Making Sense of Multidimensional Health Data to Manage Chronic Conditions: Designing to Support Episode-Driven Data Interaction

by

Shriti Raj

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2022

Doctoral Committee:

Professor Mark W. Newman, Chair
Assistant Professor Matthew Kay, Northwestern University
Associate Professor Predrag Klasnja
Professor Joyce M. Lee
Associate Professor Lena Mamykina, Columbia University

Shriti Raj

shritir@umich.edu

ORCID iD: 0000-0001-5484-9980

Dedication

This is for you, *Ma*. I would never have applied to a doctoral program had it not been for your encouragement. Thank you for making me who I am, for being my strongest supporter, and for your unconditional love. I had never imagined this moment without your physical presence in my life. But I see you... smiling and proud. I love you and miss you so much.

Acknowledgements

This dissertation would not have been possible without the continuous mentorship, support and encouragement of my advisor, Prof. Mark Newman. Thank you for sharing your expertise with me; for making me a better researcher, scholar, writer, and thinker; and for believing in me during the darkest and the brightest phases of my life. I feel extremely fortunate to have completed my doctoral journey under your guidance and to also have your mentorship as I begin my academic career.

I thank Dr. Joyce Lee for providing access to invaluable resources that made this dissertation possible. Thank you for your guidance and for your enthusiasm to continuously support my research. I am also grateful to the members of Dr. Lee's research team, Ashley Garrity, Emily Dhadphale, and Valerie Gavrila, for their valuable support in navigating the tedious research processes.

I thank my committee members Assistant Prof. Matthew Kay, Associate Prof. Predrag Klasnja, and Associate Prof. Lena Mamykina for giving their valuable time and feedback on my work, for helping me benefit from their expertise in their respective fields, and for encouraging me to push the boundaries of my research. In addition to my committee members, I am grateful to Prof. Mark Ackerman and Prof. Tiffany Veinot for their guidance and feedback in the early milestones of the program.

This dissertation would not have been possible without the contributions of my participants – patients, caregivers, and clinicians. They enthusiastically participated in the demanding research activities of my studies and trusted me with information about their lives.

To say that my doctoral journey was eventful and life changing would be an understatement. During this journey, I married my best friend, lost my mother, became a mother, and navigated a pandemic as an international student. Surviving and thriving was only made possible because of the different kinds of support I was fortunate to have received. I am grateful to my transcontinental family for always lending an ear to offer suggestions in problematic times and for visiting me in situations of desperate need. I am thankful to my parents, especially my late mother Vimla Gupta, for always putting education first and for believing in my aspirations as I began my academic journey in a different country. A big shout out to my partner, Neeraj Kumar, for enabling and encouraging me to give it my best; and for being the village I needed as a first-time mother. Thank you for your everyday endeavors to keep us sailing despite your own struggles. To my child, Avyaan – thank you for inspiring me every day. To watch you come into existence and grow into a contagiously cheerful person has been astounding. To my fur-baby, Pascal – thank you for your love. Your companionship has been invaluable during times of loneliness and stress. Your theatrics cracked me up on the gloomiest of days.

While the pandemic challenged safe international travel, my father Dr. Shriram Gupta, my mother-in-law Dr. Pushpa Kumar, my father-in-law Dr. Mohan Kumar, and my aunt Manju Chowdhury never hesitated to travel all the way from India and support my young family here. Thank you all for your help when I needed it the most. A huge thanks to my brother Varun Raj and my cousin Dr. Rahul Gupta for helping me feel at home in a place away from home; for supporting me in times of urgency; and for giving me company during school breaks.

The wonderful community of scholars and staff at the University of Michigan School of Information has been an integral part of my journey. I am thankful to my lab mates, Yung-Ju (Stanley) Chang, Rayoung Yang, Gaurav Paruthi, and Chuan-Che (Jeff) Huang, for helping me learn through their experiences and for the relaxing and fun social gatherings. I am thankful to all the members of my doctoral cohort for making the initial two years of the doctoral program enjoyable amidst the challenging coursework. A special thanks to Megh Marathe for always checking in on me; and to Linfeng Li and Youyang Hou for being the best fellow graduate student instructors. I am grateful to students who gave me an opportunity to mentor them and helped on my projects – Kelsey Toporski, Xinghui Yan, Toshi Gupta, Chloe Preble, and Cloe Kuc. My doctoral journey would not have been possible without the support of UMSI staff members, Amy Eaton and Veronica Falandino, thank you for answering all my questions and helping me navigate the requirements for this degree. Lastly, I am grateful, fortunate, and proud for having completed this journey at the University of Michigan.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

People with chronic health conditions, such as diabetes, are now able to capture large amounts of health data every day owing to improved medical and consumer sensing technology. These data, known as patient-generated data, have immense potential to inform the care of chronic conditions, both individually by patients and collaboratively by patients and clinicians. Despite the increasing ability to capture personal health data, informatics tools provide limited support to enable routine use of data for disease management. Lack of support for making sense of different types of health data challenges informed decision-making and results in missed opportunities for improving care, leading to suboptimal control and poor health consequences. Motivated by these problems, my dissertation examines the data practices and decisional needs of patients and clinicians to design novel tools for the presentation of multidimensional health data and evaluates these tools in the context of Type 1 diabetes. It employs several qualitative methods that include interviews, observations, focus groups, diary study, think aloud sessions, and user-centered design.

By examining how patients and clinicians interpret multiple streams of data from continuous glucose monitors and insulin pumps, I synthesized the episode-driven sensemaking framework, a novel framework that describes the different analytical stages through which multidimensional health data is made actionable. My work describes the four analytical stages of the episode-driven sensemaking framework that include episode detection, episode elaboration, episode classification, and episode-specific recommendation generation. I show that the episode-driven framework provides a promising basis to guide the design of tools for data-based sensemaking and decision-making as the different stages of the framework lend themselves to opportunities for combining computational and user agency in different ways.

By examining existing data review platforms, I show that the exploratory nature of these tools makes them underutilized by lay users like patients, in addition to resulting in negative experiences, such as cognitive burden, misinterpretation, and misrepresentation of reality. Given the limitations of exploratory tools, the potential of the episode-driven framework in providing a

basis for tool design, and the promise of data-driven narratives in communicating data to the lay users, I designed episode-driven data narratives to help patients review data from continuous glucose monitors and insulin pumps. An exploratory comparison of the episode-driven narratives with the commercially available data review platforms shows that the former improved data comprehension and patients' ability to make decisions from data; and lowered the cognitive load of engaging with data. Additionally, in nuanced ways, episode-driven narratives enabled user agency in making decisions for self-care.

Based on multiple studies to examine practices, and design and evaluate tools, I suggest that to support people in effectively leveraging multidimensional data for managing chronic conditions, tools must do the following - support effective problem-solving with data by creating a shared understanding of data between stakeholders, enable different types of assessments from data and help connect those assessments, and guide analytic focus using a scaffold (e.g., an episode-driven workflow) to organize and present evidence. One promising approach to implement these suggestions in the design of a tool is an episode-driven data narrative, an embodiment of the episode-driven sensemaking framework using narrative visualization techniques. By supporting the generation and presentation of episode-driven narratives from multidimensional data, tools can augment patients' abilities to effectively inform self-care of chronic conditions with their data.

Chapter 1 Introduction

Rapid improvements in consumer and medical sensing have enabled patients with chronic conditions to continuously capture data on several health indicators. Consequently, patients can now generate health data everyday outside of clinical settings. This data, called patient-generated health data, carries immense potential to improve the care of chronic conditions - by bridging the gap between home and clinic; and by enabling frequent data-driven decision-making [41,131]. While collecting data is becoming increasingly easier with advanced sensing technology, the use of data remains challenging because of the lack of tools to make this data useful and limited empirical understanding of the needs and challenges of different data consumers - patients and clinicians [5,50,100].

With the increasing adoption of self-monitoring devices, as the volume of patient-generated data continues to increase, patients and clinicians will need tools to make sense of multiple data streams and to translate data to actionable care decisions [32]. Such tools will need to be designed with the understanding of the various data consumers – patients, clinicians, and caregivers; their varying needs; their objectives – individual and collective; and their different expertise [5,166]. Given the diversity of data and its users, these tools must play different roles in response to users' needs and abilities. These roles could range from putting patients and clinicians in charge of decision-making by enabling exploratory or guided data reviews [35,117], by providing coaching and recommendations for care [47] to autonomous systems making therapy decisions on behalf of the patients [18]. This presents an opportunity to explore different classes of tools with varying levels of agency to enable value creation from patient-generated data.

Towards this opportunity, my dissertation focuses on designing systems to support sensemaking and decision-making with multidimensional health data. Broadly, it examines the ways in which patient-generated data is used or could be used by people and by systems to scaffold the ongoing management of chronic health conditions. More specifically, it investigates the analytical practices, workflows, and challenges of translating multidimensional health data to actionable insights; and articulates the design principles for tools to enable this translation. In this dissertation, I present the results of three investigations that helped me formulate a framework representing engagement with patient-generated data, which I call episode-driven sensemaking. I review existing tools to identify ways in which they limit the translation of data to care decisions by not aligning well with users' needs. This presents an opportunity for the episode-driven sensemaking framework representative of user practices to inform the design of tools. In the last phase of my dissertation, I create a novel prototype informed by the episode-driven sensemaking framework and narrative visualization techniques. I conduct a study to qualitatively compare my research prototypes with the commercial data review platforms.

In the rest of this chapter, I first define patient-generated data. I then point to the problems that further motivate my dissertation. Lastly, I provide an overview of my thesis by presenting the thesis statement followed by a summary of different chapters in this dissertation.

1.1 PATIENT-GENERATED DATA (PGD)

Continuous self-monitoring is an important self-care behavior for successfully understanding and managing chronic health conditions including but not limited to diabetes, cardiovascular diseases, irritable bowel syndrome, and affective disorders [126]. The development and commercialization of sensor-rich devices has enabled patients to increasingly collect different types of data about themselves. Majority of this data is collected outside clinical settings; it helps patients understand health-related needs and inform actions to meet those needs in between clinic visits. This data has been termed as patient-generated health data or patient-generated data and it not only includes biological markers captured from standard and prescribed medical devices but also data from commercial tools, such as phones and consumer wearables, and data from self-reports (paper logs, mobile applications for journaling). Patient-generated health data (PGHD) or patient-generated data (PGD), have been formally defined as "*health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information—created, recorded, gathered, or inferred by or from patients or their designees (i.e., care partners or those who assist them) to help address a health concern*" [158]. According to Shapiro et al.'s white paper, PGD is different from data generated in clinical settings or during patient-provider interaction in two ways – patients are responsible for collecting this data, and they direct the dissemination of this data to healthcare providers and other relevant stakeholders. PGD could include – vital signs (blood glucose, heart rate blood pressure), lifestyle data (diet, exercise, hydration, activity), perceived quality of life data (mood, pain, sleep, social activity), data unrelated to health that increases providers' understanding of the patient on a "personalized individual basis" [158].

Personal health data captured outside of clinical settings, and information and communication technologies to make this data continuously available to both patients and providers, offer an opportunity for closely monitoring disease management in and outside the clinic. It enables capturing critical events or opportunities for improvement that go unnoticed in the absence of data [49]. Considering the importance of continuous care for chronic disease management, the potential to improve control by making informed decisions based on this continuously available data is being considered valuable by the medical community [151]. This potential combined with the need to design systems to facilitate the exchange and use of PGD has also attracted the interest of researchers from the HCI community. Consequently, PGD and its use has opened research opportunities for multiple research communities.

1.2 CHALLENGES OF UNDERSTANDING THE USE OF PGD

Despite the improved capability to capture patient-generated data, practices and tools to harness the potential of these data remain underexplored. There is a limited understanding of how patient-generated data is used or could be used by patients and clinicians in the context of making care decisions and how tools could support them. There are three key challenges towards developing this understanding.

First, there are different types of data including clinical and contextual data that patients may track. While patients and providers understand how to use standard clinical outcome measures to inform care, the use of contextual information that accompanies this data is neither well understood nor sufficiently supported by tools [178]. Second, there are multiple stakeholders that care for a patient – patients themselves, their clinicians, and their caregivers. Distinct consumers of data are guided by different perceptions of the disease and different expertise. Such differences could also reflect in their decisions of collecting and interpreting data and informing care. These differences in turn could challenge patient-provider collaboration, without which effective data-driven disease management is difficult to achieve. It is important to develop an empirical understanding of the common objectives and data practices to design technology that is inclusive of the needs and practices related to individual and collaborative management of chronic conditions and that also bridges the gaps in how different consumers understand data.

Lastly, it is difficult to understand current practices and envision how they might evolve in the context of commercially available tools meant to enable the use of PGD. Real life engagement with these tools remains low – both in and outside the clinic, especially by the patients because the tools are complex [182]. Despite patients being the primary users of the data collection technology (e.g., CGM, insulin pumps, artificial pancreas systems), the accompanying data visualization software tools are primarily designed for clinicians. As a result, patients find it extremely challenging to adopt these tools in everyday management of their disease. This in turn makes it difficult for researchers to explore their practices of reviewing data. Thus, there is a need to develop tools that patients, as non-expert data consumers, can use and that would provide a basis to further understand data practices of patients and guidelines for designing systems [166].

Motivated by the above-mentioned needs for understanding practices and designing tools, my dissertation examines the ways in which PGD can be leveraged by people and by systems. Specifically, it investigates how patients and clinicians make sense of PGD to inform the management of chronic condition; it provides guidelines for designing tools to support the use of PGD; and develops and evaluates approaches for engaging with PGD to inform care. Different types of tools and approaches could support ways in which PGD is incorporated in the ongoing care of chronic conditions. In the scope of this dissertation, I focus on tools for structured sensemaking with multidimensional health data for problem-solving and decision-making. I use interviews, observations, contextual inquiry, and user-centered design and evaluation to understand practices of engaging with data, build tools anchored in those practices, and evaluate those tools.

1.3 THESIS STATEMENT

For data-driven management of chronic conditions, tools need to enable patients and their clinicians in ongoing sensemaking with multidimensional patient-generated data, both individually and collaboratively. One of the shared objectives of sensemaking is problem-solving (identifying problems and solutions from data) from multiple streams of data that include clinical and contextual data. For informing care through multidimensional data, such tools need to support four types of data-driven assessments - data sufficiency, clinical outcomes, patient behaviors, and regimen quality. Such tools should guide analytic focus by organizing and presenting relevant evidence using a practice-based workflow, an example of which is the episode-driven sensemaking workflow where episodes as phases of suboptimal management guide data analysis and reasoning. Tools supporting episode-driven sensemaking with data should – 1) identify and present an episode along with its characteristics, 2) filter evidence based on the episode identified to surface relevant evidence (contextual information and examples of the episode), and 3) suggest interpretive frames that can help explain the episode with the available evidence. A promising technique to implement an episode-driven sensemaking tool for interacting with data is data-driven storytelling. As compared to the commercially available exploratory tools, episode-driven visual data narratives simplify engagement with multidimensional health data; improve comprehension of the data; improve the understanding of potential self-care actions indicated by the data; and impose a lower cognitive burden on the patients. Enabling easy and effective sensemaking with data improves patients' experience of data review and their ability to make decisions, which can potentially increase patient engagement with data and can eventually improve self-management and overall control.

1.3.1 Mapping Thesis Claims to Chapters

Thesis Claim	**Study/Chapter**
For data-driven management of chronic conditions, tools need to enable patients and their clinicians in ongoing sensemaking with multidimensional patient-generated data, both individually and collaboratively. One of the shared objectives of sensemaking is problem-solving (identifying problems and solutions from data) from multiple streams of data that include clinical and contextual data.	Chapter 4
For informing care through multidimensional data, such tools need to support four types of data-driven assessments - data sufficiency, clinical outcomes, patient behaviors, and regimen quality.	Chapter 6
Such tools should guide analytic focus by organizing and presenting relevant evidence using a practice-based workflow, an example of which	Chapter 5 and 7

is the episode-driven sensemaking workflow where episodes as phases of suboptimal management guide data analysis and reasoning. Tools supporting episode-driven sensemaking with data should – 1) identify and present an episode along with its characteristics, 2) filter evidence based on the episode identified to surface relevant evidence (contextual information and examples of the episode), and 3) suggest interpretive frames that can help explain the episode with the available evidence.	
A promising technique to implement an episode-driven sensemaking tool for interacting with data is data-driven storytelling. As compared to the commercially available exploratory tools, episode-driven visual data narratives simplify engagement with multidimensional health data; improve comprehension of the data; improve the understanding of potential self-care actions indicated by the data; and impose a lower cognitive burden on the patients. Enabling easy and effective sensemaking with data improves patients' experience of data review and their ability to make decisions, which can potentially increase patient engagement with data and can eventually improve self-management and overall control.	Chapter 8

Table 1: Thesis claims and the corresponding chapters

1.4 SUMMARY OF THESIS CHAPTERS

1.4.1 Related Work

In chapter 2, I provide an overview of the existing models and approaches for the use of personal health data and point to their strengths and limitations. In chapter 3, I synthesize related work relevant for understanding – a) *practices* of using personal health data, b) *design* of systems supporting the use of personal data, c) *theories* to understand sensemaking with data, and d) *techniques* from the field of information visualization to design displays of data. I draw from literature on personal informatics, chronic illness management, patient-generated data, sensemaking, and information visualization. Specifically, I identify the limitations of existing tools for using personally generated data and describe the emerging but unsupported data practices. I then point to sensemaking theories that provide a lens for better understanding data practices. Lastly, I review prior work in information visualization to identify narrative visualizations and data-driven storytelling as potential techniques for designing displays to enable sensemaking and communication of insights from multi-dimensional health data for patients as non-expert users.

1.4.2 Establishing the Need to Better Understand Analytical Practices

In chapter 4, I report on an exploratory study that provides an initial comparative understanding of how PGD is used in individual and collaborative settings, that is by patients alone and by patients and clinicians together. The study shows that patients and clinicians engage in sensemaking with PGD for the shared objective of problem-solving. Despite having a shared purpose, making sense of the data to identify and solve problems is challenged by the differences in perceptions between patients and clinicians and by the complexity of commercially available tools. In the context of sensemaking and problem-solving as models of chronic disease management, I argue that patients and clinicians execute problem-solving and sensemaking through retrospective data reviews, the analytical processes of which are not considered by these models of disease management. Given the central role of PGD in driving chronic disease management, this study establishes the need to better understand, and articulate data practices and design tools aligned with those practices to support shared objectives and shared understanding of the data.

1.4.3 Understanding How Patients Engage with Multidimensional Data

In chapter 5, I report on a study to understand how patients might interpret their data using simpler displays without the involvement of clinicians, given that they find it hard to use the commercially available displays. For this study, I developed context-enhanced visualizations of PGD that were reviewed by patients and their caregivers. The study demonstrates that using simple displays of clinical data augmented with contextual data, patients were able to make sense of multiple streams of data – by identifying trends, explaining those using the contextual data, and thinking about self-care actions and consequences. Inspired by the similarity of the analytical activities of patients with the data-frame theory of sensemaking, I demonstrate that patients make sense of the data by mapping it to explanatory frames that they are aware of or that the data depicts.

1.4.4 Understanding how Patients and Clinicians Together Engage with Multidimensional Data

In chapter 6 and 7, I investigate how data review and interpretation happen in the collaborative context using commercially available tools. I report on a study where I observed how patients and clinicians engage in biweekly data review sessions as a part of a telemedicine program over a period of 6 months. This study helped identify the analytical workflow employed during the review session, the information requirements to carry out different analytical activities of the workflow, the decisions made, and the challenges encountered. I report these results in two phases. In chapter 6, I report the information needs corresponding to the different types of assessments performed during the data review. These include assessments of data sufficiency, outcomes, patient behaviors, and regimen quality. I then review the commercially available diabetes data platforms in the context of these assessment tasks and the information needs and point to their limitations. I discuss implications for improving the design of existing tools to better align them with user needs. These implications include the need to surface behavior related information (patient and/or system behavior), the need to link different types of information, and the need to provide quick access to relevant information.

In chapter 7, I report on the same study (as chapter 6) to describe the analytical workflow of the data review sessions, the types of decisions that result, and the challenges of making decisions. These analytical activities are episode-driven (where episode is a phase of suboptimal management indicated by outcomes or patient and system behavior) and primarily involve detection, elaboration, classification, and resolution of episodes. Based on the observations from this study, I synthesize a descriptive framework of episode-driven sensemaking. Using this framework as a basis, I provide design requirements for technology to facilitate guided and structured review of data. Collaborative data practices as identified in this study resemble the analytical activities conducted by patients (Chapter 5). One key difference in how patients interpreted the data was around the explanatory frames used. The frames that patients employed were primarily sought or developed from the data. The frames employed by clinicians in the collaborative data reviews were clinical heuristics, templates that map outcome-cause-corrective actions. These heuristics did not emerge from the data through exploration but rather directed what data was important to look at (i.e., data that served as evidence matching the heuristics). They simplified data interpretation and decision-making as they helped improve analytical focus by reducing the need for unnecessary data exploration.

1.4.5 Designing Episode-Driven Narratives and Assessing Them with Patients

In chapter 8, I report on an exploratory study that compared how patients interpret their data using commercially available displays and using the episode-driven data narratives that I created. In this study, I conducted two sessions each with 6 participants. In the first session, participants reviewed data using the commercial PDF reports of continuous glucose monitor and insulin pump data. In the second session scheduled two weeks apart, participants reviewed the same data as the previous session using my research prototype (the episode-driven narratives). In both the sessions participants completed a questionnaire to assess their comprehension and decision-making with data. Additionally, they took an assessment of task load and were interviewed using a semi-structured interview protocol. I also obtained an assessment of diabetes numeracy and graph literacy for each participant. The study found that the average scores obtained on the questionnaires completed based on the prototypes was higher than when completed based on the commercial PDF reports. The average task load of completing the questionnaires with the prototypes was lower than completing the questionnaires with the PDF reports. Two participants with relatively higher diabetes numeracy and graph literacy (highest that was scored) also obtained the maximum score on both the platforms and had lower task load scores with the prototype. From the interviews, the study found that the episode-driven narratives enabled better navigation of data, better interpretation of patterns, and improved comfort and confidence in coming up with plans for actions. The study also identified four dimensions across which participants' needs for information varied. Based on these findings, I discuss implications for designing interfaces for data-driven sensemaking and decision-support tools for patients to make personal health data actionable.

1.4.6 Contributions and Future Work

In chapter 9, I bring together the key learnings from these studies and present opportunities for future work. The contributions of this dissertation are as follows – 1) a novel framework of episode-driven sensemaking with personal health data that provides a basis to design tools for data-driven management of health, 2) empirical understanding of nuanced data practices of patients and clinicians, 3) principles for designing data interfaces to make personal health data actionable, 4) a novel prototype to enable interpretation of multidimensional data from medical devices, and 5) directions for future research. For future work, I discuss the opportunity to investigate how patients as users of automated therapy systems, such as the artificial pancreas, and clinicians as users of automated diagnostic systems, experience the automation provided by these systems. In the light of growing evidence that the black box nature of such systems restricts their use, there is an opportunity to understand how data interfaces can be designed to make these systems (and their automated outputs in the form of recommendations and diagnosis) intelligible to their users. Learnings from this dissertation and prior work on explainable artificial intelligence, especially in the context of clinical decision support tools, provide a foundation for designing novel data interfaces to improve system intelligibility. Additionally, I discuss future work to extend the prototypes from this dissertation to create a functional mobile app for a field study.

Chapter 2 Models of Personal Informatics, Chronic Disease Management, and Data-Driven Reasoning

This chapter reviews frameworks related to collection and use of personal data. One group of models reviewed includes data tracking and reflection by data enthusiasts while the other group involves the use of data by patients managing chronic conditions. It is important to understand these frameworks because they provide a lens to interpret studies in this domain, with some of the reviewed studies using these frameworks to guide their data analysis [37].

2.1 MODELS OF PERSONAL INFORMATICS

The term "personal informatics" was coined by Li et al. They define personal informatics systems as systems that "*help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge*". They proposed the stage-based model of personal informatics that describes five stages involved in collecting and reviewing personal data – preparation, collection, integration, reflection, and action (Figure 1). Each of these stages has its own challenges, as described in Table 2. Barriers in one stage affect the subsequent stages.

For example, not being able to collect sufficient data, creates issues when reflecting on sparse data. Similarly, not making correct decisions about what to track renders data useless later. This implies that while designing personal informatics systems, it is important to consider all these stages together instead of focusing on one stage. Additionally, these stages are iterative, that is users could change their mind about what data to collect, and what tool to use as they progress through the stages and switch between them as needed. Lastly, the study points to the need to balance system driven and user driven approaches to support different stages and the need to make such systems multi-faceted to support reflection on different aspects of a user's life (e.g., understanding multiple factors that affect a specific behavior) [106].

Stage	**Description**	**Barrier**

Preparation	Deciding what data to collect and how to collect guided by the motivation behind tracking	Wrong decisions about what to track and how to track
Collection	Capturing different types of data with varying frequency	Forgetting, lack of tools to record data, accuracy issues, understanding what data is worth recording, keeping up the motivation to collect data
Integration	Organizing, combining and transforming data for reflection	Aggregating data from multiple sources, normalizing across different formats
Reflection	Interacting with visualizations or looking at the list of collected data in short term or long term	Difficulty of interpretation, not having enough data, specially lack of context data, not finding data useful, data evoking emotions that make reflection difficult
Action	Understanding what to do based on insights obtained from data	Unsupported by tools

Table 2: Barriers associated with different stages in personal informatics.

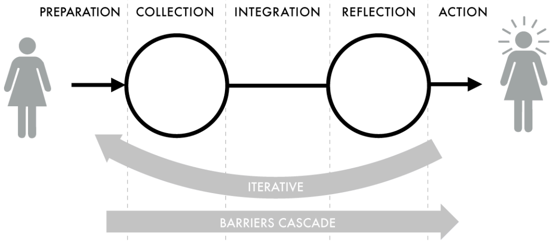


Figure 1: The stage-based model of personal informatics proposed by Li et al. (Li, Dey, and Forlizzi 2010)

Epstein et al. further extended this model to incorporate lived experiences of tracking that involve lapsing when tracking and switching tools or goals for tracking. In describing the reasons for lapsing, Epstein et al. note that people at times temporarily stop tracking because they might already know the data. For example, one participant in the study ran the same loop everyday so he did not need to track the distance. Another reason to pause tracking was that they do not want to track during a specific time. For example, vacation was one time window during which participants did not want to track [55]. In a similar study, Rooksby et al. described varying needs to track - at times people track a lot while at other times they would track little and that different stages of personal informatics (collection-integration-reflection) proposed by Li et al. occur simultaneously and overlap [147].

2.2 CHRONIC DISEASE MANAGEMENT FRAMEWORKS

Models of chronic disease management describe how disease is or should be successfully managed by patients. This section describes two such models - problem-solving and sensemaking. The problem-solving model was proposed by Hill-Briggs and accounts for how patients with chronic conditions adhere to prescribed regimens while identifying and overcoming barriers to daily self-management of their disease. According to this model (Figure 2), problem-solving for successful chronic disease management includes four components: a) problem-solving process, b) problem-solving orientation (self-efficacy to solve problems), c) disease specific knowledge, and d) transfer of past experiences. The problem-solving process includes identifying a problem, generating alternative solutions, choosing a solution to implement, implementing the solution, and evaluating if the solution worked as expected [74].

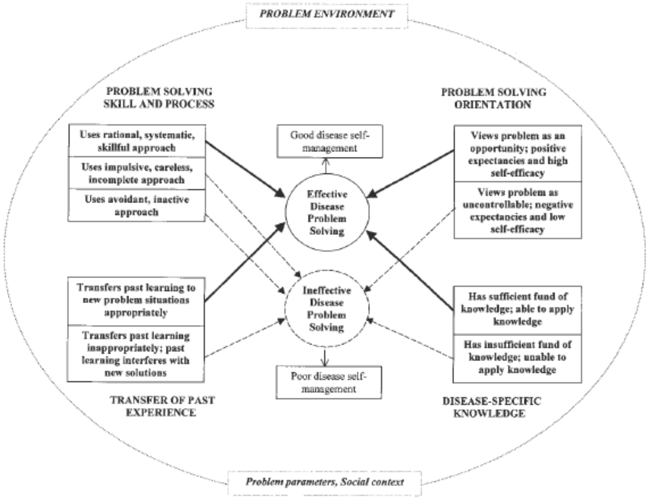


Figure 2: The problem-solving model proposed by Hill-Briggs (Hill-Briggs 2003)

The sensemaking framework was more recently proposed by Mamykina et al. to account for the continuous flow of information and experiences while managing chronic conditions (Figure 3). Mamykina et al. argue that diabetes management requires switching between two modes of operation – habitual mode and sensemaking mode. The habitual mode involves performing routine actions that happen in the absence of breakdowns. The sensemaking mode is triggered by breakdowns when routine actions need to be replaced by experimental actions. Drawing upon Weick's characterization of sensemaking as a retrospective activity in response to knowledge gaps [177], this framework describes three sensemaking activities that patients engage in to continuously tackle self-care needs – perceiving new health related information signifying a gap in understanding, situating new information within one's knowledge and experiences to either construct a new model or activate old models that explain the current situation, and an action to test new models or a routine action in response to the situation. These three activities have been called perception, inference and action. [122].

These models have two limitations. First, while problem-solving and sensemaking both involve the use of patient-generated data, these models do not clearly account for how different steps of problem-solving and sensemaking could be realized using PGD. A few recent works, described in the subsequent chapters, have started to unpack how the constituent activities of these frameworks could be supported through patient-generated data [37,118]. Second, both these models describe individual disease management and do not account for the collaboration involved in chronic care (collaboration with providers and caregivers).

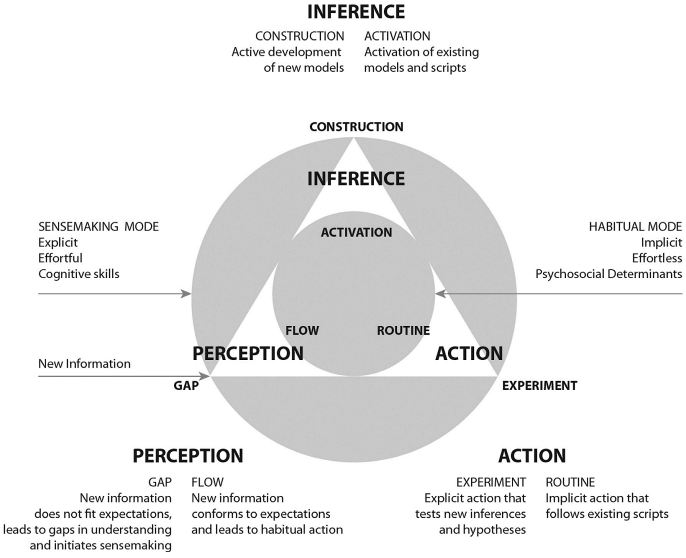


Figure 3: Mamykina et al.'s sensemaking framework (Mamykina, Smaldone, and Bakken 2015)

2.3 APPROACHES FOR DATA-DRIVEN UNDERSTANDING OF DISEASE

Two frameworks describe different ways in which patients can use their data to understand their disease - self-experimentation framework and the personal discovery framework. The self-experimentation framework was proposed by Karkar et al. to guide patients with irritable bowel syndrome to test food triggers. While the authors describe the self-experimentation process to consist of three steps – generate hypothesis, test hypothesis, and target behavior change interventions, they focus only on supporting hypothesis testing. They describe hypothesis testing to involve selecting dependent variables (symptoms, health outcomes), and independent variables (causes, triggers), and testing the independent variables by measuring the dependent variables (Figure 4). [87].

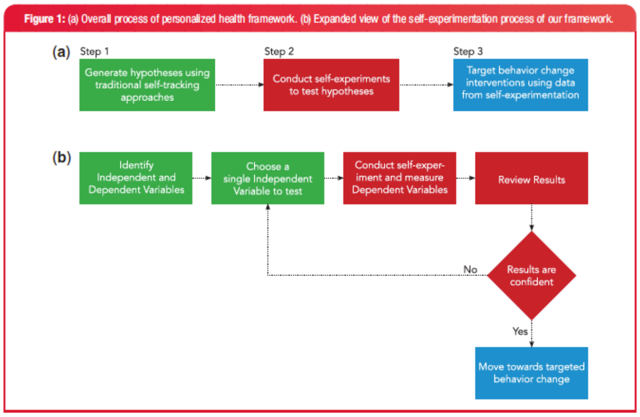


Figure 4: Karkar et al.'s self-experimentation framework for hypothesis testing (Karkar et al. 2016)

In understanding how patients make personal discoveries using their data, Mamykina et al. describe 4 steps - feature selection, hypothesis formulation, hypothesis evaluation, and goal specification (Figure 5). Feature selection does not only involve choosing an activity, but also selecting a feature of that activity to track (e.g., properties of physical activity – duration, time of day, type of activity). Figure 2.5 describes the details of the other steps. These steps emerged from the observation of a 4-week diabetes self-management and education program created to educate patients with best practices to personal discovery through self-monitored data. Although this study did not aim to investigate collaborative practices, the authors report that the help of educators was crucial in guiding patients to follow the abovementioned steps towards insight discovery [118].

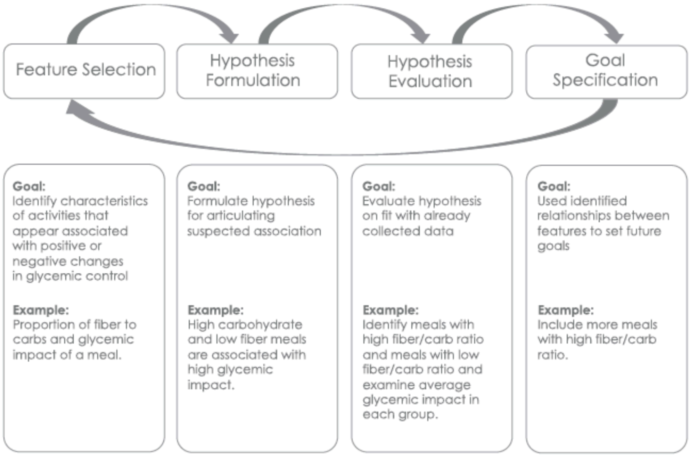


Fig. 1. Stages of discovery with personal data.

