they keep a diary, it also gives them accountability.

Outside of the patient-provider context, people also want to share data with others. Of 106 survey participants who had used apps, 23 mentioned that apps lacked support for social sharing. Similarly, 9 of 13 interview participants expressed a desire to be able to share data with friends or family (*I1, I3, I5, I6, I7, I9, I11, I12, I13*) and 5 of 13 with a boss (*I1, I2, I7, I11, I12*). Some already shared their data, despite the lack of support. For example, *I9* tracked her migraines on a Google calendar she shared with her husband “*so he could see when I did have them and how I was feeling*”.

Data sharing was motivated in part by a hope that people would better acknowledge their symptoms (e.g., family, friends, coworkers, providers). 3 participants shared data to prove their symptoms existed and were as bad as they thought (*I1, I5, I12*). *I1* started getting migraines when she was a teenager, and felt that “*nobody believed me. They thought I was just being a melodramatic teenager with headaches who didn’t want to go to class or something*”. She therefore started tracking to “[*be*] able to say, ‘No, look, here’s documentation of the last six months of my life that have been horrible. Please help me’”. Similarly, 3 interview participants mentioned they shared in order to prove to people that they were actually taking action to try to manage symptoms (*I1, I11, I12*). For example, *I12* shared his data with his boss, “*just to show them, ‘Hey, I’ve been struggling with this ... and this is the work that we’ve been doing with my neurologist, we’re working towards it. We’re actively working on stuff’*”. Overall, these goals illustrate myriad motivations and needs for sharing migraine data, each requiring different design considerations.

Tracking for Goals that Evolve Over Time

Whether, what, when, and how people want to track often depends on their current understanding of their migraine and whether they consider their situation stable. 10 of 13 migraine interview participants discussed how their tracking changed over time (*I1, I2, I3, I6, I7, I9, I10, I11, I12, I13*). Participants often initially tracked intensively with a goal of better understanding migraines (i.e., Goal 1), then transitioned to more minimal tracking with goals focused on management (i.e., Goal 2, Goal 3). For example, *I2* originally had a goal of determining her symptom contributors. She therefore spent three months using a custom spreadsheet to track everything she thought might be relevant (e.g., activities, environmental factors, foods, mood, sleep). After identifying some high-likelihood

As goals change, transitions in tracking strategies are often motivated by tracking burdens. To be prepared to answer questions that may arise, some initially wanted to track as much information as possible about symptoms, treatments, and potential contributors. However, such intensive tracking is generally unsustainable, and 7 of 13 interview participants found their tracking experiences negative (*I1, I2, I5, I6, I11, I12, I13*). All 6 providers also mentioned negative patient tracking experiences.

In addition to direct burdens, participants described frustration with negative emotions in tracking. *I11* explained, “*You’re supposed to enter [information] everyday, which I totally get the validity of and rationalization behind, but it’s just one more thing about the day that’s dedicated to my disease and it isn’t what I want to be doing with my time*”. *P4, S107, and S140* further described a concern that tracking can worsen a person’s overall experience with their condition, as *S140* described:

when I track my pain closely, I pay more attention to it, and I’m less happy. I stopped my migraine diary after a couple of years and I honestly think ceasing to consciously observe my pain level every minute of the day made me feel better.

Some participants feared that migraine tracking could itself be a contributing factor. *I7* commented: “*The more I was focusing, tracking, journaling, the more migraines I was getting*”. When participants suspected this link, they often wanted to abandon tracking, despite potential benefits.

Given this combination of burden and negative experience, all 6 providers felt patients should track according to their goals, rather than indefinitely. *P3* explained “*we recommend they do ongoing tracking, as long as they’re working with us. But once you’re in a stable point, it’s probably less important to keep tracking unless it’s helpful for the patient, because it can become a burden*”. *I6* similarly motivated a need to support lapsing, resumption, and abandonment in migraine tracking:

Having it so you don’t feel like when you have a gap in your data, you have just start from zero [is helpful]. Whatever [app designers] can do to kind of make it easy to fill in those gaps makes it less intimidating to get back in there.

Despite concerns for minimizing tracking and its burdens, participants also described appropriate tracking as an important part of maintaining motivation (i.e., Goal 4). 4 interview participants (*I1, I2, I6, I7*) described tracking as empowering, especially over the long term, as *I1* explained:

Just being able to look back at that however many years of journals and calendars stacked up and see that I'm not in the pain that I used to be and that I can live my life and understand the choices that I've made without being scared that I'm gonna be incapacitated the next day. That's a reward and I hope more people can get to that point.

I2 similarly found migraine-related tracking a positive experience: “when you look at the big list of symptoms, that’s not fun, but knowing that that data gets results is empowering”.

From a provider perspective, *P3* also described modifying a patient’s tracking according to motivational needs. If a patient was not seeing a benefit, she would have them “flip their tracking, and have them track good moments”. Although such approaches do not remove concerns for burden, encouraging people to see positive aspects can help them achieve tracking that supports their goals.

Tracking for Varying Personal Experience

Within and across tracking goals, varying personal differences require the ability customize tracking to the individual. 6 of 13 interview participants (*I1, I2, I3, I4, I10, I11*) and 5 of 6 provider participants (*P2, P3, P4, P5, P6*) discussed the wide range of potential experiences with migraine. 72 of 106 survey participants who had used apps and 11 of 13 interview participants expressed a desire for apps to better support customization (*I1, I2, I3, I4, I5, I6, I7, I9, I10, I11, I12*).

A key dimension of customization is *what* a person tracks. For example, *I1* explained she did not like using apps to record her migraines because she “[doesn’t] get all the symptoms that other people do”. *I11* explained a similar frustration with journaling data that she knew was not relevant for her:

I have been doing this for 20 years now. There are some things I can just cross off my list ... say, whether I ate onions or something? ... I know this isn't a trigger. It's not a problem for me, can I just stop answering questions about onions?

Although customization can allow adding or removing factors to track, defaults and explicit support for particular factors will still affect what people consider potentially relevant. *I3* explained how an app's defaults influenced what factors she believed could be important:

It makes sense that they would have a limited option of things to choose from, and then allow you to add more things. But it would be nice if that list was a little bit longer, because sometimes you're not sure if something you're feeling or something you did is relevant to the migraine. Or it may not occur to you, if you don't see it in a list, that it would be relevant.

A second key dimension is *when* a person tracks, specifically whether they prefer tracking *during* or *after* migraine symptoms. Of 13 interview participants, 7 reported tracking either during or after a migraine (*I1, I2, I4, I6, I7, I9, I19*), 4 reported only tracking after a migraine (*I3, I8, I11, I13*), and 2 reported only tracking during a migraine (*I5, I12*). Some people cannot track during a migraine: 36 survey participants mentioned having trouble using apps because interacting with a screen during a migraine was painful. *I4* described how her migraines impacted her cognitive ability to the point that data tracked during a migraine was often useless: “*The cloudy head makes it very difficult to put words into focus, I guess, and if I [tracked] during a migraine I’d look back and say, ‘What the hell was I talking about?’*”. In contrast, *I5* always wanted to record that she was taking medication during a migraine, before actually taking it. She explained that tracking during a migraine was necessary because it helped her remember to record the data:

Usually ... when I’m starting to feel better, there are things [I had been unable to do] that I would want to do. It’s often that I’d forget to write [the fact that I took the medication] down ... Record keeping wise, it was more accurate for me to write it down before I took the medication.

Because *I5* needs to track during migraines to ensure reliable data, she was wary of apps, as she feared they would worsen her migraines. She therefore needs to be able to record data without interacting with a screen, thus highlighting a third key dimension of flexibility in *how* people track.

Data in Communication and Collaboration with Providers

As I mentioned, people with migraine often wanted to share tracked data with their providers. Of 47 survey participants who reported bringing data to a health consultation, 22 felt that having concrete data was helpful. Similarly, 12 of 13 interview participants (all except *I8*) said tracking helped their consultations with providers. 10 of 13 interview participants mentioned tracking, and the resulting data, helped them remember their experiences (*I2, I4, I6, I7, I8, I9, I10, I11, I12*), and 6 felt it helped them more effectively communicate experiences to a provider (*I1, I2, I6, I7, I9, I10*). *I6* explained:

Really what I needed was a tool to talk to my doctor about, because it's really hard to just go in there and be like, "I have them a lot." When you speak in vague terms, you get vague treatment. That's essentially what I needed, was an empowerment tool to go in and say, "This is frequency. I need to know what treatment's doing. I need to know what's working, what's not".

I7 similarly saw benefits of tracking, even when data was not directly shared with her provider:

I'll bring out the report that I printed out in with me, and even if [my health provider] doesn't ask to see it, I feel like it's good for me to have it there in case I forget anything. Mostly it's just helping me check on myself so that I don't forget something important, or exaggerate something wildly, when I'm like, actually it wasn't that bad. Yeah, it's just like having a set of checks and balances for me, by me.

Providers also appreciated their patients self-tracking. All 6 providers we interviewed reported they had recommended a patient track their migraines. 4 of the 6 provider participants said self-tracked data can help communication and collaboration (*P1, P3, P5, P6*). All 6 providers said they wanted to be able to see a patient's data in some circumstances. *P3* explained:

I think it's helpful for everyone to be on the same page, and the same understanding. It's a way to avoid misunderstanding, because a lot of time people will be talking about their headaches. Let's say they're talking about their migraines, but really they're just

talking about their most severe migraines, and you're making this assumption that the five days a month are the only headaches they're having. When really they're having daily headache, it's just these five migraines that are really bothering them ... I feel like [having the data] takes those assumptions out of the way.

However, when patients and providers attempt to collaborate around data in a clinic appointment, they often struggle to interpret it. Of 47 survey participants that reported sharing data with their provider, 6 noted the volume of data made it difficult. S136 explained: “*Because it was so dense and not in an easily digestible form like a chart, I don't think it was very helpful in our appointment*”. All 6 providers reported they had struggled to interpret data in a clinic appointment, and 4 additionally said they did not want to see patient-generated data in some circumstances (P1, P2, P5, P6). These circumstances included patients bringing an overwhelming volume of data and patients bringing data the provider is unprepared to or uncomfortable with interpreting (e.g., food-related data).

Although consistent with prior findings that patient-provider collaboration to interpret data can be difficult (e.g., [42,215]), migraine data magnifies these difficulties due to: 1) the many factors that are potentially relevant to an individual's symptoms, and 2) the fact that those factors are generally siloed in their own app or other tracking method. S198 explained that compensating for siloes is time-consuming and burdensome: “*I usually write things out ahead of time because the reports I want aren't quickly accessible or I have to wade through info and pick out what I need*”. 6 of 13 interview participants similarly described challenges in integrating their data (I1, I6, I7, I9, I10, I11), with I11 saying, “*Having to coordinate all the apps together and then when I'm at my headache appointment, I feel like I have to take 20 minutes before I go and ... copy all of the data out long-hand so that I have it all pieced together*”. Some participants attempted to address such integration challenges by using calendars or custom spreadsheets so they could track exactly what they wanted. However, 5 of 6 providers (P1, P2, P3, P4, P5) expressed that such custom spreadsheets were particularly difficult to interpret during a clinical consultation. Current approaches and tools for capturing and integrating the wide range of potentially-relevant migraine factors can therefore create challenges in effectively communicating and collaborating with providers around that data.

Of 47 survey participants who had shared data with their provider, 13 felt their provider did not engage with their data as much as they hoped. A common breakdown was for patients to expect providers to directly review data. In part because of the challenges discussed above, providers often preferred patients interpret data themselves, on their own time. *P2* described trying to avoid misunderstanding by telling her patients, “*this is for you. This is you trying to do the detective work of figuring out what’s going on here. You’re going to be the best person to figure this out. I’m not going to be the best person to figure this out*”. Consistent with prior research finding patients and providers value the complementary expertise they each bring to interpretation of patient-generated data [215], 3 providers described patients as experts in their own data (*P1, P2, P4*).

In addition to aligning migraine *tracking* goals, successful collaboration around migraine requires agreeing about overall *management* goals. *I2* described a misalignment with her provider wherein both sought to use data for understanding (i.e., Goal 1), but within different overall management goals:

[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 to me”.

The complexity of migraine data and the variety of goals people bring to that data therefore present additional needs in self-tracking and patient-provider collaboration.

5.1.3 Discussion

Our findings illustrate how people navigate their goals for tracking migraine-related data, building upon prior research in migraine [190], personal informatics (e.g., [39, 66, 109, 133]), and patient-provider collaboration with patient-generated data (e.g., [42, 215, 251]). With our focus on the role of self-tracked data in migraine management, some of our results provide additional insights and detail within themes identified by Park and Chen [190]. For example, they characterize trigger identification as an important challenge, within which we provide greater insight into specific barriers and design opportunities (e.g., emphasizing an accumulation of contributing factors rather than any single trigger, data integration challenges presented by using multiple apps to track different

migraine-related data). I also surface data-centric insights that are distinct from the themes identified by Park and Chen (e.g., the importance of many different “before and after” questions in migraine management, roles for data in motivation and accountability in patient-provider collaborations, design challenges resulting from a person’s migraine-related goals evolving over time).

Informed by our findings, I discuss additional design opportunities for personal informatics tools to support people with migraine, organized into themes of: 1) facilitating learning to avoid common pitfalls; 2) supporting customization and flexibility; 3) accounting for burden, negativity, and lapsing; and 4) supporting management with uncertainty. These design opportunities also warrant exploration in other conditions with unpredictable, intermittent, and poorly-understood symptoms.

Facilitating Learning to Avoid Common Pitfalls

People often track to develop personalized understanding (i.e., Goal 1), but lack necessary knowledge to identify potentially-relevant factors. Absent guidance on how to systematically answer questions, many people reach conclusions that are dubious, that lead them to needlessly avoid things they enjoy, or that are otherwise problematic (e.g., *PI*’s concern for dysfunctional eating facilitated by a spurious association resulting from an individual relating gluten to their symptoms after a single episode). Park and Chen emphasized a need for education regarding potential triggers of migraine symptoms, and integrating learning into health self-tracking tools is an important theme of prior research (e.g., in making healthy decisions throughout pregnancy [193], in making healthy interpretations and decisions around sleep tracking [201], in maintaining a healthy diet while identifying and avoiding foods that trigger symptoms in irritable bowel syndrome [215]). Our research extends these ideas by emphasizing key priorities for migraine learning in the context of self-tracking and resulting data.

Prior to using self-tracking data to robustly answer questions about their migraine, individuals must address two prerequisites: 1) controlling acute medication usage (i.e., over-the-counter and prescription pain relievers), and 2) understanding that migraine symptoms generally result from an accumulation of factors, rather than a single trigger. I have noted the first is important because overuse of acute medication often causes symptoms, leading to a cycle in which a person takes more medication, leading to still more migraines [142]. Tools should prioritize support for learning

about medication overuse, and could support tracking an acute medication “budget”, but must also ensure this support does not create unnecessary burdens for people who already understand and have addressed this aspect of their migraine management. Designs should similarly emphasize accumulation of multiple factors, rather than overly emphasizing efforts to identify a single “trigger”. An emphasis on accumulation also enables learning about health behavior change (e.g., healthier eating, increased physical activity, better sleep), which our providers described encouraging as a means to improve an individual’s overall resilience to factors that contribute to their symptoms.

Failing to address these prerequisites can easily lead to data being misinterpreted (e.g., attributing migraines that result from medication overuse to other spurious correlations, identification of spurious factors that happen to occur at a time when other factors accumulate). Growing interest in techniques for individualized understanding (e.g., [53, 54, 109, 110, 132]) fosters a desire to extend approaches from specific domains into general-purpose tools. Our results balance this desire by highlighting a need to ensure tools account for condition-specific challenges and pitfalls. A general-purpose tool that does not account for medication overuse, or which emphasizes individual factors instead of accounting for their accumulation, could easily confound or mislead in the context of migraine.

Supporting Customization and Flexibility

Within and across tracking goals, tools need to support customization and flexibility in what, when, and how people track. In deciding what to track, appropriate default options are important, in part because they influence what people consider potentially relevant. But the wide variety of relevant factors and personal experiences mean many people will not want to track all default options and may have other data they do want to track. Additional challenges are presented by data that is siloed in different tracking tools, or data that must be abandoned when a person changes tracking tools due to a change in tracking goals. Similarly, people with migraine need flexibility in *when* they track, so tracking can be adapted to their needs and symptoms. Tracking designs often emphasize tracking “in the moment”, but such a requirement is inappropriate for many people with migraine.

People with migraine often turn to paper-based tracking rather than tracking via an app. In addition to allowing greater flexibility, paper-based tracking does not require interacting with a screen during a migraine, which participants described as important. Unfortunately, resulting data is difficult to integrate, analyze, and interpret. Hybrid solutions, such as tracking on paper and later digitizing data, could allow benefits from both methods. Paper forms could be customized and printed according to an individual's current needs. Research in low-resource environments has examined using phones in digitization of paper forms (e.g., [37,58,189]), and technology-centric research has explored augmented paper (e.g., [11,86,136,256]), but potential needs and opportunities for hybrid paper-digital tracking are largely unexplored in personal informatics research.

Support for customization must also consider provider needs, including their goals, limited time, expectations that patients lead the interpretation, and difficulties when interpreting data in formats that vary across individual patients. One possibility may be to define common or important questions (e.g., the various “before and after” questions), then develop tools for quickly curating a person's data into a more standard format for answering these questions. Such tools might be used prior to a clinic appointment, either because a person intends to share the resulting data and visualizations with their provider or because they want to review the data themselves prior to the appointment.

Accounting for Burden, Negativity, and Lapsing

Tools also need to account for burden and negativity. Tracking can impose many burdens [51,63], and participants described additional negativity (e.g., that tracking was a constant reminder of difficulties of their health condition, that stress associated with tracking could contribute to symptoms). Even when a person's information goals suggest “more data is better”, tools must support people in balancing this desire for data against their lived experience with tracking [66,208]. Current tools instead often assume continuous tracking and provide little or no explicit support for lapsing [46,63,66]. Research in other domains has noted tools often fail to support people as their goals evolve (e.g., healthy eating [40], financial tracking [117], menstrual tracking [65]). We see a similar need to support lapsing and evolving goals in migraine. For example, a person with a stable understanding of their migraine may stop tracking, later resume symptom-only tracking as

part of monitoring (i.e., Goal 3), and much later resume tracking behavioral data as part of asking a new question (i.e., Goal 1). Designs should support people in determining whether and what to track and should minimize any burdens or negative experiences in resumption. Designs will also need to account for such lapses in analyses and visualizations (e.g., in designing how to support “before and after” comparisons in data that includes lapsing and intermittent tracking, in historical visualizations that include many gaps and variation in what data is tracked over time). As tools explore how to support sharing migraine-related data with family, friends, and coworkers, any negative consequences of sharing will also need to be examined and addressed (e.g., embarrassment of the person with migraine, discomfort of the people with whom they share).

Supporting Management with Uncertainty

Diagnostic self-tracking often presents pitfalls to scientific rigor [39], a challenge that has motivated prior research in other domains to support self-experimentation and techniques for individualized understanding (e.g., [53, 54, 109, 110, 132]). The “before and after” comparisons found in migraine generally do not constitute rigorous experiments (e.g., they do not account for confounds), but more rigorous methods can often be impractical or unacceptable, for multiple reasons. For one, designing an experiment to robustly test a single factor can be difficult, because of 1) the wide variety of potentially relevant factors, 2) a person’s inability to control some factors (e.g., weather), and 3) symptoms generally resulting from an accumulation of factors rather than any single factor. In addition, the debilitating nature of symptoms (i.e., being unable to do the things a person wants or needs to do when they have a migraine) can mean a person is unwilling to risk inducing symptoms for the sake of a more rigorous experiment. Similarly, if a person makes a change (e.g., a new medication), and that change seems to improve symptoms, they may be unwilling to discontinue the change. When symptoms seem to be improved by changes with other benefits (e.g., increased physical activity), discontinuing for the sake of an experiment seems even more unreasonable. People with migraine will therefore often have some uncertainty regarding their personal contributors.

Personal informatics tools may be able to help people navigate this uncertainty by better differentiating the goals people bring to migraine tracking. For example, when tracking is intended to answer questions about a person's migraine (i.e., Goal 1), tools can maximize transparency in analyses, communicate underlying uncertainty, and explain potential confounds. Such tools might adopt existing representations of uncertainty for non-experts (e.g., [116]). Understanding that uncertainty, people may still want to implement a change if it has the potential to improve symptoms (e.g., a person may decide to entirely remove coffee from their diet, while understanding that correlational data behind the decision is limited). Similarly, representations of uncertainty could help people make better decisions in migraine prediction (i.e., Goal 2), analogous to results in other domains [106]. For example, tools could support a person in cost-benefit analyses of avoiding specific contributing factors (e.g., to decide whether to have a coffee today) based on their accumulation of risk and their control of those factors. Similar to how *I6* used an app to monitor the weather and then acted upon other factors she could control, tools might use a combination of self-report, modeling, and sensing across multiple factors to predict risk of symptoms (e.g., level of physical activity, menstrual patterns, sleep sensing, weather). Such tools would need to communicate their own uncertainty, but might reduce burdens and allow people to more effectively prevent symptoms. Data could also be retroactively integrated across goals (e.g., data collected as part of answering questions about a person's migraine could also later be used as part of symptom prediction and prevention). Finally, in some cases, uncertainty may be best addressed by seeking additional expert support. Tools should therefore provide support for patient-provider collaboration throughout a person's tracking, including in determining goals, in interpreting the resulting data, and in deciding on next steps.

5.1.4 Study Conclusion

We aimed to bridge a gap between the challenges people with migraine face in managing their condition, underlying self-tracking challenges in personal informatics, and patient-provider collaboration with patient-generated data. We surveyed 279 people with migraine and then conducted semi-structured interviews with 13 survey respondents and 6 health providers. We found four distinct goals that people bring to tracking and data: 1) answering questions about migraines,

2) predicting and preventing migraines, 3) monitoring and managing migraines over time, and 4) enabling motivation and social recognition. We also found needs resulting from an individual's goals evolving over time, their varied personal experiences, and their communication and collaboration with providers. I further discussed these goals and needs in terms of opportunities for personal informatics tools to facilitate learning to avoid common pitfalls, to support customization and flexibility, to account for burden, negativity, and lapsing, and to support management with uncertainty. Our research motivates design and development of new personal informatics tools for people with migraine, as well as additional research examining how these results extend to other conditions with unpredictable, intermittent, and poorly-understood symptoms.