Figure 5: Mamykina et al.'s framework for data-driven personal discovery (Mamykina et al. 2017)

2.4 CONCLUSION

The frameworks conceptualizing chronic condition management (problem-solving and sensemaking) provide a prescriptive model of how sensemaking should happen to inform chronic disease management. They do not explain sensemaking or problem-solving as they happen through continuously generated personal health data. Towards this gap, the data-frame theory of sensemaking provides the understanding of how data can facilitate sensemaking in different domains. My dissertation brings these two frameworks together to better understand and articulate data-driven sensemaking practices for chronic disease management and to design tools to support the use of personal health data.

The approaches for enabling data-driven understanding of the disease describe specific ways in which personal data or patient-generated data can be leveraged by patients or self-trackers for different objectives, such as self-experimentation and self-discovery. While these objectives become salient at different times during the management of chronic conditions, they do not target the need for regular data-driven assessment of chronic health conditions or regular problem-solving. Based on an empirical understanding of data practices, my dissertation formulates an episode-driven approach for engagement with personal health data to enable regular data reviews. Additionally, it creates novel data interfaces implementing this approach to enable sensemaking and decision-making with multidimensional data.

Lastly, these frameworks do not account for collaborative practices around patient-generated data. Since collaborative investigation with data can guide patients in discovering insights as demonstrated by Mamykina et al.'s study [118], there is value in understanding how patient-generated data can be incorporated in routine care to enable patient-provider collaboration to improve care and meet the care goals. Specifically, there is an opportunity to involve providers in guiding self-experimentation for objectives, such as hypothesis testing for diagnosis, experimentation to understand optimal treatment, and other forms of structured investigation with data. Such exploration has the potential to inform system design and can also help extend the above reviewed conceptualizations to depict collaborative practices of collecting and reviewing data. This dissertation contributes a framework for episode-driven sensemaking that is descriptive of individual and collaborative practices of using PGD to inform care.

Chapter 3 Related Work

In this chapter, I synthesize related work that is relevant for understanding *practices* of using personal health data, *design* of systems supporting the use of personal health data, *theories* to understand sensemaking with data, and *techniques* from the field of information visualizations to design displays of personal health data. Specifically, I review literature on *personal informatics* and *patient-generated data* to identify the limitations of personal informatics tools and to describe the emerging but unsupported practices of using patient-generated data. I then point to the literature from the field of *sensemaking* to describe the theoretical lens for better understanding practices of using data. Lastly, I review prior work in *information visualization* to identify techniques for designing displays and explaining them to enable sensemaking and communication of insights from multi-dimensional health data.

3.1 CHAPTER SUMMARY

A wide variety of personal informatics systems have been studied by prior work including systems for free form reflection, systems with glanceable displays for feedback, visual analytics systems, and more recently, systems for structured investigation with data through systematic data collection and review (e.g., self-experimentation, problem-solving, behavior change plan creation). Despite immense research in the domain of personal informatics, there is a lack of guidelines for designing personal data visualizations. One of the reasons for limited guidelines is perhaps that studies provide little rationale behind their selection of visualization strategies. Additionally, even though researchers have employed a variety of visualizations in different types of systems, there is limited understanding about the effect of these visualizations in improving comprehension or decision-making, the cognitive burden their use imposes, and the overall experiences of users. A few other limitations of different types of systems include lack of support for interpreting multidimensional data with more focus on exploratory visualizations than explanatory visualizations, disregard for the processes through which data-driven reflection unfolds, and lack of support to interpret system generated insights for their translation to real actions. There is a need for research to better understand existing data practices and workflows, and to create data interfaces that scaffold the process of translating data to decisions.

Systems leveraging patient-generated data include clinical decision support tools and telemonitoring systems. Their use involves multiple challenges, many of which have also been reported in the use of personal informatics systems - complexity of data presentation, misrepresentation and misinterpretation of lived experiences, difficulty achieving shared understanding of the data because of the exploratory nature of tools, and the black box nature of automated insight generation. Additionally, the use of these tools is primarily provider-driven with patient involvement being limited to data collection. There is a need for designing patient-facing tools to support individual use of data with the potential for provider involvement to support collaborative use. Better understanding of individual and collaborative practices around the use of patient-generated data can provide a basis to design tools that align with data practices within different settings.

Sensemaking theories provide a lens to understand the data interpretation activities of users. Sensemaking has also been described as an approach for chronic condition management. The sensemaking framework for chronic condition management provides a prescriptive model of how sensemaking should happen to inform chronic disease management but it does not explain sensemaking as it happens through continuously generated personal health data. Towards this gap, Klein et al.'s data-frame theory of sensemaking provides an understanding of how data can facilitate sensemaking in different domains. I bring these two theories together and contribute a descriptive understanding of data-driven sensemaking practices for chronic disease management, that is how personal health data from wearable devices informs self-care.

Research in information visualization has identified the need to design visualizations for lay users, specifically focusing on personal data and mobile interfaces. Data-driven storytelling using narrative visualizations (e.g., slideshows, comics, posters, etc.) has emerged as a powerful communication technique to engage non-expert consumers of data. In a similar direction, research on patient-facing visualizations has called for designing information rich displays or infographics that communicate the meaning of the data upfront without relying on the data literacy skills of the patients. Data-driven storytelling is a promising technique to explore in the context of presenting health data to lay users as the objective of data stories is to package and present information in a way that makes it easy for the user to understand, retain, and use information for decision-making. Lastly, patient-facing displays of data focus primarily on single variables, creating an opportunity to explore the design of multidimensional data interfaces. Visual storytelling provides ways to combine different types of data into a meaningful narrative, which makes it a suitable technique to explore for engagement with multidimensional health data that is underexplored by prior work.

Research on explanability of AI-generated insights has established the usefulness of supplementing insights (goal recommendation, stress prediction) with explanations to improve users' experience of the system. However, the role of explanations is understudied in the context of chronic condition self-management systems. I draw from this body of work to create explanatory interfaces and understand my participants' experiences with these interfaces.

Informed by the limitations of existing tools and the understanding of data practices in individual and collaborative settings, my dissertation contributes a descriptive model of data practices, contributes novel designs of data interfaces using explanatory visualization techniques, and contributes design principles for creating patient-facing data interfaces and explanations, and more broadly for systems to enable the use of personal health data and making it actionable.

3.2 PERSONAL INFORMATICS AND HEALTH SELF-MANAGEMENT SYSTEMS

3.2.1 Systems Presenting Multiple Data Streams for Free Form Reflection

To help people reflect on multiple streams of personal data, earlier studies in personal informatics and chronic condition management have explored systems to support free form reflection by having users collect relevant data (considering collection and reflection happen in tandem), and/or by presenting their data using a variety of visualizations. Data collection and presentation approaches primarily include:

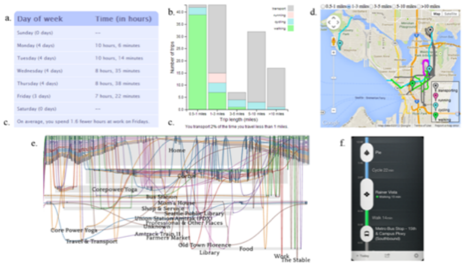
- capture and display of contextual data in addition to biomedical data – images, location, social context, mood, and personally defined tags [121,134,159,164,188],
- visualizing several subsets of self-tracked data (called visual cuts) to provide more than one perspective on data [56],

- translating lifestyle data into slideshow based stories to communicate the temporal relevance of data (what data becomes important at what time) [137],
- presentation of correlations using graphs and natural language to communicate insights from multiple data streams [85,108,169],
- presentation of multiple streams as stacked time series [90], and
- presentation of outcomes categorized by certain factors affecting them [33,108,188].

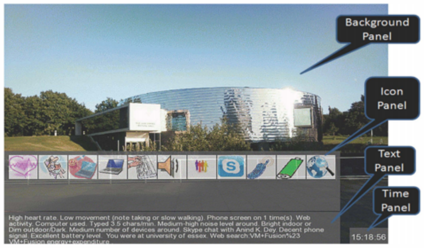
Most of these systems are exploratory in nature except for systems presenting correlations using natural languages. Users prefer to view both automated insights and the raw data behind those insights, but studies have not explored ways to systematically organize and present both types of information. Explanatory interfaces can be employed to present natural language insights and explain them using actual data.

3.2.2 Systems with Glanceable Displays

Another major class of tools have been designed to target behavior change by improving awareness through feedback on key metrics associated with a behavior (e.g., step count to improve physical activity). Different types of visual metaphors are used to provide feedback - showing activity data using a virtual fish [110] and a virtual garden [43], eco-feedback technologies to provide feedback on transportation habits using virtual trees [62] and on water consumption [63]. While these feedback techniques have shown to be effective for changing user behavior, they are not designed to help understand multi-dimensional data. Additionally, given their focus on assessing user performance, they are perhaps more suitable for supporting general wellbeing as opposed to self-management of chronic conditions as people with chronic conditions are emotionally involved in the disease management process and receiving continuous feedback on their performance could have negative emotional consequences.



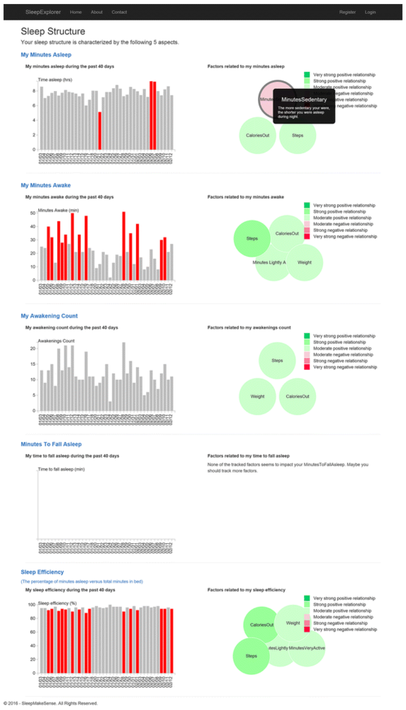
a) Visual Cuts [56]



High heart rate. Low movement (note taking or slow walking). Phone screen on 1 time(s). Web activity. Computer used. Typed 3.5 chars/min. Medium-high noise level around. Bright indoor or Dim outdoor/Dark. Medium number of devices around. Skype chat with Anind K. Dey. Decent phone signal. Excellent battery level. You were at university of essex. Web search:VM+Fusion%23 VM+Fusion energy+expenditure.

Figure 2. Example of a story event multi-media object

b) Lifestyle Stories [137]



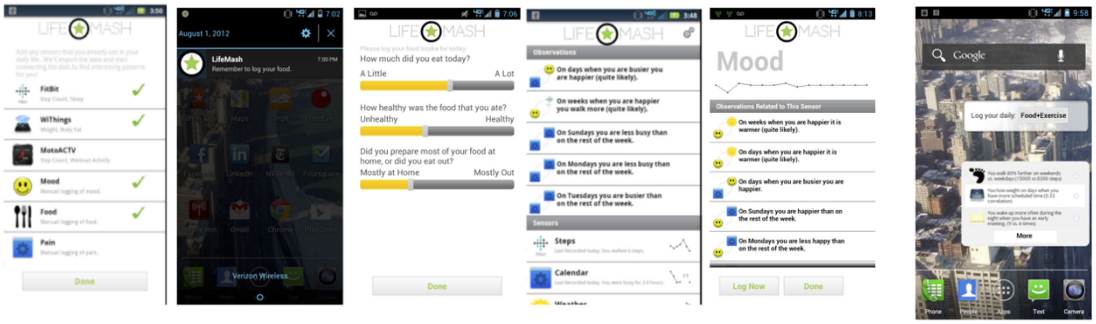
c) Sleep Explorer [108]

Fig. 6. Early Screenshots of Personalized Graphical Feedback Derived from IoT Device Collection and Analysis

1) My Day (Zhang et al., 2018)

Figure 2: MAHI website (screenshot of the actual site usage). The columns include: 1) record number, 2) date and time of capture, 3) blood glucose value, 4) picture(s), 5) audio, 6) participant's comments posted directly to the website, 7) educator's comments posted directly to the website.

e) MAHI [120]



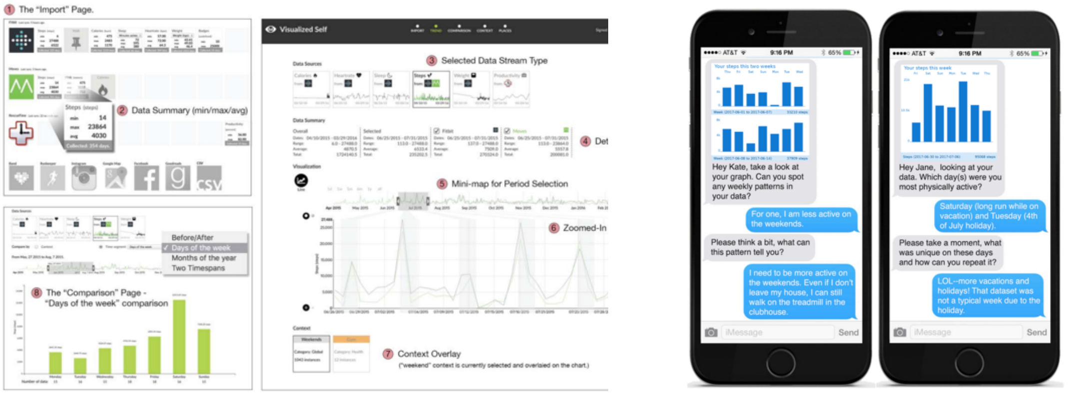
f) Health Mashups [11]

Figure 6: Different types of tools to enable reflection on multidimensional data

3.2.3 Exploratory Systems to Support the Process of Reflection

In the context of above-described systems, the lack of understanding of reflection as a phenomenon has been noted in prior work. Specifically, studies have identified what insights users want but have not understood different ways in which users could obtain those insights from their data [10]. A few studies have explored how patients reflect on their data to inform action. For example, Mamykina et al. developed MoDD, a web-based tool built on a knowledge base of a set of common problematic glycemic patterns, tips and behavioral triggers for these patterns, and guidance on action to resolve problematic patterns. They found that the problem-solving process (identifying a problem, devising alternate solutions, implementing a solution, and evaluating the results [74]) was one way of reasoning when diabetes patients reflected on their data [117]. In a 4-week study, patients with type 2 diabetes were asked to use MoDD and follow these steps – a) pick a problematic glycemic pattern that they wanted to fix (e.g., high BG number after breakfast), b) pick a causal behavioral pattern from a knowledge base created by researchers (e.g., lack of protein for breakfast), c) select an alternative behavior and set an action-oriented goal every week (e.g., have eggs for breakfast), and d) implement new behavior and monitor changes in BG number to see if the goal is being achieved. MoDD's objective was to educate patients on commonly known problems and reasons for those problems by engaging them with BG data and with a knowledge base of educational information. MoDD was an exploratory tool that had limitations, such as patients generating unreliable insights and the burden of exploration. The episode-driven data review as a basis for tool design aims to facilitate more structured engagement with diabetes data.

Another study found that patients seek cause and effect relationships from their data through four phases – feature selection, hypothesis generation, feature evaluation, and goal specification [118]. To fill the phenomenological gap [10] in the understanding of reflection, a few studies have explored techniques such as visual data exploration [35] (exploring data through interactive visualizations) and conversational agents [97]. The systems described above were designed to be primarily exploratory, making it less clear as to what extent lay users may benefit from such systems, given that exploratory data review can be burdensome and time taking.



Hey Kate, take a look at your graph. Can you spot any weekly patterns in your data?

For one, I am less active on the weekends.

Please think a bit, what can this pattern tell you?

I need to be more active on the weekends. Even if I don't leave my house, I can still walk on the treadmill in the clubhouse.

Hey Jane, looking at your data. Which day(s) were you most physically active?

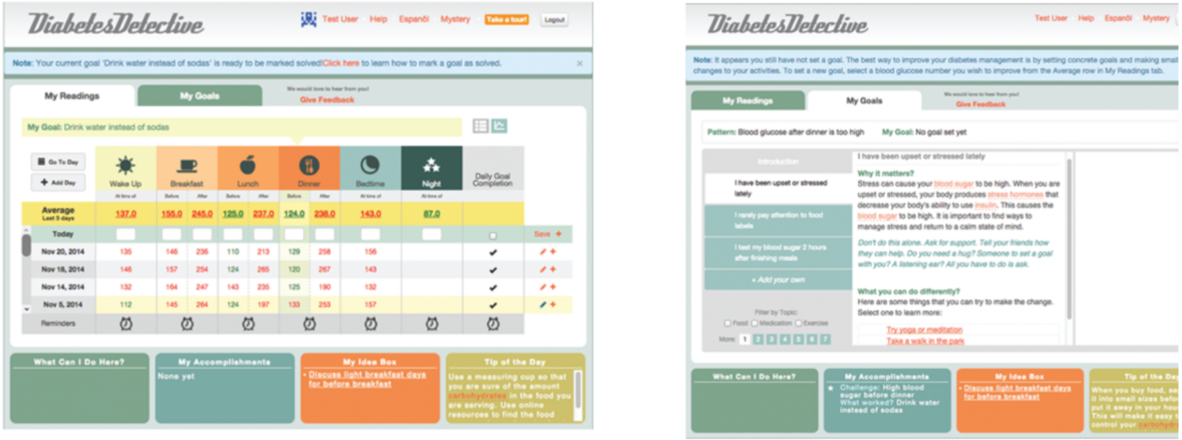
Saturday (long run while on vacation) and Tuesday (4th of July holiday).

Please take a moment, what was unique on these days and how can you repeat it?

LOL--more vacations and holidays! That dataset was not a typical week due to the holiday.

a) Visualized self [35]

b) Reflection Companion [97]



Note: Your current goal 'Drink water instead of sodas' is ready to be marked solved!Click here to learn how to mark a goal as solved.

My Goal: Drink water instead of sodas

Note: It appears you still have not set a goal. The best way to improve your diabetes management is by setting concrete goals and making small changes to your activities. To set a new goal, select a blood glucose number you wish to improve from the Average row in My Readings tab.

Pattern: Blood glucose after dinner is too high My Goal: No goal set yet

I have been upset or stressed lately

Why it matters?
Stress can cause your blood sugar to be high. When you are upset or stressed, your body produces stress hormones that decrease your body's ability to use insulin. This causes the blood sugar to be high. It is important to find ways to manage stress and return to a calm state of mind.

Don't do this alone. Ask for support. Tell your friends how they can help. Do you need a hug? Someone to set a goal with you? A listening ear? All you have to do is ask.

What you can do differently?
Here are some things that you can try to make the change. Select one to learn more:
Try yoga or meditation
Take a walk in the park

c) MoDD: Display of BG readings in different labelled time windows along with before and after annotations for meals [117]

d) MoDD: 4b – Display for goal setting in response to a pattern. Goal setting involves going through standard educational information about different factors - how they affect blood sugar numbers and what can be done to control them. [117]

Figure 7: Exploratory systems for reflection

3.2.4 Systems to Support Structured Investigation Through Data

More recently, tools are being designed to support structured investigation with data to obtain and validate insights, as opposed to promoting free form reflection. These tools help the users follow specific steps for different objectives – simulating problem-solving [117], identifying triggers for symptoms [86], tracking to obtain recommendations and evaluating those recommendations [47], and creating and evaluating behavior change plans based on a guiding framework [104]. Some of these tools automate most of the steps and guide the user through each of these steps, while the others instruct the user to follow the steps manually by using informational support from the tool. This shift from free form reflection to structured use of data to test hypothesis or evaluate behavior change plans and recommendations has offered improvement in guiding patients to gather data and understand associations across different types of data. However, interpretation related issues (e.g., misinterpretation, handling of counterintuitive insights) were found in the use of these systems indicating the need to better guide users in interpreting the results of automatically generated insights obtained by systematic data collection [86,104,117]. Taken together, these studies offer evidence that using personally collected data to conduct structured investigation with the aim of understanding problematic contextual factors, understanding steps towards obtaining a desired outcome in this regard, implementing those steps, and evaluating the consequences is promising. However, tools need to explore different ways of supporting this process to aid data interpretation and actionability not only from data that is systematically collected but also from data that is casually monitored using personal sensing devices.

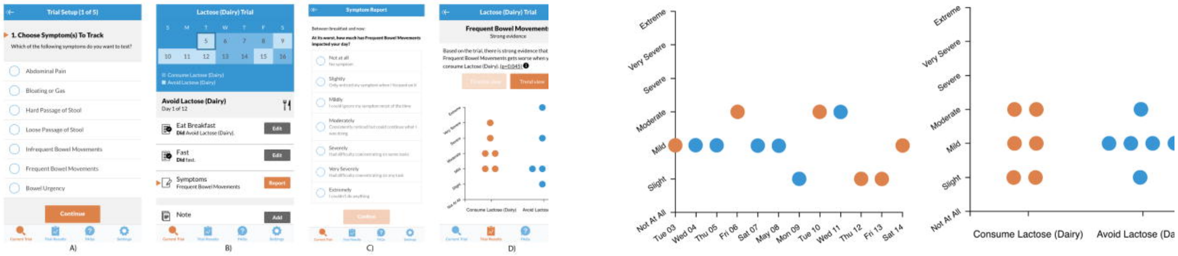


Figure 8: Tummy trials system for identifying food triggers through self-experiments (Karkar et al. 2017)

3.2.5 Key Findings from Prior Work

Although the design space of tools to support interpretation of contextualized health data remains largely underexplored, evaluation of the above-described personal informatics tools points to opportunities for studies and requirements for system design in the context of this research gap. I describe them next.

Different views of data provide different insights and are relevant under different conditions

When working with multiple types of data streams, different views of data can provide different insights - high level in terms of aggregate data, experiences and external contexts, or low level in terms of the actual data. The relevance of these insights for the users depends on their preferences and needs and since these change over time, tools should support flexibility to switch between different levels of reflection and should be sensitive to the context of information need when presenting insights to the users [35,56,169].

Automated insights may not be self-explanatory and need to be explained to users

Insights generated by automated analyses of data might not always be representative of users' lived experiences and might be misinterpreted by them (e.g., oversimplification of correlations), indicating the need for tools to not only provide insights but provide information or guidance for users to draw meaning from them [85,108]. This finding is also echoed in the literature on the use of intelligent systems and has resulted in much work on explainable artificial intelligence [176]. Several models and techniques for improving intelligibility of how systems generate specific insights or results have been studied. The presentation and explanation of health-related data-driven insights can potentially borrow from this body of work and contribute to it. The most relevant to this work is the study of intelligent clinical decision-support systems. Notable findings from the use of such systems include their inability to work with data in the wild, clinicians' need to better understand system behavior and decision-making, and decontextualized decisions that disregard the role that contextual data plays in making situated and relational health-related decisions [123,186]. There is a need to design explanatory interfaces to better communicate automatically generated insights.

There is a need to systematically explore visualization strategies for lay users

For chronic condition management, studies of commercially available mobile applications have noted that users do not always find their data presentations helpful [88,114]. Existing data presentations do not address the cognitive burden imposed by their use, resulting in disengagement or underutilization of the data [182]. In a similar direction, Nunes et al. reviewed self-care tools for chronic condition management explored by researchers to identify that while visualizations are heavily used to communicate data, the displays explored in different studies were very diverse, perhaps signifying a lack of guidelines for designing effective visualizations in this domain [132]. They suggest that the visualization strategy chosen could affect how patients understand their data and indicate the need for researchers to systematically explore different visualization strategies. However, existing studies on personal informatics systems provide limited explanation of their visualization design choices. As a part of this systematic exploration, there is a need for personal informatics researchers to draw from the field of information visualization to identify techniques to better communicate and present health data. Information visualization literature has also noted the need to improve the design of personal data visualizations, as described in section on information visualization research later in this chapter [79].

3.2.6 3.2.6 Open Challenges for Presenting and Communicating Data-Driven Insights

In this section, I discuss open challenges in the context of personal informatics and data visualizations that are relevant for this dissertation.

Are more insights and complex visualizations always better?

While data visualizations can be helpful in communicating several types of insights, depending on the type and the amount of data, they could become increasingly complex [155] and overwhelming for the non-expert data users. Additionally, certain visualizations might not scale well, especially across several device types and could place a high cognitive demand on their users, more than what lay users might prefer. There is a similar problem with systems that present automated insights beyond visualizations, such as natural language summaries. Increase in the amount of data could mean more insights, not all of which the users might find novel or worth their attention (e.g., obvious insights [169]. How can we help users and/or systems identify insights or data that is worthy of attention without exhausting users' cognitive resources in unnecessary data exploration? And what visualization techniques might be suitable to design for lightweight yet informative engagement with data? Answering these questions is necessary as designers envision to support users in sustainably drawing value from data over a lifetime of chronic condition management. My dissertation contributes an episode-driven approach for compiling meaningful modules of information from multivariate data and designs novel explanatory interfaces to present this information.

What information other than "insights" should be presented to aid interpretation?

In chronic conditions, outcome measures of interest are often affected by several factors together. Systems simply presenting monitored data or insights without the knowledge required to interpret the two (e.g., standard and/or personalized effect of a certain factor on a specific measure) may reinforce patient's biases [120] and may result in users drawing unreliable inferences [119]. Even automated mechanisms to identify insights from data may result in insights that are not representative of users' lived experiences [85,86]. This could potentially result in misinterpretation or mistrust in the system, indicating the need for tools to not only provide insights but to communicate the provenance of those insights with evidence and supplement insights with expert knowledge through templates or through active involvement of experts where domain expertise is important for interpretation.

What does it take to make the data actionable?

Several studies have resulted in implications for improving actionability of self-tracked data [38,155]. One reason perhaps for limited actionability afforded by these systems is that they promote reflective thinking without necessarily understanding how reflection happens to make data actionable [10]. The design of data representations explored in personal informatics research disregard the analytical processes involved in translating data to an insight and insight to action. For example, if two streams of data result in counter insights suggesting different actions in relation to an outcome, which one should the user opt for [85]? Existing approaches fall short of helping users answer such nuanced questions, which would entail supporting information presentation and processing in ways beyond what is supported by current tools – e.g., presenting data to compare the two scenarios and to show which one is more problematic and/or more likely to happen. While visual analytics systems that allow for data exploration could help some users [35], they also add to the burden of reflection and leave behind non-expert users. How then can visualizations or sequence of visualizations help improve actionability in scenarios similar to the above without requiring immense time, attention, and skills from the user? Towards this, system design could borrow from mechanisms used in visual analytics systems for guiding data experts in generating insights. These guidance mechanisms provide varying extent of support and include orienting, directing and prescribing [26]

What skills should the user have?

Data visualizations of multi-dimensional data might require skills for interpretation that patients as non-expert users may not have. This points to the need to design visualizations without necessarily requiring the users to be skilled at reading and interpreting data [102]. To design systems that do not rely on a user's literacy skills, designers need to strike a balance between communicative and analytic aspects of the visualizations, that is communicating insights upfront versus relying on user's visual analysis of the data to extract the insights.

3.3 PATIENT-GENERATED DATA (PGD)

In this section, I describe research on collaborative use of patient-generated data, the analytical workflows involved in making sense of patient-generated data, and decision-support tools built using patient-generated data. By providing an episode-driven framework for sensemaking, my dissertation contributes to the growing understanding of different workflows and approaches of using patient-generated data.

3.3.1 Collaborative Use

Chung et al.'s work demonstrates that patients and clinicians collaborate across all the stages of the personal informatics model (preparation, collection, integration, reflection, and action) by creating boundary negotiating artifacts that enable collaboration among data users with different expertise [37]. Mentis et al. demonstrate that patients and clinicians co-construct the meaning of data to enable a shared understanding by creating certain views of the data that fit specific objectives [127]. However, developing a shared understanding is challenged when patients' and clinicians' goals and motivations for tracking data are misaligned. For example, one study found that clinicians were motivated to understand the barriers that patients face in achieving health goals and lifestyle related information whereas patients initiated tracking for individual curiosity, self-awareness, improving self-management skills, for collaborating with clinicians, and answering their questions [194]. As a result, patients may collect data that clinicians find irrelevant. There is a need to design systems that enable patients and clinicians to leverage data in mutually beneficial ways. Towards designing such systems there is a need to identify common data practices.

3.3.2 Workflows of Data Use

One of the ways to scaffold collaborative use of data is by understanding workflows that are employed in making sense of the data and in translating it to actions. Prior work on understanding the retrospective review of PGD has reported different workflows that patients and clinicians create in individual and collaborative use of the data. When reviewing PGD, clinicians across different healthcare settings have expressed the need for frameworks to organize evidence. Consequently, they create certain workflows for using data to arrive at a diagnosis. For example, West et al. identified a common diagnostic workflow across different clinician roles that included gathering evidence, evaluating evidence, generating hypotheses, identifying knowledge gaps when testing the hypothesis, refining the hypothesis, and creating a safe care pathway [178]. In another work, West et al. provide a 6-stage data use workflow consisting of the following steps – aligning patient and clinician objectives, evaluating data quality, judging data quality, rearranging the data, interpreting the data, and deciding on a plan or action [179]. Similarly, Kim et al. explored the use of a clinician-facing interface for data review and found that clinicians first skimmed the data, then asked questions about the data, and concluded the session by deciding and recording the goals for the patient [94]. When making personal discoveries using their data in guidance with diabetes educators, patients were seen to follow four steps - feature selection, hypothesis formulation, hypothesis evaluation, and goal specification [118]. Understanding such analytical processes is necessary for providing a systematic approach for data review, which can better enable patients and clinicians in making sense of PGD [117]. While the understanding of these workflows is emerging, existing tools do not support these workflows.

3.3.3 Telemonitoring Systems

Tools to support long term collaborative use of patient-generated data primarily include telemedicine systems with patient-facing and provider-facing components. Specifically, transmission of patient-generated data via telemedicine systems has since long been a basis to bridge the gap between home and clinic. Monitoring comprehensive patient data including physiological measures, contextual information, and self-reported symptoms through telemedicine systems has resulted in improved support for patients in different ways – by intervening more frequently as compared to standard model of care where clinic visits are multiple months apart [167], by identifying opportunities for improvement [58], by predicting an adverse event and alerting the patient beforehand [101], and by generating recommendations and reminders [8].

While these systems support management, they do not enable patients towards independently managing their condition. The use of these systems is primarily provider-driven with patients only involved in data collection. The responsibility of interpreting data is largely on the providers, who then direct patients to take relevant action. There is an opportunity to design patient-facing tools to improve patient agency in day-to-day self-management.

3.3.4 Visualization Systems

More recently, researchers have explored interactive visualizations to enable collaborative interpretation in clinic by patients and providers. For example, Schroeder et al. created visualizations from food journals of IBS patients and provide an understanding of how patients and providers interpret data together using the same set of visualizations. While such visualizations could support collaborative interpretation of data, similar issues as those discussed previously, such as complexity of visualizations undermining insight generation, and inaccurate representativeness of lived experiences, challenge their use [155]. Additionally, they may not succeed at creating a shared understanding of the data because they are open to exploratory analysis and interpretation by both patients and clinicians, resulting in the two having a different understanding of the data. The issues pointed here present the need to better understand how data is translated to insights for designing tools to create a shared understanding in collaborative use of data – that is, unpacking sensemaking activities and supporting those activities through system features.

3.3.5 Clinical Decision-Support Tools

Patient data has been used to design clinical decision-support tools that automate knowledge-based tasks and processes to aid decision-making in different ways - by providing treatment recommendations or options based on patient data, by providing prognosis or prediction of the course of patient condition, by identifying potential problems, by visualizing and interpreting patient data with reference to guidelines, and by providing guideline-based considerations for decision-making [187]. While most decision support systems are created for clinical experts in different specialties [22,30,75,187], researchers have highlighted the success of engaging patients with such systems [148]. Systems that provide advice concurrently to patients and clinicians have a higher chance of succeeding. For example, a shared electronic decision-support system for primary care of diabetes improved the process of care and clinical markers of the quality of diabetes care [75]. Yet, most of these tools are still primarily designed to support clinicians, with patients only involved in data entry [166].

Much research on decision-making by patients with chronic conditions has paid attention to situated decision-making. Patient-facing tools for decision-making are cognitively burdensome and provide limited actionable information [88]. Several models of situated self-management have been proposed, which include problem-solving [74], sensemaking [122], and fluid contextual reasoning [89]. These models do not account for data practices involved in self-management, which is vital for reasoning and decision-making both within individual and collaborative contexts.

3.4 SENSEMAKING FROM DATA

In this section, I review theories on sensemaking that offer an approach to understand how complex data is interpreted. Give the large volume of research on sensemaking, I describe theories specifically relevant to my dissertation. These include conceptualizations of sensemaking from HCI, psychology, and health informatics. Additionally, I provide examples of application of these theories in understanding execution of data-driven tasks in different domains. As noted in Chapter 1, my dissertation uses two of these theories to understand data practices of patients and clinicians, offers a data-driven model of sensemaking for chronic condition management, and informs the design of prototypes to be evaluated in the proposed study.

3.4.1 Different Perspectives on Sensemaking

Different perspectives on sensemaking have been described based on the practices of knowledge workers in different domains [95,96,138,139,149]. Two of the prominent perspectives on sensemaking include the "representation construction model of sensemaking" [138,149] and the "data-frame theory of sensemaking" [95,96]. In the HCI literature, Russel et al. articulate the representation construction model by describing sensemaking as the process of searching for external representations and encoding data in these representations to answer questions related to a specific task [149]. Building on this conceptualization, Pirolli and Card describe the sensemaking process of intelligence analysts as consisting of two interconnected loops – the information foraging loop and the sensemaking loop [138]. Information foraging involves seeking information, searching and filtering information, and reading and extracting the information into a schema. In this view, sensemaking involves iterative construction of a schema that best represents the data using the information from the foraging phase. These models of sensemaking focuses on creating external knowledge representations to enable efficient use of knowledge.

The data-frame theory of sensemaking stems from psychology and describes sensemaking as involving backward-looking mental processes to explain the past and forward-looking processes to anticipate the future [139]. This theory defines sensemaking as a natural cognitive process that involves fitting data (the “interpreted signal of events”) into a frame (an explanatory structure) or, conversely, fitting a frame around available data in an attempt to continuously improve the frame while also filtering data based on that frame [96]. Here, sensemaking is directed at performing functions including but not limited to problem detection and identification, anticipatory thinking, understanding how to act in a situation, forming associations, and projecting into the future. According to this theory, both experts and novices have the same process of reasoning, which makes the data frame model applicable to all kinds of data users irrespective of their proficiency with data and their domain of use. Unlike in the representation construction model, the focus is not on creating external knowledge representations but understanding the mental processes at play while performing functions noted above.