5.2 INVESTIGATING GOAL-DIRECTED SELF-TRACKING

Our formative work revealed distinct, personalized goals people often have for their migraine tracking. To begin to examine how tools could support an individual's goals, we first reanalyzed our formative study data to identify *what*, *when*, and *how* people pursuing each goal type described tracking and wanting to track. Based on that reanalysis, we developed *goal-directed self-tracking*, a novel method to scaffold the process of *deciding* what, when, and how to track toward a specific goal, together with goal-appropriate analyses and visualizations of the resulting data. Using a paper prototype to present our design ideas for goal-directed self-tracking, we then conducted interviews with 14 people with migraine and 5 health providers to elicit initial feedback on our ideas [217].

5.2.1 Development of Goal-Directed Self-Tracking

I first discuss my reanalysis and new findings from our formative study data 5.1. I then describe my development of *goal-directed self-tracking*. Finally, I discuss our iterative design process for a goal-directed self-tracking system and present the paper prototype design we used in our interviews.

Goal Category (<i>Why</i>)	Participants	Types of Data (<i>What</i>)	Tracking Frequency (<i>When</i>)
Answering Questions	12 M (all but <i>I9</i>); 3 P (<i>P2, P4, P6</i>)	Symptoms (12 M, 3 P), Treatments (4 M, 3 P), Contributors (12 M, 3 P)	Daily (8 M, 1 P), Post Symptoms (4 M, 2 P)
Predicting and Preventing	3 M (<i>I1, I6, I7</i>)	Symptoms (3 M), Treatments (3 M), Contributors (3 M)	Daily (3 M)
Monitoring Over Time	12 M (all but <i>I8</i>); 3 P (<i>P1, P3, P5</i>)	Symptoms (12 M, 3 P), Treatments (6 M, 2 P), Contributors (6 M, 2 P)	Daily (4 M), Every 2-3 Days (1 M), Post Symptoms (7 M, 3 P)

Table 5.3: *What and when* participants described tracking and wanting to track to pursue their stated migraine tracking goal categories. *M*=participants with migraine, *P*=provider participants.

Characterizing Differences in Tracking Routines Between Migraine Tracking Goals

To investigate how a system could explicitly support the goals people have for migraine tracking, we first reanalyzed our formative study interview data. In contrast to our previous bottom-up approach focused on general practices and challenges in migraine tracking, we focused on the particular tracking habits people described as they pursued the different categories of goals. For each interview participant, we re-examined the original interview transcript to identify the categories of goals the participant described pursuing, as well as *what* and *when* they described tracking and wanting to track toward each stated goal (Table 5.3). We therefore expand on our previously-identified goal categories to characterize resulting differences in tracking routines (i.e., *what*, *when*, and *how* to track). Given the limited number of participants, these goal categories may not be comprehensive of every possible migraine-related tracking goal. However, the distinct categories allow us to investigate how to explicitly distinguish and design for the goals people bring to self-tracking for chronic condition management. Many participants described multiple goals, due to goal evolution

Supported Tracking Goals and Subgoals

1. Learning about their migraines (*usually medium effort: regular tracking of many data types*)
 - a. Learning what factors may affect their symptoms
(*i.e., identifying potential contributors*)
 - b. Learning the frequency of their symptoms or related factors
(*an alias for Goal 3 to support an alternate framing*)
 - c. Learning how a change affects their symptoms
(*i.e., a before/after test*)
 2. Predicting whether they are at risk for a migraine (*usually high effort: daily tracking of maximum data types*)
 3. Monitoring migraine-related data (*usually low effort: infrequent tracking of fewer data types*)
-

Table 5.4: Hierarchy of goals supported in the paper prototype for a goal-directed self-tracking system.

over time or simultaneous goals. Here, I briefly summarize the four distinct *goal categories* identified in our prior work, together with our new findings of *what*, *when*, and *how* participants wanted to track in pursuit of those goal categories. Table 5.4 summarizes our final set of goals and subgoals.

Answering Questions about their Migraines

“You basically turn your life into a control of an experiment and try to change one thing at a time and see what works.” (I11)

Many participants tracked to answer specific questions about their migraines (e.g., which factors contributed to or alleviated symptoms). Our reanalysis revealed participants pursuing this goal often tracked *more types of data* and *more frequently* than those pursuing a *monitoring* goal. We also found two subgoals: *identifying contributors*, pursued by eight participants (i.e., discovering what factors affect symptoms), and *before/after tests*, pursued by six (i.e., examining whether symptoms improved after a change, such as starting a new medication or eliminating a hypothesized contributor). In contrast to recent work investigating self-experimentation (e.g., [18, 53, 54, 109, 110, 132]), *before/after tests* are susceptible to confounds. However, self-experimentation is often infeasible for migraine, because: 1) self-experimentation requires contributors to be controllable and

actionable [110], and 2) due to the severity of symptoms, people are often reluctant to risk inducing a migraine in an experiment [214]. People with migraine therefore often pursue less scientifically rigorous, but more feasible, methods when attempting to identify contributors.

Predicting and Preventing Migraines

“[When I know barometric pressure is going to drop] I make sure ... that everything else in my body is as healthy as possible so that hopefully I can ride through that weather front without any other [contributors].” (I06)

After identifying personal contributors, some participants tracked exposure to those personal contributors to predict the likelihood of symptoms so they could prepare for or attempt to prevent those symptoms. Based on our reanalysis, this goal category required the *most effort*, with all participants who pursued it tracking symptoms, treatments, and contributors on a daily basis. No provider participants described a *predicting* goal, perhaps because providers generally focus on overall symptom frequency rather than the consequences of symptoms on a particular day.

Monitoring Migraines Over Time

“[my provider saw my symptom calendar] and was like, ‘That’s too many headaches ... if you’re still getting six or seven headaches a month in a couple of months, we’re gonna talk about doing something else’.” (I02)

Long-term tracking is important in monitoring a person’s condition over time (e.g., to notice changes in symptoms, to ensure management strategies remain effective). In our reanalysis, we found participants tracking towards a monitoring goal described tracking *less data, less often*. Although four tracked daily, two were also pursuing *predicting* goals, which may explain their more frequent tracking. Participants sometimes framed this goal as *learning the frequency* of migraines. To ensure we explicitly support goals in the way people find most intuitive, we therefore added a third subgoal to *answering questions: learning the frequency of migraines and related factors*.

Enabling Motivation and Social Recognition

“[A patient] might think, ‘I’m doing so horrible, I’m doing so horrible’, and then you look at it, and you’re like ‘you actually cut back your severity by 50%’ ... you kind of shift your perspective as you go on, and the new horrible becomes just the same horrible, even though it’s improved. So, it’s helpful to show that improvement more than anything.” (P3)

Participants also described seeking support as they develop personalized management strategies. Successful management often requires difficult lifestyle changes (e.g., in diet, exercise, sleep), so this goal category can help foster motivation and support needed to commit to those changes. However, our reanalysis revealed this goal was consistently *secondary* to one of the other categories. For example, many people with migraine described primary goals of *monitoring*, but also wanted to show their data to friends, family, or coworkers to share how they were doing. Similarly, provider participants were often primarily concerned with *monitoring* a patient’s symptoms and treatments to inform future treatment recommendations, with a secondary goal of *providing motivation* by showing symptom improvement. Our reanalysis therefore suggests that motivation and sharing should be available *across* tracking goals (e.g., by including shareable data visualizations, by supporting reflection on progress toward goals, by recommending and encouraging behavior changes that can reduce migraines).

Characterizing Data Elements Tracked Across Migraine Tracking Goals

Although the *types* of data participants tracked (i.e., symptoms, treatments, contributors) often depended on their goals, specific data *elements* (i.e., the *particular* symptoms, treatments, and contributors they tracked) depended more on the individual. For example, the specific contributors participants tracked varied according to their personal characteristics (e.g., people who do not menstruate do not track their menstrual cycle), their habits (e.g., people who do not consume caffeine do not track caffeine), and their hypotheses (e.g., people who suspect particular foods or drinks correlate with their symptoms often track those specific foods and drinks). Similarly, the specific symptoms participants tracked depended on personal experiences with migraine (e.g., only some people experience aura, “brain fog”, or fatigue). Our reanalysis therefore suggests that the particular

data elements people want to track are orthogonal to their goals: individuals tracking the same data element may have different goals, and individuals with the same goal may want to track different data elements. The one characteristic of data elements that did seem to correlate with goal type is the *granularity* of data: people who wanted to *learn* about an element often tracked more details than people *monitoring* that element. For example, people who wanted to learn about their migraines often wanted to track details about symptom severity and duration, while people who wanted to monitor often only recorded that they had a migraine. However, even the granularity sometimes depended more on an individual's experiences: participants who felt their migraines varied often described tracking more details, even if their primary goal was to monitor. Table 5.5 provides examples of data elements and granularities participants tracked for each data type.

Scaffolding Tracking towards Individualized Goals via Goal-Directed Self-Tracking

Based on the goals and sub-goals (corresponding to different tracking needs) and data elements (corresponding to personalized characteristics, habits, and hypotheses) identified in our reanalysis, we next formulated design requirements for *goal-directed self-tracking* tools. These tools should scaffold the process of deciding *what*, *when*, and *how* to track toward one's goals and provide goal-appropriate support for *interpreting* resulting data. We refer to this new method of explicitly eliciting and supporting an individual's goals throughout all stages of self-tracking as *goal-directed self-tracking*.

Although many participants pursuing a given goal category described tracking similar types of data with similar frequency, some had divergent tracking practices (e.g., the majority of participants with *monitoring* goals tracked only when symptoms occurred, but some chose to track more regularly). A requirement for goal-directed self-tracking tools is therefore to *recommend* what, when, and how to track toward goals, while providing *flexibility* to support people in tracking however they feel will be useful for their goals. People can therefore receive guidance in configuring appropriate tracking routines for their goals while avoiding pitfalls in which tools: 1) do not support tracking data a person wants, or 2) require tracking data a person does not want, both of which can lead to abandonment [63,66]. Tools should then also provide goal-appropriate analyses and visualizations, helping people and their health providers avoid pitfalls in interpreting collected data [39,42,64].

Data Type	Example Data Element	Example Granularity
Symptoms	Migraine	Binary Duration Location Severity from 1-10 MIDAS score (i.e., a validated scale to characterize severity [229])
	Aura	Binary
	Nausea	Binary
Treatments	Medications	Binary (i.e., for specific medications) Number of pills
	Supplements	Binary (i.e., for specific supplements)
	Yoga	Binary
	Mindfulness	Unspecified
	Exercise	Number of steps Fitbit reports Minutes
Contributors	Caffeine	Binary Number of milligrams
	Menstruation	Using a menstrual tracking app
	Diet	Everything they ate Whether they ate a specific thing (e.g., corn, dairy, wheat, tannins, nitrates)
	Sleep quality	Using a Fitbit Scale from 1-10

Table 5.5: Example data elements and granularities formative study participants described tracking for each data type. These factors were generally orthogonal to a participant’s goal.

Participants also described goals that varied over time, consistent with existing models of self-tracking [66, 179]. A second requirement for goal-directed self-tracking tools is therefore to support an individual's specific circumstances and tracking needs over time. Tools can represent a person's tracking needs as a set goals the person has explicitly selected and configured, and the person's total tracking routine is then the combination of tracking they have configured for each goal. As a person's understanding and goals evolve over time, tools can support adding, modifying, or removing goals. A goal-directed self-tracking tool can thus be adjusted to a person's evolving needs over time, rather than requiring the person seek out new tools as their needs evolve.

System Design

To investigate our ideas for goal-directed self-tracking, we created preliminary designs for a goal-directed self-tracking system that supports the categories of migraine tracking goals identified in our formative study 5.1. The design was initially expressed through a flowchart representing goal selection, configuration within each goal, and post-configuration features. We iterated on the design in pilot studies with four people with migraine, identifying problems with flow or clarity as well as gaps in supported features. Pilot testing revealed a need to enhance each goal category with educational information, examples, and decision-making support, as individuals may not know what goals are possible or understand which best describe their personal goals. We also added a way to express a goal in a person's own words *before* selecting and configuring based on a goal category, to encourage open-ended consideration of *why* a person wants to track migraine-related information.

After our initial iterations, we designed a paper prototype to elicit needs, ideas, and feedback (see Figure 5.2). Such low-fidelity prototypes can elicit valuable feedback about content and sequence of steps before investing in higher-fidelity prototypes [209]. We conducted pilot studies with five colleagues in human-centered design and engineering, computer science, and medicine to iterate on the prototype, again focusing on improving clarity and flow while adding or expanding on features.

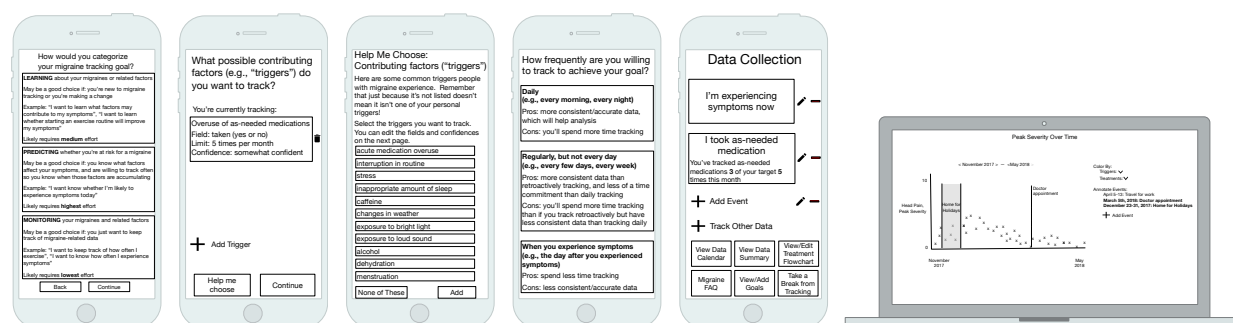


Figure 5.2: Our paper prototype for a goal-directed self-tracking system walks people through selecting a goal, configuring tracking towards that goal, and collecting and interpreting data for that goal.

Initial Goal Selection

The first step in the paper prototype is to add a goal. A person begins by stating why they want to track migraine-related data in their own words. They then select the category that best describes their goal from the list of supported goal categories (shown in Table 5.4). As described in Section 5.2.1, *Goal 1b* and *Goal 3* are two common ways we found formative study participants expressing the same underlying goal. To help people select the most appropriate goal, each category includes a brief explanation, specific examples, and an indication of likely relative effort.

Goal Configuration

After selecting a goal category, the prototype walks the person through selecting *what* they want to track for that particular goal. For most categories, they configure any *symptoms*, *contributors*, and *treatments* they want to track. Data types are ordered and recommended based on their importance to that particular goal category. For example, for monitoring (*Goal 3*) goals, *symptoms* are the only data type the prototype assumes the person will be tracking, and are always configured first. For identifying contributors (*Goal 1a*), the prototype assumes the person will be tracking potential *contributors* and configures those first. For goals related to learning how a change affects

symptoms (*Goal 1c*), people only configure the changes they want to make and the symptoms they want to examine with respect to those changes (i.e., without configuring other *treatments* and *contributors*, which would be irrelevant to a goal focused on specific changes).

A person can configure data manually or via *Help Me Choose* pages that list common migraine-related data. They can also configure semi-automated tracking routines [38], including data automatically synced from other devices and services (e.g., Fitbit, local weather). When configuring *contributors*, the app recommends tracking *overuse of as-needed medication*, a common and often-overlooked migraine contributing factor [142]. We included this recommendation because experts often recommend tracking as-needed medication overuse, as controlling medication use is often necessary to reduce migraine symptoms [214]. A person also configures *how* to track by specifying fields (e.g., binaries, categorical scales, numbers, text). They finally select *when* to track, configuring notifications to either prompt them to track at fixed intervals (e.g., daily, weekly) or to remind them to track additional information after any indication of symptoms. Such post-symptom tracking enables people to quickly indicate presence of symptoms without completing the full tracking routine, helping them minimize screen time while remembering to add necessary information for their goal after any light sensitivity has passed (e.g., the day after indicating symptoms). If the person has multiple migraine tracking goals, they can add additional goals after configuring the first.

Post-Configuration App Features

Although we focus on goal-directed tracking configuration, we also wanted to explore possible post-configuration features. After a goal is configured, the prototype includes:

quick trackers on the landing page to allow quick indication of symptoms or medication use;

data collection, customized according to the configuration;

data views, including a calendar and data summary, to quickly review the data;

a *treatment flowchart*, added because pilot study participants expressed difficulty deciding what treatments (other than as-needed medications) to pursue when experiencing symptoms;

an *FAQ page*, to provide general information about migraine;

a *goal configuration page*, for adding or removing goals and viewing past goals; and

a *lapsing* page that enables taking explicit breaks from tracking, with optional support for recording a reason and/or scheduling a resumption date.

Finally, we included several web-based goal-specific analyses and visualizations to support appropriate interpretation and collaboration: calendars and graphs over time for monitoring (*Goals 1b and 3*); charts that state and illustrate correlations between symptoms and potential contributors for learning contributors and predicting (*Goals 1a and 2*); and charts that examine symptoms before and after a change for learning how a change affects symptoms (*Goal 1c*).

5.2.2 Study Methods

We conducted interviews using our paper prototype to elicit feedback on our scaffolding of goal-directed self-tracking. Our study was approved by the UW Institutional Review Board.

Participants with Migraine

We recruited 14 participants with migraine via a website dedicated to connecting health-focused research teams to members of the community, as well as via email lists and fliers posted in medical centers. All participants self-identified as having migraine, and all either tracked data related to their migraines or had considered doing so. 12 participants were female, a gender distribution that reflects migraine diagnosis [200]. Participant ages ranged from 20 to 61 ($\bar{x}=34$).

We designed a semi-structured interview protocol to investigate whether and how our system could support an individual's migraine-related goals. We first asked about migraine and tracking experiences, then worked through the prototype to configure a tracking routine based on the participant's selected goal. Finally, we asked about the system as a whole, including whether and how they anticipated it would support their goals and how they would feel about using it collaboratively with a

Provider	Areas of Practice	Position
P1	Headache Clinic	Nurse Practitioner
P2	Headache Clinic, Neurology	Fellow
P3	Headache Clinic, Neurology	Neurologist
P4	Neighborhood Clinic	Physician Lead for Clinical Informatics
P5	Headache Clinic	Fellow

Table 5.6: Affiliations and positions of provider participants in the paper prototype study.

provider. Following the blank page technique [231], participants were provided with blank phone and laptop “screens” on which they could sketch their ideas. Sessions were approximately 90 minutes, compensated with a \$30 Amazon gift card. Here, I refer to these participants as *M#*.

A third-party service transcribed the audio recordings. To analyze the data, we took a bottom-up approach. A group of eight researchers (myself, Ravi Karkar, Sean A. Munson, Ridley Jones, Yoojung Kim, Sam Kolovson, Esther Chien, Daniel Epstein) first divided and read transcripts, extracting relevant notes. The researchers included a mix of people familiar with the project and people who were further removed, to solicit a range of perspectives in the analysis. Three researchers (myself, Daniel Epstein, and Ravi Karkar) then created an affinity diagram [95], iterating on themes via novel findings from the interview transcripts as well as discussion with members of the research team.

Provider Participants

We recruited 5 health providers via colleagues of the medical providers on our research team and snowball sampling (see Table 5.6 for health provider details). All had experience working with people with migraine, and all regularly recommend their patients with migraine self-track. The protocol followed a similar structure in which we asked about experiences working with people with migraine before walking through the prototype. However, instead of choosing a single goal, we discussed all goal categories to understand what they thought of the set we included. We investigated

whether providers believed the system would help people with migraine and whether providers believed it would help them collaborate with and treat their patients. Sessions were approximately 60 minutes, compensated with a \$50 Amazon gift card. I refer to the providers as *P#*.

Data analysis consisted of a similar process of conducting an affinity diagram analysis with interview transcripts, using the categories and notes from the sessions with people with migraine, adding notes that reflected provider perspectives, and iterating on the affinity diagram. As we finalized the affinity diagram, we reexamined transcripts to identify any additional data that supported or contradicted the emergent themes. Our final affinity diagram included 59 notes in 6 categories.

Limitations

Participants skewed WEIRD (Western, Educated, Industrialized, Rich, and Democratic; [93]), likely because our interviews were conducted in person with our paper prototype at our US institution. Our institution is in a city, and all participants lived in an urban environment; people in rural areas may have different experiences and perspectives [13, 151]. Care must therefore be taken when attempting to generalize our findings. Our inclusion criteria also required participants to have tracked migraine-related data or seriously considered doing so, to ensure their feedback was grounded in their experiences with tracking or consideration of tracking. Four participants had never tracked, but all had considered doing so. We therefore cannot characterize whether and how people who had not previously considered tracking would want to use such a system. Finally, our methods directly examined the *deciding and selecting* stage of the lived informatics model [66] by investigating participant migraine-related goals and their configuration of tracking routines based on those goals. We did not directly observe later tracking stages, as participants did not track according to their configured routines. However, configuration of goal-directed self-tracking prompted participants to discuss anticipating and planning for later stages of tracking. Our results therefore inform future in-the-wild deployments to further examine how goal-directed self-tracking supports goals through later stages of tracking. Similarly, our paper prototype allowed rapid examination of how its design might influence configuration of goals and tracking routines and how people envision using it, but deployment of a functional prototype would be required to examine ongoing use or non-use.

5.2.3 Results

All participants felt the design supported configuring tracking *exactly* and *only* what they needed for their migraine-related goals, and all said they would use a goal-directed self-tracking system if implemented. In this section, I describe benefits and pitfalls participants discussed encountering and anticipating while interacting with the prototype. Our results reveal the potential for such a system to: 1) elicit different types and hierarchies of management and tracking goals; 2) help people consider and prepare for all stages of self-tracking; and 3) contribute additional expertise in patient-provider collaboration. These attributes could help address disconnects between the goals people have for data in their chronic condition management and the tools they use to such goals. I report how many participants expressed particular perspectives, but due to the open-ended nature of our protocol, not every participant was explicitly prompted about every topic. Additional participants may have agreed with any discussed perspectives if asked. To avoid implying any individual did not hold a particular view by omission, I do not indicate which participants discussed which topics.

Eliciting Different Types and Hierarchies of Goals

All participants thought the goal categories encompassed their migraine tracking goals. People generally liked the idea of goal-directed self-tracking. Four participants with migraine and four providers described beginning with a goal as motivating; *M5* liked the idea of “*starting off with the goal and having something in your mind that you’re looking forward to or trying to achieve*”. Nine participants with migraine selected a primary goal of learning (eight with the subgoal of identifying contributors (*Goal 1a*) and one of before/after tests (*Goal 1c*)), three chose predicting (*Goal 2*), and two monitoring (*Goal 3*). Providers were primarily interested in monitoring (*Goal 3*). Eliciting goals as the first step in configuration also helped people develop and express multiple types and hierarchies of migraine-related goals. Specifically, they articulated differences between *management* versus *tracking* goals, *immediate* versus *future* goals, goals for *awareness* versus *taking action*, and *personal* versus *provider* goals. This section discusses these nuances.