3.4.2 Sensemaking in Chronic Disease Management

In the case of chronic disease management, sensemaking has been described as one of the approaches to inform action by using continuously generated information and patient’s past experiences. Mamykina et al. describe sensemaking in diabetes management as a cyclic process consisting of perception, inference, and action [122]. Patients continuously perceive new information and match it against their knowledge and experiences to activate an old mental model or construct a new model explaining the situation. This is followed by a routine action dictated by the old mental model or an experimental action governed by the new mental model to tackle the situation. While Mamykina et al. provide a strong basis for the sensemaking perspective as an approach to use data to inform action for disease management, they stop short of accounting for how different steps of sensemaking can be realized using the data that the patients generate and how tools can be designed to support sensemaking through data.

3.4.3 Application of Sensemaking Theories

Various sensemaking conceptualizations have been applied to understand the execution of tasks in several domains, such as data analysis by researchers, information visualization, education, and chronic disease management [20,70,105,122]. For example, data analysis has been described as a sensemaking task wherein theory is matched against facts, discrepancies are observed, and the theory is updated accordingly [70]. Sensemaking can thus account for both exploratory and confirmatory data analysis. Exploratory analysis requires constructing schemas from data and confirmatory analysis requires collecting data based on a schema that already exists [70]. Similarly, drawing on the data-frame theory, Lee et al. characterized the cognitive activities of novice users of visualizations that involved seeing the visualization for the first time, constructing a frame, exploring visualization to find insights using the constructed frame, identifying abnormal data and questioning the frame, and failing to construct a frame [105].

3.5 VISUALIZATIONS OF PERSONAL (HEALTH) DATA

In this section, I review literature on information visualizations to identify techniques to engage non-expert users with their data. Additionally, I review literature on visualizations in healthcare. I articulate promising techniques and research directions for designing patient-facing visualizations of personal health data, which I have used to inform the design of the prototypes that I am proposing to evaluate.

3.5.1 Moving Beyond Expert Users – Designing Personal Data Visualizations for Non-Experts

Historically, while data visualizations have been primarily used to support knowledge workers and experts in data analysis and exploration, HCI researchers have been increasingly directing their attention to support non-expert users. Two particularly relevant research directions noted in this regard include the need to better support personal data visualizations and the need to design for mobile interfaces for situated access to data [102]. Lee et al. note that there is limited work at the intersection of personal informatics and data visualizations and call for better ways of supporting on-the-go access to data through different types of visualizations. A review of personal informatics visualizations by Huang et al. provides four design dimensions of personal data visualizations – data, context, interaction, and insight [79]. They outlined several challenges for designing personal data visualizations that include designing to fit in personal routines with at-a-glance interaction, supporting recall of relevant context for data interpretation and reasoning, choosing a baseline for comparison, diversifying design perspectives by letting people design their own visualizations, and integrating automated analysis. Despite a decade of personal informatics research [106], many of these challenges remain.

At a high level, there are two types of visualizations – exploratory and explanatory. Exploratory data visualizations primarily support analytic functions and are more suitable for expert users. Explanatory visualizations primarily support communicative functions and are more suitable for lay users. It is not uncommon for designers to combine both exploratory and explanatory elements in data interfaces allowing for a visualization to serve both communicative and analytic objectives. In the next section, I describe data-driven storytelling, an explanatory visualization technique for distilling informative content from a dataset and presenting it.

3.5.2 Data-Driven Storytelling - A Technique to Engage Non-Expert Consumers of Data

As the objectives of using visualizations, their applications, and the target audiences expand, visualizations are not only concerned with supporting data exploration and analysis but also supporting communication. Data-driven storytelling and narrative visualizations have emerged as powerful techniques to engage non-expert consumers of data [28]. Communicative application of visualizations has been particularly prevalent in the field of journalism where narrative visualizations are being used to communicate news events backed by evidence/data (e.g., climate change, election results, scientific findings). Given the growing use of data-driven stories and narratives for communication, several reviews of publicly available stories have been conducted to provide a systematic understanding of the design space of narrative visualizations [156], of their communicative and analytic components [81], and of the major and recurring techniques used for storytelling [163]. Although stories only allow for limited exploration (contrary to the finding that exploration can improve engagement), controlled exploration techniques help maintain user focus on the message being conveyed and could still provide simple analytic functions, if needed, to enable users in obtaining insights not necessarily explicitly communicated [163].

To enable authoring of stories, researchers have focused on understanding the processes of story creation from data. For example, Lee et al. proposed the visual data storytelling process for transforming data into stories – finding insights, making a story, and communicating the story [103]. Hullman et al. analyzed professionally created slideshow style stories to understand different strategies for sequencing visualizations and provide implications for automated generation of sequences [82]. Other research has focused on specific aspects of story creation – effect of layout of visual and textual components and their linking [193] (e.g., slideshow improved comprehension, linking increased user engagement, better recall in slideshow + linking condition, users preferred this combination), and the effect of visualizing prior beliefs on persuasiveness of narratives [73]. There is limited application of data-driven storytelling in the context of presenting personal health data, presenting an opportunity to explore this technique.

3.6 VISUALIZATION IN HEALTHCARE RESEARCH

3.6.1 Lack of Focus on Presenting Multiple Data Streams and Supporting Workflow-Based Engagement

Research on visualizations of patient data has primarily focused on visualizing electronic health records and patient-reported outcomes for use by clinicians [92] and visualizing risk for communication with patients [2,66]. Reviews of patient-facing visualizations provide formats that are useful and well understood [171], and the high-level goals of the visualizations [113]. Specifically, there is much focus on visualizing individual variables to communicate meaning (e.g., through comparison with population average) and risk but limited focus on designing to present multiple data streams and on designing to support analytical processes of reasoning with data or making decisions with data [171]. For chronic condition management specifically, personal health data is often used for an objective, which could be self-discovery, sensemaking, self-experimentation, cause-effect analysis, and problem-solving. Such engagement involves a workflow-based engagement with data including different steps where more than one number or type of data need to be interpreted and meanings from these need to be connected. However, prior work on patient-facing displays within healthcare research provides limited understanding of designing for workflow-based engagement with data [121,188]. The limited number of studies that do exist do not report the extent to which their designs were informed by state-of-the-art information visualization techniques or the process that was followed for creating the designs. With regards to approaches for designing and evaluating visualizations, research has identified the need to employ user-centered design processes to design for both patients and clinicians [6].

3.6.2 Increasing Focus on Designing Information Rich Interfaces Over Standard Displays

More recently, there is an increasing focus on improving patient-engagement with consumer-generated health data through visualizations [3], more so for patients with chronic conditions [113]. Within this context, researchers have called for improving the design of visualizations to accommodate for audiences with different health literacy skills by going beyond traditional and simple graphical presentations of data to design information-rich graphics or infographics [78]. Contrary to the practices of using simple visualizations for patients, researchers have called for designing patient-facing infographics that contextualize the data presented and allow for extracting meaning when literally interpreted [4] to make the designs independent of literacy skills.

3.6.3 Opportunity to Leverage Data-Driven Storytelling in Patient-Facing Tools

While the increasing emphasis on designing information rich presentations or infographics (as opposed to simple graphs) to engage patients with diverse data literacy skills is a promising change, more research is needed to go beyond presenting single variables, that is presenting multiple streams of data for easy comprehension and interpretable. Towards supporting patients in the interpretation of multiple streams of data, visualizations may need to serve both communicative and analytic functions with different objectives - communicative to make it easier for the users to get the point; analytic to help work around data issues as they might compromise the extent and quality of automated insight generation [99], limiting what could be communicated with the available data and hence, requiring the user to step in to approximate missing data and draw inferences. Data-driven stories, while more powerful for communication, could also enable controlled exploration for users to infer insights where automated insight generation fails.

3.6.4 Information Visualizations for Diabetes Data

In the case of diabetes, much HCI research has focused on presenting data to enable free form reflection, as described in previously. However, these studies do not specify the rationale for the design of visualizations they used, resulting in diverse types of visualizations with little knowledge of visualization strategies that work for different types of users or objectives [133]. Recently, information visualization researchers have started exploring how to support clinicians in using data from medical devices. For example, Zhang et al. proposed hierarchical task abstractions for task analysis and identified tasks performed by clinicians in using data from CGMs and insulin pumps. Given the challenges of making nuanced temporal inferences from multiple streams of data, they designed IDMVis, a tool to support temporal folding and alignment [189]. They evaluated this tool through think aloud and interviews to find that IDMVis helped identify trends, data quality issues, and variability to make treatment decisions. Another work reviewed diabetes visualizations [190] describing different types of views used in data displays – tabular, juxtaposed and superimposed; and calls for the need to understand how well do these views support sensemaking with data and understand analytic strategies that can help improve focus when reviewing visualizations of multiple streams of data. Lastly, there has been research on interfaces to present forecasts or predictions of blood sugar values that describes the ways in which different types of patients perceive the usefulness and presentation of forecasts [51,52]. Specifically, patients preferred direct feedback and simple yet information rich designs [51]. These studies have begun to establish design guidelines for creating health data displays. My dissertation adds to this body of work by articulating design guidelines for tools to support engagement with personal health data using mobile interfaces.

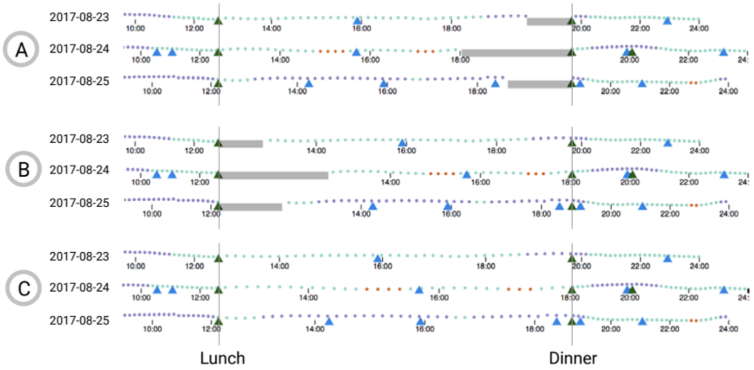
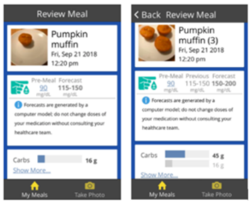


Fig. 6. Showing dual-event alignment by lunch and dinner using (A) left-justified, (B) right-justified, and (C) stretch time scaling.

a) IDMVis [189]



Forecasts are generated by a computer model; do not change doses of your medication without consulting your healthcare team.

Figure 1: Personalized BG forecasts: on the left, the user photographs one pumpkin muffin and views the forecast; on the right, the user increases the amount of pumpkin muffins and views comparison between the two forecasts.

Figure 1. Screenshots of visualizations presented. From left to right visualizations are: (a) gradient number line, (b) segmented number line, (c) speed dial, (d) the traffic light, (e) cartoon, (f) single line glucose curve, and (g) multiple line glucose curve.

c) Visualizations of BG predictions [51].

b) GlucoOracle

[52]

Figure 9: Recent work on diabetes data visualizations

3.7 EXPLAINABILITY OF HEALTH INSIGHTS

More recently, HCI research has focused on studying explainability of AI-generated health-related insights (predictions, recommendations), given the need to improve the delivery of these insights and make them more actionable [48]. Lack of explanations has shown to limit how users utilize the information presented by a system. For instance, the study of an application for nutrition-related goal recommendations for diabetes patients found that often recommendations were not actionable enough for the participants, perhaps because the recommendations did not provide explanations (e.g., what was recommended and why) [64]. In contrast, another study found that recommending fitness goal along with explanations of how the goal was generated improved users' trust in the recommendations and enhanced their commitment to the goal [184]. While providing explanations can improve users' experiences, different types of explanations could be perceived differently by users. For instance, a study on evaluating different types of explanations to improve users' understanding of context-aware systems found that among why, why not, how to, and what if types of explanations, the most useful explanations were about "why the system behaves in a certain way" followed by "why the system did not behave a certain way" [109]. While having an explanation as compared to not having one is considered useful, explanations can also have negative or unintended consequences on how users understand system-generated insights. For instance, providing detailed information might result in unintentionally drawing users' attention to less relevant information, which could result in interpretations by users that the system was not aiming to facilitate [160]. This is coherent with research in psychology that shows how drawing user attention to different explanatory features can impact reasoning [112]. These studies provide a rich understanding of how explanations might or might not add value. However, the role of explanations is relatively understudied in the context of chronic condition management where the risks of misinterpretation are high and detailed explanations are perhaps necessary. My dissertation adds to this body of work by understanding the use of explanations to make care decisions with diabetes data. It shows the value of contextualizing examples of data specific to an insight with domain knowledge to improve comprehension and actionability of information presented

Chapter 4 Understanding Individual and Collaborative Problem-Solving with Patient-Generated Data

Chronic condition management switches between phases of individual and collaborative management. While patients largely manage alone, they need to seek clinicians' help periodically. To design tools that support the use of health data for chronic condition management, it is thus, important to understand the use of data across these contexts to identify shared objectives, practices, and challenges. To develop this understanding, I conducted a formative exploration of individual and collaborative use of patient-generated data (PGD). This study was published in CSCW 2017. I led the data analysis, literature review, and paper writing along with a part of the data collection.

4.1 SUMMARY

Background: Despite increasing use of patient-generated data (PGD) by patients alone and in collaboration with providers, there is limited understanding of the data practices of these two consumers across different settings – individual use at home, and collaborative use during clinic visits.

Objective: The objective of this study was to obtain a comparative understanding of how patients and providers use patient-generated data to understand and resolve disease related issues in individual and collaborative contexts. Towards this objective, the study asked the following research questions:

- How is patient-generated data interpreted in individual and collaborative settings?
- What are the differences in patient and provider perceptions in interpreting data and using technology for diabetes management?

Methods: Interviews with 14 patients and 4 providers, 12 hours of observations of clinician and diabetes educator sessions with patients where patient data was reviewed, and a focus group

with 4 providers where they walked through patient data using commercially available data reports.

Results: In understanding how PGD is interpreted in individual and collaborative settings, the study found that both patients and providers use data with the purpose of identifying and solving disease related problems. During clinic visits, this happens collaboratively between patients, providers and caregivers. In between visits, it is primarily done by patients and caregivers. Although problem-solving diabetes issues is an important goal, there are differences in the use of diabetes data and related technology between patients and providers, which challenge problem-solving with data. Patients and providers differ in three ways when trying to problem-solve with PGD – a) the same data gives different insights to patients and providers (what might be a problem for the patients might not be seen as a problem by the provider), b) they differ on the types of problems to focus on while reviewing data, and c) they use different representations of the data in identifying problems. I draw upon the data related activities of patients and providers to reflect on the theoretical frameworks of problem-solving and sensemaking for chronic disease management [74,122]. I propose that collaborative sensemaking as an approach to problem-solving provides a useful framework to understand the use of PGD. Lastly, I provide design directions for informatics tools to support both individual and collaborative use of PGD for problem-solving and sensemaking.

4.2 MOTIVATION

As described in Chapter 1, people with chronic conditions are increasingly generating health data and sharing it with their providers with the objective of informing self-care and improving management [158]. Although patient-provider collaboration in this context holds promise [121], two concerns warrant further investigation of challenges in current practices across individual and collaborative contexts. First, informatics tools provide limited support in helping patients and providers reflect on health data to convert raw data into insights relevant to a patient's disease in and across different contexts – clinical and personal [29,114,165]. Second, patients and providers use different explanatory models to make sense of the disease that could have implications for overall disease management including how patients and providers interpret patient-generated data – individually or collaboratively [111]. This highlights an opportunity to further investigate how patients' personal understanding of their health comes together with clinicians' formal knowledge of the disease in the use of patient-generated data and the potential challenges introduced by collaborative interpretation of data.

4.3 METHODS

Towards this investigation, I used a combination of observations, interviews with patients, caregivers and providers, and a focus group with providers to gather data to understand the use of patient-generated data in collaborative and individual settings. The study was conducted at a pediatric endocrinology clinic. Data from pediatric patients includes both patients' and caregivers' point of view.

4.3.1 Interviews and Focus Groups

Patients and providers were recruited through the pediatric endocrinology clinic at a large teaching hospital using a reputational case selection method [154]. One of the lead researchers, a clinician at the facility (Dr. Joyce Lee), approached patients and clinicians. Interview and focus group data was collected by a group of researchers at the same facility (Dr. Lee's team). Fourteen patients and two clinicians were interviewed by this group.

Patient Interviews

Five interviews were conducted with the patient and the primary caregiver who accompanied the patient to clinic, one interview was conducted with only the caregiver, and eight interviews were conducted with the patient. Each interview lasted approximately an hour and followed a semi-structured interview protocol. Interviews focused on understanding the use of technology (glucometers, continuous glucose monitors (CGM), insulin pumps, data displays) to manage diabetes, practices of collecting data, making decisions based on that data, the challenges associated with using data, and frequency of engaging with data including the use of data in clinic appointments or with diabetes educators. Patients received $20 in compensation.

Clinician Interviews and Focus Group.

Two clinicians were interviewed by the previously mentioned research group (Dr. Lee's team). The focus group consisted of four providers (different from the ones interviewed) and was conducted by the same research group. The provider interviews and focus groups elicited providers' perspectives on the utility of commercial visualizations available for interpreting diabetes data, types of information providers use to make decisions, and providers' expectations from patients in using their data. The provider interviews and focus group asked providers to walk through problem scenarios using commercial visualizations typically used during clinic appointments (e.g., Figure 11). Each provider interview was about an hour long, and the focus group lasted 90 minutes. Providers received $25 in compensation. All interactions were audio recorded and transcribed. Table 3 and Table 4 provide more details on the participants.

4.3.2 Observations

Clinicians and diabetes educators for observations were recruited through the same facility. The observations were conducted by me. To understand how data is used in collaboration, 12 hours of clinic sessions and diabetes educator calls were observed, which included 5 clinic sessions and 6 phone call sessions. Each clinic session lasted about an hour (50 minutes to 70 minutes) and the duration of phone calls ranged between 7 minutes to 30 minutes. I took notes about the questions raised by patients and providers, the information shared by patients, and the decision reached during the session. Providers were informally interviewed to request more explanation. Two diabetes educators were formally interviewed to understand the problems patients report, the data they share, the questions patients or providers ask of each other, and the challenges of working with data to make decisions.

4.3.3 Data Analysis

I conducted the data analysis using a mix of structural coding and in vivo coding [150]. Guided by the interview questions, I analyzed the patient interview data to code for types of data collected and used, purpose of data use (types of problems identified from data), frequency of data use for different purposes, challenges of engaging with data (pattern extraction, understanding visualizations, aggregating data, remembering data), and data representations used for reviewing data. Similarly, I analyzed provider interviews and focus group data to code for data representations used by providers, types of information used by providers, types of problems they look for when reviewing the data, patient practices, and expectations from patients in using data (what data to use, when to use, in what format).

ID	Age	Gender	CGM	Pump

P1	22	M	Yes	No
P2	23	F	No	Yes
P3+	13	M	No	No
P4	29	F	Yes	Yes
P5	25	M	Yes	Yes
P6	16	F	No	No
P7+	13	F	No	Yes
P8+	14	F	No	Yes
P9+	12	F	No	Yes
P10+	15	M	No	Yes
P11+	12	F	Yes	Yes
P12	16	M	Yes	Yes
P13	12	M	Yes	Yes
P14	9	M	Yes	Yes

Table 3: Summary of patients who participated in interviews

ID	Gender	Title	Involvement
C1	F	Assistant Professor of Pediatrics and Communicable Diseases	Interview
C2	M	Associate Professor of Pediatrics and Communicable Diseases	Interview
C3	F	Certified Diabetes Educator	Interview and observation
C4	F	Certified Diabetes Educator	Interview and observation
C5	F	Adult Endocrinologist	Focus group
C6	M	Co-Director Diabetes Center	Focus group

C7	F	Program Director, Adult Diabetes Education	Focus group
C8	F	Pediatric and Adult Endocrinologist	Focus group

Table 4: Summary of providers who participated in the study

Coded data was further analyzed to compare and understand the differences between patients and providers with regards to the types of problems, and the ways in which they identify problems from the data. I analyzed the observation data guided by the categories that emerged from the analysis of interview data. Data from each clinic appointment and educator phone call was the unit of analysis for observation data. In the second round of analysis, coded data were grouped under themes using affinity diagrams[1]. Throughout data analysis, I tried to triangulate evidence from different sources of data – interviews, observations, and focus groups. I led the data analysis and continuously discussed the emerging themes with the other members of the research team.

4.4 FINDINGS

In this section, I first provide an overview of the data collection and reflection practices of the patients. I then describe how data was used by patients at home followed by its use in the clinic where patients and providers together used the data. It is important to understand both these use cases because health information is used in chronic care cycles that switch between individual use of information and collaborative use of information [29]. Lastly, I describe the challenges of using data that emerged by comparing the use of data by patients and by providers.

4.4.1 Data Collection and Reflection Practices

Out of the 14 patients that were interviewed, 8 patients made paper logs to track their data. Figure 10 shows a paper log template used by patients. Out of the remaining 6 patients, 2 patients used different logging tools. P4 used Beeminder[2] and your.flowingdata[3], and P8 used an app called Glucose Buddy[4]. The rest of the 4 patients had used paper logs previously but were not doing so at the time of the interview.

6 of the 8 patients that made paper logs had a continuous glucose monitor (CGM) or an insulin pump or both but they created paper logs despite the availability of data in these devices. Paper-

---

[1] Affinity Diagram. https://en.wikipedia.org/wiki/Affinity_diagram

[2] Beeminder. https://www.beeminder.com/.

[3] Your.FlowingData. http://your.flowingdata.com/.

[4] Glucose Buddy. http://www.glucosebuddy.com/.

based logging was preferred for multiple reasons. First, while the devices allowed for real time engagement with specific data points, accessing aggregate data from these devices was difficult because it required patients to have a technical set up that allowed downloading data to a computer. Second, most of the patients used a combination of multiple devices (CGM, pump, extra meters) that did not allow cross-device aggregation. Manually created logs helped bring all the information in one place. Patients and caregivers reported writing blood sugars, insulin dose, the response of insulin dose, food, exercise, site changes for sensors, birthday parties, nights out with friends or family, and any other information that they perceived as having the potential to affect diabetes management. Table 3 shows the types of diabetes technology that patients used. Similar challenges, usefulness, and feasibility of using paper logs has also been previously noted for patients with irritable bowel syndrome [195].

All the patients reported actively reviewing their data except three of them, two of which used to review their data in the past but had given it up because they had stopped logging information. One patient (P2) using CGM did not feel the need to review her data. For her just having a mental awareness of today's and yesterday's data was enough for her self-care needs. Two of the adult participants had created their own visualizations to review data (P1, P4). Patients who maintained paper logs preferred to use the logs for review in combination with the CGM. Only two participants reported downloading data from devices on a regular basis (P12, P13), in addition to keeping paper logs. For all the adolescent patients, their parents reviewed the data.

During the clinic appointments, patients brought in the paper logs, glucometers and pumps to share data with the providers. During phone calls with diabetes educators, they scanned and uploaded these logs to patient portal and/or gave providers direct access to pump and CGM web dashboards by sharing the login credentials.

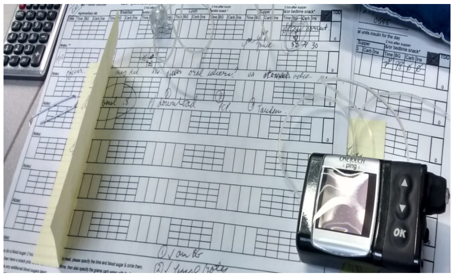


Figure 10: An example of a paper log template used by multiple participants

4.4.2 Individual Use of Data

In between clinic visits, patients used their data in five typical ways – understanding trends to take corrective action, ensuring absence of problems, understanding the effect of treatment, understanding the effect of contextual factors, and responding to goals.

Understanding trends.

Majority of the patients used data to identify trends in their blood glucose numbers, "P14 - *So the ones I look at the most would be patterns and daily trends*." These trends helped them understand the need for taking corrective actions, such as adjusting insulin dose or eating behavior. For example, one participant described the use of trends to decide if he needs to make corrections to insulin dose, *"I might correct for the numbers that I have at the moment and then check probably half an hour to an hour later and I am still going up then I will correct for that number."* (P12)

As reported in the above quote, the patient discerned trends over short cycles of time to take insulin. Patients reported looking for trends in time windows ranging between a couple of hours in the day to a month worth of data. I also found patients switching between different levels of engagement. For example, P7's mother mentioned she would usually review data every couple of weeks but at times depending on the self-care need, she would review every day's data in the evening, *"Absolutely, yeah. And sometimes it [data review] will be every evening she'll have highs or some things."*

Ensuring absence of problems.

Making sure everything was going smoothly without problems was another purpose for which most of the participants reported using their data frequently. Participants had varying types of need to pay explicit attention to data for this purpose. For example, one of the patients who played basketball reviewed numbers from his CGM app to ensure that the sport wasn't going to adversely affect his body, *"I would check my blood sugar during basketball and then towards the end. And if I was playing for like 3 hours maybe around hour 2 my blood sugar would start going up. High blood sugars and sports don't go well because you start to get muscle cramps, fatigue, dehydrated"* (P5). Similarly, another patient checked her blood glucose number before disconnecting the pump for showering to ensure she wasn't high before she disconnected her pump. (P2)

Understanding the effect of treatment.

Patients also had to maintain an understanding of whether the treatment is working or causing issues. To develop this understanding, they engaged in reflecting over their data once they had performed changes to regimen, *"I always wrote down what did she eat, how long did the insulin last, it is an hour and a half ..., what was her sugar at that hour and a half and then what did she do afterwards"* (P8's caregiver).

This reflection happened at varying time spans and with varying frequency. The above quote describes a caregiver assessing the impact of insulin in the frame of a couple of hours. Another patient described his engagement over a day to understand the correctness of treatment, *"I fast for a day to see what my blood levels looked like to make sure my basal levels were correct"* (P5). Similarly, one of the caregivers described how her frequency of downloading the meter numbers became higher around the time when a change in treatment was made, P14's mother "*I'll download it more when we've made a recent basal adjustment [changes in insulin] to see how we did."*

Understanding contextual factors.

Being aware of the problematic contextual factors affecting diabetes was an important objective of engaging with one's data for both patients and their caregivers. For example, one caregiver explained how she looked at the data to understand different contextual factors, such as food and activity, that could be affecting blood glucose numbers, *"I'll look at the trends on the CGM against the paper and then I can see like, oh, today, he had soccer or he went climbing at the rock climbing gym or whatever, or I can see his carbs or sometimes if he spikes after breakfast, I'll go, oh, he had a bagel."* (P14's caregiver)

Responding to goals.

Goals triggered engagement with data in different ways. Patients reviewed data specific to a goal to work towards that goal, "From my perspective, I look at it primarily because I know that [he] has an issue at night, because we've really gotten it down to where we focus around the night time" (P8's caregiver). At times, not meeting the goal was a problem and it prompted them to review their data, "When I do review is when I derail on one of those goals and then I'll be like 'oh so what happened in the last week that made my blood sugar average or whatever I'm using at the time" (P5).

4.4.3 Collaborative Use of Data

Data was used by providers in collaboration with the patients with the purpose of identifying and understanding problems. During clinic appointments that happened once every three months, I observed that patients and providers co-constructed the meaning of patient's data by interpreting and reinterpreting data shared by patient. This helped them crystallize the insight or the problems signified by the data. This process of meaning construction in the clinic was described by one of the patients as follows, *"My doctor always sits down with a stack of printouts of some of the Ping software and she kind of pages through them and I mean occasionally she will say 'it looks like you are having a bunch of lows' and so we'll talk about that. Or sometimes I'll be like 'oh no that was just last week why don't you look a couple weeks back.' She'll look a couple weeks back and be like 'okay you're right that was just last week what happened?' I'll be like 'I started running again so I had to change my basal but I didn't really catch on until I had a few days of lows and then I changed my basal.'"* (P4)

As described in the above quote, the clinician extracted critical events from patient's data, which she wanted to discuss further with the patient to identify potential problems. Next, the patient plays her role in invalidating provider's concern by categorizing the occurrence of that incident as short term, thereby dismissing the need to have a discussion around that data. Additionally, they discuss the potential cause of problems. In reaching a conclusion, patients and providers use information from multiple sources (glucometer, data visualizations from the device, patient's lived experiences, and clinician's domain knowledge). There were similar instances of problem-solving during patient phone calls with certified diabetes educators. These problems include physiological issues related to diabetes, treatment adjustments, understanding causes of problems, operating medical devices, and getting supplies.

4.4.4 Challenges of Using Data for Problem-Solving

As reflected by the above use cases, identifying and solving problems was an important aspect of engaging with patient-generated data in both individual and clinical settings. I further analyzed the data to understand the challenges of using data for problem-solving and found that patients and providers differed in perceptions over three aspects –defining a problem, types of problems, and use of data representations. This challenged communication during collaborative use of data.

Data gives different insights to patients and providers.

Although patients and providers use the same data, they differ in how they interpret the data for defining problems. There were several instances when the same data was interpreted differently by patients and providers. For example, during observation of a patient call with one of the educators, a patient's caregiver called with a concern of high blood sugar trends during the night and the morning. The following vignette further describes the call:

> CDE logged in to the pump dashboard and starts looking at the data. After some time, she mentions, "it is not that bad, overall we want 50% of the readings in range." She further explains "we use this to check patterns." She then thinks aloud trying to make sense of the patterns seen, "It isn't that bad. I will call her to see what is she worried about." (C3)

In the above case, while the patient's caregiver thought that the patient was high and needed some medication (insulin dose) changes, the educator did not think that the situation was as bad. In another instance, a patient, her caregiver and the clinician were discussing patient's data during a clinic appointment. The patient had brought in paper logs with two weeks of data (blood sugar numbers, carbohydrates, and insulin) and annotations related to her menstrual periods. The following vignette describes what happened,

> The patient was concerned about running high the week before her periods started and had brought in paper logs to the appointment. The clinician arranged the paper logs sequentially on the examination chair and asked the patient to circle all the high numbers in the log. When patient was done, the clinician tried to assess the pattern of high numbers around the week that was annotated with details on patient's periods, pointing to different places on the paper log. The clinician was unable to spot a trend as claimed by the patient and asked her, "*So you said you were high during the periods or going into it?*" The patient had a puzzled expression on her face while looking at the logs and was unsure about her earlier claim. She did not answer the question that clinician asked. After a moment of silence, the clinician mentioned, "*it seems like you are high 2 to 3 days into your periods. So, should we adjust the Lantus [insulin]?*"

In this case, the patient had come to clinic with a specific concern about running high. Even though both the patient and provider were looking at two weeks of data from the logs that patient brought, the clinician was unable to see the same blood sugar trend as the patient. Moreover, when the clinician asked for clarification, the patient was unable to communicate her concern effectively perhaps because realizing that the provider had a different perception of the problem made her submit to provider authority. The patient left the appointment without getting her original concerns addressed, since the clinician did not look at the problem in the same way as the patient did.

At times, patients disagreed with providers regarding a problem “She had sort of altered consciousness and we weren’t quite sure what was going on. So, it wasn’t really a seizure, but her endocrinologist thinks it probably was. Interestingly, [during] neither of her seizures could we measure her to be particularly low” (P14’s mother).

I found two reasons for this difference in perception of problems. First, the data interpretation criteria used by patients are different from that of providers. For example, when P10’s mother was asked if she uses “bolus to basal ratio” (a metric used by providers to tailor insulin dose) for making changes to basal insulin, she described that it was hard to do so, “*Well, it’s hard to look at that [bolus to basal] with him because of the way that he eats. He doesn’t eat consistently.”* Owing to the difficulties in applying guidelines for interpreting the data, patients develop their own criteria for evaluation of data. For instance, one patient recalled, *“I remember reading somewhere that they say you should think about adjusting your basal if you’re doing more than 10% of your total daily insulin from correction boluses”* (P4). Moreover, even for riskier situations such as ketones that could lead to an emergency room visit, patients did not apply the guidelines for identifying problems, as described by one of the clinicians, *“they are already supposed to check for ketones at 300 [BG number] and a lot of families will say oh they never have ketones so they kind of never check them”* (C1). As described in the quote, “300” does not get perceived as an indicator to check for ketones, even though providers would recommend that. Thus, patients and providers have different perceptions of risk.

Second, patients might be unaware about the evaluation criteria that providers use to extract instances of problems from the data and vice versa. For example, as mentioned by providers in the interviews, they use rules, such as “more than twice a week is too many lows”, and “50% or more highs denote trends,” to check for trends in the data. Patient had diverse responses about the criteria for classifying a series of numbers as a trend. For some patients five or six days of similar numbers constituted a trend, whereas for some a month long of consistent data was what constituted trends.

Such differences in perceiving data could be problematic because it might lead to a case where patients are overly concerned even when there is no problem, or it might also lead them to ignore a problem when there is one. Additionally, collaboration over patient data is affected, as patients might perceive differences in perception as a mistake on their part, affecting their confidence to deal with their data and communicate with the provider.

Patients and providers differ on the type of problem to focus on.

All the participants typically worked through the data to figure out trends to take corrective actions, such as a dose change that would fix the trend, and to evaluate the effect of the dose change on the existing trends. For example, one of the patients explicitly stressed the importance of looking at the trends, "*Trending, knowing what direction your blood sugar is trending, is really important. Basically, I check my blood sugar every once in a while: when I wake up, before meals, before sleep. Here [showing the CGM] you can look at 3 hour trends, 6 hour trends, and 24 hour trends"* (P5). As a part of understanding their data to manage diabetes, just knowing trends in data was good enough for them.

While providers want patients to get into the habit of reviewing their data to identify trends, they also want patients to focus on individual data points that are problematic, such as high or low blood glucose numbers, to figure out the causal behavior responsible for that data. For example, one of the clinicians shared the advice he gives families, *"What I tell families is, you have to get to the point where you can explain the outliers. You have to solve the problem when it's happening or it's never going to register in your brain. I say look, any time your sugar is above X, I need you to spend at least a minute thinking about how it got there. Did you bolus for your snack? Did you undercount your last meal? Because you are going to remember at that moment in time exactly what happened"* (C2). As mentioned in this quote, the clinician specifically wants the patients to pay attention to the outliers, that is the blood glucose numbers that are not typical. Additionally, he wants the patients to solve the problem at a time when it happens to understand the cause behind the problem.