Management vs Tracking Goals

Of the fourteen participants with migraine, thirteen discussed having migraine *management* goals, often around having fewer or less severe symptoms. Four providers said they developed management goals with patients, often including implementing behavior changes to reduce symptoms. Although most participants saw the critical path between their management goals and *tracking* goals, some initially had trouble defining specific tracking goals. When asked to state her migraine tracking goal in her own words, *M4* said, “*I don’t know. I just want to know how to get rid of them faster*”. *M1* also struggled to express exactly why she wanted to track migraine-related data.

However, explicit goal categories helped people define tracking goals that complemented their management goals. When *M4* saw the goal categories, she selected the goal of identifying contributors, saying, “*the example down below, I read that and I was like, ‘this would be very useful’*”. *M1* similarly felt the examples helped her define a tracking goal. The listed categories sometimes prompted people to *modify* their tracking goals to better fit their management goals. Four participants initially described monitoring goals, but selected the category of learning, explaining that learning about their migraines and related factors would help more with their ultimate management goals. Seven people with migraine and four providers said they would want to pursue multiple goals simultaneously. Overall, listing the explicit goal categories seemed to help participants understand how pursuing specific tracking goals could help them reach their management goals.

Immediate vs Future Goals

In addition to the goals they configured during the study, the explicit list of goals led some participants to hypothesize about goals they might want to pursue in the future. Eight people with migraine and two providers said they would want to start with one goal before progressing to another, usually because they felt success of the future goal depended on fulfilling their current goal(s). *M3* explained:

The most relevant one would be learning because my goal is trying to understand what causes them. But then once I've gotten past this stage and understood what essential triggers are or anything like that, then monitoring and predicting would play a higher role than learning.

Although most people described future goals as dependent on fulfilling current goals, *M14* defined hers based on her tracking habits. Despite an ultimate goal of learning (*Goal 1*), she selected monitoring (*Goal 3*) because she wanted to “*get good*” at a lower-effort goal before “*tackling*” a higher-effort goal. She then chose to track some contributors and get daily notifications, a routine that requires more effort than many monitoring goals, because she wanted to “*see how it went*”.

Goals for Awareness vs Taking Action

Across all goal categories, participants further described how their current goals focused on *awareness* might evolve into future goals focusing on *action*. Eleven people with migraine and three providers discussed pursuing goals focused on gaining awareness about their symptoms, habits, and behaviors so they could identify specific changes they might make in the future. For example, *M7* said, “*it would be nice to see what precursors might be related to having a migraine so that I could reduce those in the future*”. Three believed the act of tracking could help them take action; they wanted to track behaviors they intended to change because they thought tracking those behaviors would remind and encourage them to commit to the changes (e.g., avoiding medication overuse, taking recommended supplements, increasing mindfulness, exercising regularly).

Personal vs Provider Goals

The prototype also seemed to elicit differences between specific tracking goals people had for themselves and those they pursued for providers. All participants with migraine described having provider-related goals (e.g., tracking because their provider asked them to, tracking so they could more accurately answer the questions their providers asked). In some cases, participants did not consider personal goals until interacting with the prototype. *M3* commented:

I've never actually thought about ... what I want to accomplish aside from just being able to illustrate to my doctor what's been going on. But I think that having either a goal that you collaborate to create or separate goals and then come together to create one is a really cool idea.

Some providers had similarly never considered patients having tracking goals. P5 stated, *"To be honest, I have never really extensively talked to a patient about what their tracking goals are. I usually tell them 'this is my goal, at least please do this' ... I'm not sure what other goals they might have"*.

In general, people with migraine and providers thought that learning about each other's goals would be helpful. P1 considered encouraging patient goals important, because *"it can be empowering to patients to better understand what's going on with their body and to be able to recognize some of those things and not just feel like their body is out of their control"*. However, all providers wanted to ensure their expertise and advice were accounted for in the goals the patient ultimately pursued. In this paper, we refer to any hypothetical goals a provider would want a patient to pursue that are distinct from the patient's initial priorities as *provider goals*. Although the majority of participants with migraine were open to pursuing provider goals, M3 felt it would be *"frustrating"* if a provider asked her to modify her goal. Even though this was the same participant who had never considered having a personal goal, once she had identified one, she said she would *"want the doctor to take [her] concerns into creating this goal more than anything else"*. Three others similarly wanted to collaborate and understand their provider's rationale for modifying their goals. All participants thought simultaneously pursuing both goals was a good compromise. However, pursuing multiple goals can introduce additional burdens, so such a compromise could ultimately create other tensions and barriers.

Goal-Directed Tracking Across Tracking Stages

We designed our prototype to elicit and support the goals people have throughout self-tracking, so we wanted to determine whether and how participants anticipated a goal-directed self-tracking system could support the different stages of tracking described in the lived informatics model [66]. This section considers participant reactions in the context of each stage of the model. Our methods

directly examined only the *deciding and selecting* stages (e.g., goal selection, configuration of a tracking routine). Participants then anticipated and planned for later stages, discussing how they envisioned using the system after configuration. These anticipated uses and barriers inform how a goal-directed self-tracking system could better support later stages of self-tracking.

Deciding and Selecting

In this first stage of the lived informatics model, people choose why and how to track. Our prototype's configuration section scaffolds this stage: people select a goal and configure *what, when, and how* to track toward that goal. Some participants knew what they wanted to track; all independently configured at least one type of data, and two manually configured all data without looking at the *Help Me Choose* pages. However, the rest viewed and selected data to track based on the prototype's suggestions. Some selected data they "*hadn't thought of before*" (M11) but wanted to investigate upon seeing. When selecting potential contributors, M3 and M6 said they would "*just say all of them*". Others felt the *Help Me Choose* pages helped them remember factors they had already been wanting to track; M11 explained, "*if I get a question I have to give an answer just off the top my brain, that's harder than when it's multiple choice. Multiple choice is like, 'Oh yeah. It's right there.' Then you can start thinking about other ones because it's there*". M9 similarly selected alcohol because she knew she wanted to track it but had forgotten until she saw it.

Regardless of why they selected data from the *Help Me Choose* pages, participants generally felt they added more data elements than they would have without those pages. M11 felt that she "*probably would have had maybe four [contributors] compared to, well, how many of them. Eight. Nine*". Part of the reason that she ended up tracking extra information was because she felt the prototype "*makes me realize I have been slacking ... You make it look so easy*". After the configuration, all participants felt their routine had them tracking *exactly* and *only* what they wanted to track for their goal; when asked whether they had configured anything they did not feel was important, everyone responded that they had not. However, M1 said, "*some days, I feel like [the configured tracking routine] would be a little daunting*". P2 was worried that patients might select too much and end up with a burdensome tracking routine, suggesting:

I might limit someone to five categories to track at a time ... someone might be very interested in like, “Oh, yeah, I’m gonna track all of these,” and then it becomes too burdensome and they stop tracking after two or three days.

Participants were particularly likely to expand their goals to investigate more data elements than they would have otherwise configured when such additional data could be automatically tracked. All participants wanted to track some data automatically. Our designs included the ability to sync with outside data sources so people who already tracked with those sources could use their data. However, *P7* did not have a Fitbit, and said, “*if it automatically went to this, I might [get one]*”. In some cases, providers were hesitant about some of the types of data we intended to support. For example, we include syncing to local weather data because weather can affect migraines. Every participant with migraine chose at least one weather-related field to sync, and *P2* and *P5* were also interested in that information. However, *P1* and *P3* felt that focusing on an uncontrollable factor may not ultimately be helpful. Similarly, *P4* did not want people focusing on blood pressure data because “*it would create an assumption that if it’s a little bit high today, that means that your headache might be explained by that, which I think would be a false assumption*”.

Tracking and Acting

In the *tracking and acting* stage of the lived informatics model, people collect, integrate, and reflect on their self-tracked data. In our interviews, tracking routines configured by participants varied widely, with participants configuring tracking of 1-7 symptoms (*median*=3.5), 2-9 treatments (*median*=4), and 0-16 contributors (*median*=6). They generally appreciated the ability to control *how* to track by specifying fields, although all wanted the tool to recommend appropriate defaults during configuration. Three planned to track only when they had symptoms; the rest planned to track daily. Providers had different levels of interest in contributors but generally wanted their patients to focus on tracking symptoms and treatments (e.g., medications, lifestyle changes).

Although participants with migraine all reported wanting to pursue the tracking routines they had configured, they anticipated their tracking would evolve over time. As described in Section 5.2.3, some wanted to pursue future goals after fulfilling their current goals. But in addition to evolution *of* goals, people expected evolution *within* goals as their experiences and understanding changed. Four participants intentionally configured more data than they thought they would ultimately want to track, with the intention of later narrowing their routine after reflecting on their data. *M6* explained:

what I'd probably do is start off with all of these and then maybe after the first two or three days realize there was a couple of questions I'm either like constantly answering the same thing or just kind of, it makes me pause like how do I answer this? ... I suggested irritability, but I might drop that one because it [might make me] irritable.

Conversely, two participants wanted to start narrow and later widen their routines. *M13* wanted to make behavior changes her provider recommended and thought if she tried to do too many at once she would become overwhelmed and give up, saying, “*if I just did one thing and then if its good, I mean [after] two or three weeks, add something else. And if that's good, add another routine*”. Three participants mentioned feeling they needed more domain expertise, such as knowledge of what dietary factors commonly *could* contribute to symptoms, before they could focus on learning about whether those factors may contribute to *their* individual symptoms. *M2* explained, “*I don't feel like I have enough information about the contributing aspects of dietary factors to know what should I be tracking*”. Overall, the prototype seemed to help participants consider what data they could effectively track, both immediately and in the future, to best fulfill their goals. However, even with that support, some participants felt they needed access to more domain expertise to collect, reflect on, and act on the data in the context of their migraine management goals.

Participants also discussed how goal-directed self-tracking would inform how they acted on their data. Thirteen participants with migraine and one provider mentioned wanting to explicitly check in with their goals to reflect on their progress, review what they had learned, and decide whether anything needed to change. The other providers wanted to discuss the goals in an appointment and have the patients independently make any changes with the app itself. Participants also

anticipated goal-related data analyses and visualizations would support reflection towards their goals. Both patients and providers liked the calendar and data summary pages, to get “*a real quick glance at what you have been experiencing*” (M1). Some participants expected the calendar would also help them identify trends; M9 commented, “[*with the calendar*] *you could correlate, maybe if you’re stressed out at work, when you worked on Monday, and then you had this headache Monday too*”. Participants also generally liked the web-based analyses and visualizations. P1 felt the visualizations would help them notice more complex and longer-term trends, “*to see symptoms over time and to see in relation to changes or in relation to triggers*”. However, M10 was afraid they would be too complicated, and M11 thought she would not be able to interpret them without help. M5 similarly emphasized simplicity, stating, “*whatever graph is the easiest to see is the one you guys should go for. One with a lot of words and lots of axis, they’re really hard to tell at a glance*.” Many participants acknowledged the tension between prioritizing interpretability and visualizing all important data for a particular tracking goal. Some felt customizability of the generated visualizations could support exploration while limiting complexity; M7 explained, “*I would want to select from a dropdown like, ‘Here are all the things you’ve tracked. Select two or a few, or select pairs to see, to generate [a visualization]’* ”. However, M3 warned that while “*being able to directly say, ‘okay, this is the only thing I want to look at’ and disregard all the other data might be helpful*”, doing so could also “*get potentially confusing*” by hiding confounds and leading people to incorrect conclusions.

Lapsing and Resuming

People often exit and re-enter the *tracking and acting* stage of the lived informatics model via *lapsing and resuming*. However, current tools generally fail to satisfactorily support lapsing and resumption (see Section 2.2). We designed our prototype to support taking intentional breaks, with the ability to optionally configure reminder notifications to come back to tracking at a particular date. To investigate whether participants anticipated circumstances in which they might want to lapse, we asked whether and when participants thought they would use such functionality. All participants anticipated scenarios in which they would lapse in tracking their full routine. However, rather than lapsing altogether, seven participants with migraine and three providers described wanting the ability

to reduce tracking to only a *subset* of their configured routine. For example, *M1* configured an extensive routine, but wanted to be able to “*answer the questions as I feel like I have the time to do*”. Similarly, *M2* ideally wanted to track the *number* of alcoholic beverages she consumed, but with an option to just indicate a *binary* of alcohol consumption “*if number of drinks isn’t something that calculable in the moment*”. As described in Section 5.2.3, people anticipated modifying their tracking routines as their goals change; these changes might include lapsing in tracking types of data that are no longer relevant. However, they also wanted to be able to prioritize certain types of data on a day-to-day basis, depending on their experiences on those days. *M2* explained:

when I would see myself taking a break would be during finals week, or a really busy application season or something. Which, due to the nature of that might actually be when I’m getting more symptoms ... I could see myself taking a break from maybe more of like the background symptoms and ... just saying, “Okay, for this week I’m just going to track these really critical symptoms.”

Although lapsing has generally been considered an unexpected and often undesired stage of tracking, participants were able to predict situations in which they might want to lapse and to suggest ways to do so while continuing to pursue their goals. By expecting and supporting this kind of *non-binary lapsing* (i.e., lapsing by reducing tracking to only a subset of their configured routine), a system could help people configure a more comprehensive tracking routine while still providing flexibility in their lived experience. How often people would engage in this kind of lapsing should be investigated, but such functionality could support people in pursuing their goals despite varying circumstances and time constraints that might otherwise lead to abandonment.

Expertise in a Goal-Directed Self-Tracking System

We intended our design of a goal-directed self-tracking system to help patients and providers communicate and collaborate around data. However, the design also created expectations that could influence patient-provider relationships. We designed our prototype to include information about migraine (e.g., lists of commonly tracked migraine-related data, a popup explaining that

medication overuse is a common contributing factor). Because the system included this information, participants described expecting the system to contribute expertise in migraine management. In addition, because the prototype was designed anticipating patient-provider collaboration in goal setting and data interpretation, participants expected that providers would be able to help patients use the system. Participants discussed potential benefits and pitfalls of these design decisions.

System Support of Patient and Provider Expertise

All participants felt a system based on our design would support communication and collaboration around their experiences and goals. *M5* explained, “*showing a graph like this is better than me trying to explain, ‘Oh, I usually have this kind of symptoms,’ or ‘I think I have four a week.’ If [providers] can see a graph I think they’ll be able to help a lot more*”. However, *M2* was afraid the system would *replace* her role in her care, rather than supporting it: “*I would want to have the opportunity to have a discussion about [my experiences] and not just base it on what they’ve seen in the charts*”. *P2* similarly wanted to discuss data with patients, explaining: “*I always verify, just because I know that I’m not always 100% accurate when I enter something...I don’t know that we need to go through and pull it up, but to say, ‘This is what your app sent over. Does this sound right to you?’*”.

People with migraine and health providers also thought having data would help corroborate their claims to each other. People with migraine felt data would provide “*solid evidence*” (*M1*) that would help their providers to understand and believe their experiences. For example, *M9* wanted her providers to know that she was following their recommendations: “*I’m like, ‘Yes, I do all those things, but I still get migraines.’ So, yeah I would like to prove [that I’m doing them]*”. Providers similarly felt that having data would support their recommendations. For example, the page for configuring *when* to track included information explaining why people might want to track at different frequencies, and *P2* said, “*I mean obviously I like my patients to track every day, but I think this is a really good screen. I’m really glad that it’s included since this way you can get reminders and something that’s realistic for the patient and they can see the pros and cons*”. *P1* felt having data would help them show patients that certain factors were *unrelated* to their migraines

because it could disprove the relationship, explaining, *“I don’t feel like it’s always as helpful to try to convince someone that it’s not a trigger if they’re really convinced it’s a trigger”*. She also thought the data could help people understand why their treatment might not be working:

Sometimes, especially if people are having a hard time committing to a daily therapy, it’s good to see exactly how many days they’re skipping or missing. Good for them to see so they know why their therapy might not be helping, and so I can see and help them out.

System Expertise in Managing Health

Although we designed our prototype anticipating people would collaborate with providers to set goals and act on data, participants generally felt a system could and should support education to help people manage their migraine independently. All felt a system would be helpful to people with migraine even without a provider, and eleven participants with migraine mentioned they had learned about migraine during the configuration. M6 explained: *“I felt like I was already learning a little more about migraine tracking in a simple way where it didn’t feel like I was reading a textbook”*.

Five participants with migraine believed the expertise embedded in a system could help them avoid seeing a doctor. For example, M2 said, *“if I was experiencing symptoms but maybe not so much so that I wanted to go into the doctor this might help me do a mini-revision of how I’m approaching things”*. The ability to independently fulfill health goals could be helpful; however, people do sometimes need medical advice. P4 mentioned, *“I don’t mind if [patients] track things ... It’s the decision making part afterwards that is the high risk area.”*. One way to discourage people from making medical decisions without support may be to explicitly suggest when to see a doctor. Five people with migraine and one provider thought a system should tell people when to seek medical advice, and five participants mentioned wanting to be able to use it to message their provider.

Health Provider Expertise in Using a System

Despite wanting the system to contain sufficient information to be used independently by people with migraine, participants generally expected that providers would help with any issues patients encountered while using a goal-directed self-tracking system. All participants with migraine and three providers mentioned expecting providers would be able to help patients use the tool. Three people with migraine and three providers mentioned wanting to collaborate to ensure the patient pursued realistic tracking and management goals. Nine participants with migraine described wanting providers to answer their questions, make suggestions, and help them interpret their data. Participants also felt the provider could verify the information a system contained. For example, *M3* explained, “*being able to have [providers] say, ‘Oh yeah, this is definitely a trigger for you. Oh, this could be. Oh, that’s most likely not.’ could save you a lot of stress and time.*”. However, *P4* was worried providers might not be able to meet patient expectations regarding their ability to help them use a system, saying “*it is incredibly impressive how many [providers] just don’t know basic computing*”.

5.2.4 Discussion

Our findings indicate the potential for goal-directed self-tracking systems to address common pitfalls in self-tracking and in patient-provider collaboration. I discuss implications of our findings for future design and implementation of goal-directed self-tracking tools, including the characterization of the system as part of the care team, the need to support evolution between and within tracking and management goals, and the dangers of *goal creep* in configuration. Although I focus on these design implications in terms of health and wellness, I expect that these themes will transfer to other domains in which people struggle to define, pursue, or collaborate around self-tracking goals.

Addressing Common Self-Tracking Pitfalls with Goal-Directed Self-Tracking

Goal-directed self-tracking aims to address common pitfalls in self-tracking, and participants discussed anticipating and navigating some of these pitfalls when interacting with our prototype. For example, people often find self-tracking overwhelming and without purpose [63]. Even when

people have concrete *qualitative* goals (e.g., losing weight, having a healthy lifestyle), they often struggle to define appropriate *quantitative* goals (e.g., meeting a certain step count, eating a specific diet), thus preventing them from developing realistic and useful tracking routines [179]. However, participants were able to navigate different types and hierarchies of goals, understand the critical path between goals, and set appropriate *tracking* goals based on their overall *management* goals. By explicitly asking people to consider *why* they wanted to track data related to their migraine, the system helped them navigate these challenges and formulate their migraine tracking goals.

Although we were only able to directly observe how a goal-directed self-tracking system could support the *deciding and selecting* stage of self-tracking, participants also expected such a system would support subsequent stages. Participants felt their configured tracking routines were personalized to their goals and would allow them to track *exactly* and *only* what they needed to track to achieve those goals. Participants were also able to anticipate and plan for scenarios in which they would want to lapse. By providing support for explicit and purposeful lapsing, our designs helped people consider when and how they could do so without sacrificing their goals, in contrast to current tools that generally fail to account for such lapsing [66]. Participants also appreciated the idea of having analyses and visualizations that could help them interpret their data appropriately in the context of their goals, although future research is needed to examine which analyses and presentations of results would best support the questions people wanted to ask about their migraine-related data and balance tensions between simplicity and expressiveness. Overall, our findings suggest potential for goal-directed self-tracking to help people overcome common pitfalls in all self-tracking stages.

Supporting Patient-Provider Collaboration

When discussing how they would interact with and collaborate using a goal-directed self-tracking system, both people with migraine and provider participants anticipated that the system would help them verify and communicate their beliefs and experiences. Participants wanted to use a system to support their claims, but they also wanted to talk to each other to verify information the system displayed. These accounts are consistent with prior research finding that self-tracked data can ground and provide objectivity in clinic consultations (e.g., [42, 65, 215]) and can foster collaboration

by emphasizing both the medical expertise a provider contributes and the detailed experiential knowledge a patient provides [14,215]. Participants also described appreciating the opportunity to consider their collaborator's goals for self-tracked data; people with migraine often wanted their provider to help ensure their goals were realistic and useful, and providers generally felt that patients who had concrete goals would feel more empowered to manage their migraine. Patient-provider collaboration can break down when collaborators have different goals or expectations [42, 199, 202], so helping people understand and align their goals can help foster collaboration.

Although our findings suggest a goal-directed self-tracking system could improve communication and collaboration, they also reveal potential tensions. Some providers worried that patients would have unrealistic expectations regarding a provider's ability to use the system and interpret the data, echoing concerns raised in prior work [215]. Both patients and providers also wanted to make sure their goals are considered and pursued, which could prompt disagreement and frustration if patient and provider goals are not easily aligned. Although we designed our system to support simultaneous goals, tracking too much data could increase burdens and lead to abandonment [63]. Tools designed to support both patient and provider goals should therefore help people explain, align, and prioritize goals. Tools could also reveal any increase of burden associated with an additional goal. For example, most predicting (*Goal 2*) goals require strictly more data than monitoring (*Goal 3*) goals, so if a patient wanted to pursue a predicting goal and a provider wanted a monitoring goal, both could likely be pursued without additional burden on the patient. Tools could then produce analyses and visualizations for both goals, allowing the patient and the provider to review and discuss data as it relates to their goals. Pre-visit notes can also help health experts focus on participant goals and questions [43], so communicating a patient's goals *before* a clinic appointment could also support communication and collaboration to align goals.

System as a Member of the Care Team

Although people with migraine and health providers felt a goal-directed self-tracking system could support collaboration, both wanted the system itself to contribute its own expertise. Participants wanted the system to include educational information so people could independently learn about both

migraine and self-tracking. Recent work in mental health has found mobile mental health apps can feel like having a constantly-available member of a care team [21, 219]. Our findings suggest similar feelings in migraine, as people considered the system an expert they could turn to and learn from.