Providers looked at the information related to CGM and insulin pump alerts to understand atypical events and how the patient is responding to them, as mentioned by an endocrinologist, "*Sometimes we will pay attention to some of the alerts and alarms they have going off. Sometimes* *they don't have enough and other times they have so many that they just ignore all of them*" (C7). Going by this expectation to understand atypical events, clinicians also asked patients about these specific events during clinic appointments. However, in their day to day engagement with data, patients might be selectively attending to these events, or even if they did attend to all the events, they might not remember details of each event. For example, one of the patients who was actively monitoring his data daily found it difficult to bring up details related to critical events that clinicians asked about, *"One thing that really frustrated me was that I would go to my endocrinologist and then 3 months later go to another one and it's hard to tell what's happening in-between there. They would look at my charts and be like you were high and then low and what happened and I don't know what happened"* (P5).

Amongst the patient participants, most of the adolescent patients tended not to pay attention to individual events of atypical blood glucose numbers. For example, even though they set alarms on their insulin pump and continuous glucose monitor to make them aware of unusually high or low numbers, they usually turn these off and ignore the atypical lows and highs of blood glucose numbers. This could happen because the patient's perception of risk or problem might depend on how they feel and not entirely on the interpretation of physiological numbers, as described by the caregiver of P13, "*P13 is having trouble like the CGM will alarm him but he won't feel it or hear it and then he will be in the 300s for several hours and I will look at it and say you've been in the 300s for 4 hours, what's going on? And he says I didn't feel it."*

In another case, the patient's mother took care of these alerts as long as she could. But every time the patient was in school, she disliked the alarms going off so she stopped using those alarms completely, as described by P8's mother*, "When I was in control, I did like those alerts. But then when she was at school, she doesn't like having it go off all the time. So, that was what caused her to cease using that."*

This further demonstrates that most patients do not make use of problem-solving opportunities (i.e., atypical events) because they might not be in a situation to do so, or they might not perceive enough risk associated with the problem. Working with trends in their data is good enough for them. Moreover, even if they do pay attention to discreet events in addition to trends as expected by the providers, they are unable to recall information associated with specific events. When it comes to sharing information with the providers about specific instances, they fail to do so.

Patients and providers use different representations of data to identify problems.

For patients, reviewing one's data is the key to maintaining control in diabetes, as mentioned by one of the clinicians, "I think reviewing data is a key to success whether you are familiar with that or not you unfortunately have to become familiar with it, you know, since it's your life" (C1). However, patients struggle in reviewing their data to draw insights because of the unsuitability of commercially available data visualizations, even though they are the primary users of the devices that are the source of these visualizations. Regarding the use of commercially available graphs, P11's mother mentioned that finding patterns is difficult and she did not think that the graphs helped much: "*I feel the patterns are very hard to identify. I feel like there's no rhyme or reason [for using the graphs]. I feel like they're just a bunch of just scatter plots everywhere, lines up, lines down. I don't know how he [clinician] makes heads or tails of it."*

Most patients relied on manually compiled paper logs to understand their data, and they also brought these to clinic. One of the patients who expressed dislike for the commercially available graphs had created her own visualizations for getting a more accurate reflection of blood glucose numbers over multiple days, *"I hate the Dexcom Moday Day graph. It's the one where they overlay all your days but the dots just stack up on each other. For me the heat maps that I've done are a vast improvement on that because they actually show you. When the dots are just overlapping each other you can't get an accurate picture of the density at any given point"* (P4). On the contrary, clinicians found these commercial graphs useful, *"The stats are very helpful, that is one of the things I use the logbook for but going back to the daily strips, its useful to help find quick trends. I like different views for different reason depending on what I am doing with the patient"* (C8). As observed in clinic sessions and as reported by all providers in the focus group, these graphs were frequently used by clinicians during patient appointments. For example, Figure 11 is a visualization that shows an overview of data from the insulin pump. Some providers liked seeing such visualizations as a printout during patient appointments while some preferred an electronic version on the computer screen. The visualizations offered multiple perspectives over the data and helped the providers explore trends over multiple days, data from individual days, and behavioral data such as "how often they are checking (BG numbers)" and "how often they are disconnecting from their pump".

While providers agreed that these dashboards are useful for them, they also acknowledged that families might not be looking at them, and engaging with them as needed,

> "*I think that these kind of things are really nice for us to look at but sometimes it prevents the family from writing down and looking at blood sugar numbers. I think that would be easier for people to see if they have it in this kind of form but this is a lot of information. Or these other sheets are a lot of information, so I don't know how easy it is for people, you know to look at them or not"* (C1).

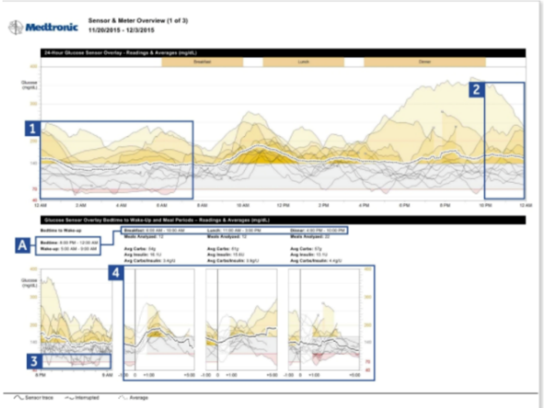


Figure 11: © Medtronic MiniMed. Inc. Example of a visualization used by providers. This is taken from a Medtronic insulin pump.

As expressed in the above quote, the clinician appreciated all the information in the graphs but also understood that there was a lot of information in the graphs, which would make the visualizations cognitively burdensome for patients, discouraging them from engaging with their data. However, there was also an expectation from the providers for patients to use these visualizations, as mentioned by one of the educators in the focus group, "*I want them to see what I am seeing so that when I do want to talk to them on the phone in between they have a good idea of the way my brain is working"* (C7).

The educator wanted patients to understand these visualizations so that patient was on the same page as her, which would make communication easier. However, patients have their own ways of obtaining insights from their data (e.g., paper logs, self-created visualizations) instead of using the visualizations providers want them to use. Patients and providers rely on different representations of the same data for extracting insights, which restricts development of an initial common ground for communication.

4.5 DISCUSSION

This study found that both patients and providers use patient-generated data to identify problems and understand those problems to reach a decision about the next step. However, in interpreting data, they have different perceptions around problem-solving: they used different data representations, they had different understanding of what is a problem and there were differences in selecting the types of problems that need attention. This makes collaboration effortful as it leads to differences in how they perceive risk of problems and the associated response to problems. While such perception differences remain unacknowledged during patient-provider interactions, they do govern what problems get discussed during such encounters. Consequently, there is an immediate impact on decisions regarding patients' treatment and behavior.

Although prior studies have acknowledged the existence of perception differences between patients and providers about patient-generated data in terms of the type of data that is useful and the overall value of data [178,194], little is known about how differences in perception affect the interpretation of patient-generated data in making disease related decisions. This study builds upon prior work to provide an understanding of *what* are the differences in patient-provider perceptions in interpreting patient-generated data and *how* they affect collaborative reflection for identifying and solving disease related problems. While a recent study demonstrated that the same set of visualizations were helpful for both patients and providers and that collaborative review of these visualizations was promising [155], this study found that this was not the case. Commercially available visualizations that are built using clinical guidelines were useful for providers but not for patients. Instead, patients had their own ways of creating representations of the same data to draw insights.

Prior work to understand collaboration with patient-generated data has identified the creation of boundary negotiating artifacts as a basis to understand and support collaborative use of PGD [37]. I propose that another basis for supporting collaborative use of PGD is problem-solving as understood through the framework of collaborative sensemaking. This approach to understand the use of PGD provides a way to not only understand collaborative use, but it consolidates both individual and collaborative practices of using PGD, as sensemaking and problem-solving can and do happen individually and collaboratively across the chronic care cycle. Sensemaking in health care has been primarily discussed as an approach for patients to take [122] or as an approach that facilitates collaboration amongst providers for information seeking activities [135]. In this study, I found that patients and providers seek to engage in collaborative sensemaking with data to problem-solve. In this discussion, I reflect on Hill-Briggs' problem-solving model and Mamykina et al.'s framework of sensemaking for diabetes management as understood through these findings. Considering the usefulness of both these frameworks, I present collaborative sensemaking as an approach to problem-solving in diabetes and provide design suggestions for tools to support problem-solving and sensemaking among patients and providers.

4.5.1 Data-Driven Problem-Solving

Problem-solving is an important skill that patients are expected to develop to tackle everyday challenges of diabetes management and adjust self-care practices in response to barriers to adherence [69]. The problem-solving model as proposed by Hill-Briggs accounts for how patients inform their actions to maintain adherence to prescribed regimens by identifying and overcoming barriers to self-management. According to this model, problem-solving for successful chronic disease management includes four components: a) problem-solving process, b) problem-solving orientation, c) disease specific knowledge, and d) transfer of past experiences [74]. In managing diabetes, an important aspect of problem-solving is the ability to engage with one's data to reflect on one's experiences using disease specific knowledge. While several interpersonal interventions have aimed to train patients in problem-solving skills [60], tools do not currently support problem-solving through engagement with data [117]. In managing their diabetes, I found that patients make use of data in different ways with the aim of identifying problems from the data (trends, contextual factors), understanding the cause of problem (contextual factors), evaluating effectiveness of their solution to problems (treatment changes), and managing in anticipation to avoid problems and minimize risks (ensuring the absence of problems, and responding to goals). Since considerable use of data, if not all, is centered around identifying and solving problems, it presents the need to support problem-identification and problem-solving through data analysis activities.

4.5.2 Data-Driven Sensemaking

In addition to problem-solving, *sensemaking* is another framework that has been considered useful in understanding how patients might inform their actions in managing diabetes based on information and experiences. Mamykina et al.'s sensemaking framework describes diabetes management as consisting of two modes: habitual mode and sensemaking mode [122]. According to this framework, both modes include three activities - perception, inference and action. Patients operate in habitual mode when information and experiences fit into already existing mental models of disease management. When new information and new experiences create gaps in understanding, patients operate in sensemaking mode to adjust their mental models. While both problem-solving and sensemaking describe a set of activities that patients perform or should perform to address self-care issues, and can be considered as complementary, they emphasize different aspects of disease management. Sensemaking focuses on informing action by generating explanatory models for breakdowns or unusual experiences during self-care, and problem-solving focuses on identifying problems or barriers to inform action in terms of selecting a solution for the problem, acting on it and evaluating the effect of the action on the problem. Both require the ability to identify problems or gaps in one's understanding before an action can be taken.

Sensemaking versus Satisficing.

In this study, I observed that in using their data to problem-solve diabetes issues, patients did not necessarily engage in sensemaking right when problems (e.g., unusual blood sugar readings) happened. That is, patients did not necessarily make use of what Mamykina et al. refer to as "teachable moments" (e.g., unusual highs or lows of blood sugar), as much as providers wanted them to. Instead, patients satisficed to handle atypical situations by doing what they perceive to be necessary to quickly fix the problems and perhaps to avoid the burden of explicit sensemaking, which could result in diabetes burnout [141]. For example, to handle unusual blood sugars, it is easier and more imperative for patients to take corrective action (i.e., take insulin or consume carbohydrates), than to also understand why those numbers happen. Thus, problem-solving in real time involved using data to the extent that was good enough to resolve the immediate symptoms (unusual blood sugars) without deeply engaging in sensemaking to find probable explanations for the problem.

Sensemaking in Retrospect.

Once patients have the time or an opportunity to go through their data, it was then that they might try to retrospectively engage in sensemaking with data to understand what factors might be causing the problem, and what that means for patients' disease management. This mostly happened during planned engagement sessions that involved the retrospective use of data at home or in clinic, as shown in the study. Understanding this episodic nature of sensemaking is important to inform the design of tools to support it because patients might not always be receptive to opportunities to perceive gaps in their understanding and to bridge those gaps because of the burden that sensemaking entails. This further demonstrates that patients might not necessarily experience or practice different activities of sensemaking (perception-inference-action [122]) distinctly or in that order. For example, perception and inference could happen retrospectively through data analysis activities. This observation echoes the lived informatics critique of the stage-based model of personal informatics, which claims that the steps of the stage-based model are not distinctly and rationally experienced in real life [147]. Similarly, the steps of sensemaking (proposed in the framework in [122]) might not necessarily be descriptive of natural behavior, but rather prescriptive of ideal behavior. Considering the burden that sensemaking entails, it is important for tools to provide support for data analysis activities for sensemaking in ways that are concordant with people's actual practices.

Data-Driven Collaborative Sensemaking.

Another aspect that the findings from this study bring to light is that sensemaking for self-management of diabetes might not only happen individually, but patients also engage in collaborative sensemaking with their providers throughout the chronic care cycle. That is, patients switch between individual and collaborative sensemaking, which points to the need to align these efforts. While Mamykina et al.'s framework accounts for individual sensemaking by patients, it does not consider the role of caregivers and providers in sensemaking for guiding and supporting self-management of diabetes. For caregivers and providers who are considerably removed from patients' lived experiences, patient-generated data is the basis for understanding patients' experiences and supporting management [91]. Hence, sensemaking through data analysis and reflection, that is understanding the data before acting on the data, becomes crucial not only for patients individually, but also for caregivers and providers if they are to support the patient. This study further unpacks the complexity that multiple stakeholders bring when engaging with patient data. Above, I reported the differences in stakeholder perceptions that challenge collaborative interpretation of data. Considering that collaboration happens in all stages of PGD use [37], there is a need to extend this framework to incorporate collaborative processes in the use of data.

Both problem-solving and sensemaking are useful frameworks to understand diabetes management, and they have mostly been considered separately. There is a potential to integrate these two frameworks as both require considerable engagement with one's data. That is, data analysis activities are a part of both problem-solving and sensemaking in the management of diabetes. I further elaborate the proposal of supporting data analysis activities for problem-solving through collaborative sensemaking.

Collaborative Sensemaking as an Approach to Problem-Solving with Data

As described in the findings, during clinic sessions, patients and providers worked through the data to identify critical events from the data, classify those events as problems or non-problems and generate explanations for the cause of problems to decide appropriate response. Sensemaking thus, primarily happened through data analysis activities. In analyzing data to first identify relevant information and then co-construct what that data means for patients' disease, providers employ and attempt to convey disease specific knowledge, which is the third component of problem-solving model [74]. Patients contribute their experiential knowledge to validate or invalidate the meaningfulness of the data in question, which relates to the past experiences component of the problem-solving model [74]. In this way patients' personal understanding of their health comes together with clinicians' formal knowledge to problem-solve using sensemaking with data. Problem-solving as understood through collaborative data-based sensemaking thus provides a useful lens to understand the practices of interpreting PGD to make disease related decisions.

4.5.3 Design Implications

The insights from this study point to design suggestions for computer-supported cooperative systems to support both individual and collaborative problem-solving with data as understood though the framework of *sensemaking*.

Promote mutual intelligibility of individual sensemaking efforts.

Sensemaking in the clinic, which is a case of collaborative sensemaking, was fraught with three challenges as this study found: patients and providers used different data representations, they had different understanding of problems, and they prioritized different types of problems that needed attention. In collaborative sensemaking, every individual engages in sensemaking effort based on their experience and knowledge, which may be conflicting with the worldviews of others involved in collaborative sensemaking [157]. As shown by this study, conflicts in problem-solving challenged collaboration between patients and providers in sensemaking with patient data. This points to the need to better understand how such conflicts can be resolved.

One of the ways to resolve this conflict follows from the findings. I observed that patients often use criteria developed based on their experiences to understand their diabetes data. These criteria might not entirely align with clinical guidelines but are still used by the patients for their own benefit. Consequently, misunderstandings and disagreements arise when they engage in collaborative sensemaking with their providers. Such disagreements could be resolved by making them aware of each other's ways of looking at the data. Tools to support awareness of each other's data-related work (findings, hypothesis, evidence) have been studied for data analysts [116]. Such tools could potentially be adapted to support mutual understanding of sensemaking efforts between patients and providers. Moreover, effective collaboration requires both patients and providers to trust each other's use of patient data [155]. For this to happen, each party needs to know how the other is using the data throughout the care cycle so that differences in perceptions can be minimized. Such tools can bring to the foreground the tacit knowledge used by both parties in interpreting patient-generated data, thereby promoting trust.

Support problem identification for sensemaking.

I observed that not being able to assertively identify problems from the data discouraged patients while communicating with clinicians. This can also affect patient's problem-solving orientation [74]. Patients capture considerable amounts of data that cannot possibly be comprehensively explored, given the short time of clinic visits. It becomes important to help patients and providers retrieve meaningful and useful information sooner and without putting in more effort. To support sensemaking with information, as a first step, tools need to support identification and selection of relevant problems or gaps. Patient generated data, such as physiological numbers, can be used to computationally identify problems in a simple manner [21]. Machine learning approaches, such as anomaly detection [27], are more advanced techniques that could help.

Use problems as a basis for goal-setting.

I observed that patients and providers collaborated in clinic to identify problems from the data, and those problems became the basis for further use of data to understand causal factors behind the problem, the current state of treatment, and the potential treatment changes that might help. Prior work has identified the need to create goals for collaborative use of PGD by having goal-based data collection, curation and visualization [37]. This study suggests that problems identified from the data can be used as goals for data review during clinic sessions. Systems can support patients in selecting problems that they might want to discuss in the clinic visit. For example, a system that can identify possible problems, could prompt the patient to bookmark relevant events for review in the clinic and gather more data related to those events. A bookmarking feature supplemented with annotations and context data can augment clinical communication by helping the patient recall the problem and the context in which that problem happened. Because problem-solving is an ongoing process throughout the chronic care cycle, systems can also involve providers, especially diabetes educators, to guide patients in selecting the problems that they need to focus on in between clinic visits. This would further streamline patients' efforts in collecting and using their data for sensemaking and problem-solving activities.

Support need driven sensemaking with data.

Patients engaged with their data in planned and unplanned ways, as reported in the findings. In trying to make sense of their data, they needed to access and review data from varying time windows (e.g., data for a day, data for a month) for varying needs. For example, sensemaking with data was prompted by the need to understand the effect of treatment or by the realization of not meeting a goal. Such needs to review one's data could arise multiple times during the day or could happen over a larger frame of time. Informatics tools should support tailoring the amount of data (e.g., time range) that patients want to explore and the frequency with which they want to review this data.

4.6 CONCLUSION

In understanding how patient-generated data is interpreted individually by patients and collaboratively by patients and providers, this study found that patients and providers use patient-generated data for the shared purpose of identifying and understanding disease related problems.

In exploring differences in the interpretation of data and the use of diabetes technology, the study found three themes in how patients and providers differ, which challenge collaborative use of data for problem-solving – differences in understanding what is a problem, what types of problems to focus on, and differences in the use of data representations. Drawing on these insights, I reflect on two specific conceptualizations of disease management behavior (frameworks of *sensemaking* and *problem-solving*) as they relate to data analysis activities of patients and providers and suggest opportunities for extending these frameworks. I propose *collaborative sensemaking* as an approach to problem-solving with patient-generated data. Given the perception differences between patients and providers, this study suggests that informatics tools need to support mutual intelligibility of sensemaking efforts with data to facilitate effective collaboration, and they need to support problem-identification for effective sensemaking.

This study contributes in three important ways. First, while prior studies provide individual accounts of data use by patients or by providers, this study contributes knowledge of nuanced challenges in the collaborative use of data by comparing patient and provider perspectives. Second, it reflects on theoretical frameworks of *problem-solving* and *sensemaking* to uncover opportunities for extension and integration of these frameworks for chronic disease management. Third, it offers practical guidelines for designing systems to support patient-provider collaboration in using data to drive disease-related decisions.

Towards designing better systems, it is important to understand data analysis practices in individual and collaborative settings in the context of real tools. As noted in this chapter, patients find the commercially available data reports complex and do not review them at home. It is thus difficult to study practices of patients in the context of these tools. In the next chapter, I describe a study in which I developed context-enhanced visualizations of data and used them as probes to understand how patients make sense of their data on their own without clinician involvement. In the subsequent chapters, I investigate collaborative data practices as they happen using commercially available tools.

Chapter 5 Clinical Data in Context: Towards Sensemaking Tools for Interpreting Personal Health Data

A key limitation of existing tools is the lack of support for making sense of multidimensional data (e.g., understanding contextual factors affecting outcomes). In this chapter, I report on a study wherein I built displays of multi-dimensional data (including clinical and contextual data) to investigate how patients and their caregivers might engage in sensemaking. The findings from this investigation show that patients and caregivers identify trends and build explanations of those trends to identify self-care actions. Additionally, I show that the sensemaking activities performed by patients closely aligns with the sensemaking processes described by the data-frame theory of sensemaking (described in Chapter 3), which provides an informative framework to understand sensemaking activities with multidimensional data. Lastly, I provide suggestions for designing tools to support sensemaking with multidimensional data.

5.1 SUMMARY

Background: Establishing associations between outcome measures and contextual factors is one of the primary reflection needs of patients in interpreting personal health data. Prior research has also highlighted the need for systems to include contextual information to enable better data analysis and interpretation [5,39]. However, it remains largely unexplored how disease-related outcome measures augmented with contextual data could be analyzed and interpreted by patients and to what extent their practices could be computationally supported.

Objective: The objective of this study was to understand different ways in which patients engage with their data using displays of contextual and clinical data. Specifically, the study asks the following research questions:

- What are the data-based sensemaking activities that patients and caregivers engage in for using contextual and biomedical data?
- What are the challenges they face in engaging with their data?

- How can tools be designed to support data-based sensemaking activities?

Methods: interviews with patients and caregivers using context-enhanced visualizations of clinical data as probes to facilitate data-based sensemaking activities.

Results: Participants performed four analytical activities when interpreting clinical data augmented with contextual data – finding context-based trends and explaining them, triangulating multiple factors, suggesting context-specific actions, and hypothesizing about alternate contextual factors affecting outcomes. They faced two challenges – the inability to identify clear trends challenged action planning and counterintuitive insights compromised trust in data. I interpret these findings using the data-frame theory of sensemaking [18] to show how participants' information needs evolve and emerge through various sensemaking processes. Situating these findings within the existing perspectives on sensemaking, I show that sensemaking could not only result in action, as described in the sensemaking framework for diabetes management [27], but it could also result in further exploration and information needs. I argue that contextual data, which is burdensome to collect and interpret, should be explored using a sensemaking approach as sensemaking could drive information needs and hence, meaningfully guide the collection of contextual data in response to those information needs. Noting the value of sensemaking approach in exploring contextual data, I provide suggestions for designing sensemaking tools to improve awareness of contextual factors affecting patients and to support patients' agency in making sense of data by enabling user feedback on computational insights, enabling cross validation between several factors or data streams, by supporting anticipation of future conditions, and by incorporating expert knowledge.

5.2 MOTIVATION

Even though reflecting on one's data to problem-solve or make sense of disease related experiences is important to inform treatment and disease management behavior, sensor-rich tracking tools provide limited support to help patients and providers reflect on data [115,165]. Specifically, while both providers and tools can support patients in interpreting standard clinical measures, interpreting contextual information (e.g., lifestyle data) is considered beyond the expertise of providers and is not yet supported by tools that patients use.

As a part of understanding chronic conditions through personally generated data, establishing associations between outcome measures and contextual factors is one of the primary reflection needs of patients [118]. However, collecting and making sense of multiple streams of contextual data affecting an outcome measure is challenging [36], more so in the absence of appropriate tools. While prior research has highlighted the need for contextual information to enable better data analysis and interpretation [36,155], it remains largely unexplored how disease-related outcome measures augmented with contextual data could be analyzed and interpreted by patients and to what extent their practices could be computationally supported. This presents an opportunity to understand how patients could interpret clinical data augmented with contextual data, which this study investigates.

5.3 METHODS

The objective of this study was to understand different ways in which patients engage with their data using displays of contextual and biomedical data. The study used a multi-method approach. I first interviewed patients and their caregivers to understand the current practices of collecting and reflecting on their data. I then asked the participants to collect data over 3 weeks using DReflect, a context-enhanced diary application developed by me (Figure 12). In the 4th week of the study, I interviewed patients and caregivers using context-enhanced visualizations created from their data as probes to facilitate data analysis and interpretation activities. I chose to focus on Type 1 Diabetes patients because this condition requires continuous monitoring of several clinical factors, including blood glucose (BG), insulin dosages, and carbohydrate intake. These measures are known to be impacted by contextual factors (exercise, mood, sleep, routine, stress) [72,152], which makes it important to understand how patients interpret clinical and contextual data and how tools can be designed to help them understand the role of these factors.

5.3.1 Patient and Caregiver Recruitment

Patients and their caregivers were recruited from a pediatric endocrinology clinic at a large teaching hospital. Patients willing to participate were screened to identify if they met the recruitment criteria in accordance with criteria-based sampling. The inclusion criteria were: Type 1 Diabetes patients that were 13-17 years of age, English speaking, willing to carry their own mobile phones or phones provided by the study team always while awake for the duration of the 3-week diary study, with adult caregivers willing to fill out diary entries for three weeks. I excluded new patients that had been diagnosed less than 6 months prior to the start of the study because they might have not yet formed stable management practices for using data. I recruited 16 patient-caregiver dyads out of which 15 completed the study (Table 5) and one dropped out after the initial interview because of time constraints. Patients and caregivers who completed the entire study received $75 in compensation. All the caregivers participating in the study were mothers of the patients.

5.3.2 Initial Interview

The initial interview was semi-structured in nature and lasted an hour. I first gathered demographic information and basic details about the patient's management (devices used for management, prescribed blood glucose range, prescribed target blood glucose number, prescribed insulin regimen, insulin to carbohydrate ratio and correction factors). Next, I demonstrated the use of DReflect (Figure 12), the android application I had developed for data collection, and explained the tasks expected of them during the 21 days of the diary study. Participants were then asked to label and save specific locations that they were going to frequently visit during the three-week study period (e.g., Home, School, Work, Grandma's, Friend, Soccer, Horseback Riding). Additionally, patients were asked to specify their wake-up time and bedtime so that the application could determine an appropriate time window for sending notifications. In the second half of the interview, I followed a semi-structured interview protocol where I asked patients and their caregivers to recall the previous day and describe their routine. I then asked them about their engagement with diabetes-related devices and the data from those devices (how often they log data, how often they review data, and how often they change insulin dose based on their data). This interview helped me understand the patient's living context and informed the exit interview questions at the end of the study.

ID	**Age**	**Sex**	**Diagnosis**	**Blood glucose monitoring**	**Insulin Administration**	**Ethnicity**
P-01	14	F	Jan 2011	Continuous Glucose Monitor (CGM) + Glucometer	Pump	White
P-02	17	M	Apr 2013	CGM + Glucometer	Pump	White

P-03	15	F	Jul 2007	Glucometer	Pen	White
P-04	17	F	Jul 2013	Glucometer	Pump	Asian
P-05	16	F	Jul 2011	Glucometer	Pump	White
P-06	16	M	Sep 2013	CGM + Glucometer	Pump	White
P-07	16	F	Jan 2007	Glucometer	Pen	White
P-08	14	M	Jul 2015	CGM + Glucometer	Pen	White
P-09	17	M	Oct 2007	CGM + Glucometer	Pump	White
P-10	13	M	Jun 2010	CGM + Glucometer	Pump	White
P-11	15	F	May 2009	CGM + Glucometer	Pump	White
P-12	13	M	Oct 2009	Glucometer	Pen	White
P-13	16	M	Oct 2011	Glucometer	Pump	White
P-14	13	F	Aug 2010	CGM + Glucometer	Pump	White
P-15	13	F	Dec 2016	Glucometer	Pen	White

Table 5: Summary of patients and caregivers who participated in the study

5.3.3 Diary Study

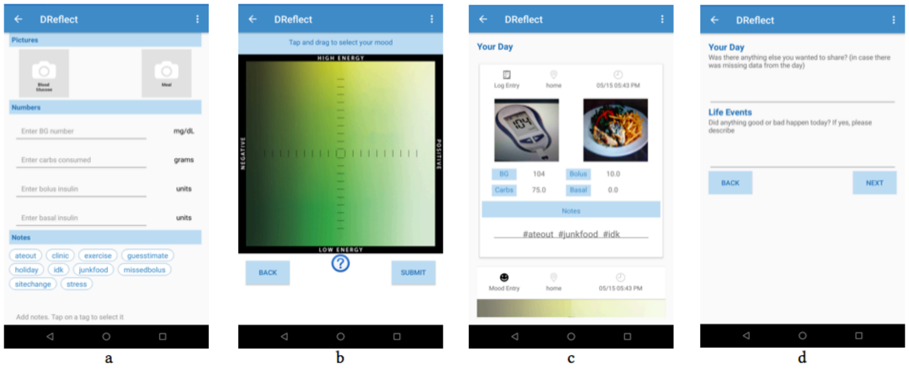
Patients. Patients were asked to log their meals (via captured images and estimated carbohydrates), blood glucose number (via text input and images of glucometer), basal and bolus insulin[5] (via text input), mood (using mood map [130]), and contextual factors affecting diabetes management and routine through hashtags and notes (Figure 15). While they had the freedom to create their own tags, I also gave them ten prepopulated tags that were developed by the study team's clinical expert based on her knowledge of typical factors that affect patients (#ateout, #junkfood, #exercise, #holiday, #idk, #missedbolus, #feelinghigh, #guesstimate, #sitechange, and #stress). The data collection application, DReflect (Figure 12), also tracked participants' location and, where possible, assigned location labels defined in the initial interview (Home, School). Lastly, patients were notified to complete an end of day diary that first showed a

---

[5] In Type 1 diabetes, the pancreas stops producing insulin, which leads to abnormal blood glucose levels. As a result, managing diabetes involves taking insulin and balancing it with food, exercise, and other factors. Insulin is administered using a combination of basal insulin and bolus insulin. Basal insulin is the long-acting insulin that manages blood glucose throughout the day. Bolus insulin is the short-acting insulin taken before meals to cover for carbohydrates in the food. (http://www.jdrf.org/about/what-is-t1d/facts/)

context-enhanced summary of data reported during the day and then asked them to report critical incidents for that day (diabetes and non-diabetes related), challenges of managing diabetes for the day, and other things that they might wish to share or forgot to report during the day. To ensure compliance with self-reporting, I also notified participants if they missed entering meal logs during a day and asked them why the entries were missed. I expected three log entries from participants each day and if they did not provide any data until 5pm, they were notified. In doing so, my aim was not to assess adherence to self-care activities but to gather sufficient data each day to enable end-of-day reflection on data.

Caregivers. I asked caregivers to fill out diary entries for 21 days. Each day at a time of their choosing I emailed them a link to an online questionnaire. They were asked to report about their day, the routine of their family, unusual events in their day and their child's day, and their perspective on the challenges of managing diabetes for that day.



BG 104
Bolus 10.0
Carbs 75.0
Basal 0.0

#ateout #junkfood #idk

Was there anything else you wanted to share? (in case there was missing data from the day)

Did anything good or bad happen today? If yes, please describe

Figure 12: The DReflect system for tracking Type 1 Diabetes data - a) Screen to add a meal log (BG number, food, insulin, notes), b) Screen to report mood, c) End-of-Day-Diary screen showing data recorded during the day, and d) An additional diary screen with reflection questions.

5.3.4 Data Analysis Phase 1 and Exit Interview Documents

After the diary study phase, data (diabetes logs, diary entries, and initial interview) were analyzed to understand critical incidents, such as breakdowns in management, atypical circumstances, unusual outcome measures, and missing data for each participant. During analysis, free text diary responses from parents' and patients' self-reports were cross-checked to assess any discrepancies. Insights from this analysis were used to prepare exit interview documents for patients and their caregivers. The exit interview document consisted of two parts. The first part displayed data visualizations (context-enhanced biomedical measures) to participants comprising of raw and aggregate data (Figure 13 - Figure 18). The second part of the exit interview document consisted of summary of events and specific questions based on breakdowns observed in the data. The questions for patients and caregivers were mostly identical with a few variations depending on the diary responses received from each. In this section of the document, I also prompted the participants to address any mismatch in the parent and the patient responses from the diary entries in two ways – 1) I asked them to confirm if the summary of events seemed representative of the study days, and 2) I asked questions related to specific events that did not match in patient and caregiver reports. The format of visualizations remained the same for all the participants. The questions asked were specific to the events reported by the participants.

The data visualizations in the first part of the exit interview document were prepared by the research team after reviewing commercial visualizations offered by diabetes-related device companies, and by reviewing relevant literature that characterizes basic visualization insights (trends, comparisons, and distribution) [34]. Among all the visualizations I reviewed, I selected Tidepool[6] formats (Figure 13 - Figure 15) to serve as a model for my displays, given their simple design. Additionally, I adopted the classic "logbook" version (Figure 16) of visualizations[7], considering their extensive use in the clinic during patient appointments and hence, the familiarity of patients with this specific format. Figure 2, Figure 3, Figure 4, and Figure 5 are modeled after existing visualizations and Figure 6 and Figure 7 are the novel ones developed by us. Specifically, Figure 2 and Figure 3[8] represent an adapted version of Tidepool visualizations, to which I added mood and energy labels. The scatter plot of blood glucose numbers in Figure 4[9] was also borrowed from Tidepool to which I added hashtags. My aim in developing these visualizations was not to evaluate them for their effectiveness but to present multiple representations of data for participants to have multiple opportunities to obtain insights.

In selecting aggregate data insights to be displayed in the visualizations, I was guided by Choe et al.'s research characterizing basic visualization insights, such as time trends, averages for data summary, and comparisons [34]. Additionally, the use of contextual data (location and type of day) was inspired by Epstein et al.'s concept of visual cuts where an outcome measure is profiled

---

[6] https://tidepool.org/products/tidepool/

[7] https://www.medtronicdiabetes.com/CareLinkPDF/Logbook/

[8] https://support.tidepool.org/article/29-viewing-your-data-daily-view

[9] https://support.tidepool.org/article/89-viewing-your-data-weekly-view

by different contextual parameters (e.g., amount of physical activity against commute time, and location) [56].

In deciding the low-level features of my visualizations (e.g., bars, lines, solid color or gradient), I created multiple versions of displays using pilot data and obtained feedback from the pilot study participants (two patient-caregiver dyads). Based on this feedback, I selected vertical bar charts, scatter plots, solid colors, and abstract keywords and symbols for mood and energy, such as high and low, to represent the data. The options evaluated by pilot participants included line charts, horizontal and vertical bar charts, combination of line and bar charts, scatter plots with solid colors and gradient, and scatter plot combined with bar charts.

Although I collected data from Continuous Glucose Monitors (CGM) and insulin pumps for the three weeks by having patients download data from these devices, I did not use that data for two reasons. First, not all participants used a CGM or an insulin pump. Second, for participants who used the CGM and/or the pump, their use of these devices was discontinuous because of accuracy issues, device failures, and preference to not wear CGM and/or pump, for example during family vacations. As a result, I could not obtain data consistently across participants. Additionally, although I gave participants the option to log basal insulin in addition to bolus insulin1, the majority of the participants did not log this information. Consequently, I dropped it from the analysis and from the data visualizations.

5.3.5 Exit Interview

After the diary study, I conducted an hour-long exit interview with the patient and the caregiver separately. The aim was to understand how they use context-enhanced displays to make sense of their experiences from the 21 days. To maximize recall, all the interviews except two (P10 and P12) were conducted within a week from when the participant ended the diary study phase. Participants were sent their exit interview documents that contained visualizations and questions (described above) a day or two. prior to their exit interview to review them. All the participants confirmed in the exit interview that they were able to review their data prior to the interview.

During the interview, I asked participants to walk through the data representations while thinking aloud. Additionally, when needed, I probed them by asking what they learned from the visualizations and how the data displays helped or didn't help them understand diabetes. Next, I asked them the questions from the second part of the exit interview document. These questions focused on days that were unusual or that involved a critical incident as identified from the diary data. These questions were supplemented with day-wise displays of their data to enable recall and reflective conversation [159]. Some participants printed out the exit interview documents for use during the interview. Other participants accessed the PDF file on their computers. They walked through the visualizations one by one as directed by the interviewer.

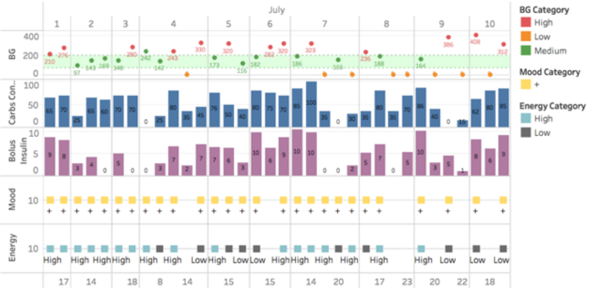


Figure 13: Multiple streams of data from each day (Blood glucose(BG), Carbs, Bolus insulin, Mood, and Energy)

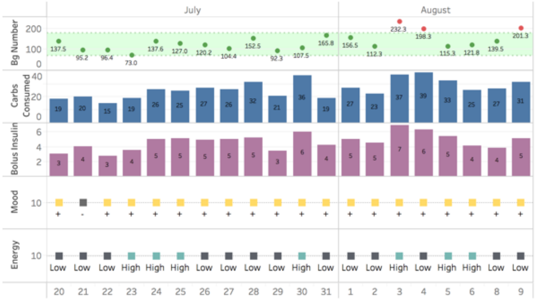


Figure 14: Average measure of BG number, Carb, Insulin, & Mood for each day over 21 days

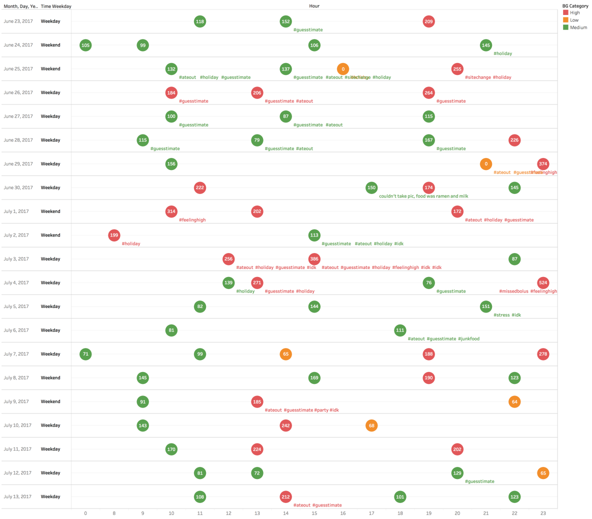


Figure 15: BG numbers displayed with hashtags & notes

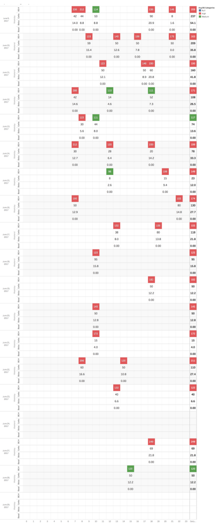


Figure 16: BG numbers displayed in format like a logbook template provided by the clinic

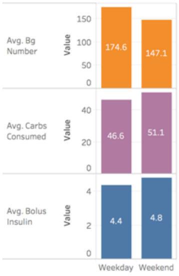


Figure 17: Average BG no., carbs, & insulin on weekdays & weekends

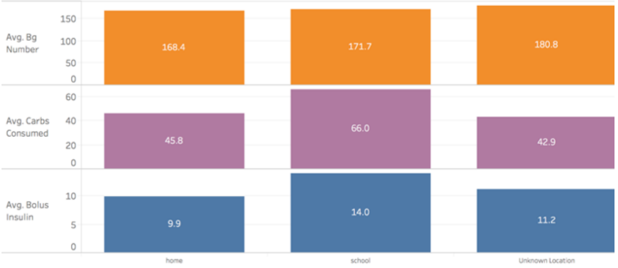


Figure 18: Average BG number, and carbs for each location

5.3.6 Data Analysis Phase 2

In this phase, the exit interviews from caregivers and patients were analyzed focusing on their analytical activities during the interview. The first iteration of data analysis was primarily guided by the interview questions and data was coded using in vivo and descriptive coding [150] to label the insights or findings patients gathered from their data, how they came up with those findings, what is challenging about using data to generate insights, and what other information could be useful. Next, the codes were iteratively reviewed to merge similar codes and disregard less prominent codes. The codes and associated interview quotes were then analyzed to group them into meaningful themes using affinity diagrams. These were discussed within the research team to further develop them.

5.4 FINDINGS

This section first describes data-based analytical activities that patients and caregivers engaged in while exploring their data during the exit interviews. Next, it describes the challenges that patients and caregivers faced while exploring and analyzing data. Lastly, it demonstrates how these analytical activities and challenges relate to the different sensemaking processes described in the data-frame theory of sensemaking to facilitate the emergence of information needs [96].

5.4.1 Data-Based Analytical Activities

When interpreting their data, I observed that participants engaged in a range of analytical activities, including four distinguishable activities that emerged as themes from the data — they observed context-based trends and explained them, they triangulated information from multiple factors, they generated context-specific self-care actions and anticipated the consequences of those actions, and they hypothesized about additional factors to explain their data.

Patients Found Context-based Trends in Outcome Measures and Explained Them

In analyzing their data, most of the patients and caregivers sought to understand trends in BG numbers in combination with other factors. While exploring time-based trends, patients and caregivers perceived the variation of BG numbers with time in different ways, such as trends in specific time periods of the day, "*mostly I think of it [trends] in time periods*." (P02 Mother), trends related to a specific time of the day over multiple days, "*A time of day, for me, has been my go to, 'Okay, let's hone in on a problem here because it's happening at the same time every day.'*" (P11 Mother), and trends within a day, "*A trend for me is when he has two to three highs right in a row.*" (P13 mother)

Participants also found trends using a combination of the primary outcome measure (BG number) and specific factors other than time, such as mood, food, insulin, sleep, type of day (e.g., weekend/weekday/vacation), location, physical activity, flu, travel, and other factors unique to each participant. For example, P12 spotted a trend relating to sleep (using the label #sleep), "*And I can also see a trend ... for example, when I wake up, it was mostly two or three hundreds [i.e., higher than recommended] and it was all #sleep so I know it was sleep*". Similarly, P06 and his mother noted that every BG number associated with "#running" was above range.

Investigating context-based trends in outcome measures prompted patients and caregivers to generate explanations when they noticed something obvious or unusual. For example, noticing a trend of difference in numbers between the first half and the second half of the day, P02's mother explained that not checking and correcting BG numbers in between meals was the reason why the patient had high numbers in the second half of the day, "*When he comes home from school he doesn't immediately look at his [Continuous Glucose Monitor] and say, oh, I better give myself the correction right now so that I don't end up high at dinner.*" Similarly, seeing how BG numbers remained steady with location, P15's mother explained why there was no variation in BG numbers or food across different locations, "*going to grandma's cottage wasn't that different than being at home on the weekend. You still have consistent meals and snacks. I might see a bigger difference if it's her being away at camp or, you know, a trip to the amusement park like I said, versus a* *normal day at home.*" She further projected that locations unlike home that are associated with more activity could affect management in a more noticeable way.

Patients Triangulated Information from Multiple Factors.

In addition to understanding the effect of individual factors on a certain measure (amount of carbs consumed, or BG number), patients connected multiple factors to obtain a more comprehensive understanding. This helped them discover new insights and validate existing knowledge (confirm or deny). For example, when looking at the variation of numbers across different locations (Fig. 7: home, grandparents' house, church/youth group, others) and different types of days (Fig. 6: weekdays and weekends), P01's mother noted, "*It reaffirms what I knew about her grandparents' house. It doesn't surprise me at all that her sugars are really high over the weekends and yet she doesn't eat many carbs, because I don't know that my parents are generally telling her, hey did you shoot [insulin] for that? Do you know how many carbs that is?*" The caregiver in this case saw that the patient has high BG numbers on weekends (type of day) and is also at her grandparents' house (location) for weekends. She triangulated information from these two contextual factors to confirm her suspicion that diabetes management does not happen as expected over the weekends at the grandparents' house, resulting in high BG numbers.

Similarly, P03's mother discovered that P03 did better when they were both away from home together. This led to the conclusion that being together during the time away from home afforded more supervision and better control, "*She was getting better control when she was not at home. [laughs] I'm thinking a lot of the "unknown location" [locations not labeled during the initial interview] was when we were on vacation. And I was with her so I was, like, on top of her all the time making sure things were checked and things were done more accurately. I realized it when I looked at this [data], I didn't really realize it before.*" They both got more time together when they were out of home for vacation, or family dinners because P03's mother was at work most of the time otherwise.

As a part of triangulation, a few patients and caregivers (P03 Parent, P11 Patient and Parent, P12 Parent) moved between displays to compile the inferences they obtained and confirm their understanding, *"I looked at graph one and I could see ... as I would go down and looked at the others, it just kind of confirmed where I was at."* (P03 Parent)

Patients Generated Suggestions for Context-based Self-care Actions and Their Consequences.

Patients suggested actions that could help them change insulin dose, eating behavior, and adherence to self-care activities to obtain better control over their BG numbers. For example, while understanding the time-based trends in BG numbers, P03's mother broke down the day into three parts (breakfast, lunch, and dinner) to identify the time windows for which insulin dose could be adjusted in response to food intake. Similarly, P12's mother investigated three types of data streams together (BG numbers, carbs, and insulin) and compared the existing patterns with what the expected patterns in these streams should be to identify corrective action, "*if the carbs pattern matches the blood glucose pattern then the insulin needs to be higher to bring the blood glucoses down. What it tells me when I look at it that way is that his carb ratio [a patient-specific value that determines how much bolus insulin to give based on carbohydrates consumed] is not high enough.*" She concluded that the similarity between carb pattern and glucose pattern implies that the carb ratio should change to keep the BG numbers steady.

While the action suggestions in the abovementioned cases naturally followed from the observation of trends and aimed at fixing only the current problem, there were other, more intricate suggestions that also considered the interaction between multiple factors and the potential impact on the anticipated consequences of a self-suggested action. For example, P06's mother pointed out that P06 was often high after running and explained how a temporary dose of insulin might help bring down the numbers but she also had to be cautious about the additive effects of insulin and physical activity to avoid low BG numbers, "*We've talked about maybe doing that temp basal after he gets done running, but typically he comes right down after running. He'll do a correction and he comes down. So, running a temp basal makes me a little nervous because I don't want him going [too] low later.*"

Patients Hypothesized About Alternate Contextual Factors to Explain the Data.

When participants could not explain the outcome measures (BG numbers) using factors present in the displays, they generated ideas about alternate factors. For example, P08's mother noted the need to understand sleep patterns with the BG number patterns, especially during a trip in the summer vacation when the sleep schedule was unusual, *"on [the trip], he slept a lot, he slept late in the day, he was high. So, I'm wondering what impact that has.* Similarly, P02 thought of physical activity to consider as a factor to explain his numbers for a given day, *"I think one of the things that would be helpful for me because I have such an erratic schedule would be like activity level. That would probably be helpful for myself because I am not consistently being active."* Another factor that was considered was specific food choices, "*I would love to see this with food choices, like… Something that would help to really hone in on, 'Does that really have a major impact on the numbers?'."* (P11 parent)

5.4.2 Data Exploration Challenges

Even though patients and their caregivers meaningfully engaged with patients' data, they faced two challenges – the inability to identify clear trends challenged action planning and counterintuitive insights were hard to explain and compromised trust in data.

Inability to Identify Clear Trends Challenged Action Planning

Particular characteristics of the data patients generated influenced whether participants were able to identify any trends. For example, caregivers of P02 and P08 were unable to find trends in the data, which made it difficult to obtain any insights, "*considering most of the boxes are red, it's hard to really see a true pattern because he's high, like, what? 85-90 percent of the time here so it's really hard to glean anything from that other than he's high a lot.*" (P08 Mother). While having high numbers most of the times can be seen as a "pattern" in itself, participants and caregivers did not perceive such general tendencies of the data as useful for action planning. As a result, participants were unable to think of the next steps, "*It was so hard to figure out what to do next because it was just all over the place, so random.*" (P11 Mother). In such cases, patients and caregivers felt the need to call the clinic to make dose changes or understand the next steps, "*so what we would do is when it's random like that we call the clinic and we say, 'Tell me the first place to start because there's nothing clear for me to do*'" (P12 Mother).

Similarly, when a contextual factor resulted in both desirable and undesirable trends, patients found it difficult to understand its true effect and needed more information. For example, P04's mom could not understand the effect of eating out (#ateout) and guessing carbs (#guesstimate) on P04's BG numbers (Figure 15), "*I'm just looking at the red [numbers] talking about guessing and eating out. But then even [for] her green ones [number], there's guessing and eating out and she was green."* She felt the need for more information to quantify the effect of the same factor on in range and out of range BG numbers and the accuracy of guesses with different types of food eaten outside home.

Counterintuitive Insights Were Hard to Explain and Compromised Trust in Data.

When engagement with data resulted in counterintuitive insights, participants found it difficult to explain those. For example, P06's mother was surprised to observe that the patient was not so well controlled at home, as she thought that the home environment provided an ideal set of self-management "tools." "*I think I would have expected him to have better numbers at home so I'm kind of surprised. Only because we have the tools at home.* " (P06 parent).

In another case, P04 and her mother found it difficult to convince themselves of the truthfulness of a counterintuitive insight that they got from the data when they saw that P04's weekday averages were higher than weekend averages, "*I don't think this is a completely accurate representation of... you would think that the BG would be higher on the weekend, but it was actually lower. But I think maybe if you were to do the next 21 days, things could be flipped, you know, the weekday would be lower and the weekend higher.*" (P04 Mother).

5.4.3 Understanding the Emergence of Information Needs from the Data-Frame Perspective

As I conducted data analysis, I was struck by the connections between the findings and the processes described in the data-frame theory of sensemaking [96]. While I did not set out to apply this theory to analyze my data, the synergy between my findings and the data-frame theory of sensemaking inspired me to further interpret the findings from the perspective of this theory. Doing so helped me reflect on and add to the existing perspective on sensemaking for disease self-management [122]. It also helped me demonstrate the value of the sensemaking approach for exploring contextual data, considering that such data is hard to capture and analyze computationally. I discuss these points in section 5. In what follows, I first describe the important aspects of the data-frame theory and then demonstrate its relevance to the findings.

The data-frame theory describes sensemaking in terms of two entities: data and frames. Data is defined as the "interpreted signal of events". Frames are "explanatory structures" that describe the relationship between different data elements. A frame could be a story describing events in a chronological sequence, a map showing spatial layout, a script describing the steps of a routine job, or a plan showing the sequence of future actions.

In this theory, sensemaking has been defined as the process of fitting data into a frame to establish relationships between different data elements and, hence, explain the data [96]. It is these relationships that, in turn, assign relevance to different data elements. In addition to identifying a frame, sensemaking also involves defining what counts as data based on the emerging frame. The process of constructing a frame begins with a few salient data elements called “anchors”. Multiple anchors can be compiled to create a frame that explains an event. If new data not fitting into the frame are perceived, the sensemaking process modifies the frame or creates a new frame to fit the new data element(s). In the course of constructing a frame, different lower-level sensemaking processes may occur, relating to how the data and the frame mutually inform each other. These include seeking a frame, elaborating the frame, preserving the frame, questioning the frame, re-framing, and comparing frames [96].

From the perspective of this theory, the findings show how the information needs of patients emerge and evolve through these sensemaking processes, as they refine their understanding and awareness. With regards to disease management, the majority of the “frames” that I refer to here are cause-effect relationships and conditions in which these relationships manifest, leading to narratives describing when and how one or more factors (either clinical or contextual) affected an outcome measure. Here, contextual factors (e.g., location, mood, hashtags and notes), and clinical factors (e.g., insulin, carbohydrates) served as anchors to create a frame (i.e. an explanatory structure) describing different ways in which the factors affect or could affect the outcome measures (i.e. the BG numbers).

Seeking a Frame.

Seeking or constructing a frame involves deriving a frame from the data by identifying plausible anchors (factors likely to affect BG numbers) and then filtering the data based on the emerging frame. In this study, patients and caregivers constructed frames by identifying context-based trends and then seeking information to explain those trends, i.e., by using one or more factors as anchors to explain the variations in outcome measures. For example, P04 considered sleep to be an anchor to explain high numbers after waking up but she could not find consistent data to support this frame and eventually rejected it, "*sometimes if I like wake up later than usual, then like my blood sugar might run like a little bit high, but it's not really consistent, so there's not really any consistent difference*." In another instance, P11 recalled one of the days from the study to understand a set of BG number variations wherein she considered multiple anchors to explain what affected the numbers on that day, *"On August 5th, I was on a road trip so I was sitting in the car and eating a lot. And it was earlier in the morning so, like, tired mood, plus also sitting and eating made my number go really high."*

Elaborating the Frame.

According to the data-frame theory, once sensemakers have a frame, they try to extend or strengthen this frame by seeking and inferring more data to fit the frame. In this study, participants started with frames based on their lived experiences with diabetes. For example, as previously described, P01's caregiver held the frame that P01 tends to have poor control when she is visiting grandparents' because nobody reminds her to take care of herself. Seeing data coherent with one of her frames, she extended the frame to fit other relevant anchors by triangulation. In this case, the caregiver triangulated BG number variations by two factors – location and type of day. The distribution of BG numbers by weekdays and weekends demonstrated poor management on weekends. Based on this data, in addition to associating poor management with the grandparents' house (location), P01's mother added "weekends" (type of day) as another anchor to her explanation. The caregiver elaborated the frame initially described by "grandparents' house is associated with poor management of diabetes" to "her diabetes is poorly managed at grandparents' house and on weekends. P01 is at grandparents' house on weekends."

Preserving the Frame.

At times, sensemakers commit to a frame and then, when they encounter data that does not fit the frame, they distort the data to fit the frame or find an explanation to discard the data while keeping the frame. For example, P04 and her caregiver believed that she tends to go higher on weekends. However, her data from three weeks showed otherwise (i.e. higher on weekdays). On seeing the data, P04 questioned the calculation of averages, "*I feel like it might just be the, like, imbalance between how many weekdays there are and how many days there are in the weekend*". P04's caregiver did not think it was a true representation, "*I don't think [the difference between* *weekdays and weekends] has much to do with anything.*" The caregiver and the patient explained away the data that did not align with their frame of how weekends and weekdays affected BG numbers.

Comparing Frames.

Another form of sensemaking is identifying alternate frames or sharpening differences between two frames. In reviewing their data, patients and caregivers came up with alternate frames using hypothetical anchors (not present in the displays). For example, as described above patients and caregivers hypothesized and sought information about multiple factors not present in the display to explain the data (sleep, physical activity, and type of food). In addition to identifying alternate frames, they also compared frames to make explicit the distinction between them. For example, observing less difference in numbers between weekdays and weekends, P08's caregiver considered two frames, one based on summer vacation and the other on school routine. During summer vacation, the variation between weekend and weekday numbers was less because every day was the same. But during the schoolyear, she projected that this difference was going to be higher because weekdays were different from weekends. Similarly, P15 compared two different locations for similar effect in summer vacation, as described previously.

Re-framing.

Reframing could involve establishing new anchors, looking at discarded data, or reinterpreting the data. For understanding changes to BG numbers in patients with type 1 diabetes, the two most salient anchors to explain fluctuations in BG numbers are the amount of insulin taken and the amount of carbohydrates consumed. However, other factors, such as physical activity, or the specific type of food could also affect BG numbers. P11's caregiver looked at the patient's BG numbers along with carbohydrates and insulin data. In contrast to general beliefs about the determinants of BG changes, the caregiver concluded that the low BG numbers were not caused by insulin or by carbohydrate consumption, "*It seems like when she had the lows it didn't necessarily mean, like, over-bolusing or, like, a high carb consumption.*" She then thought of other factors (exercise and food type) that could be causing such variations in the BG numbers to re-frame the data using other anchors, "*I'm assuming it's the exercise thing. Or perhaps the food choices that I'm not sure, probably both play into it.*"

Questioning the Frame.

Sensemakers question a frame when they encounter inconsistent data or detect an anomaly, or when their expectations in the context of a particular frame are violated. When patients and caregivers derived counterintuitive insights as described above, they questioned the anchors they were considering. For example, P04's caregiver wanted to understand if eating out and guessing the amount of carbohydrates in food was affecting the patient's numbers. Looking at the days associated with #ateout #guesstimate, she could see both red (out of range) numbers and green (within range) numbers. She was left questioning if eating out and guessing were relevant anchors to explain the numbers, "*so it's like, it didn't even matter whether she was guessing or eating out.*" She sought more information on how the distribution of BG numbers is affected by these events of eating out and guessing carbs.

5.5 DISCUSSION

In this study, I explored how patients and caregivers made sense of multiple streams of data collected by the patients. While prior work identifies insights that patients want from their data, I provide an understanding of the interpretation activities that result in these insights. I observed four analytical activities in which participants engaged while interpreting their data – finding context-based trends and explaining them, triangulating multiple factors, suggesting context-specific actions, and hypothesizing about alternate contextual factors affecting outcomes. In pursuing these analyses, participants faced two challenges – inability to identify clear trends challenged action planning and counterintuitive insights compromised trust in data. I interpret the findings using the data-frame theory of sensemaking [96] to show how participants' information needs emerge and evolve through different sensemaking processes described in this theory (e.g., seeking a frame by identifying trends and factors causing them, elaborating a frame by triangulating multiple factors).

In this section, I compare sensemaking as observed in this study with prior conceptualizations to demonstrate that sensemaking can not only inform disease-management action, but it can also result in the identification of new information needs and drive further exploration (as described in the interpretation of findings through the data-frame theory of sensemaking). This makes sensemaking a valuable approach for facilitating the emergence and discovery of patients' information needs, which can guide the collection of further data that are relevant to fulfill those information needs. I thus argue that contextual information should be explored using a sensemaking approach. Such data, which come with their own challenges of capture and analysis, need not all be collected upfront, as sensemaking could drive information needs and hence, guide the collection of data that would be needed to fulfill those information needs.

5.5.1 Sensemaking for Guiding Action

Existing sensemaking perspectives on chronic disease self-management describe disease management behavior in terms of continuously perceiving and inferring information and experiences to inform action [122]. For example, in the "sensemaking" mode of diabetes management described by Mamykina et al., patients may construct new explanatory models and take an experimental action instead of a routine action to accommodate new experiences. Here, developing explanatory models is essential for action. In a similar direction, the findings show that when participants were able to explain their data, they were more easily able to identify the need for an action (i.e., if a self-care action was required or not). Additionally, in a few cases, they were also able to identify the action they should take.

Sensemaking to Identify Information Needs for Guiding Data Collection and Exploration

There were also instances when the participants were not entirely able to explain their data (i.e. construct a frame) and finish their exploration. For example, the challenges that participants faced while reviewing their data exemplify cases when they were unable to understand or explain the data. While this did not result in a suggestion for action, it prompted them to think of alternate anchors not present in the data display that could meaningfully explain the data. Because one of the key aspects of sensemaking is to identify and define relevant data, even when such data is not available, sensemaking can proceed with access to limited data (p. 122) [96]. For health-related tracking, this could suggest an episodic approach to the collection of contextual data. That is, contextual data might not need to be tracked and reflected upon continuously (e.g., tracking needs are governed by patient condition [134]) and patients might not always know what data to collect. Instead, the need for more relevant data could emerge as and when patients make sense of the available data. Thus, sensemaking could help patients understand the collection and use of contextual data for health-related tracking without necessitating the availability of comprehensive data upfront.

This also highlights that a possible outcome of sensemaking is exploration by identifying information needs, as also described in the data-frame theory of sensemaking (i.e. when data-frame congruence does not happen [96]). This observation differs from Mamykina, et al.'s, sensemaking-based disease management model that considers action as a necessary outcome of sensemaking in the perception-inference-action cycle [122]. Instead, my observations align with the data-frame theory of sensemaking to suggest that in addition to action, sensemaking could drive information needs for meaningful exploration. These information needs are continuously shaped and reshaped through the iterative application of several sensemaking processes described above that include seeking a frame, elaborating a frame, preserving a frame, comparing frames, re-framing, and questioning a frame. The identification and refinement of information needs in turn can potentially help patients identify the data they need to collect in response to the information needs.

5.6 DESIGNING TOOLS TO SUPPORT SENSEMAKING

My findings and my reflection on different sensemaking perspectives are suggestive of design implications for sensemaking tools to improve patients' awareness of factors affecting outcomes and to support their agency as sensemakers.

5.6.1 Improving Awareness of Factors Affecting Outcomes

A recent body of work has explored the use of self-experimentation tools that help patients collect relevant data to test self-generated hypotheses and establish associations between factors and outcomes (e.g., dairy worsens irritable bowel symptoms). To form hypotheses, such tools rely on patient's own knowledge of what factors are affecting their symptoms or on clinicians to guide patients in selecting independent and dependent variables for the experiment. While disease specific outcome measures can be universal, we cannot assume that all the factors affecting these measures are necessarily universal among patients. Additionally, there could be factors or combination of factors unique to each patient that the patient or their clinicians might not be entirely aware of. This suggests the need to support a pre-experimentation step to explore the factors that are worth testing. In the pre-experimentation phase, patients could gain an improved understanding of the relevant factors or information needs through retrospective sensemaking, similar to what I observed in this study. My findings show that when patients engaged in sensemaking, they identified information needs to continue the exploration in order to eventually explain their data, as described above. They hypothesized about alternate factors not present in the data displays to explain their data. In this way, sensemaking resulted in better awareness of relevant factors affecting outcomes and the understanding of data that needs to be collected in response to an emerging hypothesis and information needs. Existing tools do not entirely support an informed selection of factors to test among the many that could be affecting patients' condition [86]. These tools can be extended by including a pre-experimentation phase where patients collect multiple streams of data including factors and outcome measures, rank these factors through retrospective analysis and sensemaking, and conduct experiments based on the prominence of the factors. Additionally, the system can suggest potential factors to the patient to collect data on and further explore that factor.

5.6.2 Supporting the Sensemaker's Agency

Automation undermining individual agency is of growing concern in the field of personal informatics [45]. While computational support is necessary when dealing with multiple streams of data as it is challenging for the users to make sense of it [36], it is also important to understand the nuanced role of human intelligence and cognitive processes that need to be scaffolded. My findings are suggestive of three ways to support the sensemaker's agency.

Enable Feedback on Computational Insights

The data analysis activities of patients revealed that when they obtained counterintuitive insights, they did not always believe them, rather questioning how the data was analyzed or how the results might change if they collected another data sample. This shows that even simple computational insights gathered from data might not truly reflect the lived experiences of patients, which was also demonstrated by prior research on self-experimentation (patients did not entirely believe the conclusion reached by the self-experimentation application [86]). This could happen because of reasons such as, data was collected over a short period of time, or the data collection happened over a time that was not representative of typical days.

The possibility that these insights might not resonate with patients' lived experiences suggests that tools that provide insights drawn from patient-generated data need to support manual assessment of the insights. One way in which this could be accomplished is by giving patients the ability to provide feedback on the system-generated insights. For example, participants could rate the insights on a scale of 1 to 10 to indicate how representative of their experiences an insight is.

Similarly, tools that generate and test health recommendations, such as the sleep recommendations generated by SleepCoacher [47], can be extended by letting users prioritize system-generated recommendations for implementation and experimentation. Obtaining user feedback on system-generated information can help researchers understand the performance of experimentation or analysis techniques over time to assess if the techniques need to change to provide better recommendations or there are genuine variations in data that the user is unaware of.

Support Insight Validation Through Triangulation

Our findings show that patients and caregivers triangulated insights from multiple factors. While it is important and perhaps easier to understand the effect of single factors on an outcome measure, multiple factors operate together in everyday context of disease management, as reflected by participants' experiences. Hence, it becomes important to understand the cumulative effect of these factors in several combinations that they manifest. Cross verification between factors is one way in which participants increased their understanding of the cumulative and nuanced effect of these factors on both the outcome measure and the disease management behavior. Tools that help patients collect multiple streams of data need to provide ways in which triangulation can be performed as it could potentially increase users' confidence in self-generated insights. For example, self-experimentation tools [87] could be extended to use multiple experiments to cross verify insights from each of them instead of relying on one experiment for a conclusion.

Support Anticipation of Future Conditions and Consequences

Some of the participants not only wanted to understand the past and the present from their data, they also projected insights for different contextual conditions and anticipated the consequences of self-care actions. Tools should support simulation to enable users in understanding the impact of a different contextual condition that is less apparent in a data snapshot. For example, context prediction algorithms can be used to infer future contexts from the past contexts [125]. Additionally, such prediction could be supplemented with information about uncertainty for users to exercise discretion in assigning relevance to these predictions.

Enabling Clinician Involvement

When understanding self-care actions and their consequences, it will be essential for tools to enable patients to involve clinicians when needed. For example, patients could invite clinicians to review a self-care action that resulted from a patient's analysis of their data. Clinicians could then validate the self-care action suggested by the patient while also providing information about what consequences should the patient be aware of, what other data could help them make a better decision or recommending other self-care actions. Such information can also be automatically sourced from disease-specific knowledge repositories built over time in collaboration with clinicians (e.g., diabetes [42]) and made available to patients through standard templates if they do not wish to seek clinician support [117].

5.7 LIMITATIONS

Our study participants were adults and adolescents. In the exit interviews, both these participants separately engaged in sensemaking with patients' data. While the simpler analytical activities (e.g., finding a trend) could be generalized for both these populations as I saw both patients and caregivers engage in these activities, the complex analytical activities (e.g., triangulation, hypothesizing, suggesting action) would perhaps be more natural for adults than adolescents because of differences in education and cognitive abilities. Additionally, one of the factors affecting data interpretation could be data literacy, which I did not assess in this study. However, I used simple representations of data, which is perhaps why I did not find any of the participants struggling to engage with their data.

We acknowledge that the granularity of data collected by my self-tracking tool, DReflect, was less than what could have been obtained by continuous glucose monitors (CGM). However, my aim in this study was to obtain a volume of self-reports sufficient to engage participants in self-reflection at the end of each day during the diary study phase. Additionally, the variety of CGMs used by the participants posed a technical challenge in obtaining this data because it would have required introducing participants to a new tool suite (e.g., Tidepool[10]) that they were unfamiliar with.

Regarding the visualizations I presented to participants, I want to note that having granular data could have resulted in more informative visualizations. However, my aim in this study was to present simple visualizations to participants, considering that the study involved working with adolescents. Some of the visualizations are a lot simpler than the others. For example, the aggregate

---

[10] https://tidepool.org/products/tidepool/

visualizations (Figure 17 and Figure 18) show simple insights, such as comparison of BG numbers between weekends and weekdays, and comparison of BG numbers across locations. While these displays might seem simple, they provide a contextual profile of patient's disease control and aim at improving awareness of the contexts in which the patient needs to improve disease management. The majority of the participants found these displays useful for improving awareness.

5.8 CONCLUSION

Given the limited understanding of how patients could interpret clinical data with contextual data, I explored interpretation activities of patients with Type 1 diabetes and their caregivers using context-enhanced visualizations I created using patient data collected through a three-week diary study. I found four different analytical activities performed by participants as they interpreted contextual and clinical data - finding context-based trends and explaining them, triangulating multiple factors, suggesting context-specific actions, and hypothesizing about alternate contextual factors affecting outcomes. I also noted two challenges that participants faced in engaging with their data – inability to identify clear trends challenged action planning and counterintuitive insights compromised trust in data. Interpreting these findings using the data-frame theory of sensemaking [96] shows how information needs of participants were shaped by the various sensemaking processes, which I elaborate in the paper. Situating these findings within two existing perspectives on sensemaking (the data-frame theory and Mamykina et al.'s sensemaking framework for chronic disease self-management), I show that sensemaking could not only result in action as described by Mamykina et al.'s sensemaking model for chronic disease self-management [122], but it could also generate new information needs and guide exploration. Considering the challenges of capturing and computationally analyzing different types of contextual data, I argue that the sensemaking approach should be used for exploring contextual data as it can result in information needs that could further guide the collection of relevant data in response to those needs. Lastly, noting the value of sensemaking approach for using contextual data, I provide design suggestions for sensemaking tools to improve awareness of contextual factors affecting patients and to support patients' agency in making sense of data.