In some cases, our design intentionally provided this expertise. For example, we included a pop-up that recommended tracking acute medication overuse, a common contributing factor that many people find counter-intuitive [214]. However, some content was *not* designed with the intention of teaching people about migraine, and could result in misunderstandings and knowledge gaps if interpreted as educational or personalized. For example, the *Help Me Choose* pages list *common* migraine-related data. However, migraine is personalized, so some factors will be irrelevant for some people with migraine and no list can characterize all potential factors. We included a note that additional factors may be relevant, but did not remind people that some listed factors may not apply to them. A system should therefore differentiate content designed to *provide* expertise (e.g., FAQ pages, informative pop-ups) from content that *requires* outside expertise (e.g., configuration pages that assume people have knowledge of personally relevant data). In addition, although a system could better support some areas of expertise (e.g., realistic management goals), others cannot be realistically incorporated (e.g., answers to the wide variety of health-related questions an individual might have). A system therefore should clearly convey when it cannot provide sufficient expertise, so people and health providers know when they need to instead leverage their own expertise or defer to each other's.

Another key area in which a goal-directed self-tracking system should provide expertise is in self-tracking itself. Our designs conveyed some tracking-related recommendations based on past research in self-tracking. For example, when participants configured tracking frequency, the system lists pros and cons, acknowledging burdens and encouraging the minimum tracking necessary to achieve their specific goals. However, our findings reveal that both patients and providers could benefit from additional scaffolding of tracking expertise. For example, many wanted to track more data than may be feasible for long-term tracking, and people generally wanted the system to have smart defaults for *how* to track because they did not know what would reasonably balance burden and usefulness. Adding this kind of tracking expertise in the tool itself could help guide both people with migraine and health providers in decisions about self-tracking data and goals.

Supporting Evolution Within and Between Goals

Self-tracking goals are varied and tend to evolve, both in migraine [214] and in other domains (e.g., healthy eating [40], financial tracking [117], menstrual tracking [65]). For example, we previously found that people with migraine often started tracking toward a data-intensive *learning* goal before changing to a less-burdensome *monitoring* goal [214]. Many would also transition back to a *learning* goal if their symptoms or management approaches changed. Current apps generally fail to support this evolution of goals, forcing people to add or change tracking tools as their goals change. Informed by this prior work, we designed our prototype to: 1) include different categories of tracking goals; 2) facilitate differentiation between *current* and *future* goals (e.g., by explaining that successful prediction requires knowledge of one's personal contributors); and 3) support evolution between goals. However, participants identified types and hierarchies of goals that our design did not explicitly support, resulting in challenges when people attempted to express their migraine-related goals. For example, some participants initially expressed *management* goals but then struggled to define corresponding *tracking* goals. Some also worried about their ability to independently define management or tracking goals that were *realistic* or *feasible*. Additional support for defining and differentiating *management* and *tracking* goals, and for helping people define realistic subgoals, may help people better understand and navigate these different types and hierarchies of goals.

We also found the need to support evolution *within* goals, not just *between* them. For example, people often mentioned wanting to be able to modify their tracking routine as they reflected on their data and gained understanding. In addition, our configuration assumed people would define and then pursue a goal, but some participants discussed the possibility of having *retroactive* goals. For example, I discussed in Section 5.2.3 that although *M14* ultimately wanted to learn about her migraines, she selected a monitoring goal (*Goal 3*) because monitoring goals require lower effort. She then additionally chose to track some contributors and wanted to attempt daily tracking. Supposing she did track consistently enough to support learning goals (*Goal 1*), she may ultimately want to view her collected data through learning-specific analyses and visualizations, despite not explicitly selecting that goal type. If a system supported retroactive changes in goals, she could use

her existing collected data in the manner most appropriate for her goals at any given time. Similarly, our design supported *before/after tests* so people can examine their symptoms before and after making a change, but the current design assumes the change has not yet happened. By allowing people to indicate a change that happened in the past, a system could support *before/after tests* even if a person forgets to explicitly pursue that goal, or retroactively realizes they made a change. A system might even be able to automatically detect when a person's collected data could support goals they did not previously explicitly configure, giving them the option of examining their data in light of other possible goals. Considering goal-directed self-tracking not only as a method to enable simple progressions from a goal to data to analyses and visualizations, but instead as a framework for reasoning about relationships between goals and data elements, could therefore enable systems that better support data re-use, simultaneous goals, and evolution between different goal categories.

Awareness of Goal Creep During Configuration

Although goal-directed self-tracking helped participants define appropriate tracking goals, *Help Me Choose* pages also often prompted participants to track more data than they might have otherwise configured. Similarly, although participants proposed configuring a tracking routine that ensured *both* patient and provider goals were met, such compromises would likely introduce additional tracking. We do not want to *discourage* people from tracking data they think is necessary for their goals, but attempting to track too many things is a common barrier that can lead to abandonment [39, 51, 63]. Although people may naturally want to add to their tracking routine as they learn, doing so could distract from their primary goals, creating a *goal creep* that makes their goals too broad to be feasibly pursued. This *goal creep* seemed particularly likely when people configured automatically-tracked data, which they seemed to consider “free” without associated burdens. However, even automatically-tracked data can introduce burdens (e.g., the necessity to charge and wear a Fitbit, the battery life consequences of location tracking) [66, 90]. To balance the tendency people may have for adding data as they learn what could relate to their migraines, a

system should help people understand the burdens associated with that data. For example, a system could estimate how long a tracking routine would take each day and help a person prioritize and deprioritize data elements until they have a routine they are comfortable pursuing.

Deprioritizing data does not require abandoning it altogether. Many participants wanted a system to support *non-binary lapsing*, in which they could decide whether to track specific data elements on a given day depending on how they were feeling that day. A key opportunity is exploring how a system can support non-binary lapsing that encourages people to track data necessary for their goals. One possibility is dynamically defining subsets of a person's full tracking routine, allowing them to pursue as complex of a routine as their goal requires while minimizing tracking burdens on a day-to-day basis (e.g., narrowing their tracking to focus on symptoms during particularly stressful time periods). Designing for longer-term use, a system could also help people set *future* goals, starting with a limited tracking routine (e.g., as *P2* suggested) and prompting them to add other data elements they are interested in investigating over time. Conversely, a system could also reduce tracking burden by encouraging lapsing after a tracking goal is achieved. A system could thus support the positive aspects of *goal creep* (e.g., adjusting tracking routines based on educational information or greater understanding developed over time) while avoiding potential pitfalls (e.g., adding burdensome or irrelevant data to track based on overly ambitious or broad recommendations).

Goal-Directed Self-Tracking in Other Health Domains

Although we focus on migraine, the principles of goal-directed self-tracking likely extend to domains within and beyond chronic condition management. For example, a review of commonly-used diabetes self-management apps found many implicitly support some goals while failing to support others [112]. Some designs were best suited to longer-term data collection to identify trends or to use in a medical appointment, others better supported daily overviews, and a common goal of situated decision support was not well-supported by any of the apps. A goal-directed self-tracking tool in this domain could help people better understand and navigate these goals with their data. Similarly, people with Parkinson's disease often self-track for a variety of goals that may require different data and representations (e.g., to monitor and plan for the long-term

progression, to understand effects of new medications and identify any need for adjustment, to combat denial by providing an objective record) [162]. A recent study of the goals people have for self-experimentation, a form of self-tracking in which people perform n-of-1 experiments to answer specific questions, also identified distinct goals corresponding to different data collection, analysis, and visualization needs that could be supported by goal-directed self-tracking tools [216].

Goal-directed self-tracking could also reduce challenges people encounter when application designers assume certain goals. For example, some menstrual tracking apps assume a pregnancy goal, which can be frustrating or painful for people not pursuing that goal [65]. Similarly, people with chronic fatigue syndrome often use activity tracking apps with a goal of limiting their physical activity; however, those apps generally encourage them to be *more* active, shaming them for activity levels that match their goals [56]. In general, domains in which people self-track for distinct purposes that require different data elements, granularity, frequency, or analysis and visualization may benefit from explicitly expressing and pursuing goals via a goal-directed self-tracking system. Tracking not focused on a goal, or in domains where there is less variety in or evolution of goals (e.g., documentary or fetishized tracking [208]), is less likely to benefit from goal-directed self-tracking.

One important consideration in developing a goal-directed self-tracking system is to determine the set of goals that can be feasibly supported within the domain or tool. Expecting designers to implement every feature needed to support every potential goal is unreasonable, as is requiring people to take the time to explicitly customize all such features. Some goals that people would like to pursue are also infeasible given current technology and knowledge limitations (e.g., no current system could collect and model data to prevent every migraine). In addition, designing support for a specific goal can still be valuable; for example, people with IBS who think a certain nutrient may cause their symptoms and want to perform an self-experiment to establish causality would benefit from TummyTrials [109], a system designed to support exactly that goal. Designers of such a system should not be expected to necessarily include support for all other IBS-related goals. Designs that focus on a subset of goals can nonetheless benefit from a goal-directed perspective by better communicating the set of goals the tool can and cannot support. Where feasible, a tool even could suggest alternative resources for unsupported goals. This support could help people understand

the limitations of a system, decide whether the system is a match for their goals, and configure and use the system toward their goals. Such a goal-directed perspective might also help app designers attract people who will actually benefit, leading to higher overall satisfaction with a tool.

5.2.5 Study Conclusion

We developed *goal-directed self-tracking*, a new method to help people and health providers express and pursue the goals they have for chronic condition management. Goal-directed self-tracking is designed to support people in deciding *what*, *when*, and *how* to track toward specific goals and to provide analyses and visualizations of the resulting data that support those goals. We reanalyzed data characterizing migraine tracking goals, iteratively designed a paper prototype for a system to support goal-directed self-tracking for migraine management, and used that prototype to conduct interviews with people with migraine and health providers. We found that such a system could elicit many different types and hierarchies of management and tracking goals, help people prepare for all stages of self-tracking, and contribute additional expertise in patient-provider collaboration. We also found pitfalls people may encounter when relying on a system's expertise or attempting to define and navigate appropriate tracking goals. I presented these findings together with design implications for future implementations of goal-directed self-tracking systems, including: 1) the importance of considering all sources of expertise when designing a system that may be considered a member of a patient's care team; 2) the necessity of supporting evolution between and within different types and hierarchies of goals; and 3) the challenges of *goal creep* as people configure self-tracking routines and potential approaches to avoiding its negative aspects, such as enabling *non-binary lapsing* and recommending that some goals be deferred for future pursuit. Finally, I described how goal-directed self-tracking represents an important design and research opportunity in and beyond chronic condition management. We believe designers of future self-tracking tools that aim to support the specific goals people may have for their data could benefit from this explicit goal-directed perspective.

5.3 SUMMARY

Goal-directed self-tracking aims to address common self-tracking pitfalls by emphasizing the reasons a person has for collecting data, rather than focusing only on the data itself. When self-trackers fail to fully consider their goals before defining and pursuing their tracking routines, they often 1) attempt to track too much data and quickly burn out; 2) fail to track the data they need to track to answer the questions they have; and 3) struggle to interpret their data to answer their questions [39]. With goal-directed self-tracking, we hope to avoid these pitfalls by helping people define their goals; collect *exactly* and *only* the data they need to achieve those goals; and analyze and visualize the resulting data to support those goals. To develop this new method, we first completed formative work to investigate self-tracking by people with migraine, identifying distinct types of goals people often have for their migraine tracking and management. These findings directly informed our development of goal-directed self-tracking to explicitly elicit, represent, and support an individual's personalized goals. Our preliminary evaluation of our ideas suggest that goal-directed self-tracking has the potential to help people collaborate with their health providers and develop and pursue useful and actionable goals towards personalized understanding and management of their condition.

My work in migraine therefore supported people and their providers in *expressing* and *pursuing* their multiple, distinct, and evolving goals, helping them *overcome barriers to the personalized management* of their migraine by tailoring every stage of self-tracking to their goals.

Chapter 6

DISCUSSION AND CONCLUSION

My dissertation research has focused on helping people and their health providers better understand and manage their chronic conditions. I first described my research in mental health (Chapter 3), in which I investigated how a mobile DBT app could help people engage with their therapy. I found that the app supported common goals that people and health providers often have in the mental health space (e.g., improved depression, anxiety, and coping skill use). I also identified opportunities for personalized goals to better support an individual's mental health management. I then described my research in irritable bowel syndrome (Chapter 4), in which I investigated how tools could support personalized trigger identification. Our tools helped people and health providers overcome barriers to this common goal. However, individuals often wanted to pursue more nuanced and personalized goals they identified throughout the management of their condition. Finally, in Chapter 5, I described how my research culminated in the development of *goal-directed self-tracking*, a novel method designed to elicit, represent, and support an individual's goals for their self-tracking. I investigated goal-directed self-tracking in the context of migraine, finding potential for the method to support people and health providers in explicitly defining, expressing, and pursuing their personal migraine-related goals. Together, this research demonstrates my thesis:

Tools that support people and providers in expressing and pursuing their multiple, distinct, and evolving goals can help them overcome barriers to personalized management of complex, chronic health conditions.

In this chapter, I examine how the ideas and lessons learned from this work might inform future research, design, and development in personal informatics and chronic condition management. I first review next steps in goal-directed self-tracking: immediate work that would further validate

and support the method across tracking stages (Section 6.1). I then discuss future challenges for goal-directed self-tracking: higher-level research questions revealed or reiterated during the development of the method (Section 6.2). Finally, I summarize my main contributions (Section 6.3).

6.1 NEXT STEPS IN GOAL-DIRECTED SELF-TRACKING

Our initial evaluation of goal-directed self-tracking (see Section 5.2.3) revealed the potential for the method to support every tracking stage across a person’s long-term migraine management. However, it was only able to directly examine how goal-directed self-tracking can support the *deciding and selecting* stage. Validating our ideas for *all* stages of self-tracking requires implementation of a goal-directed self-tracking system to further investigate benefits and pitfalls.

6.1.1 Supporting Later Tracking Stages with Goal-Directed Self-Tracking

As a proof-of-concept for supporting additional stages of self-tracking through goal-directed self-tracking, I led the initial implementation of my proposed goal-directed self-tracking system. This implementation will enable investigations into how goal-directed self-tracking can help people collect, interpret, and act on migraine-related data to pursue their migraine-related goals. Although my designs (described in Section 5.2.1) proposed both mobile and web-based components, participants in the design study generally wanted to focus on the mobile components, often finding the web-based visualizations less accessible and potentially confusing. I therefore focused on the mobile app component, updating my designs based on my findings from the paper prototype study. To allow for participants using any kind of phone, I used the Ionic framework [77], a JavaScript-based framework for platform-agnostic mobile app development. I implemented the majority of the front-end functionality described in this section; Liwei Jiang is leading the implementation of the backend (e.g., the database, the system notifications), as well as interface improvements. We worked with Neha Kaura and Natra Zeryihun to iterate on and improve the visual and interaction design.

System Design and Development

In this section, I discuss my design decisions in my implementation of the goal-directed self-tracking system proposed in Section 5.2.1. I focus on changes I made based on my findings from the paper prototype study (see Section 5.2.3). For example, the implementation enables easy addition of multiple goals; identification of goal-based recommendations; non-binary lapsing in tracking routines; and evolution both *between* and *within* goals. I worked with Melissa Schorn, a collaborator at the UW Headache Clinic, to iterate on and evaluate these changes. Figures 6.1 and 6.2 include examples of different pages in the app. These visual designs were developed by Natra Zeryihun.

Initial Goal Selection

The biggest change to the *goal selection* stage is the ability to select multiple goals at the same time. In my initial designs, people would configure data for a single goal, adding additional goals separately to explicitly indicate which data was associated with which goals. However, participants found this distinction unintuitive; they generally thought about and expressed their goals collectively, rather than separately. Which data applies to which goals can also often be inferred from the goals themselves. In the new designs, people can therefore select and configure their goals together.

I also added three subgoals to the *monitor* goal category: *monitor for my own knowledge*, *monitor for my doctor*, and *monitor for my insurance*. These goals are designed to support people who may not know exactly what their doctors or insurance companies likely want or require (e.g., to be able to prescribe or pay for certain medications). They also aim to support people who do not currently have a personal goal and are only tracking due to a health professional's recommendation.

Finally, instead of starting with the text field to express a goal in one's own words, people now optionally include one *after* selecting their goal categories. Participants often found the initial text box too open-ended and lacking examples, and some worried about their ability to choose a feasible goal. The text box explains that the goal will not affect configuration, but is intended to help the person remember any more specific goals and share them with a doctor if they wish.

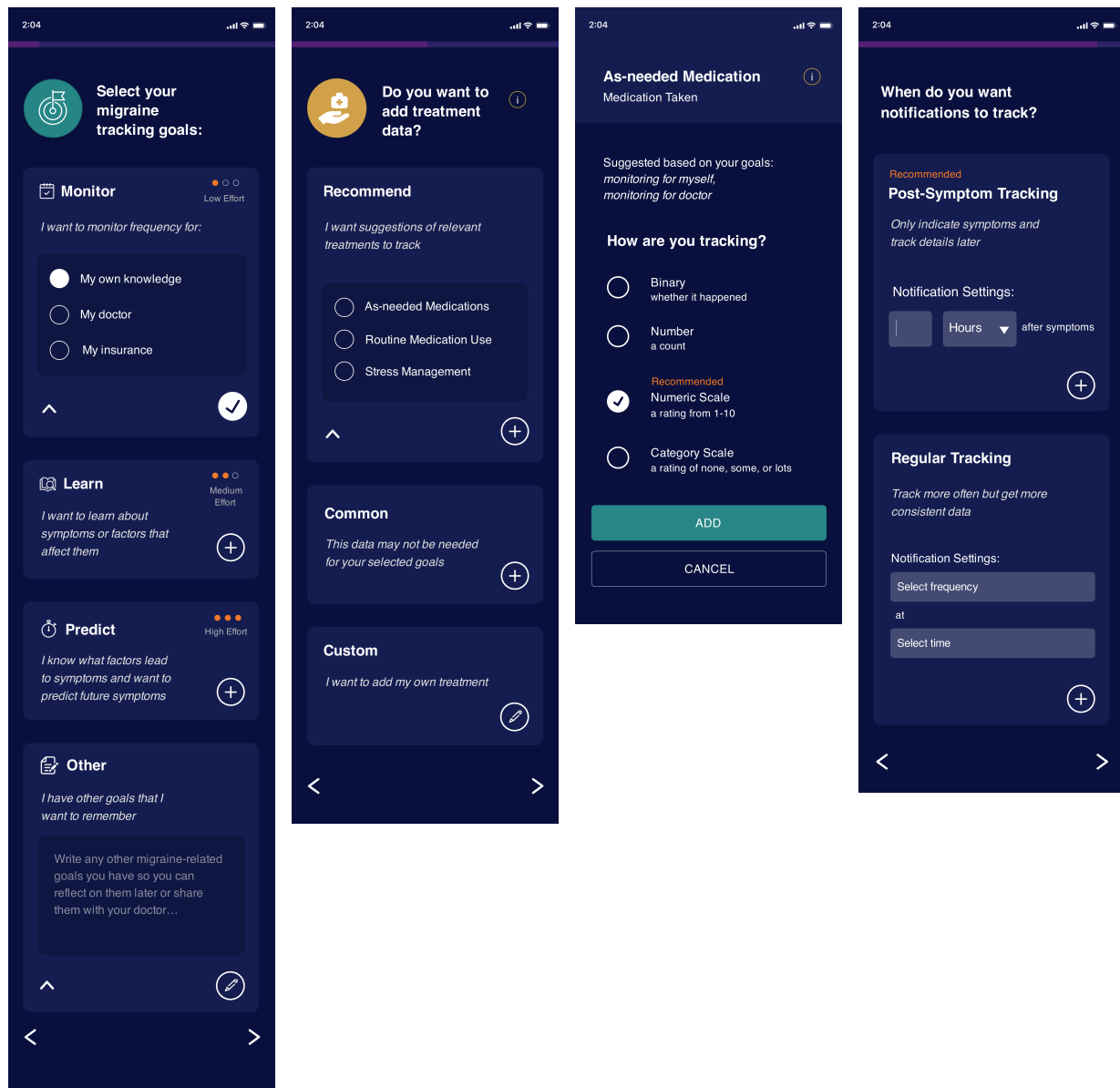


Figure 6.1: Pages from the configuration section of our goal-directed self-tracking app for migraine management. People first indicate *why* they want to track, selecting their goals and subgoals and optionally recording a written goal in their own words. They then configure *what* changes, symptoms, contributors, treatments, or other data to track for their goals, either by selecting from goal-specific recommendations or commonly-tracked data (i.e., data not recommended for their goals) or by adding custom data elements. When selecting an element, they can customize *how* they want to track it (or just continue with our defaults). Finally, they choose *when* to receive reminder notifications.

Goal Configuration

After selecting their goals, a person next configures the data they want to track towards those goals. I again separate the data into explicit *data types*: *changes*, seen only when the goal of learning how a change affects symptoms is selected (*Goal 1c*; see Table 5.4); *symptoms*; *treatments*; *contributors*; and *other information*. Treatments are no longer separated into *as-needed* and *regular*, as the distinction was sometimes confusing. The data type ordering is now consistent regardless of the selected goals, based on provider participant recommendations from the paper prototype study: they worried people would want to select less data as they continued in the configuration process and wanted to make sure they prioritized any nuances they needed in their symptoms and treatments before moving onto other data types. I added the *other* category to capture any data people might want to track that did not fit into any other category; for example, our headache clinic collaborators thought some people would want to track how effective an as-needed treatment had been.

Within each data type page, the original designs emphasized adding custom data, de-emphasizing lists of commonly-tracked data with the expectation that people would generally have a good idea of what they wanted to track. However, most participants wanted to review and select data elements from those lists. The new designs therefore immediately surface lists of data elements for each data type. I separate the lists into two categories: goal-specific *recommended* data elements and *common* data elements. I developed the lists of data elements from a combination of online research, data from my past migraine studies, and information from the Headache Clinic. I then worked with Melissa Shorn, an A.R.N.P at the Headache Clinic, to develop the goal-specific recommendations. For example, if someone selects the *monitor for my insurance* goal, the *recommended symptoms* will include symptoms insurance companies need to continue to fund certain drugs (e.g., non-migraine headaches, migraine impact on disability). If they instead select the *monitor for my doctor* goal, the *recommended symptoms* will instead include symptoms that doctors often need for their diagnosis and treatment recommendations (e.g., peak migraine severity, fatigue, pain duration). Any data element that is not recommended for the person's currently-selected goals appears in the *common* data element list. People can also still add any *custom* data elements they need.

When a person enters each data type page, the *recommended* data elements are expanded, and the *common* ones are collapsed with a message that the elements may not be needed for their selected goals. We also present a pop-up discouraging tracking of too many data elements at once if a person selects more than five on a page. These features were added to discourage *goal creep*, which could result in burdensome or unsustainable tracking routines (see Section 5.2.4).