This study contributes in three different ways. First, while prior research shows the types of insights patients want from their data, this work contributes an understanding of the interpretation activities that lead to those insights. Second, I inform the current sensemaking perspective on chronic disease self-management by comparing it with the data-frame theory of sensemaking. Recognizing the complexity of interpreting contextual data, I identify a theoretical basis for understanding the use of contextual data by proposing that the sensemaking approach should be used for exploring contextual data. Third, I contribute design knowledge for sensemaking tools to support exploration of multiple data streams including clinical and contextual data.

In the next two chapters, I describe the findings from the investigation of collaborative data practices in the context of commercially available tools used during collaborative review of data by patients and clinicians. Taken together, these two studies help understand common data practices and objectives of patients and clinicians in using patient-generated data to inform care.

Chapter 6 Data-Driven Assessments and Tools for Diabetes Data Review

The previous chapter provides an understanding of individual data practices of patients and caregivers. In this chapter and the next one, I report on a study where I investigated how patients and clinicians collaboratively engage in biweekly data review sessions as a part of a telemedicine program over a period of 6 months. In addition to analyzing these data review sessions, I reviewed the commercially existing data reporting tools, some of which were used by patients and clinicians during the data review sessions.

In this chapter, I describe the different types of assessments that are performed using patient data and the information sought from the data reports for those assessments. Additionally, I describe the data reporting tools that are used and point to their limitations in the context of the assessments performed.

6.1 SUMMARY

Background: While there is much understanding of mobile apps for diabetes management, there is no review of the reporting tools that are provided by diabetes device manufacturers, such as Dexcom and Medtronic. Type 1 diabetes patients and clinicians find it challenging to use these commercially available tools. Given the lack of understanding about these tools, there is a need to assess them in relation to the information that patients and clinicians seek from them and identify the opportunities to improve their design.

Objective: The objective of this study is to – a) identify the assessments performed using patient data and the types of data that those assessments rely on, and b) understand the design space of the commercially available data reporting tools to review Type 1 diabetes data.

Methods: Analysis of transcripts of data review sessions between patients, caregivers, and providers to identify assessment tasks and information required for those tasks; a review of 6 diabetes data platforms to document what data is shown, how it is shown, and connect them to the assessment tasks performed using data; focus groups with providers.

Results: Four types of assessments are performed using patient data – data sufficiency, outcomes, patient or system behaviors, and regimen quality. These assessments rely on different types of information that are sought from the commercially available data reports, such as distribution of BG numbers, and frequency of insulin intake. The review of these data reports shows that they provide limited support for understanding patient behaviors - their primary focus being the presentation of insights from BG values. Additionally, these reports create information overload - they contain a wide variety of displays. The multiplicity of displays can be attributed to the use of several visualization formats, more than one representation of time and more than one categorization of BG values. I point to the ways in which this multiplicity of displays could challenge effective use of data through these platforms.

6.2 MOTIVATION

One of the challenges in making sense of patient-generated data is the lack of suitable displays for reviewing data. To design data interfaces aligned with actual needs and practices, it is important to understand available tools in the context of user needs and practices, which motivates the need to better understand – a) how the currently available tools are employed for making sense of patient-generated data and arriving at care decisions, b) the specific types of information that patients and clinicians seek from these tools and c) how these tools support or limit the use of data.

6.3 BACKGROUND ON TYPE 1 DIABETES

Type 1 diabetes is an autoimmune condition in which the pancreas produces little to no insulin resulting in abnormal blood glucose (BG) levels. As a result, T1D patients depend on insulin to control their blood sugar levels. T1D is usually diagnosed in children but it can develop at any age. This condition cannot be cured requiring life-long self-management. Regular self-management tasks involve monitoring blood sugar levels multiple times during the day, counting carbohydrates and taking insulin when eating, taking insulin in between meals, and adjusting the dose of insulin based on several factors affecting blood sugar levels, including the amount of carbohydrates eaten, the activity levels, and the overall routine [183]. There are two types of insulin that patients need to take based on a regimen decided in collaboration with their endocrinologists– long-acting insulin (basal insulin) and short acting insulin (bolus insulin) [196]. The two types of insulin differ in how quickly they act and how long they remain effective in reducing blood glucose levels. The long-acting insulin acts over multiple hours after taking it whereas the short acting insulin starts to work within 15 minutes of administration and is usually taken for covering food [197]. Medical devices, such as the continuous glucose monitor (CGM) and the insulin pump, automate some of the abovementioned tasks. Specifically, the insulin pump can be configured to automate the calculation of basal insulin and bolus insulin. The two configuration parameters used are insulin to carb ratio and correction factor or insulin sensitivity factor [196]. The insulin to carb ratio is the amount of insulin to be administered for every 10 grams of carbohydrates consumed. Different meals during the day could use a different insulin to carb ratio depending on what is eaten and how it affects the patient's blood glucose levels. The other configuration parameter, the correction factor, is the amount of insulin to be administered for every unit increase in blood sugar above the target blood sugar number or the amount of insulin to be deducted for every unit decrease in blood sugar below the target number. When using the insulin pump, the patients need to record two data points in their pump at the beginning of a meal - the current BG number and the carbohydrate amount. The pump then calculates the amount of insulin required by using the insulin to carb ratio to adjust for food and the correction factor to adjust for the difference between the current BG number and the target BG number [196].

The pump is also configured to enable bursts of basal insulin infusion depending on the amount of insulin a patient requires. Different amounts of basal insulin are scheduled to be delivered at different times during the day [153]. The two parameters described above, and the basal insulin settings are continuously assessed and changed in response to the patient's glycemic performance after reviewing their data. Hemoglobin A1c (also known as HbA1c or A1c) is the estimated average glucose over a certain period and is used as the marker of glycemic performance. It is assessed about every three months during a patient's clinic visit [198].

6.4 METHODS

The objective of this study was to understand how patient-generated data (for T1D) is reviewed using commercially available data reports. The study was conducted in multiple phases that involved analysis of data review sessions, review of commercial reporting platforms, and analysis of two focus groups of diabetes providers.

6.4.1 Analysis of Data Review Sessions

I first analyzed the transcripts of 71 data review sessions between patients and clinicians held remotely on BlueJeans (a video conferencing platform) as a part of a 6-month telemedicine program. The participant pool consisted of 6 adolescent patients, their caregivers, 2 endocrinologists, and one diabetes educator. The telemedicine program involved biweekly check-in sessions between caregivers of patients and the diabetes educator, and two 3-month virtual clinic visits with their endocrinologists. Each of these sessions involved a review of 14 days of patients' data, which they downloaded before every session. I coded the transcripts of the remote data review sessions for the following - topics discussed during a session, questions raised by patients or providers, and the information that was exchanged, which involved information related to BG values, insulin and carb intake, insulin pump settings, and other aspects of the patients' lives, such as physical activity, school, and vacation. This led to the identification of different types of assessments performed using data and the corresponding parameters used for the assessments.

6.4.2 Review of Commercial Diabetes Data Platforms

In the next phase, to further understand the role of the device data reports in facilitating the data review sessions, I reviewed the data reports provided by 6 diabetes platforms – Dexcom Clarity[11], Glooko Diasend[12], Tandem[13], Libre View[14], Medtronic Carelink[15], and Tidepool[16]. The review used static data reports made available by the patients. Depending on how the reports were generated, they may not have captured all of the platform's features. I analyzed one example of each kind of report. The results represent the displays that were available in those specific reports. The goal of the review was to identify how different types of data are visualized and the extent to which they provide the information required for different types of assessments. For each display in the reports generated by the abovementioned platforms, I documented the following – the type of data shown, the transformations performed on the data (e.g., calculation of descriptive statistics), the source of the data (i.e., the device), time duration over which the data was presented,

---

[11] https://www.dexcom.com/clarity

[12] https://amsldiabetes.com.au/products/glooko-diasend/

[13] https://www.tandemdiabetes.com/products/software-apps/tconnect-application

[14] https://freestylediabetes.ie/freestyle-libre/libreview-reports

[15] https://www.medtronicdiabetes.com/products/carelink-personal-diabetes-software

[16] https://www.tidepool.org/viewing-your-data

format of the visualizations, time scales used to present the data, BG categorizations used, and the assessment tasks supported by the visualizations.

6.4.3 Analysis of Provider Focus Groups

Lastly, I analyzed the transcripts of two focus groups that involved experts discussing the different types of reports, their advantages, and limitations. The focus group included experts from different clinics. Both the focus groups consisted of 5 experts with one expert serving as the moderator for both the sessions. The participants included adult and pediatric endocrinologists and certified diabetes educators. These focus groups were conducted by the pediatric endocrinology clinic of Michigan Medicine. They involved understanding the different displays that each of the providers used, the challenges of using them, and walkthrough of data reports.

6.5 FINDINGS

In this section, I describe the different types of assessments performed using patient data and the corresponding information that these assessments rely on. I then describe the characteristics of the reporting platforms that limit their effective use for reviewing patient-generated data.

6.5.1 Data-Driven Assessments

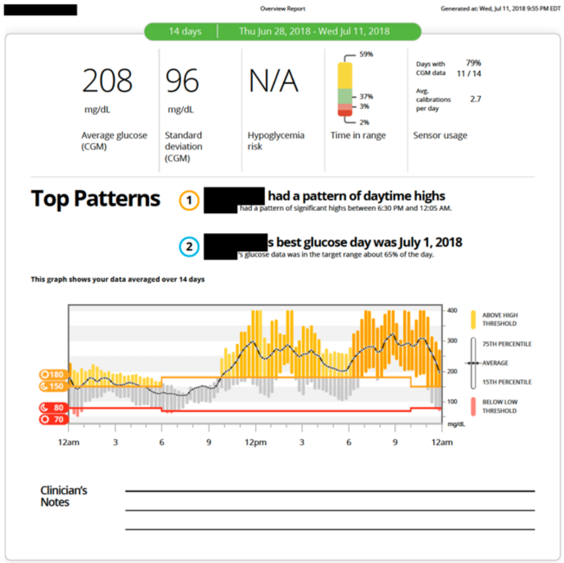
Different types of assessments performed using patient data include assessment of CGM data sufficiency, BG outcomes, patient and/or system behaviors, and regimen quality. These assessments rely on different types of information obtained from the data. In this section, I describe the different types of assessments performed, the corresponding parameters that these assessments rely on, and the displays that are used. Table 6 maps assessments to parameters along with representative quotes from the transcripts of the data review sessions. In the quotes, I refer to the sessions with endocrinologists as Endo-Patient ID-Session ID and to the sessions with the certified diabetes educators as CDE-Patient ID-Session ID.

Assessment of CGM data sufficiency

For clinicians to make care decisions from the data, there should be sufficient blood sugar readings to understand patient's management, identify potentially problematic episodes and their solutions. If the data is insufficient, clinicians are unable to assess glycemic performance and the overall management. For patients using a CGM, data sufficiency is assessed by the time for which the patient wears the CGM and the number of times the patient calibrates the CGM using a conventional glucometer. Measures for assessing data sufficiency are displayed in the “Sensor Usage” section in Figure 19. For patients that do not use a CGM but use an insulin pump, data sufficiency is assessed through the self-reported data from the insulin pump. Figure 20 shows a display of insulin pump data with the summary statistics at the bottom that provide details on how many BG values per day have been logged by the patient.

Assessment of BG outcomes

This involved assessment of the patient’s overall glycemic performance using the distribution of BG values to identify the proportion of readings that are in range and those that are not. Additionally, statistical measures such as the average, standard deviation, range, and variability of BG numbers over 14 days are used. Lastly, patterns or isolated atypical events of BG readings were discussed for the larger part of the review session. Discussion of patterns could also involve comparing patterns across different time periods or looking at meal-time patterns of BG readings. When inspecting the data day by day, providers also draw attention to single events of unusually high or low BG numbers and ask patients for reasons behind those. Outcome assessment aims to identify opportunities for intensifying management. Figure 19 shows the primary display used for assessing outcomes. The display is rendered from 14 days of CGM data. The range graph at the bottom is used to identify patterns in different time windows and the statistics at the top provide the distribution of BG numbers, the average, and the standard deviation. For patients who do not use a CGM, the logbook view shown in Figure 20 is used. The statistics section at the bottom of the logbook view provides the distribution of BG values, the 14-day average, the range, and the standard deviation. The logbook view is also used to identify patterns in BG values and unusual high or low numbers in the absence of CGM displays.



[redacted]

Overview Report

Generated at: Wed, Jul 11, 2018 9:55 PM EDT

14 days | Thu Jun 28, 2018 - Wed Jul 11, 2018

208 mg/dL Average glucose (CGM)

96 mg/dL Standard deviation (CGM)

N/A Hypoglycemia risk

Time in range

Days with CGM data 79% 11 / 14

Avg. calibrations per day 2.7

Sensor usage

Top Patterns

① [redacted] had a pattern of daytime highs

had a pattern of significant highs between 6:30 PM and 12:05 AM.

② [redacted]s best glucose day was July 1, 2018

's glucose data was in the target range about 65% of the day.

This graph shows your data averaged over 14 days

Clinician's Notes

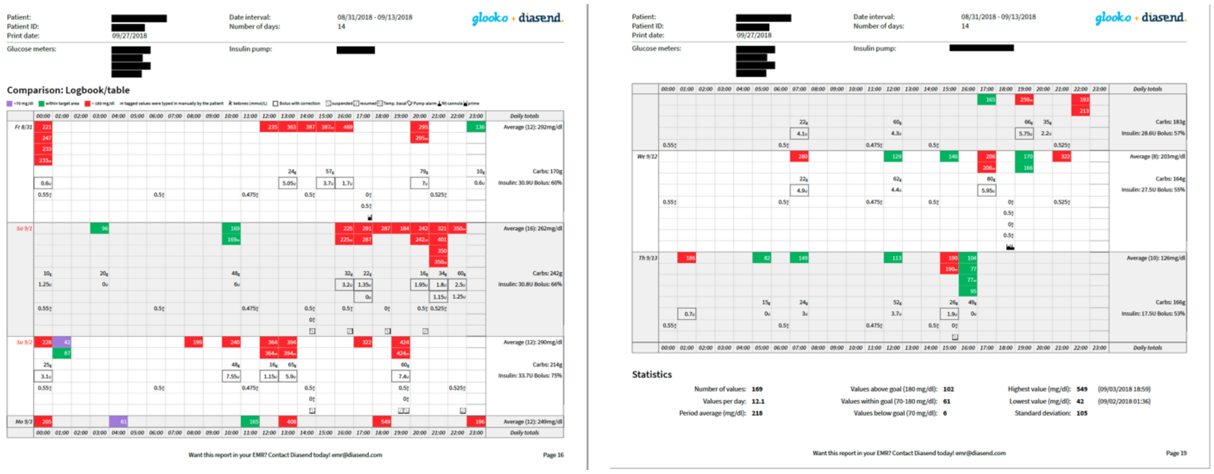
Figure 19: Dexcom Clarity's overview report showing BG statistics and the pattern chart from 14 days of CGM data.

Assessment of behaviors

Different types of patient behaviors and system behaviors were assessed to understand their impact on glycemic performance. These include assessment of manual or system intervention behavior, manual data input behavior, bolus insulin administration behavior, and food intake behavior.

Manual or system intervention behavior: These include checking for the frequency and timing of correction insulin intake, uncovered carbohydrates (carbs taken without insulin) intake and basal changes done by the patient or by the system. (e.g., an automated insulin delivery system). Insulin pumps deliver long-acting insulin, that is the basal insulin, based on a pre-configured schedule. Patients or a system, such as the artificial pancreas, can make temporary changes to the scheduled delivery or suspend the delivery. Providers seek information on different types of interventions to understand why they were done and their effect on the BG numbers. The logbook view shown in Figure 20 is used to identify these interventions. Interventions related to the basal insulin are denoted through various symbols listed at the top of the logbook, such as suspending basal insulin, resuming basal insulin, and temporary basal. Carbohydrate and correction insulin intake can be identified by inspecting the rows for carbs and short acting insulin (or bolus insulin).

Manual data input behavior: For patients to get an appropriate amount of insulin from the pump, they need to manually record BG numbers in the pump, except for users of closed-loop systems where the CGM can share BG values with the pump. Providers expect a minimum of 4 BG numbers during the day - one for each meal and one at bedtime. For 3 meals, providers expect patients to log 3 carbohydrate amounts. The logbook in Figure 20 is used to assess how often the patient is recording or missing to record data. For example, the first row in Figure 20b shows missing BG records where carb and insulin values have been logged.



a

b

Figure 20: Glooko Diasend's Logbook view - a) the first page of the logbook with legends at the top to depict various interventions, b) the last page of the logbook with summary statistics in the bottom panel.

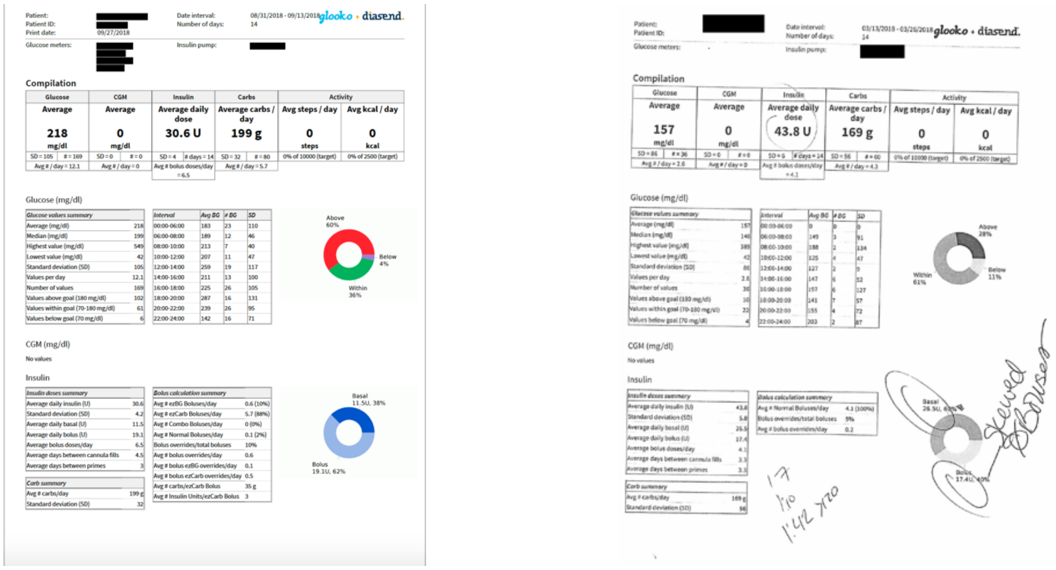
Bolus administration behavior - Providers check for total bolus insulin taken for a day and the missed boluses (meal without a corresponding insulin amount). Where possible, providers also distinguish between correction bolus (bolus taken for a high BG number) and a meal bolus (bolus taken for a meal, which includes insulin for food and could include correction insulin for a high BG number). Bolus insulin data is displayed in the logbook as one of the rows (Figure 20). The distribution of bolus insulin into different types (e.g., food bolus, correction bolus) is available in the "bolus calculation summary" section of the display shown in Figure 21.

Food intake behavior - Providers check for total carbs being consumed in a day to identify any unusual variations in carb intake over several days, such as multiple snacks instead of meals, frequent uncovered snacks, and unrealistic guesstimates of carb amounts. Data on total daily carb intake is present in the last column of the logbook display (Figure 20). Additionally, providers look for missing meals, that is days with less than three carb reports. Carb intake data is displayed in the logbook (Figure 20) as one of the rows.

Assessment of Regimen Quality

To ensure that the patient is getting the appropriate amount of insulin, providers check for regimen balance and regimen sufficiency to understand the overall regimen quality. Regimen balance refers to the ratio of short acting insulin (bolus) and long-acting insulin (basal). Information to assess regimen balance is displayed through a donut chart shown in Figure 21, annotated by the diabetes educator in one of the sessions (Figure 21b). Regimen sufficiency refers to the sufficiency of the carb ratio (the amount of carb per 1 unit of insulin or the amount of insulin per gram of carb), the correction factor (the amount of insulin taken per unit increase in BG number beyond the target), and the basal insulin. To understand the sufficiency of the carb ratio, providers use the logbook view (Figure 20) to look for BG numbers a few hours after a meal. To understand the sufficiency of the correction factor, providers look for BG numbers a few hours after correction insulin is taken, again using the logbook view (Figure 20). The following quote from one of the experts in the focus group summarizes the task of assessing regimen quality.

> "*And a couple of things that I always like to try to do is one, find a glucose that's at target (in range) when they ate carbs and see, if they didn't eat any additional carbs for the next 3 to 4 hours, were they in range or not. So looking at that top value there, with a 147, 3 hours later, they were 290 after giving coverage for just carbs, not correction—maybe a little bit for correction, but not much. But then the next day, it seemed to be okay, um, you know, the day below that. So it seems, it kind of... and then also it's really helpful if you can see, like for example, you can't see my cursor, but on Saturday 2/4, at 14:00, they were 305, did not eat any carbs... And they came down appropriately to 128. And so I try to find few examples like that to assess the carb ratio and correction factor, respectively.*" (P2-FocusGroup2)



Patient: [redacted]
Patient ID: [redacted]
Print date: 09/27/2018
Date interval: 08/31/2018 - 09/13/2018
Number of days: 14
Glucose meters: [redacted]
Insulin pump: [redacted]

Compilation

Glucose	CGM	Insulin	Carbs	Activity	
Average	Average	Average daily dose	Average carbs / day	Avg steps / day	Avg kcal / day
218 mg/dl	0 mg/dl	30.6 U	199 g	0 steps	0 kcal
SD = 105 # = 169	SD = 0 # = 0	SD = 4 # days = 14	SD = 32 # = 80	0% of 10000 (target)	0% of 2500 (target)
Avg # / day = 12.1	Avg # / day = 0	Avg # bolus doses/day = 6.5	Avg # / day = 5.7		

Glucose (mg/dl)

Glucose values summary	
Average (mg/dl)	218
Median (mg/dl)	199
Highest value (mg/dl)	549
Lowest value (mg/dl)	42
Standard deviation (SD)	105
Values per day	12.1
Number of values	169
Values above goal (180 mg/dl)	102
Values within goal (70-180 mg/dl)	61
Values below goal (70 mg/dl)	6

Interval	Avg BG	# BG	SD
00:00-06:00	183	23	110
06:00-08:00	189	12	46
08:00-10:00	213	7	40
10:00-12:00	207	11	47
12:00-14:00	259	19	117
14:00-16:00	211	13	100
16:00-18:00	225	26	105
18:00-20:00	287	16	131
20:00-22:00	239	26	95
22:00-24:00	142	16	71

CGM (mg/dl)

No values

Insulin

Insulin doses summary	
Average daily insulin (U)	30.6
Standard deviation (SD)	4.2
Average daily basal (U)	11.5
Average daily bolus (U)	19.1
Average bolus doses/day	6.5
Average days between cannula fills	4.5
Average days between primes	3

Carb summary	
Avg # carbs/day	199 g
Standard deviation (SD)	32

Bolus calculation summary	
Avg # ezBG Boluses/day	0.6 (10%)
Avg # ezCarb Boluses/day	5.7 (88%)
Avg # Combo Boluses/day	0 (0%)
Avg # Normal Boluses/day	0.1 (2%)
Bolus overrides/total boluses	10%
Avg # bolus overrides/day	0.6
Avg # bolus ezBG overrides/day	0.1
Avg # bolus ezCarb overrides/day	0.5
Avg # carbs/ezCarb Bolus	35 g
Avg # Insulin Units/ezCarb Bolus	3

a

Patient: [redacted]
Patient ID:
Date interval: 03/13/2018 - 03/26/2018
Number of days: 14
Glucose meters: -
Insulin pump: [redacted]

Compilation

Glucose	CGM	Insulin	Carbs	Activity	
Average	Average	Average daily dose	Average carbs / day	Avg steps / day	Avg kcal / day
157 mg/dl	0 mg/dl	43.8 U	169 g	0 steps	0 kcal
SD = 86 # = 36	SD = 0 # = 0	SD = 6 # days = 14	SD = 56 # = 60	0% of 10000 (target)	0% of 2500 (target)
Avg # / day = 2.6	Avg # / day = 0	Avg # bolus doses/day = 4.1	Avg # / day = 4.3		

Glucose (mg/dl)

Glucose values summary	
Average (mg/dl)	157
Median (mg/dl)	140
Highest value (mg/dl)	389
Lowest value (mg/dl)	42
Standard deviation (SD)	86
Values per day	2.6
Number of values	36
Values above goal (180 mg/dl)	10
Values within goal (70-180 mg/dl)	22
Values below goal (70 mg/dl)	4

Interval	Avg BG	# BG	SD
00:00-06:00	0	0	0
06:00-08:00	149	3	91
08:00-10:00	188	2	134
10:00-12:00	125	4	47
12:00-14:00	127	2	9
14:00-16:00	147	6	52
16:00-18:00	157	6	127
18:00-20:00	141	7	57
20:00-22:00	155	4	72
22:00-24:00	203	2	87

CGM (mg/dl)

No values

Insulin

Insulin doses summary	
Average daily insulin (U)	43.8
Standard deviation (SD)	5.8
Average daily basal (U)	26.5
Average daily bolus (U)	17.4
Average bolus doses/day	4.1
Average days between cannula fills	3.3
Average days between primes	3.1

Carb summary	
Avg # carbs/day	169 g
Standard deviation (SD)	56

Bolus calculation summary	
Avg # Normal Boluses/day	4.1 (100%)
Bolus overrides/total boluses	5%
Avg # bolus overrides/day	0.2

17
1:10
1:42 x20

Skewed Boluses

b

Figure 21: a) The Compilation report from Glooko Diasend to show averages for different types of data, and summary and distribution of BG readings and insulin intake, b) A printed reported annotated by the diabetes educator when assessing regimen quality.

S. No.	Data-Driven Assessment	Parameters Used	Quotes from Data Review Sessions
1	Assess CGM data sufficiency	Sensor wear time: the duration for which patient wears the CGM	"You're wearing the Dexcom seven days a week?" Endo-P02-01
		Calibrations: number of times the patient calibrates the CGM using a conventional glucometer	"It says that you're calibrating 4.5 times per day. Make sure you don't calibrate that thing too much because you're going to throw it off in wonky numbers." CDE-P01-01
2	Assess BG outcomes	Distribution of BG values to show proportion of readings in range, proportion of hypoglycemic readings, and proportion of hyperglycemic readings	"Her average went from 194 to 176. And then, um, her standard deviation went from 85 to 73 which is pretty awesome. And the time spent in range went from 36 to 45%" CDE-P04-03

		Average and standard deviation of BG readings over 14 days	"There's a little bit of a difference in her numbers from last week compared to this week." CDE-P04-05
			"The average was a little bit higher than normal." CDE-P02-02
		Variability and range of BG readings in different time windows over 14 days	"A little bit more variability in the evening. Um, after like 3pm she kind of has this like slow steady rise up. Not all the time" CDE-P06-02
		Patterns in BG readings during different time windows over 14 days	"Even from the beginning of June until the end of June she's pretty consistently going up with breakfast and then she goes, she crashes down, uh, pretty hard" CDE-P04-01
			"So the morning blood sugars some of them are looking high, but most of them are looking pretty good in range." CDE-P07-01
		Comparison of BG patterns between 2 time periods	"Her average is coming down I think about thirty points in the last two weeks compared to the two weeks prior." CDE-P02-10
		Unusual out of range BG reading	"She was 55, she treated her low and then she came up to 108, and then all of a sudden she's 539." CDE-P03-02
3	Assess manual or system intervention behavior	Frequency and timing of correction bolus intake	"Corrections are still the biggest struggle, um, I feel like we have to correct all the time". Endo-P03-02
		Frequency and timing of uncovered carbs intake	"I think you're eating [uncovered carbs] way too much in the middle of the night and its setting up your morning bad because you're waking up at 300 sometimes." CDE-P07-07
		Frequency and timing of interventions related to basal	"You are suspended about 22% of the time overnight". CDE-P06-05

		insulin (suspending or resuming basal insulin, making a temporary change to basal insulin)	“Why don’t we go over changes real quick so you guys don’t have to run those 120% temp basals all the time. " CDE-P02-04
			“You’re having 5 suspensions a day. you’re actually suspending more during the day than you are at night." CDE-P03-10
			"We’re using a lot of temp basals to keep things down" Endo-P06-02 parent
4	Assess manual data input behavior	BG input behavior: 4 BG checks a day with no more than 4-6 hours between checks except for overnight	"Most days you’re hitting that 3-4 times a day blood sugar check which is really, really good. Um, somedays are you know better and stronger than others.” CDE-P07-01
			"I still feel like I don’t know what her overall trend is though. Right because it’s not enough data." CDE-P03-10
			"The blood sugar input though needs to improve. So there wasn’t four times a day blood sugars entered in everyday." CDE-P02-02
		Carb input behavior	"She’s not putting any carbs in”. CDE-P03-01
			“I think we need more, uh numbers, more carbs, um, definitely need more data during the day.” CDE-P01-04
5	Assess Bolus administration behavior	Total bolus taken for each day	“There’s lots of missed meal boluses and the way you can tell is because she goes from 100 to 400." CDE-P03-02
			"How many boluses are you giving a day typically?" Endo-P01-02. Patient: it ranges three or four.

		Types of boluses being taken - correction bolus, food bolus, correction and food bolus	"You're missing lots of carb boluses so, um, you need to make sure you're putting your food in there because there's many, many times where you're not." CDE-P03-06
			"I'm trying to figure out if that's basal or correction [bolus]". CDE-P01-05
			"So just keep working at putting those numbers in, because that's one of the reasons why you're running high, is because you're not getting a correction to bring you down." CDE-P02-09
		Frequency and timing of missed bolus	"Those days that you're like three-, four hundred by one o'clock you know just because you missed those checks and bolus in the morning" CDE-P01-03
7	Assess food intake behavior	Total carbs taken for each day	"You were only bolusing like 60, 30, and 90 grams of carbs the last three days so maybe missing out on a couple boluses". CDE-P01-07
			"Are you really only eating 119 carbs or do you think you're missing out on some boluses?" CDE-P01-05
			"There's not very good carb counts in there. Even when she is eating, everything is either 30, 60, or 70 grams of carbs." CDE-P03-02
		Frequency and timing of missed carbs	"I think there's something, carbs or something that's missing out. I don't know. Because, like, there's no breakfast boluses." CDE-P03-02
8		The distribution of different types of insulin: the ratio of	"Her insulin balance is fine because she is on 23 of basal and 50 of bolus on average." CDE-P06-01

	Assess regimen quality	bolus to basal in a balanced regimen should be 60:40	"At her age you want about 40% of her insulin being from basal. And about 60 of that from bolus". CDE-P02-09
		Sufficiency of correction factor (correction factor is the amount of insulin to be administered for every unit increase in blood sugar above the target blood sugar number or the amount of insulin to be deducted for every unit decrease in blood sugar below the target number)	"You don't have a ton of examples where you're just giving a correction and you come down. But, it seems like when you're high you kind of stay high so if we give you a little bit more for your correction it may actually help." CDE-P01-03
		Sufficiency of insulin to carb ratio (insulin to carb ratio is the amount of insulin taken for every 10 grams of carbohydrate)	"The consistent pattern that I noted when I looked at this was the fact that, that midday high, I mean, I think she needs more food coverage [insulin to carb ratio] for lunch and supper." Endo-P04-01

Table 6: Different types of data-driven assessments, parameters used for those assessments, and representative quotes from the data.

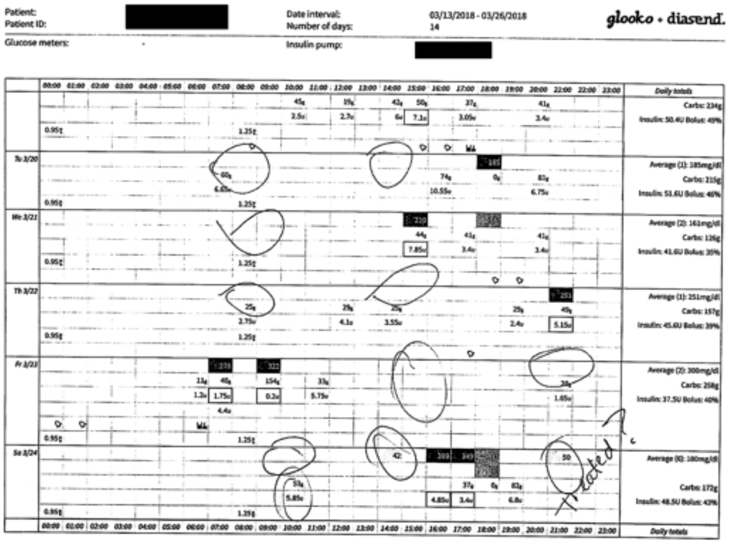
6.5.2 Reporting Platforms' Characteristics That Limit Effective Use

Here, I point out specific platform characteristics that may limit effective use of data because they do not align well with the assessment tasks described above.

Limited information on patient behaviors

While trend displays are available for BG numbers (e.g., Figure 19), different types of patient behaviors of interest, such as food intake behavior, insulin administration behavior, and data recording behavior are not visualized for patterns but only used for augmenting BG values. For example, Table 7 shows that there is no independent display for carbohydrate consumption events to provide an understanding of when a patient might have skipped a meal, had an unusually high carbohydrate meal, and what are the usual mealtimes for a patient. Such information is crucial for understanding patterns in patient's eating behavior, their consequence on BG numbers, and the corresponding insulin pump settings to align with these behaviors to keep the BG numbers in control.

In the data review sessions, I observed clinicians using the Logbook view of the insulin pump data to obtain such behavioral information by looking at each row of the logbook one by one. For example, Figure 22 shows markings on a printed logbook display explicitly added by a provider to reveal a pattern of missing BG data in the pump. In this case, the patient has missed to record multiple BG values in the insulin pump when they recorded carbs or insulin and the markings help surface the frequency and regularity with which that happened. Similarly, there is limited visual support for surfacing patterns in insulin administration behavior. While there are displays showing the overall insulin distribution between bolus and basal insulin and total insulin consumption each day that helps ascertain regimen balance (Figure 21), information on behaviors such as frequency and timing of missed bolus insulin or frequency of overcorrection with insulin, is not directly surfaced in these reports. Such information must be visually inferred using the Logbook report, requiring time and cognitive effort.



Patient:
Patient ID: [redacted]
Date interval: 03/13/2018 - 03/26/2018
Number of days: 14
Glucose meters: -
Insulin pump: [redacted]

Figure 22: Annotated Logbook of Glooko Diasend to identify the frequency and pattern of missing BG data to assess manual data input behavior and carb intake for a low BG number.

Platform	Chart Tile	Data		
		BG	Carbs	Insulin
Tandem	BG Trends - Hourly	✓		
	BG Trends - Daily	✓		
	Therapy Timeline	✓	✓	✓
	Summary of therapy	✓	✓	✓
	CGM Hourly	✓		
	Logbook	✓	✓	✓
	Activity Summary	✓		✓
LibreView-1	Snapshot	✓		✓
	Daily log	✓		✓
	Mealtime Patterns	✓	✓	✓
	Modal Day	✓		✓
	Weekly summary	✓		✓
LibreView-2	Snapshot	✓	✓	✓
	Daily log	✓	✓	✓
	Daily Patterns	✓	✓	✓
	Glucose Pattern insights	✓		
	Mealtime Patterns	✓	✓	✓
	Monthly Summary	✓		
	Weekly summary	✓	✓	✓
	Pattern chart	✓		

Dexcom Clarity	Explanation of Top patterns	✓		
	Overlay	✓		
	Daily view	✓		
	Daily stats report	✓		
	Hourly stats report	✓		
	Ambulatory glucose profile	✓		
Glooko Diasend	Glucose Standard Day	✓		
	Glucose Trend	✓		
	CGM Standard Day	✓		
	CGM Statistics	✓		
	Insulin Week			✓
	Insulin Bolus Doses			✓
	Comparison Logbook	✓	✓	✓
	Comparison Day by Day overview	✓	✓	✓
	Insulin Bolus Adherence	✓		✓
	Insulin Trend			✓
	Insulin Day by Day			✓
	Compilation	✓	✓	✓
	Comparison day by Day	✓	✓	✓
Tidepool	The basics	✓	✓	✓

	Daily views	✓	✓	✓
Medtronic	sensor and meter overview	✓	✓	✓
	logbook	✓	✓	✓
	daily detail	✓	✓	✓
	adherence	✓		✓
	therapy management dashboard	✓	✓	✓
	meal bolus wizard report	✓	✓	✓
	weekly review report	✓	✓	✓
	assessment and progress report	✓	✓	✓

Table 7: Different platforms with their constituent displays, and the type of data within displays.

Multiple disconnected and inconsistent visualizations

There are three high-level dimensions along which these data reports vary within and across platforms. These include formats of data displays (Table 8), categorization of BG numbers (Table 9), and representation of time (Table 10). Variations in the format, the categorization of BG values, and the use of time result in multiple disconnected and inconsistent displays. With multiple visualizations, lack of consistency is one of the top contributors of usability issues [61]. Next, I describe these variations and inconsistencies.

Diverse formats: The reporting platforms present each type of data using a diverse pool of displays, with the presentation of BG numbers being the most visually diverse. Data on BG levels have been presented using a large variety of formats (Table 8) - single or multiple line charts, scatterplots, area chart, calendar chart, simple table, color coded table (logbook), box plot/whisker plot, ribbon graph, range graph, pie chart, donut chart, stacked single column chart, and horizontal bar chart. Only a few of these formats are used during the data review to meet the different information needs (range graph, logbook) related to the assessment of BG outcomes, understandably so given the time constraints of a clinic visit. Providers pointed out four types of displays they may look at when reviewing the data. These include the logbook, the overview or the dashboard page, the daily views, and the pump settings page, "*if you are prioritizing, it would be the logbook (*Figure 20*), the overview page (*Figure 19*), the daily views, and the pump settings.*" (Moderator summarizing group response in Focus Group 1)

Type of Data	**Parameters Displayed**	**Format of Display**
BG	Individual data points	line chart (single), line chart (multiple color coded lines), line chart with marks overlayed on the lines, color coded table (logbook), scatter plot, calendar chart, area chart
	Averages and median	line chart (single), whisker plot with marks overlayed, table, scatter plot, calendar chart
	Standard deviation, interquartile range, maximum and minimum, variability, quartiles (10, 90, 25, 75)	ribbon/stream graph, table, range graph, box plot/whisker plot
	Number of BG readings	table, calendar chart
	Distribution of BG numbers	pie chart, single stacked bar chart, horizontal bar chart, donut chart
	Number of hypoglycemic events	table
	Likelihood of a low, variability below median, median compared to goal	Icons to show high, medium, and low
Carbs	Individual data points	labels on a horizontal timeline, labels on top of a bar, scatter plot of meal icons (time by value)
	Average	table
	Total	table

	Meal events/Mealtime windows	vertical lines on a BG line chart, marks overlayed on BG line chart, labels on a horizontal timeline
	Skipped meal (based on user configured mealtime windows)	zigzag line
	Number of carb entries	table, labels
Basal insulin	Scheduled basal delivery	area chart, solid line chart overlayed on a calendar background, table, labels on a timeline
	Temp basal	broken line
	Interventions to schedule - basal suspension and temp basal event	calendar with days of basal suspension marked by a solid circle, labels on area or line chart, table
	Average	table
	Number of basal doses	table
Bolus insulin	Individual values	bars with labels, table, line chart
	Bolus distribution	pie chart, table
	Average	table
	Total	table
	Number of bolus doses (total or average)	table, A calendar with each cell containing dots representing a bolus, labels
	Bolus events	vertical line on a BG line chart, table
	Active insulin time	curve following a bar representing bolus insulin
Basal+Bolus	Insulin distribution	pie chart, table, stacked bar chart, donut chart, table
	Insulin trend	line chart overlay
	Insulin pump settings	table, Color coded bands on a horizontal timeline

Table 8: Parameters displayed from different types of data and the display formats used.

Diverse representations of blood sugar severity: Reports varied in how they categorized BG values for presentation, giving a different sense of severity with changing categorizations within and across platforms. Across platforms, the most common encoding categorized BG numbers in 3 bins – high, low, in range. Across multiple displays within a platform, visualizations have been designed using more than one encoding. For example, LibreView used 4 encodings and Tidepool used 3 as shown in Table 9.

Platform	**BG Severity Encoding**
Tandem	High, Low, Target, Above Target, Below Target
	Target, Above Target, Below target
Libre View	High, 180-250, Target, Low
	High, Low
	High, Low, >350, Average
	High, Low, Post Meal Peak
Dexcom	High, Low, In range
	Very low, Low, In range, High, Very high
	Urgent low, Low, In range, High
Diasend	High, Low, In range
Tidepool	>250, 180-250, 70-180, 54-70, <54
	High, Low, In range
	70-180, Below 54
Medtronic	40-50, 50-70, 70-180, 180-240, 240-400
	High, Low, In range

Table 9: BG severity encoding for different platforms.

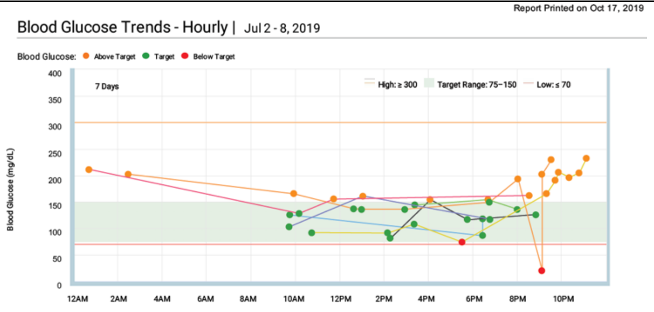
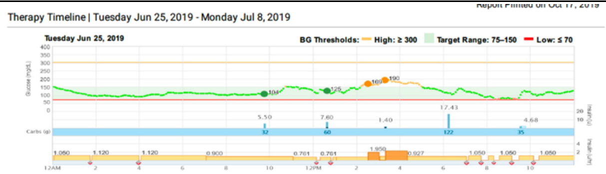
Diverse time scales: Displays also used different time scales to present data within and across the same platform. For each platform, Table 10 lists the different durations for which data was presented in the reports I analyzed and the visual representation of time that was used. Table 11 provides a more detailed example of the Tandem platform showing the displays corresponding to the various time representations from Table 10. The most common time scale used in the displays across platforms shows multiple days of data from midnight to midnight over 1-, 2-, or 3-hour intervals. This representation is referred to as the standard day. Within the majority of the platforms, the duration of the data displayed, and the representation of time varied across displays, as shown in Table 10. In our analysis, we used data reports that visualized data from 14 days, the standard time period typically reviewed by the clinicians.

Platform	**Duration of data**	**Representation in visualization**
Tandem	one week of data	midnight to midnight over 2-hour intervals
	one day each over one or multiple weeks	midnight to midnight over 2-hour intervals
	one or multiple weeks	days of the week
	14 days of data	midnight to midnight over 1- or 6-hour intervals
	1 week of data	1 average day
LibreView	one month of data	one day each
	one day each over 14 days	midnight to midnight over 1-hour intervals
	one day each over a week	midnight to midnight over 2-hour intervals
	14 days of data	midnight to midnight over 2-hour intervals
	14 days of data	6am to 6am over 6-hour intervals to capture meals
	14 days of data	event-based time windows from 3 am to 3 am
	14 days of data	one day each
Dexcom	1, 7, and 14 days of data	midnight to midnight over 1-hour intervals
	one or multiple days from 14 days of data	midnight to midnight over 1-hour intervals
Glooko Diasend	14 days of data	midnight to midnight over 1- and 3- hour intervals
	14 days of data	midnight to 6 am as one time window and 6 am to midnight over 2-hour windows
	1 day of data	midnight to midnight over 1-hour intervals
	14 days of data	one day each
Tidepool	multiple days of data	one day each
	1 day of data	midnight to midnight over 3-hour intervals

Medtronic	7 days of data	midnight to midnight over 1-hour intervals
	one day each over 7 days	midnight to midnight over 3-hour intervals
	14 days of data	6 am to 6 am over 4-time windows
	14 days of data	midnight to midnight over 1-hour intervals
	14 days of data	1 day each
	one day each over 14 days	midnight to midnight over 1-hour intervals

Table 10: Time scales used for presenting data

These variations result in multiple inconsistent views, requiring the user to reinterpret the BG encoding and the time differently for different displays. One of the providers from the focus group described the extra time involved in orienting to different formats, "*When you're seeing a bunch of patients back-to-back with diabetes you don't want to see information in different formats because then you have to orient to the format. I know that it takes maybe a second or two, but we are jealous of every second.*" (P1-FocusGroup1)

Duration of data	**Representation in visualization**	**Visualization**
one week of data	midnight to midnight over 2-hour intervals	
one day each over one or multiple weeks	midnight to midnight over 2-hour intervals	

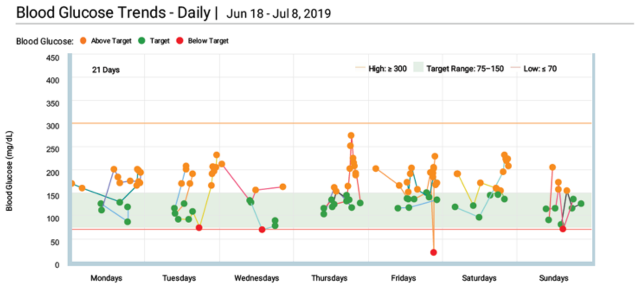
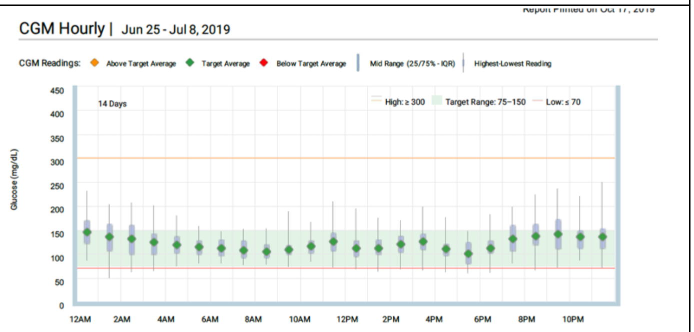
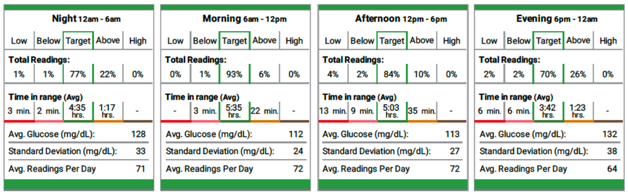
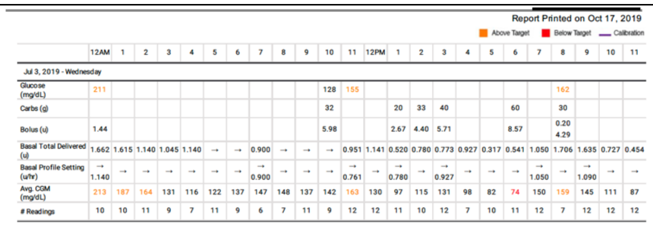
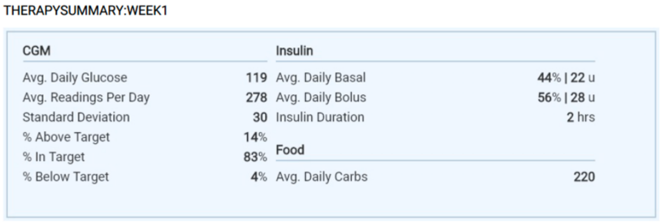
One or multiple weeks	days of the week	
14 days of data	midnight to midnight over 1-hour intervals	
	midnight to midnight over 6-hour intervals	
	midnight to midnight over 1-hour intervals	
1 week of data	1 average day	

Table 11: Different time representations in displays used by Tandem.

Night 12am - 6am

Low	Below	Target	Above	High
Total Readings:				
1%	1%	77%	22%	0%
Time in range (Avg)				
3 min.	2 min.	4:35 hrs.	1:17 hrs.	-

Avg. Glucose (mg/dL):	128
Standard Deviation (mg/dL):	33
Avg. Readings Per Day	71

Morning 6am - 12pm

Low	Below	Target	Above	High
Total Readings:				
0%	1%	93%	6%	0%
Time in range (Avg)				
-	3 min.	5:35 hrs.	22 min.	-

Avg. Glucose (mg/dL):	112
Standard Deviation (mg/dL):	24
Avg. Readings Per Day	72

Afternoon 12pm - 6pm

Low	Below	Target	Above	High
Total Readings:				
4%	2%	84%	10%	0%
Time in range (Avg)				
13 min.	9 min.	5:03 hrs.	35 min.	-

Avg. Glucose (mg/dL):	113
Standard Deviation (mg/dL):	27
Avg. Readings Per Day	72

Evening 6pm - 12am

Low	Below	Target	Above	High
Total Readings:				
2%	2%	70%	26%	0%
Time in range (Avg)				
6 min.	6 min.	3:42 hrs.	1:23 hrs.	-

Avg. Glucose (mg/dL):	132
Standard Deviation (mg/dL):	38
Avg. Readings Per Day	64

Report Printed on Oct 17, 2019

🟧 Above Target 🟥 Below Target — Calibration

	12AM	1	2	3	4	5	6	7	8	9	10	11	12PM	1	2	3	4	5	6	7	8	9	10	11
Jul 3, 2019 - Wednesday																								
Glucose (mg/dL)	211										128	155									162			
Carbs (g)											32			20	33	40			60		30			
Bolus (u)	1.44										5.98			2.67	4.40	5.71			8.57		0.20 4.29			
Basal Total Delivered (u)	1.662	1.615	1.140	1.045	1.140	→	→	0.900	→	→	→	0.951	1.141	0.520	0.780	0.773	0.927	0.317	0.541	1.050	1.706	1.635	0.727	0.454
Basal Profile Setting (u/hr)	→ 1.140	→	→	→	→	→	→	→ 0.900	→	→	→	→ 0.761	→	→ 0.780	→	→ 0.927	→	→	→	→ 1.050	→	→ 1.090	→	→
Avg. CGM (mg/dL)	213	187	164	131	116	122	137	147	148	137	142	163	130	97	115	131	98	82	74	150	159	145	111	87
# Readings	10	10	11	9	7	11	9	6	7	11	9	12	12	11	10	12	7	10	11	12	7	12	12	12

THERAPYSUMMARY:WEEK1

CGM		Insulin	
Avg. Daily Glucose	119	Avg. Daily Basal	44% | 22 u
Avg. Readings Per Day	278	Avg. Daily Bolus	56% | 28 u
Standard Deviation	30	Insulin Duration	2 hrs
% Above Target	14%		
% In Target	83%	**Food**	
% Below Target	4%	Avg. Daily Carbs	220

6.6 DISCUSSION

In this chapter, I describe the assessments performed using patient data, the data parameters that these assessments rely on, and the characteristics of available platforms that could potentially limit or challenge the use of data – multiple disconnected and inconsistent displays, and limited support for identifying behavioral insights. These limitations have implications for data review and technology design that I discuss in this section. Specifically, they point to the need to make data review more efficient and effective by designing platforms with the consideration of the constraints (e.g., short clinic visits) under which these displays are reviewed. Specific data visualization guidelines and techniques can lower the effort and time involved in visual data analysis by better guiding user attention and by better aligning with user needs. Particularly, the design of multiple views and visualizations need to ensure that the visualizations are consistent, the relationships between them or their connectedness are self-evident, and they help focus user attention for efficiency [175]. Next, I provide suggestions for improving the design of these reporting platforms.

6.6.1 Providing Quick Access to Relevant Information

The diversity of displays available in each of the reviewed platforms can challenge data review in two ways. First, in the absence of appropriate markers of relevance or priority, there is an additional burden of first selecting relevant displays, that is identifying what to look at before being able to identify specific insights. For example, which BG patterns among the hourly, daily, and weekly should be explored and/or addressed? Should a “high” BG number and an “above target” BG number be treated differently? Second, the availability of varying types of displays could also compromise the data interpretation standards that these platforms attempt to enable as different clinicians might attend to different displays and could come up with different care decisions using the different views on the same data [175]. While the flexibility offered by multiple views could add value for clinicians, the resulting complexity may also discourage them from using patient data [12] and may introduce the possibility of excluding important information contained in one or multiple displays. Patients as the non-expert users of data may be easily overwhelmed by the information overload these reports create, resulting in underutilization of data [182].

These issues point to the need to make data review simpler by enabling easier access to the information required. One of the ways to enable quick access to information in the displays is to provide natural language summaries briefly describing the potential insights in the displays (using data to text technology [136]). This would alleviate the need for visually scanning and processing of the displays putting lesser demand on user attention. Natural language generation approaches have been used to augment the interpretation of visual displays in settings such as the ICU

[67,128], automated generation of interventional messages for a mental health community [83], and summarizing electronic health records [140]. Similar approaches could be applied here to summarize each display for enhanced readability and indexing of the different displays by the user. Among the existing tools, only Dexcom Clarity's overview report shown in Figure 19 provides a text summary of patterns visible in the trend chart.

6.6.2 Linking Information Across Displays

As described in the findings, data review involved connecting multiple types of information from different displays. In order to be able to successfully link a BG pattern from the CGM data with causal behaviors of carb and insulin intake from the insulin pump data, providers make use of different displays with one display showing the patterns (Figure 19), and the other display showing multiple data streams for individual days (Figure 20). These displays usually varied in the formats they use, in how they represent time and in how they encode BG numbers. Connecting information from these two platforms required finding the pattern shown in one display, locating it in the other display and then focusing on a specific area of the latter display to extract causal events. The reporting platforms do not make relationships (e.g., outcome in one display, causes in another display) between views apparent, requiring cognitive processing that could be avoided with a better design. There is no indication of how these representations are connected or relevant to each other (e.g., connection by cause-effect relationship, event-consequence relationship, temporal relationship) inducing sensemaking burden to understand when and how to connect information.

These issues point to the need to enable smoother transition between displays by making them visually consistent and by linking them in ways that help maintain orientation towards specific information of interest and help reduce the cognitive cost of switching between displays. Here, guidelines and techniques documented by information visualization researchers on making multiple views consistent could offer guidance [143,175]. Specifically, highlighting techniques [107,162] can be used to surface related sets of information across displays or provide a visual linking across displays to reveal their connectedness. One specific technique, called focus plus context [40,145], involves focusing on a particular information element to draw user attention and reducing the visibility of surrounding information. Such a technique can be used to draw attention to patterns in the logbook. For example, the logbook view requires users to visually align and aggregate red (out of range) and green (in range) cells to obtain a sense of BG patterns in the insulin pump data. This visual processing could be simplified by employing techniques [107] to highlight a specific region of the display where a pattern exists. Highlighting could involve using a different color for that region, or a zoomed in view of that region. Understanding what to highlight needs to be informed by T1D community standards for data interpretation [9,12,46] and empirical understanding of practices developed through research [189,190].

6.6.3 Surfacing Behavioral Insights

The review of the data reporting platforms revealed that the primary focus of these reports is blood glucose data with multiple dedicated displays. However, understanding and assessing patient behaviors from data on carb and insulin recorded in the insulin pump forms one of the major assessment tasks in these reviews. With limited support to understand patient behaviors, clinicians have to perform several low-level tasks (sub-setting, aligning, filtering, and counting) [16] without any visual aid. To improve the performance of these tasks, the visualizations need to include insights on patient behaviors, such as when is a patient skipping insulin or meals, and when is a patient overcorrecting with insulin or carbs. Future research should explore ways in which behaviors can be visualized to facilitate analysis of outcomes or risks. For example, computational behavior modeling techniques [7] can be used to identify behavioral rhythms, which can then be visualized to show behaviors of interest.

6.7 CONCLUSION

In this chapter, I review the commercially available displays in relation to the assessment tasks performed when reviewing diabetes data. I point to the limitations of the commercially available data reporting platforms. These platforms create an information overload with limited guidance for users to navigate that information. There is a need for these platforms to support quicker access to relevant information, to link information across displays for sensemaking, and to surface behavioral insights upfront that are relevant for contextualizing the BG outcomes. There is an opportunity for exploring data visualization techniques to design consistent and connected displays and improve the ways in which the information needs of the users could be met. Towards this opportunity, I discuss guidelines established by information visualization researchers to address the above-identified limitations of existing tools.

As noted in chapter 3 on related work, it is not only important to understand types of information required to make decisions from the data, but also the analytical activities and workflows involved in the use of data to better conceptualize tools to support how the data is or could be used. In this direction, the next chapter describes the nuanced analytical activities involved in making sense of the data from the analysis of the data review sessions and provides a descriptive model of data practices.

Chapter 7 Episode-Driven Sensemaking: A Framework for Navigating Multidimensional Patient-Generated Data

In chapter 4, I note that a shared and central objective for patients and clinicians in sensemaking with patient-generated data is problem-solving. Sensemaking and problem-solving are primarily data-driven and may not only happen in real time but through periodic retrospective analysis of self-monitored data. The existing frameworks of sensemaking and problem-solving provide little understanding of the data analysis activities involved, which makes them prescriptive of the ideal high-level behaviors rather than descriptive of the data practices. Moreover, they do not account for collaborative data interpretation practices.

In response to the limitations noted above, in this chapter, I describe the results from a deeper investigation of data analysis practices of patients and clinicians to interpret multi-dimensional data. Through this investigation, I synthesize a framework called episode-driven sensemaking to describe the different analytical activities involved in making decisions from multidimensional data. Specifically, I found that patients and clinicians use episodes (periods of suboptimal management indicated by outcomes or patient behaviors) to maintain analytic focus and perform four sensemaking tasks in the context of episodes - episode detection, episode elaboration, episode classification, and episode-specific recommendation generation. I also point to the challenges encountered in these steps. Lastly, I provide design guidelines for tools to support episode-driven sensemaking with personal health data. For this study, I led the data collection, data analysis, and paper writing.

7.1 SUMMARY

Background: To effectively manage chronic health conditions, patient and clinicians need to frequently engage with patient-generated data. However, existing informatics tools provide limited support for leveraging data to make care decisions. Limited understanding of how patient-generated data is translated to actionable information restricts the development of effective tools. There is a need to understand the practices of reasoning and decision-making with data to inform the design of better tools.

Objectives: To articulate design directions for informatics tools to support reasoning and decision-making with multiple streams of patient-generated data informed by user practices and challenges with the existing tools. I investigate data-driven decision-making in the context of chronic disease management, particularly focusing on the use of data to inform the management of Type 1 diabetes.

Methods: I conducted an observational qualitative study of a telemedicine program that involved biweekly data review sessions (n=71) between Type 1 diabetes patients, their caregivers, and their clinicians to assess and inform diabetes management using patient-generated data. I used qualitative methods to analyze video transcripts of the sessions to identify different analytical tasks performed in making decisions from data, the sequence of these tasks, and the challenges associated with them.

Results: To assess and inform chronic condition management, the analysis of multiple data streams involved observation of episodes of suboptimal management from time series data, and a careful assessment of these episodes using different types of data identify episode-specific interventions. The episodes of suboptimal management served two important roles. First, they enabled identification of interventions for improvement. Second, they facilitated analytic focus to guide the analysis of multidimensional data by helping invoke relevant decision-making heuristics and directing attention to relevant data streams. I show that episode-driven analysis, interpretation, and decision-making with multidimensional patient-generated data involved four stages – 1) episode detection, 2) episode elaboration, 3) episode classification, and 4) episode-specific recommendation generation. I present opportunities to augment human analysis, judgment, and decision-making with informatics tools in different stages of episode-driven data analysis by – 1) detecting episodes of suboptimal outcomes and behaviors and enabling user feedback on computationally detected episodes, 2) enabling episode-driven data filtering and presentation for episode elaboration, 3) quantifying and visualizing episode characteristics for episode classification and for shared understanding between patients and clinicians, and 4) presenting episode-specific recommendations with representative evidence to explain the recommendations for intelligibility.

7.2 MOTIVATION

The continuous availability of patient-generated data (PGD) from sensing devices has created new opportunities for realizing the vision of data-driven health. PGD has the potential to bridge the gap between clinic visits, enable deeper insight into a patient's condition, and facilitate communication and shared decision-making between patients and clinicians [41,131]. However, before the vision for data-driven health can be realized, significant challenges need to be understood and addressed to enable the efficient use of PGD within the constraints that patients and clinicians have [5]. For system design, these challenges include providing appropriate data displays and reports, and supporting decision-making [5]. Current tools for PGD use provide minimal decision-support [50], which makes it challenging for patients and clinicians to make decisions from the data [12]. The lack of suitable tools could be attributed to the limited understanding of the role and impact of PGD in data-driven decision-making by patients and clinicians [100]. Hence, recent research has indicated the need to better understand how PGD is translated into actionable information within both individual and collaborative contexts of use [5,166]. This study investigates how PGD is collaboratively reviewed on an ongoing basis to arrive at one or more care decisions.

7.3 METHODS

7.3.1 Objective and Study Setting

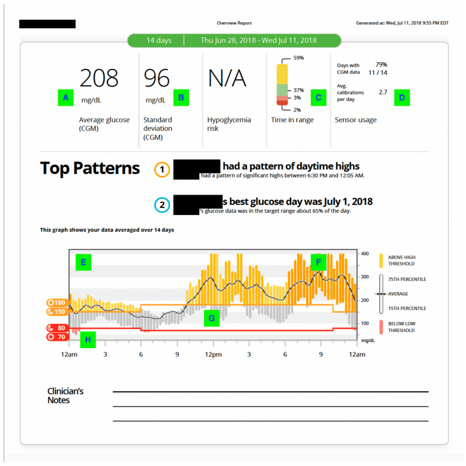
The aim of this study was to understand how patients and clinicians together review patients' data to make care decisions. For this purpose, I observed video recordings of data review sessions between adolescent patients, caregivers, and their clinicians. The data review sessions were held as a part of a telemedicine care program run by the pediatric endocrinology clinic of Michigan Medicine. Telemedicine care involved two types of data review sessions - biweekly sessions with the patient's caregiver guided by a certified diabetes educator (CDE) and quarterly meetings with an endocrinologist. As a part of the program, caregivers uploaded two weeks of their patient's device data (CGM, insulin pump, glucometers) before a scheduled session. During their appointment, they met the provider over a video call to review the data together. The aim of these sessions was to use the patient's data to identify opportunities for making changes to the care regimen or patient behaviors and improve control. Table 12 shows the differences between the traditional care provided at the clinic and the modified care delivered via telemedicine.

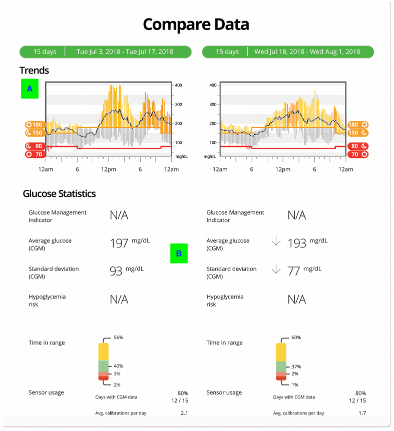
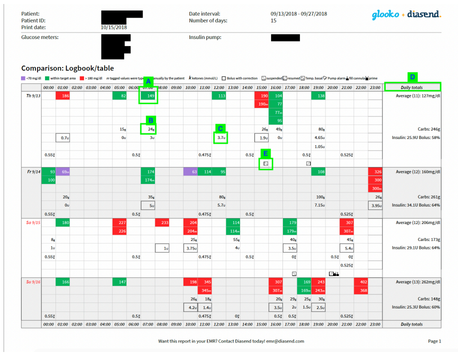
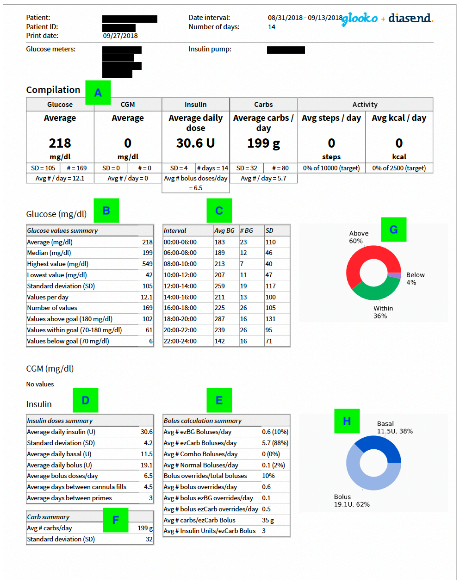
Traditional Care	**Telemedicine Care**
Clinic visit every 3 months	Tele-visit with Endocrinologist every 3 months
Diabetes educators review data on demand by patients – patients share numbers, and they talk with the educator over a phone call	Diabetes educators review data every 2 weeks at a scheduled time – patients upload numbers prior to the scheduled session, and they talk with the educator over a video call
Blood test at the hospital on the visit day to determine A1C	Blood test to determine A1C at a convenient location prior to the scheduled quarterly session

Table 12: Difference between the traditional care model and the telemedicine care model.

7.3.2 Tools Used

Table 13 shows the primary displays that were used for data review. These included the trend reports from continuous glucose monitor data, the logbook view of the insulin pump data, and summary dashboards of data from both devices.

S. No.	**Report**	**Image**	**Description**
1	Dexcom Clarity summary and trend report	 [redacted] Overview Report Generated at: Wed, Jul 11, 2018 9:55 PM EDT 14 days | Thu Jun 28, 2018 - Wed Jul 11, 2018 Top Patterns ① [redacted] had a pattern of daytime highs had a pattern of significant highs between 6:30 PM and 12:05 AM. ② [redacted]s best glucose day was July 1, 2018 's glucose data was in the target range about 65% of the day. This graph shows your data averaged over 14 days Clinician's Notes 	An overview of the CGM data. The top section displays global indicators of management that include average BG (A), standard deviation (B), time in range (C), and CGM wear time (D). The bottom section shows a range graph of CGM readings from 14 days aggregated by time of day (E). Each vertical bar in this graph displays the 15%-75% range of BG values in a 15-minute window. Yellow bars indicate BG numbers that are out of range (F), grey bars indicate BG numbers within rage (G), and red points indicate patterns of low BG numbers (H). This graph represents a standard day and shows the diurnal variations in BG levels based on data from 14 days.