Upon selecting a data element from one of the lists or configuring a custom data element, a person can modify how they track that element (i.e., as a binary, number, numeric or categorical scale, time, time range, or note). Based on participant suggestions in the paper prototype study, we also developed intelligent defaults for how to track each listed data element.

The final step in the configuration process is selecting whether and when to receive tracking notifications. Similarly to the paper prototype, people can select notifications for *regular* tracking, which sends notification on a daily, weekly, or monthly schedule, and/or *post-symptom* tracking, which sends notifications a set number of hours or days after symptom indication. Recommended notification frequency is again based on the goal (e.g., regular tracking is suggested for predicting goals (*Goal 2*), post-symptom tracking is suggested for monitoring goals (*Goal 3*)).

Post-Configuration App Features

After configuring goals, the home page displays the quick trackers (which now include a binary for symptom tracking and a counter for as-needed medication tracking, if the person is tracking as-needed medications). The current date is shown by default, but people can also navigate to past dates to input past data. The home page also includes entries into tracking pages for each data type the person is tracking. We separate data types into their own pages instead of just having a single entry into the full tracking routine to better support *non-binary lapsing* (see Section 5.2.3). For example, if someone only wants to track symptoms on a given day, they can directly navigate to that single page. Within the tracking page for each data type, people can either go back to the home page or progress to the next data type (to enable easy tracking of the full tracking routine when desired).

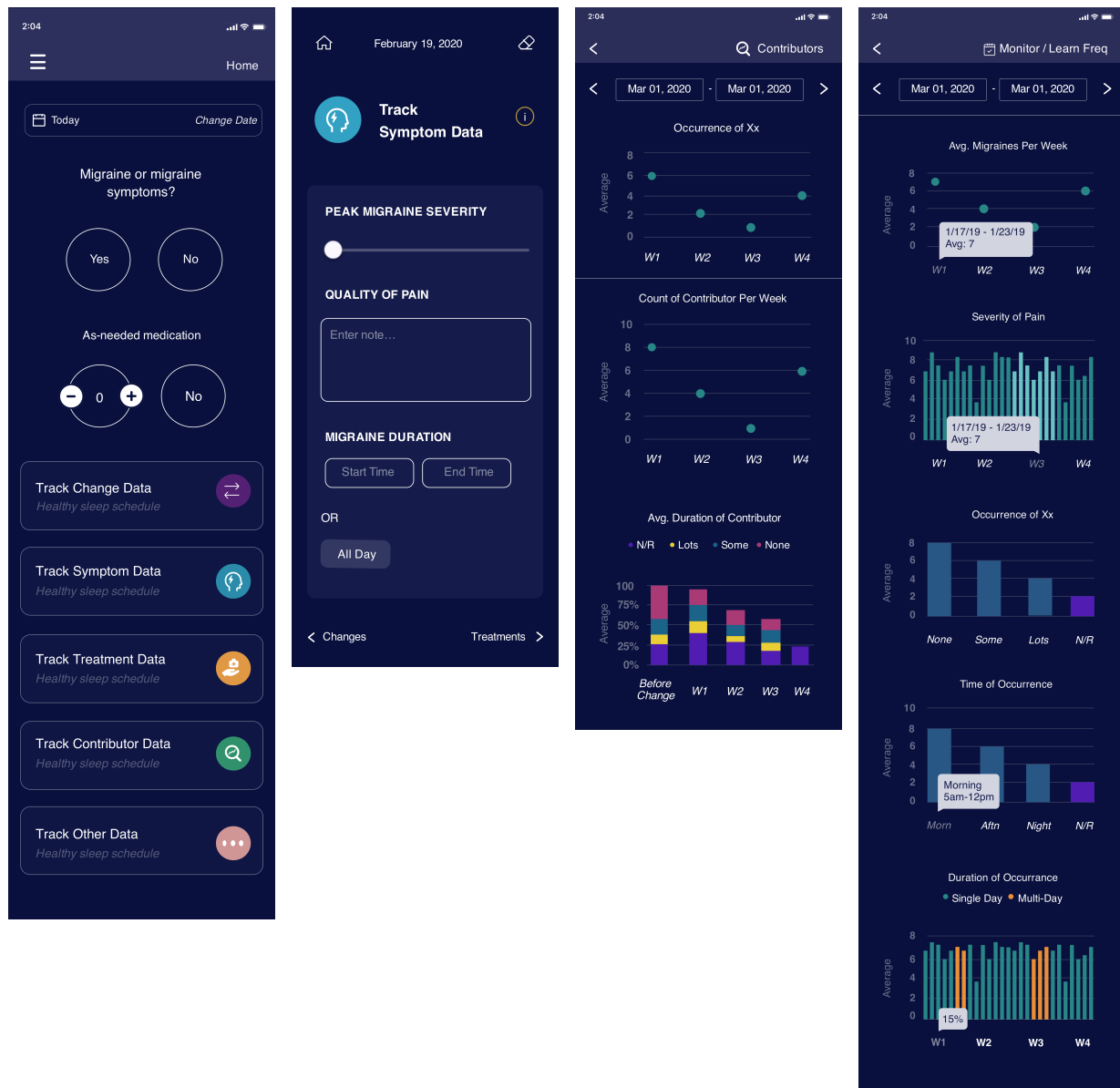


Figure 6.2: Pages from the post-configuration section of our goal-directed self-tracking app for migraine management. The home page includes functionality to quickly track symptoms (and medications, if configured), as well as entries into the full tracking routines for each data type. The app also includes distinct visualization pages for goal-directed reflection.

In addition to the tracking functionality, the app also contains many of the features proposed in the paper prototype, including an *about* page, to provide information about migraine; a *data summary* page, to summarize data in a specific timeframe (e.g., to easily tell a doctor the number

of migraines experienced in the past month); a *data calendar*, to show migraine and non-migraine days and allow easy review and editing of data reported on specific dates; pages to view and modify the current *goals* and *tracking routine*, separated to better support evolution both within and between goals; and a *lapsing* page, to enable explicit breaks from tracking (e.g., breaks of a planned duration, indefinite breaks, breaks with check-in notifications after a specified amount of time).

The current system does not include web-based visualizations; many participants did not want to have to use a computer to interpret their data, and developing compelling visualizations without any real-world data to base them on can be difficult. We therefore considered web-based visualizations out-of-scope for the current study. We do provide some simple goal-specific visualizations within the app itself. For monitoring (*Goals 1b and 3*; see Table 5.4), we visualize the tracked data over time, as well as high-level trends where available (e.g., the times of day reported for time data, the number of times each category was reported for categorical scale data). For predicting and learning contributors (*Goals 1a and 2*), we fix a graph of symptoms over time at the top of the page with graphs of tracked contributors scrollable beneath it, so people can compare how their symptoms and contributors fluctuate (inspired by the visualization approach of small multiples [242]). Finally, for learning how a change affects symptoms (*Goal 1c*), we similarly fix a symptom graph at the top with graphs of tracked changes scrollable below (all showing the average before the change, where available, as well as the weeks since the change occurred). We show the most recent weeks by default, but people can also navigate to past weeks. The feasibility study may include additional hand-developed visualizations to investigate how to best support people in interpreting their data for each goal. The data participants collect in the feasibility study will also enable development of more advanced algorithms and visualizations for goal-appropriate analyses.

The current version of the system does not include the *treatment flowchart*. Although many participants liked the idea, it was not necessary for the examination of goal-directed self-tracking. Developing compelling, personalized treatment guidance would also require additional investigations with people with migraine and health providers to ensure safety and usefulness. The flowchart was therefore also deemed out-of-scope for this implementation. Future research should examine how to help people select an appropriate treatment when experiencing migraine symptoms.

Planned Evaluation with People with Migraine and Health Providers

When the implementation is sufficiently complete, the next step will be to evaluate the system with people with migraine and health providers. The research questions will ask how goal-directed self-tracking can help people collect, interpret, and act on migraine related data and achieve their migraine-related goals. With this focus, the study will directly examine the later stages of self-tracking (i.e., in contrast to the paper prototype study [217], which focused on the *deciding and selecting* stage). The study will include both pre- and post-interviews. Pre-interviews will consist of questions about the participant's goals, and a patient and provider will work together to ensure the app is appropriately configured for the study. They will also probe for benefits and pitfalls of the configuration process (e.g., successful and failed defaults, confusing steps or descriptions). Participants will then use the app, likely for approximately one month. Post-interviews will likely resemble the approach I took with my IBS visualizations (see 4.2): people with migraine and health providers will be interviewed both individually and collaboratively to evaluate how the system could support the goals pursued by people with migraine and those emphasized by health providers. The exit interviews may also include additional visualizations to explore better support for goal-specific interpretation. A population of people with migraine that does not interact with health providers could also be recruited to enable comparison of the process when completed alone and when done in collaboration with providers. Participants will again be recruited with help from medical collaborators. Inspired by *technology probe* studies [98], the evaluation will consist of two phases. First, the app will be deployed with a limited number of participants (e.g., 1-2 providers, 2-4 people with migraine), providing information about any unexpected issues and enabling iteration on the methods, tools, and protocols. A more substantial feasibility evaluation will follow.

6.1.2 Supporting Longer-Term Goal-Directed Self-Tracking

The feasibility study will enable direct examination of the *tracking and acting* stages of self-tracking. However, goal-directed self-tracking tools should also support long-term migraine management, throughout evolving experiences and goals. For example, people may need to be able to review

and revisit past goals, and reuse any applicable past data for their future goals. People also may want guidance around goal completion and future goal recommendations. Future studies should investigate how to better support these functionalities focused on longer-term use.

One way to support this longer-term interpretation and actionability may be through Bayesian network structure learning [126]. Bayesian networks can represent correlational relationships without requiring independent variables. They can therefore model the complex relationships between different potential contributors and between contributors and symptoms. Bayesian networks are also robust to missing data (e.g., when a person fails to track, when historical data is re-used and some data elements were not previously tracked). These features make Bayesian networks a strong fit for modeling needs that evolve with a person's tracking goals. The structure learning algorithm can construct an individual's personalized Bayesian network between their contributors and symptoms, helping people: 1) identify factors that may contribute to or protect against their migraines; 2) answer more nuanced questions about their contributors (e.g., how much they can be exposed to the contributor before experiencing symptoms, how severe their symptoms will be if exposed to the contributor, how contributor interactions affect symptoms); 3) formulate useful and actionable future goals (e.g., collecting more data on a particular factor to increase confidence on whether it contributes to symptoms); and 4) reuse their data as their goals evolve.

Alex Okeson is currently investigating how Bayesian network structure learning could support people in understanding relationships between chronic condition symptoms and potentially-related factors. Incorporating such an algorithm into goal-directed self-tracking systems could enable more actionable and understandable data interpretation and better support long-term use.

6.1.3 Broader Examinations of Goal-Directed Self-Tracking

After investigating how goal-directed self-tracking could provide more comprehensive support across a person's long-term management, future research might involve larger deployments targeting health outcomes (e.g., via R21 and R01 proposals). Future work could also explore commercialization, open-source dissemination, and other potential paths to reaching more people—both within and beyond the headache clinic—to support their migraine management goals. And finally, although

this work has focused on migraine management, goal-directed self-tracking likely applies to many self-tracking domains, both within and beyond health. In addition to continuing to examine migraine management, future studies could also investigate this anticipated generalizability.

6.2 FUTURE CHALLENGES FOR GOAL-DIRECTED SELF-TRACKING

The research I suggested in Section 6.1 would enable examinations of how goal-directed self-tracking can elicit, represent, and support an individual's tracking-related goals throughout all tracking stages. However, more comprehensive technological support of an individual's health management would also require investigations of higher-level research questions in the personalized health domain. Although many of these questions have been posed in previous work, my development of goal-directed self-tracking has reiterated and emphasized their importance. In this section, I discuss opportunities for additional research in supporting holistic health management, patient-provider collaboration with patient-generated data, and artificial intelligence to further personalize health technologies.

6.2.1 Supporting Holistic Health Management

Goal-directed self-tracking aims to support people in pursuing their personal health tracking goals. However, some aspects of health and wellness management are not explicitly supported in my current designs. For example, as I discussed in Section 5.2.3, participants in my paper prototype study described a difference between *tracking* and *management* goals. A system that holistically supports health management would need to ensure an individual has defined appropriate management goals, and that their tracking goals will support those management goals. In our investigation of Bayesian analyses to answer the questions an individual wants to investigate via self-experimentation [216], we discussed ideas and concerns around eliciting testable (or estimable) questions; these concepts likely also apply to goal elicitation. For example, in some cases, a recommendation system could suggest goals that people with similar health conditions have asked in the past (e.g., suggesting someone with IBS ask the same goals that others with IBS have asked). However, as health management goals are so personal, different individuals will want to investigate different aspects of

their health. Goal-directed self-tracking systems will need to have functionality to elicit feasible and useful tracking and management goals for any individual, either through community outsourcing, collaboration with health providers, intelligent recommendations, or some combination of methods. As I discussed in Section 5.2.4, supporting any goal any individual could ever have for their health and wellness tracking and management is likely infeasible, so systems will need to provide clarity around what goals can and cannot be pursued. They should also ensure that the goals they do support are as inclusive as possible (e.g., by avoiding making assumptions about a person's identity).

In addition to *eliciting* feasible and useful tracking and management goals, a system that holistically supports a person's health management would also need to explicitly *support* both types of goals. For example, my current goal-directed self-tracking designs support people in answering myriad questions related to their health (i.e., their personal *tracking* goals). However, a person's health management goals often include not only learning the answers to their questions, but also making changes based on those answers [110]. Whether and how people could use the information gained in a goal-directed self-tracking system to improve their health and/or their quality of life is therefore a significant open question. Goal-directed self-tracking should be *empowering* to the individual using the system, so it needs to avoid any prescriptive approaches that push behavior changes a person does not want to make. For example, if a person is tracking towards the goal of identifying potential contributors and a certain factor (e.g., caffeine) is found to be a contributor, a seemingly-intuitive assumption is that the person will want to avoid that factor moving forward. For some people, that assumption might be correct (e.g., if they do not enjoy caffeine and do not feel like they need it to function). But if the person values having that factor in their life, they may only want to avoid it during periods in which: 1) avoiding symptoms is particularly important (e.g., if they have a job interview), or 2) avoiding other contributors is infeasible (e.g., during highly stressful periods). A system should support any changes a person wants to make at any time.

To support appropriate action based on information gained from goal-directed self-tracking, the system would also need to ensure the person can correctly interpret that information. Even with well-designed, goal-appropriate data analyses and visualizations, people with low data literacy will likely struggle with interpretation. As I discussed in Section 5.1.3, a system must also convey

any uncertainties or confounds associated with the data. To support interpretation, a system would likely need to include educational information about data analysis and visualization interpretation in general, as well as any necessary details concerning the specific health context (see Section 4.2.4). More research is needed to investigate how to incorporate such information to ensure that everyone can fully understand their self-tracked data and benefit from goal-directed self-tracking systems.

Finally, future work should address health management challenges at the interpersonal and societal levels. For example, people with chronic conditions often need to communicate symptoms to others in their lives (e.g., bosses, partners, children); exactly what information they want to communicate, and when and how they want to do so, is an open question. In addition, much of my work to-date has focused on supporting individuals in discovering what *they* can do to manage their health conditions; however, the attitudes and structures of society also affect an individual's health and wellness. Inspired by the social model of disability [184], additional work could investigate how societal changes could eliminate challenges and barriers people with chronic conditions face today. For example, certain types of environmental factors are common migraine contributors (e.g., bright lights, strong smells, loud sounds). Enabling people with migraine to identify other contributors could help them change their behaviors to avoid migraines even when exposed to those factors, but doing so can be highly burdensome. Instead, introducing expectations that certain environments should avoid those factors (e.g., workplaces, public spaces) could maximize quality of life for people with migraine. Exactly what role technology could and should play to facilitate the identification and communication of these challenges, needs, and opportunities should be investigated.

6.2.2 Supporting Patient-Provider Collaboration with Patient-Generated Data

Goal-directed self-tracking aims to support patient-provider collaboration by eliminating challenges patients and providers face when their goals for self-tracked data do not align. It does so by explicitly eliciting, representing, and supporting goals, which patients and providers could review together to ensure goal alignment. However, additional work may be needed to examine when and how those goals should be communicated and conciliated. For example, providers might not have time to configure goals and routines with their patients. The system could therefore allow a provider

to “prescribe” a configuration, which the patient could augment or modify on their own before discussing any changes with the provider during an appointment. Such a “prescription” could allow additional scaffolding of domain-specific expertise into a goal-directed self-tracking system, which could also be beneficial for anyone who lacks access to health specialists. However, allowing such “prescriptions” also risks prioritization of provider goals over patient goals, deemphasizing the expertise patients bring to the collaboration (e.g., in the lived experience of a health condition, in their own habits and behaviors). Patients need to feel like their tracking routine is their own, so they can exert the effort they feel comfortable exerting and pursue the goals they want to pursue. They also need to have equal voice throughout the collaboration, so they feel empowered to suggest and reject priorities and ideas. Future research should investigate how to best communicate and reconcile patient and provider goals in practice, while supporting patient engagement and provider workflows.

Systems also need to facilitate expectation management, both for patients and providers. Patients often overestimate the time and knowledge providers have to interpret their data, expecting them to be able to review the data and answer their questions. A system may be able to support providers by displaying standardized visualizations and summaries across all patients, so providers do not need to attempt to interpret a different data representation for each patient. It could also communicate what questions the data *cannot* answer. Providers often underestimate how burdensome tracking can be, expecting patients to pursue infeasible tracking routines. A system could support patients by enabling automatic tracking of any data that can be sensed. It could also communicate whether a tracking routine is realistic. By automatically supporting any needs it can and explicitly communicating limitations, a system could help minimize unrealistic expectations that often result in mutual frustration.

Finally, systems should also emphasize patient agency. Many patients in my studies described having different management goals than their providers believed they should have. A goal-directed self-tracking system could exacerbate these issues. For example, some providers may encourage the use of a daily medication if a goal-directed self-tracking system reveals it to be effective, while a patient could be uncomfortable with its cost or potential side effects. Similarly, some providers might encourage people to eliminate a factor if a goal-directed self-tracking system reveals it to be a contributor, while the patient might not be willing to do so. As predictive technology becomes more

advanced, people may be able to more easily identify decisions they can make to avoid symptoms—and then may be blamed if they do not make those decisions. Exactly what information the system shares with a health provider and how it does so could therefore influence the care an individual can receive. Careful research is needed to investigate how to ensure these systems *increase* patient agency and empowerment, despite the potential of this information to do the opposite.

6.2.3 Supporting Artificial Intelligence to Further Personalize Health Technologies

Finally, my focus on better personalizing technologies to support health and wellness would be incomplete without discussing the role artificial intelligence (AI) could and should play in future health-related systems. Recent research has motivated efforts to ensure intelligent systems are usable and ethical (e.g., [2, 114, 257]), and I see many parallels between supporting self-tracking towards an individual's health-related goals and incorporating artificial intelligence into health technologies. I presented data-driven design opportunities for using AI to support the development of positive coping skills in Section 3.2.3. Here, I take a higher-level look at how human-centered research could support the incorporation of artificial intelligence into health-related systems. Specifically, I believe that insights provided by artificial intelligence in health and wellness technologies must be *actionable*, *personalized*, and *interpretable* to ensure the resulting systems are feasible, ethical, and useful.

First, for a machine learning algorithm to be feasible and useful, it needs to generate *actionable* insights. Ensuring actionability likely requires formative research with stakeholders in the specific health context to understand what insights are needed for that context. For example, DBT often includes self-tracking of negative moods. At first look, one might therefore think people would benefit from algorithms that can automatically track those moods. However, the act of self-tracking those moods can itself promote beneficial mindfulness and awareness. Instead of just automatically tracking without supporting people in understanding and acting on the data, a better system might sense a negative mood and recommend an in-the-moment intervention to ensure people recognize their mood and know how to employ a positive coping mechanism to improve it. Without this human-centered perspective that emphasizes actionability for a health context, insights generated by machine learning algorithms are unlikely to support people and providers in that context in practice.

Although ensuring actionability within a given context is important, my work has also emphasized the need for tools to also be *personalized* (i.e., tailored to an individual's preferences, characteristics, and in-the-moment context). For example, the DBT system that detects distress and suggests an actionable skill sounds helpful in theory. However, some common DBT skills may be inappropriate or infeasible for particular individuals (e.g., suggesting that someone who cannot walk should go for a walk) or specific contexts (e.g., suggesting that someone who is at work should take a cold shower). Recommendations must therefore be personalized to the individual and their current environment. Similarly, some people may not be comfortable with an app automatically sensing their distress. People should therefore be able to personalize the system's behavior to instead elicit and rely on self-reports of distress. Research into tools that support personal health will therefore need to examine how to take this wide range of personal and contextual data into account, both in the machine learning models those tools employ *and* in the interfaces they present.

Finally, *interpretability* is critical for data and artificial intelligence in health. Interpretable AI is an open area of research in general, as interpretability can help people understand how an algorithm came to its conclusion and verify that the conclusion is a reasonable one. However, interpretability is particularly important in health, because the goal of health systems should not be to just tell people what to do. Instead, these systems should support people in identifying and implementing strategies to independently manage their health by making decisions that are most positive for them in a given context. These decisions often involve a more holistic understanding than is feasible for a system to sense on its own. For example, caffeine is common trigger for people with IBS, but even with knowledge that drinking a cup of coffee is likely to cause symptoms, sometimes people need to do so to be sufficiently alert for important events. Without interpretability, a system might exacerbate pitfalls in which participants rely too much on a system to contribute necessary expertise and end up making decisions that might not actually be ideal for them personally. A system that supports interpretability could help people understand likely consequences of their actions, empowering them to implement beneficial choices for their lives in any scenario they might encounter.

Overall, intelligent systems have the potential to be transformative in supporting personalized health, because they could help people collect and interpret data and identify patterns they may otherwise overlook. But for these systems to be feasible, ethical, and useful, they cannot be “one size fits all”, assuming what everyone in a given context wants, consents to, and needs. Careful research is needed to ensure these systems acknowledge and adjust to people’s goals, needs, and constraints.

6.3 CONCLUSIONS

Self-tracking is an increasingly common practice, but current systems often fail to explicitly support an individual’s personal tracking and management goals. In my dissertation research, I investigated how to help people and their health providers express and pursue their personal goals to better manage chronic health conditions. I investigated three complex, chronic health contexts in which symptom management depends on discovering, understanding, and acting on personalized symptom contributors. My research in mental health supported the common goal of improving personal moods, behaviors, and relationships and revealed opportunities to better support people in learning and implementing personalized coping skills. My research in irritable bowel syndrome supported the common goal of identifying personal food triggers and uncovered nuances in the goals people and their health providers may develop for their personal trigger identification. Finally, my research in migraine management investigated explicitly eliciting, representing, and supporting an individual’s tracking goals via *goal-directed self-tracking*. Together, these projects contribute novel understanding, methods, and tools for supporting people with chronic conditions and their health providers in *expressing* and *pursuing* their *multiple, distinct, and evolving* goals.

BIBLIOGRAPHY

- [1] Rikke Aarhus, Stinne Aaløkke Ballegaard, and Thomas Riisgaard Hansen. 2009. The eDiary: Bridging Home and Hospital through Healthcare Technology. In *The European Conference on Computer-Supported Cooperative Work*. Springer, 63–83. https://doi.org/10.1007/978-1-84882-854-4_4
- [2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300233>
- [3] Jessica S Ancker, Holly O Witteman, Baria Hafeez, Thierry Provencher, Mary Van De Graaf, and Esther Wei. 2015. “You Get Reminded You’re a Sick Person”: Personal Data Tracking and Patients with Multiple Chronic Conditions. *Journal of Medical Internet Research* 17, 8 (2015). <https://doi.org/10.2196/jmir.4209>
- [4] Tariq Andersen, Pernille Bjørn, Finn Kensing, and Jonas Moll. 2011. Designing for Collaborative interpretation in Telemonitoring: Re-Introducing Patients As Diagnostic Agents. *International Journal of Medical Informatics* 80, 8 (2011), E112–E126. <https://doi.org/10.1016/j.ijmedinf.2010.09.010>
- [5] Diane Andress-Rothrock, Wesley King, and John Rothrock. 2010. An Analysis of Migraine Triggers in a Clinic-Based Population. *Headache: The Journal of Head and Face Pain* 50, 8 (2010), 1366–1370. <https://doi.org/10.1111/j.1526-4610.2010.01753.x>
- [6] David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. 1997. The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology* 31, 3 (1997), 597–606. <https://doi.org/10.1177/0038038597031003015>
- [7] American Gastroenterology Association. 2002. American Gastroenterological Association Medical Position Statement: Irritable Bowel Syndrome. *Gastroenterology* 123, 6 (2002), 2105. <https://doi.org/10.1053/gast.2002.37095b>
- [8] America Psychiatric Association. 1994. Diagnostic and Statistical Manual of Mental Disorders. (1994). <https://doi.org/10.1176/ajp.152.8.1228>

- [9] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub. <https://doi.org/10.1176/appi.books.9780890425596>
- [10] Glen Aukerman, Doug Knutson, and William F Miser. 2002. Management of the Acute Migraine Headache. *American Family Physician* 66, 11 (2002), 2123–2130. <https://www.ncbi.nlm.nih.gov/pubmed/12484694>
- [11] Daniel Avrahami, Scott E Hudson, Thomas P Moran, and Brian D Williams. 2001. Guided Gesture Support in the Paper PDA. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/502348.502385>
- [12] Amid Ayobi, Paul Marshall, Anna L Cox, and Yunan Chen. 2017. Quantifying The Body and Caring for The Mind: Self-Tracking in Multiple Sclerosis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3025869>
- [13] James A Bahensky, Mirou Jaana, and Marcia M Ward. 2008. Health Care Information Technology in Rural America: Electronic Medical Record Adoption Status in Meeting the National Agenda. *The Journal of Rural Health* 24, 2 (2008), 101–105. <https://doi.org/10.1111/j.1748-0361.2008.00145.x>
- [14] Stinne Aaløkke Ballegaard, Thomas Riisgaard Hansen, and Morten Kyng. 2008. Healthcare in Everyday Life: Designing Healthcare Services for Daily Life. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/1357054.1357336>
- [15] Vicente Baos, Francisco J Ester, Avilés Castellanos, Gonzalo Nocea, Marià T Caloto, and William C Gerth. 2005. Use of a Structured Migraine Diary Improves Patient and Physician Communication About Migraine Disability and Treatment Outcomes. *International Journal of Clinical Practice* 59, 3 (2005), 281–286. <https://doi.org/10.1111/j.1742-1241.2005.00469.x>
- [16] Andrea M Barbarin, Laura R Saslow, Mark S Ackerman, and Tiffany C Veinot. 2018. Toward Health Information Technology that Supports Overweight/Obese Women in Addressing Emotion-and Stress-Related Eating. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY, USA. <https://doi.acm.org/10.1145/3173574.3173895>
- [17] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2470654.2481364>

- [18] Colin Barr, Maria Marois, Ida Sim, Christopher H Schmid, Barth Wilsey, Deborah Ward, Naihua Duan, Ron D Hays, Joshua Selsky, Joseph Servadio, and others. 2015. The PREEMPT study-Evaluating Smartphone-Assisted n-of-1 Trials in Patients with Chronic Pain: Study Protocol for a Randomized Controlled Trial. *Trials* 16, 1 (2015), 67. <https://doi.org/10/b2h5>
- [19] Eric PS Baumer, Sherri Jean Katz, Jill E Freeman, Phil Adams, Amy L Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M Olson, and Geri K Gay. 2012. Prescriptive Persuasion and Open-Ended Social Awareness: Expanding The Design Space of Mobile Health. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. <https://doi.org/10.1145/2145204.2145279>
- [20] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- [21] Katherine Berry, Amy Salter, Rohan Morris, Susannah James, and Sandra Bucci. 2018. Assessing Therapeutic Alliance in the Context of mHealth Interventions for Mental Health Problems: Development of the Mobile Agnew Relationship Measure (mARM) Questionnaire. *Journal of medical Internet research* 20, 4 (2018). <https://doi.org/10.2196/jmir.8252>
- [22] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the Computer-Patient Working Alliance in Automated Health Behavior Change Interventions. *Patient Education and Counseling* 59, 1 (2005), 21-30. <https://doi.org/10.1016/j.pec.2004.09.008>
- [23] Timothy W Bickmore, Kathryn Puskar, Elizabeth A Schlenk, Laura M Pfeifer, and Susan M Sereika. 2010. Maintaining Reality: Relational Agents for Antipsychotic Medication Adherence. *Interacting With Computers* 22 (2010), 276-288. <https://doi.org/10.1016/j.intcom.2010.02.001>
- [24] Amelia J Birney, Rebecca Gunn, Jeremy K Russell, and Dennis V Ary. 2016. Moodhacker Mobile Web App with Email for Adults to Self-Manage Mild-to-Moderate Depression: Randomized Controlled Trial. In *JMIR mHealth and Uhealth*. <https://doi.org/10.2196/mhealth.4231>
- [25] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. <https://doi.org/10.1109/TVCG.2011.185>

- [26] Anthony S Bryk and Stephen W Raudenbush. 1992. *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage. <https://isbnsearch.org/isbn/9780761919049>
- [27] Sorrel Burden. 2001. Dietary Treatment of Irritable Bowel Syndrome: Current Evidence and Guidelines for Future Practice. *Journal of Human Nutrition and Dietetics* 14, 3 (2001), 231–241. <https://www.ncbi.nlm.nih.gov/pubmed/11424515>
- [28] Nigel E Bush, Steven K Dobscha, Rosa Crumpton, Lauren M Denneson, Julia E Hoffman, Aysha Crain, Risa Cromer, and Julie T Kinn. 2015. A Virtual Hope Box Smartphone App as an Accessory to Therapy: Proof-of-Concept in a Clinical Sample of Veterans. *Suicide & Life-Threatening Behavior* 45, 1 (2015), 1-9. <https://doi.org/10.1111/sltb.12103>
- [29] Nigel E Bush, Derek J Smolenski, Lauren M Denneson, Holly B Williams, Elissa K Thomas, and Steven K Dobscha. 2017. A Virtual Hope Box: Randomized Controlled Trial of a Smartphone App for Emotional Regulation and Coping With Distress. *Psychiatric Services* 68, 4 (2017), 330-336. <https://doi.org/10.1176/appi.ps.201600283>
- [30] Christine Buttorff, Teague Ruder, and Melissa Bauman. 2017. *Multiple Chronic Conditions in the United States*. RAND Santa Monica, CA. <https://www.rand.org/pubs/tools/TL221.html>
- [31] Ayşe G Büyüktür, Mark S Ackerman, Mark W Newman, and Pei-Yao Hung. 2017. Design considerations for Semi-Automated Tracking: Self-Care Plans in Spinal Cord Injury. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. <https://doi.org/10.1145/3154862.3154870>
- [32] Brianne Campbell, Kelly Caine, Kay Connelly, Tom Doub, and April Bragg. 2015. Cell Phone Ownership and Use Among Mental Health Outpatients in the USA. *Personal and Ubiquitous Computing* 19, 2 (2015), 367–378. <https://doi.org/10.1007/s00779-014-0822-z>
- [33] Caroline Canavan, Joe West, and Timothy Richard Card. 2014. The Epidemiology of Irritable Bowel Syndrome. *Clinical Epidemiology* 6 (2014), 71–80. <https://doi.org/10.2147/CLEP.S40245>
- [34] JE Carnahan and Nugent CA. 1975. The Effects of Self-Monitoring By Patients on The Control of Hypertension. *The American Journal of the Medical Sciences* 269, 1 (1975), 69-73. <https://www.ncbi.nlm.nih.gov/pubmed/1130437>
- [35] Nutrition Coordinating Center. 2013a. NDSR User Manual. (2013).

- [36] Nutrition Coordinating Center. 2013b. Nutrition Data System for Research (NDSR). Nutrition Coordinating Center, University of Minnesota, Minneapolis, MN. (2013).
- [37] Kuang Chen, Akshay Kannan, Yoriyasu Yano, Joseph M Hellerstein, and Tapan S Parikh. 2012. Shreddr: Pipelined Paper Digitization for Low-Resource Organizations. In *Proceedings of the ACM Symposium on Computing for Development*. <https://doi.org/10.1145/2160601.2160605>
- [38] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A Epstein, Felicia Cordeiro, Matthew Kay, Gregory D Abowd, Tanzeem Choudhury, James Fogarty, and others. 2017. Semi-Automated Tracking: a Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing* 16, 1 (2017), 74–84. <https://doi.org/10.1109/MPRV.2017.18>
- [39] 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 ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/2556288.2557372>
- [40] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3025747>
- [41] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, and Sean A Munson. 2015. More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management. *Journal of Medical Internet Research* 17, 8 (2015), E203. <https://doi.org/10.2196/jmir.4364>
- [42] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine K Zia, James Fogarty, Julie A Kientz, and Sean A Munson. 2016. Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*. <https://doi.acm.org/10.1145/2818048.2819926>
- [43] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A Munson. 2019. Identifying and Planning for Individualized Change: Patient-Provider Collaboration Using Lightweight Food Diaries in Healthy Eating and Irritable Bowel Syndrome. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. <https://doi.org/10.1145/3314394>

- [44] Christine Clar, Katharine D Barnard, Ewen Cummins, Pamela Royle, and Norman Waugh. 2010. Self-Monitoring of Blood Glucose in Type 2 Diabetes: Systematic Review. *Health Technology Assessment* 14, 12 (2010), 1-140. <https://www.doi.org/10.3310/hta14120>
- [45] Janine Clarke, Judith Proudfoot, Mary-Rose Birch, Alexis E Whitton, Gordon Parker, Vijaya Manicavasagar, Virginia Harrison, Helen Christensen, and Dusan Hadzi-Pavlovic. 2014. Effects of Mental Health Self-Efficacy On Outcomes of a Mobile Phone and Web Intervention for Mild-to-Moderate Depression, Anxiety and Stress: Secondary Analysis of a Randomised Controlled Trial. In *BMC Psychiatry*. <https://doi.org/10.1186/s12888-014-0272-1>
- [46] James Clawson, Jessica Pater, Andrew D Miller, Elizabeth D Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating The Abandonment of Personal Health-Tracking Technologies on Craigslist. In *Proceedings of the ACM Conference on Ubiquitous Computing*. <https://doi.org/10.1145/2750858.2807554>
- [47] Sunny Consolvo, Predrag Klasnja, David W McDonald, and James A Landay. 2009. Goal-Setting Considerations for Persuasive Technologies that Encourage Physical Activity. In *Proceedings of the ACM Conference on Persuasive Technology*. <https://doi.org/10.1145/1541948.1541960>
- [48] Sunny Consolvo, Predrag Klasnja, David W McDonald, James A Landay, and others. 2014. Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. *Foundations and Trends in Human-Computer Interaction* 6, 3-4 (2014), 167-315. <https://doi.org/10.1561/11000000040>
- [49] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony Lamarca, Louis Legrand, Ryan Libby, Ian Smith, and James A Landay. 2008. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1357054.1357335>
- [50] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015a. Rethinking The Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2702123.2702154>
- [51] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. 2015b. Barriers and Negative Nudges: Exploring Challenges in Food Journaling. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2702123.2702155>

- [52] Riccardo Dalle Grave, Simona Calugi, Enrico Molinari, Maria Letizia Petroni, Mario Bondi, Angelo Compare, and Giulio Marchesini. 2005. Weight Loss Expectations in Obese Patients and Treatment Attrition: An Observational Multicenter Study. *Obesity Research* 13, 11 (2005), 1961–1969. <https://doi.org/10.1038/oby.2005.241>
- [53] Nediya Daskalova, Karthik Desingh, Jin Young Kim, Lixiang Zhang, Alexandra Papoutsaki, and Jeff Huang. 2017. Lessons Learned from Two Cohorts of Personal Informatics Self-Experiments. In *Proceedings of the ACM Conference on Ubiquitous Computing*. <https://doi.org/10.1145/3130911>
- [54] Nediya Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoach: A Personalized Automated Self-Experimentation System for Sleep Recommendations. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. <https://doi.org/bwt4>
- [55] Tabby Davies, Simon L Jones, and Ryan M Kelly. 2019a. Patient Perspectives on Self-Management Technologies for Chronic Fatigue Syndrome. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300452>
- [56] Tabby Davies, Simon L Jones, and Ryan M Kelly. 2019b. Patient Perspectives on Self-Management Technologies for Chronic Fatigue Syndrome. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Article 222, 13 pages. <https://doi.acm.org/10.1145/3290605.3300452>
- [57] Mary Jo Deering, Erin Siminerio, and Scott Weinstein. 2013. Issue Brief: Patient-Generated Health Data and Health IT. *Office of the National Coordinator for Health Information Technology* (2013).
- [58] Nicola Dell, Jessica Crawford, Nathan Breit, Timóteo Chaluco, Aida Coelho, Joseph McCord, and Gaetano Borriello. 2013. Integrating ODK Scan into The Community Health Worker Supply Chain in Mozambique. In *Proceedings of the ACM International Conference on Information and Communication Technologies and Development*. <https://doi.org/10.1145/2516604.2516611>
- [59] Pooja M Desai, Matthew E Levine, David J Albers, and Lena Mamykina. 2018. Pictures Worth a Thousand Words: Reflections on Visualizing Personal Blood Glucose Forecasts for Individuals with Type 2 Diabetes. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174112>

- [60] Gavin Doherty, David Coyle, and John Sharry. 2012. Engagement With Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2207676.2208602>
- [61] Stefan Rennick Egglestone, Sarah Knowles, Gill Toms, Penny Bee, Karina Lovell, and Peter Bower. 2016. Health Technologies ‘In the Wild’: Experiences of Engagement With Computerised CBT. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2858036.2858128>
- [62] Sigrid Elsenbruch. 2011. Abdominal Pain in Irritable Bowel Syndrome: A Review of Putative Psychological, Neural and Neuro-Immune Mechanisms. *Brain, Behavior, and Immunity* 25, 3 (2011), 386–394. <https://doi.org/10.1016/j.bbi.2010.11.010>
- [63] Daniel A Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A Munson. 2016. Beyond Abandonment To Next Steps: Understanding and Designing for Life After Personal Informatics Tool Use. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2858036.2858045>
- [64] Daniel A Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean A Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. In *Proceedings of the ACM Conference on Designing Interactive Systems*. <https://doi.acm.org/10.1145/2598510.2598558>
- [65] Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean A Munson. 2017. Examining Menstrual Tracking To Inform The Design of Personal Informatics Tools. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3025635>
- [66] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. <https://doi.org/10.1145/2750858.2804250>
- [67] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. I Always Assumed that I Wasn’t Really that Close to [Her]: Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/2702123.2702556>
- [68] Deborah Estrin and Ida Sim. 2010. Open mHealth Architecture: An Engine for Health Care Innovation. *Science* 330, 6005 (2010), 759–760. <https://doi.org/10.1126/science.1196187>

- [69] Shanti Eswaran, Jan Tack, and William D Chey. 2011. Food: The Forgotten Factor in the Irritable Bowel Syndrome. *Gastroenterology Clinics of North America* 40, 1 (2011), 141–162. <https://doi.org/10.1016/j.gtc.2010.12.012>
- [70] Aojg Farmer, Oj Gibson, L Tarassenko, and A Neil. 2005. A Systematic Review of Telemedicine Interventions To Support Blood Glucose Self-Monitoring in Diabetes. *Diabetic Medicine* 22, 10 (2005), 1372-1378. <https://doi.org/10.1111/j.1464-5491.2005.01627.x>
- [71] Martin Fitzgerald and Tracy McClelland. 2017. What Makes a Mobile App Successful in Supporting Health Behaviour Change? *Health Education Journal* 76, 3 (2017), 373–381. <https://dx.doi.org/10.2196%2Fmental.4984>
- [72] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using A Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. In *JMIR Mental Health*. <https://doi.org/10.2196/mental.7785>
- [73] National Institute for Clinical Excellence and Great Britain. 2002. *Guidance on the Use of Computerised Cognitive Behavioural Therapy for Anxiety and Depression*. National Institute for Clinical Excellence.
- [74] Centers for Disease Control and Prevention. 2015. 10 Leading Causes of Death By Age Group, United States, 2015. (2015). https://www.cdc.gov/injury/wisqars/pdf/leading_causes_of_death_by_age_group_2015-a.pdf
- [75] The Robert Wood Johnson Foundation. 2014. Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health. *Health Data Exploration Project* (2014).
- [76] Susannah Fox and Maeve Duggan. 2013. *Tracking for Health*. Pew Research Center. <https://www.pewresearch.org/internet/2013/01/28/tracking-for-health/>
- [77] Ionic Framework. . <https://ionicframework.com/>
- [78] RH Friedman, LE Kazis, A Jette, MB Smith, J Stollerman, J Torgerson, and Carey K. 1996. A Telecommunications System for Monitoring and Counseling Patients with Hypertension: Impact on Medication Adherence and Blood Pressure Control. *American Journal of Hypertension* 9, 4 (1996), 285-92. <https://www.ncbi.nlm.nih.gov/pubmed/8722429>
- [79] Susan N Friel, Frances R Curcio, and George W Bright. 2001. Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education* (2001), 124–158. <http://doi.org/10.2307/749671>

- [80] Lina Gega, Isaac Marks, and David Mataix-Cols. 2004. Computer-Aided CBT Self-Help for Anxiety and Depressive Disorders: Experience of a London Clinic and Future Directions. *Journal of Clinical Psychology* 60, 2 (2004), 147-57. <https://doi.org/10.1002/jclp.10241>
- [81] Peter R Gibson and Susan J Shepherd. 2012. Food Choice as a Key Management Strategy for Functional Gastrointestinal Symptoms. *The American Journal of Gastroenterology* 107, 5 (2012), 657–666. <https://doi.org/10.1038/ajg.2012.49>
- [82] Itzhak Gilat, Yishai Tobin, and Golan Shahar. 2011. Offering Support to Suicidal Individuals in An Online Support Group. *Archives of Suicide Research : Official Journal of the International Academy for Suicide Research* 15, 3 (2011), 195-206. <https://doi.org/10.1080/13811118.2011.589675>
- [83] Marianne Goodman, Uday Patil, Lauren Steffel, Jennifer Avedon, Scott Sasso, Joseph Triebwasser, and Barbara Stanley. 2010. Treatment Utilization By Gender in Patients With Borderline Personality Disorder. *Journal of Psychiatric Practice* 16, 3 (2010), 155-63. <https://doi.org/10.1097/01.pra.0000375711.47337.27>
- [84] Erik Grönvall and Nervo Verdezoto. 2013. Understanding Challenges and Opportunities of Preventive Blood Pressure Self-Monitoring At Home. In *Proceedings of the European Conference on Cognitive Ergonomics*. <https://doi.org/10.1145/2501907.2501962>
- [85] James J Gross and Ricardo F Muñoz. 1995. Emotion Regulation and Mental Health. *Clinical Psychology: Science and Practice* (1995), 151–164. <https://doi.org/10.1111/j.1468-2850.1995.tb00036.x>
- [86] François Guimbretière. 2003. Paper Augmented Digital Documents. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/964696.964702>
- [87] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. 2016. Fostering Engagement with Personal Informatics Systems. In *Proceedings of the ACM Conference on Designing Interactive Systems*. <https://doi.org/10.1145/2901790.2901803>
- [88] Alben Halpert, Christine B Dalton, Olafur Palsson, Carolyn Morris, Yuming Hu, Shrikant Bangdiwala, Jane Hankins, Nancy Norton, and Douglas Drossman. 2007. What Patients Know about Irritable Bowel Syndrome (IBS) and What They Would Like to Know. National Survey on Patient Educational Needs in IBS and development and validation of the Patient Educational Needs Questionnaire (PEQ). *The American Journal of Gastroenterology* 102, 9 (2007), 1972. <https://doi.org/10.1111/j.1572-0241.2007.01254.x>

- [89] Wendy Hardeman, Julie Houghton, Kathleen Lane, Andy Jones, and Felix Naughton. 2019. A Systematic Review of Just-In-Time Adaptive Interventions (JITAIs) to Promote Physical Activity. *International Journal of Behavioral Nutrition and Physical Activity* 16, 1 (2019), 31. <https://doi.org/10.1186/s12966-019-0792-7>
- [90] Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity Tracking: Barriers, Workarounds and Customisation. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. <https://doi.org/10.1145/2750858.2805832>
- [91] Jeffrey Heer and Maneesh Agrawala. 2008. Design Considerations for Collaborative Visual Analytics. *Information Visualization* 7, 1 (2008), 49–62. <https://doi.acm.org/10.1145/1391107.1391112>
- [92] Reetta Heinonen, Riitta Luoto, Pirjo Lindfors, and Clas-Håkan Nygård. 2012. Usability and Feasibility of Mobile Phone Diaries in an Experimental Physical Exercise Study. *Telemedicine and E-Health* 18, 2 (2012), 115–119. <https://www.doi.org/10.1089/tmj.2011.0087>
- [93] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The Weirdest People in the World? *Behavioral and Brain Sciences* 33, 2-3 (2010), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- [94] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What Does All This Data Mean For My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. *Human-Computer Interaction* 32, 5-6 (2017), 208–267. <https://doi.org/10.1080/07370024.2016.1277724>
- [95] Karen Holtzblatt, Jessamyn Burns Wendell, and Shelley Wood. 2004. *Rapid Contextual Design: a How-To Guide to Key Techniques for User-Centered Design*. Elsevier. <https://dl.acm.org/doi/book/10.5555/2891404>
- [96] Cheng-Kang Hsieh, Hongsuda Tangmunarunkit, Faisal Alquaddoomi, John Jenkins, Jinha Kang, Cameron Ketcham, Brent Longstaff, Joshua Selsky, Betta Dawson, Dallas Swendeman, and others. 2013. Lifestreams: A Modular Sense-Making Toolset for Identifying Important Patterns from Everyday Life. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*. <https://doi.org/10.1145/2517351.2517368>
- [97] Anna Huguet, Sanjay Rao, Patrick J Mcgrath, Lori Wozney, Mike Wheaton, Jill Conrod, and Sharlene Rozario. 2016. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. *PLOS One* 11, 5 (2016), E0154248. <https://doi.org/10.1371/journal.pone.0154248>

- [98] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology Probes: Inspiring Design For and With Families. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/642611.642616>
- [99] Alfred Inselberg. 1997. Multidimensional Detective. In *Information Visualization, 1997*. 100–107. <https://dl.acm.org/citation.cfm?id=857631>
- [100] Neil S Jacobson and Paula Truax. 1991. Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology* 59, 1 (1991), 12. <https://www.ncbi.nlm.nih.gov/pubmed/2002127>
- [101] Luis G Jaimes, Martin Llofriú, and Andrew Raij. 2015. Preventer, A Selection Mechanism for Just-In-Time Preventive Interventions. *IEEE Transactions on Affective Computing* 7, 3 (2015), 243–257. <https://doi.org/10.1109/TAFFC.2015.2490062>
- [102] Anne E Jamieson, Paula C Fletcher, and Margaret A Schneider. 2007. Seeking Control through the Determination of Diet: A Qualitative Investigation of Women with Irritable Bowel Syndrome and Inflammatory Bowel Disease. *Clinical Nurse Specialist* 21, 3 (2007), 152–160. <https://doi.org/10.1097/01.NUR.0000270015.97457.9c>
- [103] Susan L Janson, Kelly Wong McGrath, Jack K Covington, Su-Chun Cheng, and Homer A Boushey. 2009. Individualized Asthma Self-Management Improves Medication Adherence and Markers of Asthma Control. *Journal of Allergy and Clinical Immunology* 123, 4 (2009), 840–846. <https://doi.org/10.1016/j.jaci.2009.01.053>
- [104] Sooyeon Jeong and Cynthia Lynn Breazeal. 2016. Improving Smartphone Users’ Affect and Wellbeing With Personalized Positive Psychology Interventions. In *Proceedings of the International Conference On Human Agent Interaction*. <https://doi.org/10.1145/2974804.2974831>
- [105] Mehul Jhaveri and Elizabeth Lee. 2007. Performance of Electronic Diaries in Diabetes Clinical Trials Measured Through Overall Satisfaction of Site Coordinators. *Journal of Diabetes Science and Technology* 1, 4 (2007), 522–530. <https://www.doi.org/10.1177/193229680700100410>
- [106] Susan L Joslyn and Jared E LeClerc. 2012. Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *Journal of Experimental Psychology: Applied* 18, 1 (2012), 126. <https://doi.org/10.1037/a0025185>

- [107] Adrienne S Juarascio, Megan N Parker, Madeline A Lagacey, and Kathryn M Godfrey. 2018. Just-In-Time Adaptive Interventions: A Novel Approach for Enhancing Skill Utilization and Acquisition in Cognitive Behavioral Therapy for Eating Disorders. *International Journal of Eating Disorders* 51, 8 (2018), 826–830. <https://doi.org/10.1002/eat.22924>
- [108] Eiman Kanjo, Luluah Al-Husain, and Alan Chamberlain. 2015. Emotions in Context: Examining Pervasive Affective Sensing Systems, Applications, and Analyses. *Personal and Ubiquitous Computing* 19, 7 (2015), 1197–1212. <https://doi.org/10.1007/s00779-015-0842-3>
- [109] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilaradaga, and Jasmine Zia. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation To Detect Individualized Food Triggers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3025480>
- [110] Ravi Karkar, Jasmine Zia, Roger Vilaradaga, 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). <https://doi.org/10.1093/jamia/ocv150>
- [111] Dmitri S Katz, Blaine A Price, Simon Holland, and Nicholas Sheep Dalton. 2018a. Data, Data Everywhere, and Still Too Hard to Link: Insights from User Interactions with Diabetes Apps. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174077>
- [112] Dmitri S Katz, Blaine A Price, Simon Holland, and Nicholas Sheep Dalton. 2018b. Data, Data Everywhere, and Still Too Hard to Link: Insights from User Interactions with Diabetes Apps. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174077>
- [113] Dmitri S Katz, Blaine A Price, Simon Holland, and Nicholas Sheep Dalton. 2018c. Designing for Diabetes Decision Support Systems with Fluid Contextual Reasoning. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174199>
- [114] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists’ Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3313831.3376219>

- [115] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. 2012. Lullaby: A Capture & Access System for Understanding The Sleep Environment. In *Proceedings of the ACM Conference on Ubiquitous Computing*. <https://doi.org/10.1145/2370216.2370253>
- [116] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) Is My Bus?: User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2858036.2858558>
- [117] Joseph (Jofish) Kaye, Mary Mccuiston, Rebecca Gulotta, and David A Shamma. 2014. Money Talks: Tracking Personal Finances. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2556288.2556975>
- [118] Christina Kelley, Bongshin Lee, and Lauren Wilcox. Self-Tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 629–641. <https://doi.org/10.1145/3025453.3025750>
- [119] Leslie Kelman. 2007. The Triggers or Precipitants of the Acute Migraine Attack. *Cephalalgia* 27, 5 (2007), 394–402. <https://doi.org/10.1111/j.1468-2982.2007.01303.x>
- [120] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017). <https://doi.org/10.1145/3130930>
- [121] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to Evaluate Technologies for Health Behavior Change in HCI Research. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1978942.1979396>
- [122] Predrag Klasnja and Wanda Pratt. 2012. Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions. *Journal of Biomedical Informatics* 45, 1 (2012), 184–198. <https://doi.org/10.1016/j.jbi.2011.08.017>
- [123] Kenneth A Kobak, James C Mundt, and Betsy Kennard. 2015. Integrating Technology Into Cognitive Behavior Therapy for Adolescent Depression: A Pilot Study. In *Annals of General Psychiatry*. <https://doi.org/10.1186/s12991-015-0077-8>
- [124] Rafal Kocielnik and Gary Hsieh. 2017. Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*. <https://doi.org/10.1145/2998181.2998324>

- [125] Robert Kohn, Shekhar Saxena, Itzhak Levav, and Benedetto Saraceno. 2004. The Treatment Gap in Mental Health Care. *Bulletin of the World health Organization* 82, 11 (2004), 858–866. <https://doi.org/S0042-96862004001100011>
- [126] Daphne Koller and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT press. https://books.google.com/books/about/Probabilistic_Graphical_Models.html?id=dOruCwAAQBAJ&hl=en
- [127] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine* 16, 9 (2001), 606–13. <https://www.ncbi.nlm.nih.gov/pubmed/11556941>
- [128] Uri Ladabaum, Erin Boyd, Wei K Zhao, Ajitha Mannalithara, Annie Sharabidze, Gurkirpal Singh, Elaine Chung, and Theodore R Levin. 2012. Diagnosis, Comorbidities, and Management of Irritable Bowel Syndrome in Patients in a Large Health Maintenance Organization. *Clinical Gastroenterology and Hepatology* 10, 1 (2012), 37–45. <https://doi.org/10.1016/j.cgh.2011.08.015>
- [129] Michael J Lambert and Allen E Bergin. 1994. The Effectiveness of Psychotherapy. *Handbook of Psychotherapy and Behavior Change* 4 (1994), 143–189.
- [130] Mars Lan, Lauren Samy, Nabil Alshurafa, Myung-Kyung Suh, Hassan Ghasemzadeh, Aurelia Macabasco-O’Connell, and Majid Sarrafzadeh. 2012. WANDA: An End-to-End Remote Health Monitoring and Analytics System for Heart Failure Patients. In *Proceedings of the ACM Conference on Wireless Health*. <https://doi.acm.org/10.1145/2448096.2448105>
- [131] Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. <https://doi.org/10.1145/2750858.2804288>
- [132] Jisoo Lee, Erin Walker, Winslow Burleson, Matthew Kay, Matthew Buman, and Eric B Hekler. 2017. Self-Experimentation for Behavior Change: Design and Formative Evaluation of Two Approaches. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3026038>
- [133] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/1753326.1753409>
- [134] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. In *Proceedings of the International Conference on Ubiquitous Computing*. 405–414. <https://doi.acm.org/10.1145/2030112.2030166>

- [135] Ian Li, Anind K Dey, and Jodi Forlizzi. 2012. Using Context to Reveal Factors that Affect Physical Activity. *ACM Transactions on Computer-Human Interaction* 19, 1 (2012), 7. <https://doi.org/10.1145/2147783.2147790>
- [136] Chunyuan Liao, François Guimbretière, Ken Hinckley, and Jim Hollan. 2008. Papiercraft: A Gesture-Based Command System for Interactive Paper. *ACM Transactions on Computer-Human Interaction* 14, 4 (2008), 18. <https://doi.org/10.1145/1314683.1314686>
- [137] Solveig C Ligaarden and Per G Farup. 2011. Low Intake of Vitamin B 6 is Associated with Irritable Bowel Syndrome Symptoms. *Nutrition Research* 31, 5 (2011), 356–361. <https://doi.org/10.1016/j.nutres.2011.04.001>
- [138] Elizabeth O Lillie, Bradley Patay, Joel Diamant, Brian Issell, Eric J Topol, and Nicholas J Schork. 2011. The n-of-1 Clinical Trial: the Ultimate Strategy for Individualizing Medicine? *Personalized Medicine* 8, 2 (2011), 161–173. <https://doi.org/10.2217/pme.11.7>
- [139] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish’n’Steps: Encouraging Physical Activity with an Interactive Computer Game. In *Proceedings of the ACM Conference on Ubiquitous Computing*. https://doi.org/10.1007/11853565_16
- [140] Marsha M Linehan. 1993. *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. Guilford Press. <https://isbnsearch.org/isbn/9780898621839>
- [141] Marsha M Linehan. 2014. *DBT Skills Training Manual*. Guilford Publications. <https://isbnsearch.org/isbn/9781462516995>
- [142] Richard B Lipton. 2015. Risk Factors for and Management of Medication-Overuse Headache. *Continuum: Lifelong Learning in Neurology* 21, 4, Headache (2015), 1118–1131. <https://doi.org/10.1212/CON.0000000000000216>
- [143] Richard B Lipton, Joan C Amatniek, Michel D Ferrari, and Michael Gross. 1994. Migraine. Identifying and Removing Barriers To Care. *Neurology* 44, 6 Suppl 4 (1994), S63–8. <https://www.ncbi.nlm.nih.gov/pubmed/8008228>
- [144] Richard B Lipton, Walter F Stewart, and David Simon. 1998. Medical Consultation for Migraine: Results From The American Migraine Study. *Headache: The Journal of Head and Face Pain* 38, 2 (1998), 87-96. <https://www.ncbi.nlm.nih.gov/pubmed/9529763>

- [145] Christine L Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Transactions on Management Information Systems* 4 (2013), 19:1-19:28. <https://doi.org/10.1145/2544103>
- [146] George F Longstreth, W Grant Thompson, William D Chey, Lesley A Houghton, Fermin Mearin, and Robin C Spiller. 2006. Functional Bowel Disorders. *Gastroenterology* 130, 5 (2006), 1480–1491. <https://doi.org/10.1053/j.gastro.2005.11.061>
- [147] Rebecca M Lovell and Alexander C Ford. 2012. Effect of Gender on Prevalence of Irritable Bowel Syndrome in the Community: Systematic Review and Meta-Analysis. *The American Journal of Gastroenterology* 107, 7 (2012), 991. <https://doi.org/10.1038/ajg.2012.131>
- [148] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's Only a Computer: Virtual Humans Increase Willingness to Disclose. *Computers in Human Behavior* 37 (2014), 94-100. <https://doi.org/10.1016/j.chb.2014.04.043>
- [149] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing Food Trackers with Dietitians: Identifying Design Opportunities for Food Tracker Customization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300822>
- [150] Christina Lyons, P Hopley, and J Horrocks. 2009. A Decade of Stigma and Discrimination in Mental Health: Plus Ça Change, Plus C'est La Même Chose (The More Things Change, The More They Stay The Same). *Journal of Psychiatric and Mental Health Nursing* 16, 6 (2009), 501–507. <https://doi.org/10.1111/j.1365-2850.2009.01390.x>
- [151] Haley MacLeod, Ben Jelen, Annu Prabhakar, Lora Oehlberg, Katie Siek, and Kay Connelly. 2016. Asynchronous Remote Communities (ARC) for Researching Distributed Populations. In *EAI International Conference on Pervasive Computing Technologies for Healthcare*. <https://dl.acm.org/citation.cfm?id=3021319.3021320>
- [152] Morris Maizels. 1998. The Clinician's Approach To The Management of Headache. *Western Journal of Medicine* 168, 3 (1998), 203. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1304871/>
- [153] Lena Mamykina, Elizabeth M Heitkemper, Arlene M Smaldone, Rita Kukafka, Heather J Cole-Lewis, Patricia G Davidson, Elizabeth D Mynatt, Andrea Cassells, Jonathan N Tobin, and George Hripcsak. 2017. Personal Discovery in Diabetes Self-Management: Discovering Cause and Effect Using Self-Monitoring Data. *Journal of Biomedical Informatics* 76 (2017), 1–8. <https://doi.org/10.1016/j.jbi.2017.09.013>

- [154] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/1357054.1357131>
- [155] Lena Mamykina, Elizabeth D Mynatt, and David R Kaufman. 2006. Investigating Health Management Practices of Individuals with Diabetes. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1124772.1124910>
- [156] Gabriela Marcu, Anind K Dey, Sara Kiesler, and Madhu Reddy. 2016. Time to Reflect: Supporting Health Services Over Time By Focusing on Collaborative Reflection. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*. <https://doi.acm.org/10.1145/2818048.2820079>
- [157] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2014. Capturing the Mood: Facebook and Face-to-Face Encounters in the Workplace. In *Proceedings of the ACM conference on Computer supported cooperative work & social computing*. <https://doi.org/10.1145/2531602.2531673>
- [158] Mark Matthews, Gavin Doherty, David Coyle, and John Sharry. 2008. Designing Mobile Applications to Support Mental Health Interventions. In *Handbook of research on user interface design and evaluation for mobile technology*. IGI Global, 635–656. <https://doi.org/10.4018/978-1-59904-871-0.ch038>
- [159] Richard M McFall and Constance L Hammen. 1971. Motivation, Structure, and Self-Monitoring: Role of Nonspecific Factors in Smoking Reduction. *Journal of Consulting and Clinical Psychology* 37, 1 (1971), 80. <https://doi.org/10.1037/h0031279>
- [160] Mollie McKillop, Lena Mamykina, and Noémie Elhadad. 2018. Designing in the Dark: Eliciting Self-tracking Dimensions for Understanding Enigmatic Disease. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174139>
- [161] Helena M Mentis, Anita Komlodi, Katrina Schrader, Michael Phipps, Ann Gruber-Baldini, Karen Yarbrough, and Lisa Shulman. 2017. Crafting a View of Self-Tracking Data in the Clinical Visit. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3025453.3025589>
- [162] Sonali R Mishra, Predrag Klasnja, John MacDuffie Woodburn, Eric B Hekler, Larsson Omberg, Michael Kellen, and Lara Mangravite. 2019. Supporting Coping with Parkinson’s Disease Through Self Tracking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3290605.3300337>

- [163] Debanjali Mitra, Keith L Davis, and Robert W Baran. 2011. All-Cause Health Care Charges among Managed Care Patients with Constipation and Comorbid Irritable Bowel Syndrome. *Postgraduate Medicine* 123, 3 (2011), 122–132. <https://doi.org/10.3810/pgm.2011.05.2290>
- [164] David C Mohr, Michelle Nicole Burns, Stephen M Schueller, Gregory Clarke, and Michael Klinkman. 2013. Behavioral Intervention Technologies: Evidence Review and Recommendations for Future Research in Mental Health. *General Hospital Psychiatry* 35, 4 (2013), 332–338. <https://doi.org/10.1016/j.genhosppsych.2013.03.008>
- [165] David C Mohr, Stephen M Schueller, Kathryn Noth Tomasino, Susan M Kaiser, Nameyeh Alam, Chris Karr, Jessica L Vergara, Elizabeth L Gray, Mary J Kwasny, and Emily G Lattie. 2019. Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. *Journal of Medical Internet Research* 21, 8 (2019), e13609. <https://doi.org/10.2196/13609>
- [166] David C Mohr, Kathryn Noth Tomasino, Emily G Lattie, Hannah L Palac, Mary J Kwasny, Kenneth Weingardt, Chris J Karr, Susan M Kaiser, Rebecca C Rossom, Leland R Bardsley, Lauren Caccamo, Colleen Stiles-Shields, and Stephen M Schueller. 2017. IntelliCare: An Eclectic, Skills-Based App Suite for The Treatment of Depression and Anxiety. *Journal of Medical Internet Research* 19, 1 (2017). <https://dx.doi.org/10.2196%2Fjmir.6645>
- [167] Ramin Mojtabai, Mark Olfson, Nancy A Sampson, Robert Jin, Benjamin Druss, Philip S Wang, Kenneth B Wells, Harold A Pincus, and Ronald C Kessler. 2011. Barriers to Mental Health Treatment: Results from the National Comorbidity Survey Replication. *Psychological Medicine* 41, 8 (2011), 1751–1761. <https://doi.org/10.1017/s0033291710002291>
- [168] KW Monsbakken, PO Vandvik, and PG Farup. 2006. Perceived Food Intolerance in Subjects with Irritable Bowel Syndrome—Etiology, Prevalence and Consequences. *European Journal of Clinical Nutrition* 60, 5 (2006), 667–672. <https://doi.org/10.1038/sj.ejcn.1602367>
- [169] Margaret Morris and Farzin Guilak. 2009. Mobile Heart Health: Project Highlight. *IEEE Pervasive Computing* 8, 2 (2009). <https://doi.org/10.1109/MPRV.2009.31>
- [170] Margaret E Morris and Adrian Aguilera. 2012. Mobile, Social, and Wearable Computing and the Evolution of Psychological Practice. *Professional Psychology: Research and Practice* 43, 6 (2012), 622. <https://psycnet.apa.org/doi/10.1037/a0029041>
- [171] Frederick Muench and Amit Baumel. 2017. More Than A Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions. In *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.7463>

- [172] Frederick Muench, Katherine Van Stolk-Cooke, Alexis Kuerbis, Gertraud Stadler, Amit Baumel, Sijing Shao, James R Mckay, and Jon Morgenstern. 2017. A Randomized Controlled Pilot Trial of Different Mobile Messaging Interventions for Problem Drinking Compared to Weekly Drink Tracking. In *PLOS One*. <https://doi.org/10.1371/journal.pone.0167900>
- [173] Sean A Munson and Sunny Consolvo. 2012. Exploring Goal-Setting, Rewards, Self-Monitoring, and Sharing to Motivate Physical Activity. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*. <https://doi.org/10.4108/icst.pervasivehealth.2012.248691>
- [174] Inbal Nahum-Shani, Shawna N Smith, Ambuj Tewari, Katie Witkiewitz, Linda M Collins, Bonnie Spring, and S Murphy. 2014. Just In Time Adaptive Interventions (JITAI)s: An Organizing Framework for Ongoing Health Behavior Support. *Methodology Center Technical Report* 2014 (2014), 14–126. <https://doi.org/10.1007/s12160-016-9830-8>
- [175] Felix Naughton. 2017. Delivering “Just-In-Time” Smoking Cessation Support via Mobile Phones: Current Knowledge and Future Directions. *Nicotine & Tobacco Research* 19, 3 (2017), 379–383. <https://doi.org/10.1093/ntr/ntw143>
- [176] Jasmir G Nayak, Andrea L Hartzler, Liam C Macleod, Jason P Izard, Bruce M Dalkin, and John L Gore. 2015. Relevance of Graph Literacy in the Development of Patient-Centered Communication Tools. *Patient Education and Counseling* (2015). <https://doi.org/10.1016/j.pec.2015.09.009>
- [177] Andrada D Neacsiu, Shireen L Rizvi, and Marsha M Linehan. 2010a. Dialectical Behavior Therapy Skills Use as a Mediator and Outcome of Treatment for Borderline Personality Disorder. *Behaviour Research and Therapy* 48, 9 (2010), 832-9. <https://doi.org/10.1016/j.brat.2010.05.017>
- [178] Andrada D Neacsiu, Shireen L Rizvi, Peter P Vitaliano, Thomas R Lynch, and Marsha M Linehan. 2010b. The Dialectical Behavior Therapy Ways of Coping Checklist: Development and Psychometric Properties. *Journal of clinical psychology* 66, 6 (2010), 563–582. <https://doi.org/10.1002/jclp.20685>
- [179] Jasmin Niess and Paweł W Woźniak. 2018. Supporting Meaningful Personal Fitness: The Tracker Goal Evolution Model. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3173745>
- [180] Sonya B Norman, Laura Campbell-Sills, Carla A Hitchcock, Sarah Sullivan, Alexis Rochlin, Kendall C Wilkins, and Murray B Stein. 2011. Psychometrics of a Brief Measure of anxiety to Detect Severity and Impairment: The Overall Anxiety Severity and Impairment Scale (OASIS). *Journal of Psychiatric Research* 45, 2 (2011), 262–268. <https://dx.doi.org/10.1016%2Fj.jpsychires.2010.06.011>

- [181] Sonya B Norman, Shadha Hami Cissell, Adrienne J Means-Christensen, and Murray B Stein. 2006. Development and Validation of an Overall Anxiety Severity and Impairment Scale (OASIS). *Depression and Anxiety* 23, 4 (2006), 245-9. <https://www.ncbi.nlm.nih.gov/pubmed/16688739>
- [182] National Institutes of Health: National Institute of Mental Health. 2015. Mental Illness. (2015). <https://www.nimh.nih.gov/health/statistics/mental-illness.shtml>
- [183] Mark Olfson, Ramin Mojtabai, Nancy A Sampson, Irving Hwang, Benjamin Druss, Philip S Wang, Kenneth B Wells, Harold Alan Pincus, and Ronald C Kessler. 2009. Dropout from Outpatient Mental Health Care in the United States. *Psychiatric Services* 60, 7 (2009), 898-907. <https://doi.org/10.1176/ps.2009.60.7.898>
- [184] Michael Oliver and Bob Sapey. 1999. *Social Work with Disabled People*. Macmillan International Higher Education. <https://doi.org/10.1007/978-1-349-14823-3>
- [185] Phillip Olla and Caley Shimskey. 2015. mHealth Taxonomy: a Literature Survey of Mobile Health Applications. *Health and Technology* 4, 4 (2015), 299–308. <https://dx.doi.org/10.1007/s12553-014-0093-8>
- [186] Lucille ML Ong, Johanna CJM De Haes, Alaysia M Hoos, and Frits B Lammes. 1995. Doctor-Patient Communication: A Review of the Literature. *Social Science & Medicine* 40, 7 (1995), 903–918. <https://www.ncbi.nlm.nih.gov/pubmed/7792630>
- [187] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014a. PopTherapy: Coping with Stress Through Pop-culture. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, ICST, Brussels, Belgium, Belgium, 109–117. DOI: <https://dx.doi.org/10.4108/icst.pervasivehealth.2014.255070>
- [188] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014b. PopTherapy: Coping with Stress Through Pop-Culture. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 109–117. <https://doi.org/10.4108/icst.pervasivehealth.2014.255070>
- [189] Tapan S Parikh, Paul Javid, Sasikumar K, Kaushik Ghosh, and Kentaro Toyama. 2006. Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1124772.1124857>

- [190] Sun Young Park and Yunan Chen. 2015. Individual and Social Recognition: Challenges and Opportunities in Migraine Management. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*. <https://doi.org/10.1145/2675133.2675248>
- [191] Rupa A Patel, Predrag Klasnja, Andrea Hartzler, Kenton T Unruh, and Wanda Pratt. 2012. Probing the Benefits of Real-Time Tracking During Cancer Care. In *AMIA Annual Symposium Proceedings*, Vol. 2012. 1340. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540467/>
- [192] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830. <https://dl.acm.org/doi/10.5555/1953048.2078195>
- [193] Tamara Peyton, Erika Poole, Madhu Reddy, Jennifer Kraschnewski, and Cynthia Chuang. 2014. Every Pregnancy Is Different: Designing mHealth for The Pregnancy Ecology. In *Proceedings of the Conference on Designing Interactive Systems*. <https://doi.org/10.1145/2598510.2598572>
- [194] Iris Posserud, Hans Strid, Stine Störsrud, Hans Törnblom, Ulla Svensson, Jan Tack, Lukas Van Oudenhove, and Magnus Simrén. 2013. Symptom Pattern Following a Meal Challenge Test in Patients with Irritable Bowel Syndrome and Healthy Controls. *United European Gastroenterology Journal* (2013), 358–67. <https://doi.org/10.1177/2050640613501817>
- [195] Judith Proudfoot, Janine Clarke, Mary-Rose Birch, Alexis E Whitton, Gordon Parker, Vijaya Manicavasagar, Virginia Harrison, Helen Christensen, and Dusan Hadzi-Pavlovic. 2013. Impact of a Mobile Phone and Web Program On Symptom and Functional Outcomes for People With Mild-to-Moderate Depression, Anxiety and Stress: A Randomised Controlled Trial. In *BMC Psychiatry*. <https://doi.org/10.1186/1471-244X-13-312>
- [196] Paul Quinnett, Philip Batterham, Wouter Van Ballegooijen, Ursula Whiteside, Anita Lungu, Julie Richards, Gregory E Simon, Sarah Clingan, Jaeden Siler, Lorlei Snyder, and Evette Ludman. 2014. Designing Messaging to Engage Patients in An Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation. In *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.3173>

- [197] Kavita Radhakrishnan, Kathryn Bowles, Alexandra Hanlon, Maxim Topaz, and Jesse Chittams. 2013. A Retrospective Study on Patient Characteristics and Telehealth Alerts Indicative of Key Medical Events for Heart Failure Patients at a Home Health Agency. *Telemedicine and E-Health* 19, 9 (2013), 664–670. <https://doi.org/10.1089/tmj.2012.0307>
- [198] Shriti Raj, Joyce M Lee, Ashley Garrity, and Mark W Newman. 2019. Clinical Data in Context: Towards Sensemaking Tools for Interpreting Personal Health Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 22. <https://doi.org/10.1145/3314409>
- [199] Shriti Raj, Mark W Newman, Joyce M Lee, and Mark S Ackerman. 2017. Understanding Individual and Collaborative Problem-Solving with Patient-Generated Data: Challenges and Opportunities. *Proceedings of the ACM on Human-Computer Interaction* (2017), 88:1–88:18. <https://doi.acm.org/10.1145/3134723>
- [200] Birthe Krogh Rasmussen, Rigmor Jensen, Marianne Schroll, and Jes Olesen. 1991. Epidemiology of Headache in a General Population—A Prevalence Study. *Journal of Clinical Epidemiology* 44, 11 (1991), 1147–1157. <https://www.ncbi.nlm.nih.gov/pubmed/1941010>
- [201] Ruth Ravichandran, Sang-Wha Sien, Shwetak N Patel, Julie A Kientz, and Laura R Pina. 2017. Making Sense of Sleep Sensors: How Sleep Sensing Technologies Support and Undermine Sleep Health. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3025453.3025557>
- [202] Meghan J Reading and Jacqueline A Merrill. 2018. Converging and Diverging Needs between Patients and Providers who are Collecting and Using Patient-Generated Health Data: An Integrative Review. *Journal of the American Medical Informatics Association* 25, 6 (2018), 759–771. <https://doi.org/10.1093/jamia/ocy006>
- [203] Suzanne M Rebro, Ruth E Patterson, Alan R Kristal, and Carrie L Cheney. 1998. The Effect of Keeping Food Records on Eating Patterns. *Journal of the American Dietetic Association* 98, 10 (1998), 1163–1165. [https://doi.org/10.1016/S0002-8223\(98\)00269-7](https://doi.org/10.1016/S0002-8223(98)00269-7)
- [204] Blaine Reeder, Jym Chung, Thao Lê, Hilaire J Thompson, and George Demiris. 2014. Assessing Older Adults’ Perceptions of Sensor Data and Designing Visual Displays for Ambient Assisted Living Environments: An Exploratory Study. *Methods of Information in Medicine* 53, 3 (2014), 152. <https://doi.org/10.3414/ME13-02-0009>
- [205] William T Riley, Russell E Glasgow, Lynn Etheredge, and Amy P Abernethy. 2013. Rapid, Responsive, Relevant (R3) Research: A Call for a Rapid Learning Health Research Enterprise. *Clinical and Translational Medicine* 2, 1 (2013), 10. <https://doi.org/10.1186/2001-1326-2-10>

- [206] Shireen L Rizvi, Linda A Dimeff, Julie Skutch, David Carroll, and Marsha M Linehan. 2011. A Pilot Study of the DBT Coach: An Interactive Mobile Phone Application for Individuals With Borderline Personality Disorder and Substance Use Disorder. *Behavior Therapy* 42, 4 (2011), 589-600. <https://doi.org/10.1016/j.beth.2011.01.003>
- [207] Shireen L Rizvi, Christopher D Hughes, and Marget C Thomas. 2016. The DBT Coach Mobile Application as an Adjunct to Treatment for Suicidal and Self-Injuring Individuals With Borderline Personality Disorder: A Preliminary Evaluation and Challenges to Client Utilization. *Psychological Services* 13, 4 (2016), 380-388. <https://doi.org/10.1037/ser0000100>
- [208] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal Tracking As Lived Informatics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2556288.2557039>
- [209] Jim Rudd, Ken Stern, and Scott Isensee. 1996. Low vs. High-Fidelity Prototyping Debate. *Interactions* 3, 1 (1996), 76–85. <https://doi.org/10.1145/223500.223514>
- [210] Akane Sano, Paul Johns, and Mary Czerwinski. 2015. HealthAware: An Advice System for Stress, Sleep, Diet and Exercise. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 546–552. <https://doi.org/10.1109/ACII.2015.7344623>
- [211] Randy A Sansone and Lori A Sansone. 2011. Gender Patterns in Borderline Personality Disorder. *Innovations in Clinical Neuroscience* 8, 5 (2011), 16-20. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3115767/>
- [212] Jeff Sauro. 2011. Measuring Usability With The System Usability Scale (SUS). (2011). <https://measuringu.com/sus/>
- [213] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. 2002. Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/564376.564421>
- [214] Jessica Schroeder, Chia-Fang Chung, Daniel A Epstein, Ravi Karkar, Adele Parsons, Natalia Murinova, James Fogarty, and Sean A Munson. 2018. Examining Self-Tracking by People with Migraine: Goals, Needs, and Opportunities in a Chronic Health Condition. In *Proceedings of the Conference on Designing Interactive Systems*. 135-148. <https://doi.org/10.1145/3196709.3196738>

- [215] 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 ACM Conference on Computer-Supported Cooperative Work & Social Computing*. <https://doi.org/10.1145/2998181.2998276>
- [216] Jessica Schroeder, Ravi Karkar, James Fogarty, Julie A Kientz, Matthew Kay, and Sean A Munson. 2018. A Patient-Centered Proposal for Bayesian Analysis of Self-Experiments for Health. In *Journal of Healthcare Informatics Research*. <https://doi.org/10.1007/s41666-018-0033-x>
- [217] Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A Munson. 2019. Examining Opportunities for Goal-Directed Self-Tracking to Support Chronic Condition Management. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 3. ACM New York, NY, USA, 1–26. <https://doi.org/10.1145/3369809>
- [218] Jessica Schroeder, Jina Suh, Chelsey Wilks, Mary Czerwinski, Sean A Munson, James Fogarty, and Tim Althoff. 2020 (To Appear). Data-Driven Implications for Translating Evidence-Based Psychotherapies into Technology-Delivered Interventions. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*.
- [219] Jessica Schroeder, Chelsey Wilks, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M Linehan. 2018. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3173574.3173972>
- [220] Stephen M Schueller, Jason J Washburn, and Matthew Price. 2016. Exploring Mental Health Providers' Interest in Using Web and Mobile-Based Tools in Their Practices. *Internet Interventions* 4, 2 (2016), 145-151. <https://doi.org/10.1016/j.invent.2016.06.004>
- [221] Adrian BR Shatte, Delyse M Hutchinson, and Samantha J Teague. 2019. Machine Learning in Mental Health: A Scoping Review of Methods and Applications. *Psychological Medicine* 49, 9 (2019), 1426–1448. <https://doi.org/10.1017/S0033291719000151>
- [222] Stephen D Silberstein and The US Headache Consortium. 2000. Practice Parameter: Evidence-Based Guidelines for Migraine Headache (An Evidence-Based Review) Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 55, 6 (2000), 754–762. <https://doi.org/10.1212/WNL.55.6.754>

- [223] Magnus Simrén, A Månsson, Anna Maria Langkilde, Jan Svedlund, Hasse Abrahamsson, Ulf Bengtsson, and Einar S Björnsson. 2001. Food-Related Gastrointestinal Symptoms in the Irritable Bowel Syndrome. *Digestion* 63, 2 (2001), 108–115. <https://www.ncbi.nlm.nih.gov/pubmed/11244249>
- [224] Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/506443.506619>
- [225] Glen D Solomon and Kimber L Price. 1997. Burden of Migraine: A Review of its Socioeconomic Impact. *Pharmacoeconomics* 11, 1 (1997), 1-10. <https://doi.org/10.2165/00019053-199700111-00003>
- [226] Donna Spruijt-Metz, Cheng KF Wen, Gillian O'Reilly, Ming Li, Sangwon Lee, BA Emken, Urbashi Mitra, Murali Annaram, Gisele Ragusa, and Shrikanth Narayanan. 2015. Innovations in the Use of Interactive Technology to Support Weight Management. *Current Obesity Reports* 4, 4 (2015), 510–519. <https://doi.org/10.1007/s13679-015-0183-6>
- [227] Timothy J Steiner, Lars Jacob Stovner, and Gretchen L Birbeck. 2013. Migraine: The Seventh Disabler. *Headache: The Journal of Head and Face Pain* 53, 2 (2013), 227–229. <https://doi.org/10.1186/1129-2377-14-1>
- [228] Moira A Stewart. 1995. Effective Physician-Patient Communication and Health Outcomes: A Review. *CMAJ: Canadian Medical Association Journal* 152, 9 (1995), 1423. <https://www.ncbi.nlm.nih.gov/pubmed/7728691>
- [229] Walter F Stewart, Richard B Lipton, J Whyte, A Dowson, K Kolodner, JNet al Liberman, and J Sawyer. 1999. An International Study to Assess Reliability of the Migraine Disability Assessment (MIDAS) Score. *Neurology* 53, 5 (1999), 988–988. <https://doi.org/10.1212/WNL.53.5.988>
- [230] Colleen Stiles-Shields, Enid Montague, Emily G Lattie, Mary J Kwasny, and David C Mohr. 2017. What Might Get in the Way: Barriers to the Use of apps for Depression. *Digital Health* 3 (2017). <https://doi.org/10.1177/2F2055207617713827>
- [231] Brian Still and John Morris. 2010. The Blank-Page Technique: Reinvigorating Paper Prototyping in Usability Testing. *IEEE Transactions on Professional Communication* 53, 2 (2010), 144–157. <https://doi.org/10.1109/TPC.2010.2046100>
- [232] Cristiano Storni. 2011. Complexity in an Uncertain and Cosmopolitan World. Rethinking Personal Health Technology in Diabetes with The Tag-It-Yourself. *PsychNology Journal* 9, 2 (2011), 165–185.

- [233] Anselm Strauss and Juliet Corbin. 1967. *Discovery of Grounded Theory*.
- [234] Victor J Strecher, Matthew Kreuter, Dirk-Jan Den Boer, Sarah Kobrin, Harm J Hospers, and Celette S Skinner. 1994. The Effects of Computer-Tailored Smoking Cessation Messages in Family Practice Settings. *Journal of Family Practice* 39, 3 (1994), 262–270. <https://www.ncbi.nlm.nih.gov/pubmed/8077905>
- [235] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More Text Please! Understanding and Supporting the Use of Visualization for Clinical Text Overview. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3173996>
- [236] Kirsten Swearingen and Rashmi Sinha. 2001. Beyond Algorithms: An HCI Perspective on Recommender Systems. In *ACM SIGIR 2001 Workshop on Recommender Systems*, Vol. 13. 1–11.
- [237] Paul C Tang, J Marc Overhage, Albert Solomon Chan, Nancy L Brown, Bahar Aghighi, Martin P Entwistle, Siu Lui Hui, Shauna M Hyde, Linda H Klieman, Charlotte J Mitchell, Anthony J Perkins, Lubna S Qureshi, Tanya A Waltmyer, Leigha J Winters, and Charles Y Young. 2013. Online Disease Management of Diabetes: Engaging And Motivating Patients Online With Enhanced Resources-Diabetes (EMPOWER-D), a Randomized Controlled Trial. *Journal of the American Medical Informatics Association* 20, 3 (2013), 526–534. <https://doi.org/10.1136/amiajnl-2012-001263>
- [238] Frances E Thompson and Amy F Subar. 2008. Dietary Assessment Methodology. *Nutrition in the Prevention and Treatment of Disease* 2 (2008), 3–39. <https://doi.org/10.1016/B978-012193155-1/50003-9>
- [239] Josephine Raun Thomsen, Peter Gall Krogh, Jacob Albæk Schnedler, and Hanne Linnet. 2018. Interactive Interior and Proxemics Thresholds: Empowering Participants in Sensitive Conversations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3173642>
- [240] Erik H Trainer, Ban Al-Ani, and David F Redmiles. 2011. Impact of Collaborative Traces on Trustworthiness. In *Proceedings of the International Workshop on Cooperative and Human Aspects of Software Engineering*. 40–47. <https://doi.acm.org/10.1145/1984642.1984649>
- [241] Andreas K Triantafyllidis, Carmelo Velardo, Dario Salvi, Syed Ahmar Shah, Vassilis G Koutkias, and Lionel Tarassenko. 2015. A Survey of Mobile Phone Sensing, Self-Reporting, and Social Sharing for pervasive Healthcare. *IEEE Journal of Biomedical and Health Informatics* 21, 1 (2015), 218–227. <https://doi.org/10.1109/JBHI.2015.2483902>

- [242] Edward R Tufte, Nora Hillman Goeler, and Richard Benson. 1990. *Envisioning Information*. Vol. 126. Graphics Press Cheshire, CT. <https://isbnsearch.org/isbn/9780961392116>
- [243] SM Turner, DC Beidel, SA Spaulding, and JM Brown. 1995. The Practice of Behavior Therapy: A National Survey of Cost and Methods. *Behaviour Therapist* 18 (1995), 1–1.
- [244] Julio Vega, Samuel Couth, Ellen Poliakoff, Sonja Kotz, Matthew Sullivan, Caroline Jay, Markel Vigo, and Simon Harper. 2018. Back to Analogue: Self-Reporting for Parkinson’s Disease. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3173574.3173648>
- [245] Geert Verbeke and Geert Molenberghs. 1997. *Linear Mixed Models in Practice: a SAS-Oriented Approach*. Vol. 126. Springer-Verlag New York, LLC, New York, NY.
- [246] John T Vessey and Kenneth I Howard. 1993. Who Seeks Psychotherapy? *Psychotherapy: Theory, Research, Practice, Training* 30, 4 (1993), 546. <https://doi.org/10.1037/0033-3204.30.4.546>
- [247] David Walker. 2004. Comptroller General’s Forum on Health Care: Unsustainable Trends Necessitate Comprehensive and Fundamental Reforms to Control Spending and Improve Value. *Washington, DC: US Government Accountability Office* (2004).
- [248] Bruce E Wampold and Zac E Imel. 2015. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. Routledge. <https://isbnsearch.org/isbn/0805857087>
- [249] Philip S Wang, Michael Lane, Mark Olfson, Harold A Pincus, Kenneth B Wells, and Ronald C Kessler. 2005. Twelve-Month Use of Mental Health Services in the United States: Results From The National Comorbidity Survey Replication. *Archives of General Psychiatry* 62, 6 (2005), 629–640. <https://doi.org/10.1001/archpsyc.62.6.629>
- [250] Thomas N Ward. 2012. Migraine Diagnosis and Pathophysiology. *Continuum: Lifelong Learning in Neurology* 18, 4, Headache (2012), 753–763. <https://doi.org/10.1212/01.CON.0000418640.07405.31>
- [251] Peter West, Richard Giordano, Max Van Kleek, and Nigel Shadbolt. 2016. The Quantified Patient in the Doctor’s Office: Challenges & Opportunities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/2858036.2858445>
- [252] Peter West, Max Van Kleek, Richard Giordano, Mark Weal, and Nigel Shadbolt. 2017. Information Quality Challenges of Patient-Generated Data in Clinical Practice. *Frontiers in Public Health* 5 (2017), 284. <https://dx.doi.org/10.3389%2Fpubh.2017.00284>

- [253] Peter West, Max Van Kleek, Richard Giordano, Mark J Weal, and Nigel Shadbolt. 2018. Common Barriers to the Use of Patient-Generated Data Across Clinical Settings. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.acm.org/10.1145/3173574.3174058>
- [254] Kathleen M Wright, Oscar A Cabrera, Paul D Bliese, Amy B Adler, Charles W Hoge, and Carl A Castro. 2009. Stigma and Barriers to Care in Soldiers Postcombat. *Psychological Services* 6, 2 (2009), 108. <https://doi.org/10.1037/a0012620>
- [255] Chih-Hsiang Yang, Jaclyn P Maher, and David E Conroy. 2015. Implementation of Behavior Change Techniques in Mobile Applications for Physical Activity. *American Journal of Preventive Medicine* 48, 4 (2015), 452–455. <https://doi.org/10.1016/j.amepre.2014.10.010>
- [256] Ron Yeh, Chunyuan Liao, Scott Klemmer, François Guimbretière, Brian Lee, Boyko Kakaradov, Jeannie Stamberger, and Andreas Paepcke. 2006. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1124772.1124859>
- [257] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300509>
- [258] Kathryn Yorkston, Kurt Johnson, Estelle Klasner, Dagmar Amtmann, Carrie Kuehn, and Brian Dudgeon. 2003. Getting the Work Done: A Qualitative Study of Individuals with Multiple Sclerosis. *Disability and Rehabilitation* 25, 8 (2003), 369-379. <https://www.doi.org/10.1080/0963828031000090506>
- [259] Haining Zhu, Joanna Colgan, Madhu Reddy, and Eun Kyoung Choe. 2016. Sharing Patient-Generated Data in Clinical Practice: An Interview Study. *Proceedings of the American Medical Informatics Association* (2016), 1303-1312. <https://www.ncbi.nlm.nih.gov/pubmed/28269928>
- [260] Jasmine Zia, Pamela Barney, Kevin C Cain, Monica E Jarrett, and Margaret M Heitkemper. 2016a. A Comprehensive Self-Management Irritable Bowel Syndrome Program Produces Sustainable Changes in Behavior After 1 Year. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association* 14, 2 (2016), 212–219. <https://doi.org/10.1016/j.cgh.2015.09.027>

- [261] Jasmine Zia, Chia-Fang Chung, Jessica Schroeder, Sean A Munson, Julie Kientz, James Fogarty, Elizabeth Bales, Jeanette M Schenk, and Margaret Heitkemper. 2016b. The Feasibility, Usability and Perceived Clinical Utility of Traditional Paper Food and Symptom Journals for Patients with Irritable Bowel Syndrome. *Neurogastroenterology & Motility* (2016). <https://doi.org/10.1111/nmo.12935>
- [262] Jasmine Zia, Chia-Fang Chung, Kaiyuan Xu, Yi Dong, Kevin Cain, Sean A Munson, and Margaret Heitkemper. 2016c. Inter-Rater Reliability of Healthcare Provider Interpretations of Food and Gastrointestinal Symptom Paper Diaries of Patients with Irritable Bowel Syndrome. *Digestive Disease Week* (2016). [https://doi.org/10.1016/S0016-5085\(16\)33045-1](https://doi.org/10.1016/S0016-5085(16)33045-1)
- [263] Jasmine Zia, Jessica Schroeder, Sean A Munson, James Fogarty, Linda Nguyen, Pamela Barney, Margaret Heitkemper, and Uri Ladabaum. 2016d. Feasibility and Usability Pilot Study of a Novel Irritable Bowel Syndrome Food and Gastrointestinal Symptom Journal Smartphone App. *Clinical and Translational Gastroenterology* 7, 3 (2016), E147. <https://doi.org/10.1038/ctg.2016.9>