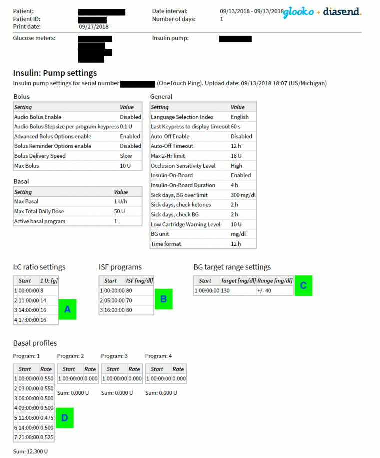
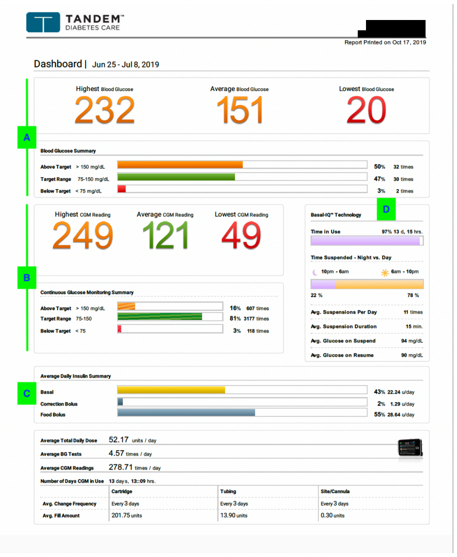
<table>
<tr>
<td>2</td>
<td>Dexcom Clarity trend comparison display</td>
<td>
Compare Data
15 days | Tue Jul 3, 2018 - Tue Jul 17, 2018
15 days | Wed Jul 18, 2018 - Wed Aug 1, 2018
Trends
A
Glucose Statistics
B
Glucose Management Indicator N/A | Glucose Management Indicator N/A
Average glucose (CGM) 197 mg/dL | Average glucose (CGM) ↓ 193 mg/dL
Standard deviation (CGM) 93 mg/dL | Standard deviation (CGM) ↓ 77 mg/dL
Hypoglycemia risk N/A | Hypoglycemia risk N/A
Time in range 56% 40% 3% 2% | Time in range 60% 37% 2% 1%
Sensor usage: Days with CGM data 80% 12 / 15; Avg. calibrations per day 2.1 | Sensor usage: Days with CGM data 80% 12 / 15; Avg. calibrations per day 1.7
 </td>
<td>A comaprison of CGM data from two different weeks using range graphs (A), and statistics (B). Statistics include glucose management indicator, average glucose, standard deviation, time in range, and CGM usage data.</td>
</tr>
<tr>
<td>3</td>
<td>Glooko Diasend Logbook</td>
<td>
Patient: [redacted]
Patient ID:
Print date: 10/15/2018
Date interval: 09/13/2018 - 09/27/2018
Number of days: 15
glooko + diasend.
Glucose meters: [redacted]
Insulin pump: [redacted]
Comparison: Logbook/table
🟪 <70 mg/dl 🟩 within target area 🟥 > 180 mg/dl m tagged values were typed manually by the patient k ketones (mmol/L) ☐ Bolus with correction suspended resumed Temp. basal Pump alarm fill cannula prime
A
B
C
D
E
Daily totals
Want this report in your EMR? Contact Diasend today! emr@diasend.com
Page 1</td>
<td>A “Logbook” view of insulin pump data. Each row represents a single day, and each column represents an hour within the day running from “00:00” to “23:00”. Individual cells show when and what types of data were recorded – BG (A), carbs (B), and insulin (C). On the extreme right, daily averages and totals are displayed for BG, carb, and insulin (D). Other types of information, such as insulin pump events, are also displayed (E).</td>
</tr>
<tr>
<td>4</td>
<td>Glooko Diasend Compilation report</td>
<td>
Patient: [redacted]
Patient ID: [redacted]
Print date: 09/27/2018
Date interval: 08/31/2018 - 09/13/2018
Number of days: 14
glooko + diasend.
Glucose meters: [redacted]
Insulin pump: [redacted]
Compilation A
<table>
<tr><th>Glucose</th><th>CGM</th><th>Insulin</th><th>Carbs</th><th colspan="2">Activity</th></tr>
<tr><td>Average
218
mg/dl</td><td>Average
0
mg/dl</td><td>Average daily dose
30.6 U</td><td>Average carbs / day
199 g</td><td>Avg steps / day
0
steps</td><td>Avg kcal / day
0
kcal</td></tr>
<tr><td>SD = 105 # = 169</td><td>SD = 0 # = 0</td><td>SD = 4 # days = 14</td><td>SD = 32 # = 80</td><td>0% of 10000 (target)</td><td>0% of 2500 (target)</td></tr>
<tr><td>Avg # / day = 12.1</td><td>Avg # / day = 0</td><td>Avg # bolus doses/day = 6.5</td><td>Avg # / day = 5.7</td><td></td><td></td></tr>
</table>
Glucose (mg/dl) B C G
<table>
<tr><th>Glucose values summary</th><th></th></tr>
<tr><td>Average (mg/dl)</td><td>218</td></tr>
<tr><td>Median (mg/dl)</td><td>199</td></tr>
<tr><td>Highest value (mg/dl)</td><td>549</td></tr>
<tr><td>Lowest value (mg/dl)</td><td>42</td></tr>
<tr><td>Standard deviation (SD)</td><td>105</td></tr>
<tr><td>Values per day</td><td>12.1</td></tr>
<tr><td>Number of values</td><td>169</td></tr>
<tr><td>Values above goal (180 mg/dl)</td><td>102</td></tr>
<tr><td>Values within goal (70-180 mg/dl)</td><td>61</td></tr>
<tr><td>Values below goal (70 mg/dl)</td><td>6</td></tr>
</table>
<table>
<tr><th>Interval</th><th>Avg BG</th><th># BG</th><th>SD</th></tr>
<tr><td>00:00-06:00</td><td>183</td><td>23</td><td>110</td></tr>
<tr><td>06:00-08:00</td><td>189</td><td>12</td><td>46</td></tr>
<tr><td>08:00-10:00</td><td>213</td><td>7</td><td>40</td></tr>
<tr><td>10:00-12:00</td><td>207</td><td>11</td><td>47</td></tr>
<tr><td>12:00-14:00</td><td>259</td><td>19</td><td>117</td></tr>
<tr><td>14:00-16:00</td><td>211</td><td>13</td><td>100</td></tr>
<tr><td>16:00-18:00</td><td>225</td><td>26</td><td>105</td></tr>
<tr><td>18:00-20:00</td><td>287</td><td>16</td><td>131</td></tr>
<tr><td>20:00-22:00</td><td>239</td><td>26</td><td>95</td></tr>
<tr><td>22:00-24:00</td><td>142</td><td>16</td><td>71</td></tr>
</table>

CGM (mg/dl)
No values
Insulin D E H
<table>
<tr><th>Insulin doses summary</th><th></th></tr>
<tr><td>Average daily insulin (U)</td><td>30.6</td></tr>
<tr><td>Standard deviation (SD)</td><td>4.2</td></tr>
<tr><td>Average daily basal (U)</td><td>11.5</td></tr>
<tr><td>Average daily bolus (U)</td><td>19.1</td></tr>
<tr><td>Average bolus doses/day</td><td>6.5</td></tr>
<tr><td>Average days between cannula fills</td><td>4.5</td></tr>
<tr><td>Average days between primes</td><td>3</td></tr>
</table>
<table>
<tr><th>Bolus calculation summary</th><th></th></tr>
<tr><td>Avg # ezBG Boluses/day</td><td>0.6 (10%)</td></tr>
<tr><td>Avg # ezCarb Boluses/day</td><td>5.7 (88%)</td></tr>
<tr><td>Avg # Combo Boluses/day</td><td>0 (0%)</td></tr>
<tr><td>Avg # Normal Boluses/day</td><td>0.1 (2%)</td></tr>
<tr><td>Bolus overrides/total boluses</td><td>10%</td></tr>
<tr><td>Avg # bolus overrides/day</td><td>0.6</td></tr>
<tr><td>Avg # bolus ezBG overrides/day</td><td>0.1</td></tr>
<tr><td>Avg # bolus ezCarb overrides/day</td><td>0.5</td></tr>
<tr><td>Avg # carbs/ezCarb Bolus</td><td>35 g</td></tr>
<tr><td>Avg # Insulin Units/ezCarb Bolus</td><td>3</td></tr>
</table>

<table>
<tr><th>Carb summary F</th><th></th></tr>
<tr><td>Avg # carbs/day</td><td>199 g</td></tr>
<tr><td>Standard deviation (SD)</td><td>32</td></tr>
</table>
</td>
<td>An overview of insulin pump data. The top section displays the average glucose, carbs, and insulin over 14 days (A).

The tables show summary statistics calculated on BG data from the CGM and from the insulin pump. CGM data is summarized for – 1) 14 days to show average, median, highest, and lowest values, etc. (B) and 2) different time windows in 14 days to show average, standard deviation, and number of BG values (C).

The insulin pump data is summarized over 14 days to show – 1) average daily insulin, standard deviation, average daily bolus and</td>
</tr>
</table>

			basal, etc. (D), 2) bolus calculation summary – number of boluses per day, etc. (E), and 3) carb summary – average carbs per day, and standard deviation (F). The donut charts show time in range (G), and a distribution of basal and bolus insulin (H) taken over 14 days.
5	Glooko Diasend Device Settings		A screen showing insulin pump settings, including the insulin to carb ratio (I:C) (A), the correction factor or insulin sensitivity factor (ISF) (B), the BG target range (C), and basal rates for different time periods of the day (D).
6	Tandem Dashboard		A dashboard showing data from both the insulin pump and the CGM. The top two sections (A and B) show summaries of BG numbers from each device including time in range, highest and lowest BG, and the average. Another section shows the distribution of insulin between basal insulin, correction insulin, and food insulin (C). A section on the right shows device behavior – time in use for basal IQ (closed loop) technology, time in suspend mode, average numbers of suspensions, etc. (D).

Patient: [redacted]
Patient ID: [redacted]
Print date: 09/27/2018
Date interval: 09/13/2018 - 09/13/2018
Number of days: 1
Glucose meters: [redacted]
Insulin pump: [redacted]

Insulin: Pump settings

Insulin pump settings for serial number [redacted] (OneTouch Ping). Upload date: 09/13/2018 18:07 (US/Michigan)

Bolus

Setting	*Value*
Audio Bolus Enable	Disabled
Audio Bolus Stepsize per program keypress	0.1 U
Advanced Bolus Options enable	Enabled
Bolus Reminder Options enable	Disabled
Bolus Delivery Speed	Slow
Max Bolus	10 U

Basal

Setting	*Value*
Max Basal	1 U/h
Max Total Daily Dose	50 U
Active basal program	1

General

Setting	*Value*
Language Selection Index	English
Last Keypress to display timeout	60 s
Auto-Off Enable	Disabled
Auto-Off Timeout	12 h
Max 2-Hr limit	18 U
Occlusion Sensitivity Level	High
Insulin-On-Board	Enabled
Insulin-On-Board Duration	4 h
Sick days, BG over limit	300 mg/dl
Sick days, check ketones	2 h
Sick days, check BG	2 h
Low Cartridge Warning Level	10 U
BG unit	mg/dl
Time format	12 h

I:C ratio settings

	Start	*1 U: [g]*
1	00:00:00	8
2	11:00:00	14
3	14:00:00	16
4	17:00:00	16

ISF programs

	Start	*ISF [mg/dl]*
1	00:00:00	80
2	05:00:00	70
3	16:00:00	80

BG target range settings

	Start	*Target [mg/dl]*	*Range [mg/dl]*
1	00:00:00	130	+/- 40

Basal profiles

Program: 1

	Start	*Rate*
1	00:00:00	0.550
2	03:00:00	0.550
3	06:00:00	0.500
4	09:00:00	0.500
5	11:00:00	0.475
6	14:00:00	0.500
7	21:00:00	0.525

Sum: 12.300 U

Program: 2

	Start	*Rate*
1	00:00:00	0.000

Sum: 0.000 U

Program: 3

	Start	*Rate*
1	00:00:00	0.000

Sum: 0.000 U

Program: 4

	Start	*Rate*
1	00:00:00	0.000

Sum: 0.000 U

[redacted]

Report Printed on Oct 17, 2019

Dashboard | Jun 25 - Jul 8, 2019

Highest Blood Glucose 232
Average Blood Glucose 151
Lowest Blood Glucose 20

Blood Glucose Summary

Above Target	> 150 mg/dL	50%	32 times
Target Range	75-150 mg/dL	47%	30 times
Below Target	< 75 mg/dL	3%	2 times

Highest CGM Reading 249
Average CGM Reading 121
Lowest CGM Reading 49

Continuous Glucose Monitoring Summary

Above Target	> 150 mg/dL	16%	607 times
Target Range	75-150	81%	3177 times
Below Target	< 75	3%	118 times

Basal-IQ™ Technology

Time in Use 97% 13 d, 15 hrs.

Time Suspended - Night vs. Day

10pm - 6am: 22 %
6am - 10pm: 78 %

Avg. Suspensions Per Day	11 times
Avg. Suspension Duration	15 min.
Avg. Glucose on Suspend	94 mg/dL
Avg. Glucose on Resume	90 mg/dL

Average Daily Insulin Summary

Basal	43%	22.24 u/day
Correction Bolus	2%	1.29 u/day
Food Bolus	55%	28.64 u/day

Average Total Daily Dose 52.17 units / day
Average BG Tests 4.57 times / day
Average CGM Readings 278.71 times / day
Number of Days CGM in Use 13 days, 13::09 hrs.

	Cartridge	Tubing	Site/Cannula
Avg. Change Frequency	Every 3 days	Every 3 days	Every 3 days
Avg. Fill Amount	201.75 units	13.90 units	0.30 units

Table 13: Different data reports used during data reviews - 1) Dexcom Clarity summary and trend report, 2) Dexcom trend comparison display, 3) Glooko Diasend Logbook, 4) Glooko Diasend Compilation report, 5) Glooko Diasend Device Settings, and 6) Tandem Dashboard.

7.3.3 Recruitment and Data Collection

Patients and their caregivers were recruited by clinic research staff and informed assent and consent was obtained from all patients and caregivers. Seven patients signed up to receive care via the telemedicine program, but one patient (P05) dropped out before the first data review session. One CDE consulted all the patients in the study. Two endocrinologists (Endos) consulted their respective patients for the duration of the study. Table 14 shows the details of all the patient participants, including the age, gender, the devices used by them, the participating caregivers, and their endocrinologist. In this paper, I refer to the meetings with the CDE as "CDE visit" and the ones with the endocrinologists as "Endo visit". A CDE visit involved only data review whereas an Endo visit followed the protocol of a standard clinic visit, such as discussing blood work, urine test results, and growth measures in addition to data review. The average length of a CDE visit was 11 minutes and that of an Endo visit was 38 minutes. One CDE visit for P03 did not take place. One Endo visit for P06 was not recorded, for which I used the endocrinologist's notes for analysis. In total, I analyzed 71 interactions that involved data review.

Patient ID	Age, Gender	Devices Used	Caregiver	Endocrinologist
P01	16, F	CGM, Pump, Glucometer	Mother	Endo 1
P02	13, F	CGM, Pump	Mother	Endo 1
P03	12, F	CGM, Pump, Glucometer	Father	Endo 1
P04	9, F	CGM, Pump, Glucometer	Father	Endo 2
P06	16, F	CGM, Pump, Glucometer	Mother	Endo 1
P07	17, M	CGM, Glucometer	Mother, Father	Endo 1

Table 14: Study participants

7.3.4 Data Analysis

The video recordings of all data review sessions were transcribed for analysis. Transcripts were analyzed in two rounds using a mix of descriptive, in-vivo and process coding [129]. In the first round of coding, each video recording transcript was coded to identify the topics discussed, the questions asked by the patients, their caregivers, and the providers, the information exchanged, the problems identified, the decisions made, and the challenges of making decisions. The coded data was then grouped into themes to identify the types of problems addressed, the types of decisions made, and the types of challenges faced. In the second round of coding, I coded the transcripts with words describing the analytical tasks being performed during the data review sessions. This led to the identification of different analytical steps in the data review process and the sequence in which they were performed. Coding was performed iteratively until I reached saturation where different steps started to appear repeatedly within and across data review sessions and no new steps were identified.

7.4 FINDINGS

7.4.1 An Overview of Retrospective Data Review

My analysis found that the review of multidimensional Type 1 diabetes (T1D) data including BG, carbs, and insulin to assess self-management involved identifying episodes of suboptimal behaviors and outcomes from the data and assessing these episodes to identify their cause, the necessity and feasibility of an intervention, and the appropriate intervention as indicated by the data. The following vignette, representative of most of the data review sessions that I observed, provides an illustrative overview of a typical data review session.

> P03 CDE visit 7
> The CDE noted that P03 had been having high BG numbers after lunch. She (P03) had also been skipping a few lunches. In the corresponding data, there were some days with no carb intake numbers at lunch (i.e., only BG and insulin implying that either the carb numbers were not entered in the pump or that the lunch was skipped altogether) and there were other days when carb numbers were recorded for lunch.
>
> Given this type of data, the CDE speculated that post lunch high BG numbers could either be because of missing carb entries (resulting in missing corresponding bolus insulin) at lunch or because of insufficient correction insulin (the only insulin taken on days when the lunch carbs were not recorded in the pump).
>
> To figure out which explanation was more reasonable, the CDE inspected data from different days – days with carb (including bolus and correction insulin) and without carb (including only correction insulin). The CDE observed that on both types of days (with and without carb) P03's BG numbers spiked after lunch time. CDE considered the days when carbs were recorded to be more concerning because even after recording carbs for lunch and taking both

insulin components, correction and bolus insulin (which was the ideal behavior), P03's post lunch BG numbers were out of range.

"*you're still going quite high after lunch. And some days you bolus, some days you don't. And maybe you're not eating and maybe that's your correction, but even when you do bolus, like on Sunday you bolused and still went sky high afterwards. Same thing with last Friday, last Thursday.*" (CDE)

This meant that the insulin treatment configured for lunch was insufficient. The CDE decided to increase insulin for lunch by changing the corresponding insulin to carb ratio for that time window.

7.4.2 Identifying Episodes of Suboptimal Outcomes and Behaviors

In the sessions I observed, data review typically began with noting outcome or behavior related events from a particular data stream (e.g., CGM or insulin pump records) to identify episodes of suboptimal management. These episodes portrayed disease-related or management-related experiences of patients that could be potentially problematic and that could be prevented. They were initially indicated by events from a single data stream – either data on outcomes (e.g., BG) or data on patient or system behaviors (e.g., food intake, insulin administration, BG monitoring).

Each episode was primarily represented by a collection of *events* (i.e., time-stamped data points) and depicted a recurring suboptimal outcome (e.g., a high BG event after lunch on several days, a low BG event at bedtime on weekends) or suboptimal behavior (e.g., missed insulin dose at lunch for several days). Episodes could also be represented by singular *events* that indicated a one-time occurrence of an unusual outcome or behavior. Different types of episodes observed included – unusual one-time out of range BG numbers, unusual one-time self-management or device behavior, a pattern of out-of-range BG numbers, and a pattern of suboptimal self-management or device behaviors.

"*The consistent pattern that I noted when I looked at this was … that midday high."* (Endo2 in P04's Endo visit 1).

> *"You're missing lots of carb boluses so, um, you need to make sure you're putting your food in [the pump] because there's many, many times where you're not."* (CDE in P03's CDE visit 6).

> *"She was 55 [BG], she treated her low and then she came up to 108, and then all of a sudden she's 539."* (CDE in P03's CDE visit 2)

Temporally, the events constituting the episodes were observed within a timeframe of two weeks (the standard time window used in clinical practice and supported by data download and analysis tools). Within the 14-day duration, different episodes could be marked by different time periods and granularity. Such time periods could be based around routines (e.g., breakfast, bedtime), a single 24-hour period, certain days of the week (e.g., weekends or Thursdays), or a very specific time range (3 am to 4 am). Observations of episodes were made possible by different time representations in the data reports. For example, a "standard day" (see Table 13 #1) representation supports looking at specific time slices (e.g., "mornings") aggregated over 14 days but not individual days. The "logbook" (see Table 13 #3) view supports looking at individual days as well as specific time slices over 14 days, albeit with more effort required. Episodes of suboptimal BG outcomes were more easily observed because many data reports color-code BG data based on deviation from the target range as defined by standard interpretation criteria. For instance, Dexcom Clarity's range graph (see Table 13 #1) displays BG data from the CGM using three colors – yellow for above range, grey for in range, and red for below range. Thus, yellow spikes stand out as a high BG pattern, with darker yellows indicating a more severe BG pattern.

The observation of episodes of suboptimal behaviors related to food intake, insulin intake, and data recording was not as straightforward as it was for BG outcomes because the relevant events were not highlighted in the data reports. They were often visually or logically inferred by the clinicians from the available data displays and then confirmed, refuted, or detailed in conversation with the patients and caregivers. For example, the pump logbook (see Table 13 #3) shows events of blood sugar check, carb entry, and insulin delivery in a grid that uses days and hours of the day as dimensions. However, it does not enable easy temporal pattern detection in these events.

When identifying episodes from the data, patients and clinicians characterized them to understand their severity or strength and eventually established whether they merit concern. One commonly noted characteristic was the frequency with which events constituting an episode took place,

> Caregiver: "*She's going a lot higher after breakfast.*"
>
> CDE: "*Like only a couple, like three days maybe out of the last six, seven.*" (P06 CDE visit 3)

> "*I have the last 14 days here and it's a handful of times [of low numbers] and its usually between 3 and 4 am"* (CDE in P07's CDE visit 7).

In addition to frequency, overall indicators of glycemic performance associated with the larger 14-day timeframe in which an episode was nested were also considered (e.g., averages, proportion of values in range),

> "*The night average is right around 150 [i.e., in range] but the day average is, is probably pushing 220-240.*" (Endo 2 in P04's Endo visit 2).

The observed episodes played an important role in guiding data review for further assessment. In what follows, I describe the steps involved in assessing these episodes of suboptimal management.

7.4.3 Assessing Episodes of Suboptimal Outcomes and Behaviors

Identified episodes and the events constituting them were carefully assessed to understand the necessity and feasibility of a care action. Not all episodes observed in the reports were considered worthy of further investigation or a corrective action.

Episodes of suboptimal BG outcomes were assessed by characterizing their severity, by examining the global indicators of management from the data (time in range, average BG, BG variability), and by understanding the potential causes of these episodes to identify the need and possibility for an intervention.

Assessing behavior-related episodes involved understanding the frequency of suboptimal behaviors, the potential or actual consequences of the behaviors, the global indicators of management as described above (even though management might not be optimal, if the overall performance is good enough, an intervention may not be needed), and the elicitation of context-appropriate details from the patient, such as data that might be relevant but was not tracked (e.g., physical activity).

Next, I describe in detail the process of assessing these episodes, which involved building narratives to explain the observed episodes, classifying the episodes based on the need and the feasibility of an intervention, and lastly identifying an intervention for episodes that were concerning and that could be prevented in the future.

Building narratives to explain the observed episodes

After episode identification, clinicians, with the help of patients and caregivers, would build narratives to connect the identified episodes with domain knowledge, with evidence from the other data streams (e.g., connecting an episode of high BG numbers with a preceding episode of suboptimal self-care behaviors or connecting a behavior-based episode with a subsequent episode of problematic BG outcomes), and with additional information exchanged between patients and clinicians. Based on the characteristics of the episode (daytime versus overnight, pre or post meal, etc.), clinicians invoked heuristics that govern the relationships between outcomes, self-care behaviors, and the treatment or intervention needed. These heuristics offered an explanatory frame to make sense of a particular episode. They helped direct attention to relevant evidence from the different data streams and identify a potential intervention. For example, a steady rise in BG numbers could be attributed to insufficient basal insulin, "*after like 3pm she kind of has this like slow steady rise up. It actually might just be a little basal that you need to tweak*" (CDE in P06's CDE visit 2), while a sudden spike would usually be related to meals. Daytime BG spikes are typically associated with meals, the insulin taken for meals, and the patient's meal-related behavior. Sufficient insulin for meals as determined by the insulin to carb ratio should not result in spikes but insufficient insulin to carb ratio results in getting less insulin and a BG spike after the meal. Daytime peaks could also be attributed to behaviors, such as undercounting carbs in the meal, taking meal insulin after the meal instead of before it, and not entering BG number for a meal resulting in lesser insulin than is needed due to the lack of correction insulin component. Administration of fast-acting insulin overnight indicates the need for more basal insulin overnight, "*if you're giving corrections all night then we just need to give her more insulin.*" (Endo in P03's Endo visit 02).

Upon invocation of such heuristics, clinicians sought evidence from the available data and additional context from the patient and the caregiver to establish the potential causes of episodes– ruling out explanations that are less likely or less evident in the data and further considering the ones that are more likely or more evident in the data. Evidence was sought for and against an emerging narrative to connect outcomes, behaviors, and regimen. The following vignette from P04's first visit with the endocrinologist demonstrates how elaboration happened,

> The clinician observed midday high BG numbers and explained that daytime highs are often caused by insulin to carb ratio issues. He shared more evidence pointing towards the insufficiency of the daytime carb ratio by showing patterns of red numbers (high BG) during the day. He then found counterevidence in the form of a few green numbers (in range BG numbers) during lunch. He thought the lunch time green numbers might be because of school activity and discounted them as counterevidence. He decided that changes needed to be made to the carb ratio to provide more insulin for lunch and supper to avoid daytime highs. (P04's Endo visit 1)

Challenge: Insufficient outcome or behavior data

Not having enough data to explain observed episodes compromised clarity about causal associations which eventually resulted in not being able to make changes to the patient's care plan,

> *"I was looking at the data and talked to [the endocrinologist] and, uh, couple of the other [CDEs] at work and I was trying to figure out what we could do. But I think we need more, uh numbers, more carbs, um, definitely need more data during the day. Because there's not enough data to support any kind of changes or anything right now."* (CDE in P01's CDE visit 4)

In one case, P03 who was on metformin, a diabetes medication to enhance the body's insulin sensitivity, had a one-time high BG number of 600 that resulted in an emergency room visit. The CDE wanted to understand if the metformin dosage was working to keep P03's BG numbers in control. She could not determine that because P03 had not been putting her data in the pump. Despite a problematic episode, a solution could not be determined because the data was not sufficient.

> *"I wouldn't change any doses today. She's missing so much right now that I mean we're not going to be able to tell if that metformin is working or not because not all the data is getting in the pump."* (CDE in P03's CDE visit 09).

Challenge: Isolating factors associated with episodes

Another challenge was related to identifying the most likely cause of an observed episode when there seemed to be multiple potential causes. Cause identification relied on device data as well as contextual information provided by the patients. Such information (e.g., activity, stress, routine - mealtimes, bedtime, awake time, menstruation) was elicited from the patient by asking questions relevant to a particular episode. For example, for an episode of low BG numbers, providers would often ask the patient about an increase in activity or increase in the amount of insulin taken. Similarly, providers would ask about the patient's routine to decide what insulin dose should be set for different times of the day. While the majority of the BG-related suboptimal episodes were attributed to food or insulin (long-acting insulin and short-acting insulin), it did require clinicians to isolate the most influential of the two,

> CDE while trying to establish the potential cause of a high BG pattern in the afternoon*, "I almost wonder if you're missing a couple carbs [food] here and there or if that's your basal [long- acting insulin]."* (CDE in P01's CDE visit 08).

With other factors in addition to food and insulin, it became even more challenging to isolate the effect of these factors on a particular episode. For example, while exploring the cause and fix for a midday high BG pattern for P04, the endocrinologist noted that on school days, variable activity can make it hard to understand the impact of food and insulin on the BG trends,

> *"The midday [pattern] is more challenging because kids in the middle of the day at school are getting more activity a lot of the time, so then we need to look at the days on the weekend where, you know, maybe it's a little more average activity all day."* (Endo in P04's Endo visit 1)

Classifying episodes of suboptimal management

Classifying episodes involved understanding the extent to which an episode of suboptimal management is or can be problematic and can be or should be fixed. This in turn was determined by whether that episode could be explained in terms of the available evidence; and the certainty with which it could be explained. Connections between an episode of suboptimal management and different data streams could indicate issues that can be and need to be fixed, issues that cannot be entirely fixed even though they need to be fixed, issues that need to be monitored but need not be fixed, and non-issues that do not require further attention. In the following conversation, the CDE points to an episode of high BG numbers from the data that seems problematic but only needs to be monitored for recurrence because it is not severe, and it does not seem to have a likely cause evident in the data.

> The CDE notes a pattern of high BG numbers in the data and asks the caregiver about it in P06's CDE visit 4, "*Only [high BG] trend that I'm seeing is last week or so. It's been going on a little bit like after school like 3-4 o'clock*." The caregiver and the patient do not understand what might be causing the pattern. The CDE explains that it's not something that requires attention right now but is something to keep an eye on, "*I mean it's not even, when I say a little bit I'm like a little bit up to 160, 170. Like it's nothing that needs to your attention right now but. Anyways, just an area to keep your eye out on. Those carbs there.*" To quickly see if a potential cause could be determined, CDE checks about food,
>
> CDE: "*Have you been eating different snacks or anything?"*
>
> Patient: "*Nothing that would affect that time*."
>
> CDE: "*Okay well you're doing awesome. I wouldn't change anything*."

While the CDE thinks that the carb intake in a particular time window might be causing the high BG pattern, she is unable to determine that with certainty because the patient thinks differently. As a result, this episode of high BG numbers was not seen as requiring an intervention.

A closer look at the data to categorize these episodes was usually carried out when there was unclear or weak evidence, which could result in patients and clinicians not being able to build narratives or having different opinions about a particular episode being problematic. It required them to revisit the evidence and closely inspect it to come to a shared understanding. The following case further demonstrates how this happened,

> In P04's CDE visit 1, the caregivers were concerned about experiencing a higher frequency of low BG numbers. When the caregivers brought up their concern about the low numbers, together with the CDE they tried to elaborate the low BG number events. The caregiver mentioned that the insulin dose was overcorrecting P04 at times,
>
> Caregiver: "*To me it seems like it's overcorrecting you know just the insulin not the activity.*"
>
> CDE: "*So, you're talking about when you give her those corrections at 5pm that's when you're noticing that she's coming down too hard, too fast?*"
>
> Caregiver: "*I don't have any specific [data/evidence].. let me see if I can find a specific…"*
>
> Because the caregiver is unable to point to specific evidence for the likely cause of low BG numbers, the CDE further asked about the patient's activity, snacks, and meals to understand better.
>
> CDE: "*What did she do Monday? Was she pretty active in the afternoon?*"
>
> Caregiver: "*Yeah, she played out on the trampoline with the kids.*"
>
> CDE: "*Yeah. I'd probably cross that low [on Monday] out just because that's more than likely that's activity. . . My worry is that when you're giving that correction in the evening and she's crashing I think that you might be correcting a high number from her two o'clock snack. So that could tank her. So maybe keep an eye on that [the high BGs] and we won't change anything right now with that one*."

In the above case, the nature of the evidence required the caregivers and the CDE to engage in deeper analysis by going into the details of the potentially problematic events of low BG numbers and comparing different narratives based on activity or insulin. The CDE concluded that one of the low BG number events was because of a one-time activity event and it did not indicate an issue at that point. Because the potential cause of the other low BG events could not be determined, the low BG numbers were not seen as fixable right away and the CDE wanted the caregivers to monitor these events going forward instead of making a change. Additionally, the CDE pointed to an episode of recurring behavior (giving insulin to fix high BGs) that was likely the cause of the low BG number events but could not determine that with certainty yet.

Classifying the episodes of suboptimal management thus explicitly happened to determine the extent to which an episode required an intervention. This particularly happened when an episode of suboptimal management did not clearly seem to need an intervention or when there wasn't a clear explanation for it to be dealt with. These in turn could happen where there was weak or mixed evidence in the data, when the episodes did not seem severe enough and when patient and clinician differed in their opinions.

Challenge: Inconsistent evidence

As described above, providers invoked domain-specific heuristics to find evidence reflecting the typical relationships between outcomes, patient behaviors, and the regimen, which eventually helped identify the change in treatment needed. However, often there was mixed evidence in the data, such as when a potential cause seemed to create divergent effects. In such cases, patients and providers had difficulty explaining the episodes and identifying a corrective intervention.

The following conversation from P01's Endo visit 1 illustrates the challenge of interpreting inconsistent evidence,

> Endo: "*Yeah so, the 18th. Like I'm looking at like a, it's five minus [two], 3pm and then you have like, a 7pm right so there's like a good three or four-hour window in which you've got coverage for your carbs right?*"
> Patient: "*Yeah*"
> Endo: "*You theoretically should have got coverage for your correction, but then you were 293 like three or four hours later. Right, so what I see is a lot of like reds [high BGs] that are followed by reds right?*"
> Patient: "*Yeah*"
> Endo: "*Yeah so, um but then if you look at the 15th on Sunday? That seems to be a little better right you had 203 [high BG] go to 144 [in range BG] in the afternoon or like 372 go to 140.*"
> Patient: "*Yeah*"

> Caregiver: "*That was the day we were flying all day.*"
> Endo: "*It was okay. Okay. And then let's even look at Sunday the 22nd. So, at 8am you didn't really come down but then at 11am you did? Oh.*"

In another case, the clinician could not explain a patient's data as every day seemed to be different for the timeframe being examined. P04 had post-breakfast and lunch spikes that made the clinician question the carb ratio for that time window of the day. However, the spikes were inconsistent. As a result, the endocrinologist made a small change to the insulin dose and decided to reassess in another week, contrary to his expectation that more aggressive changes were needed.

> *"November 7, 8 and 9 her breakfast excursion [blood glucose trace after breakfast] was beautiful 3 days in a row, and then, starting on the 12th, the breakfast excursion has been high, and I can't explain it. The lunch excursion has been, on the 15th and 16th, or the 14th, 15th it was fine, the 13th and 12th it went high. I can't explain why some of the days she spikes and some of the days she doesn't, I just really can't explain it. Um. I definitely want to switch that last carb ratio to be more aggressive. The question is how much. I'd probably just go down by 1 [e.g., from taking 1 unit for every 11g of carbs to taking 1 unit for every 10g of carbs] and look at it again in another week."* (Endo2 in P04's Endo visit 2)

Although the clinician expected that more insulin is needed to fix the post meal spikes, given the inconsistency in the numbers, he only made a small change at first.

Identifying corrective steps to fix suboptimal management

After establishing whether an episode required an intervention, participants focused on providing recommendations in response to the different types of episodes, problematic or non-problematic, and fixable or non-fixable. For the episodes classified as being problematic and feasible to correct, a fix could involve one or more of the following—a one-time change, a context-based change, or an anticipated change needed in the future. Problematic episodes that could not be fixed resulted in instructions to gather more data to help understand it better. Additionally, decisions made in a particular review session to fix a specific problematic episode did not always result in resolving it. Consequently, similar episodes were observed in the data and discussed over multiple data review sessions in several patient cases.

A one-time change of dose was an immediate change to the settings of the insulin pump and could involve changing the insulin to carb ratio, the correction factor, the basal schedule, and/or the target BG number. In the following vignette, the CDE changes the correction factor and dinner carb ratio so that the patient will get more insulin,

> The CDE notes post-dinner high BG numbers and shows them to the patient (P01), focusing on examples of highs after 5 pm continuing overnight. P01 has a BG spike after eating dinner which continues overnight despite taking correction insulin. After seeing this data, the CDE thinks more insulin could be used for dinner and for corrections after dinner. CDE increases the insulin to carb ratio at dinner and the midnight correction factor for P01 to get more insulin at dinner and overnight. (P01's CDE visit 3)

A situational change of insulin dose or behavior involves taking an action under certain conditions, such as during physical activity, work, menstruation, or when certain BG thresholds are crossed.

> P02 had been noting high BG numbers during her periods. The CDE suggested using a temporary basal increase (i.e., temporarily changing the basal insulin setting in the pump) for that time to fix the high numbers, "*You can increase [basal] by 20% and then that way you can do it for like 12 hours and it'll take an hour to take effect, but it will just give you a little bit of an increase in your total basal insulin. So that way you don't go too low, or too high. So, in the pump you would set it for 120%. So that's 20% more than what your basal is. You could probably start out for 6 or 7 hours, and then after that time period it'll go back to, revert back to whatever your, uh normal basal settings are.*" (CDE in P02's CDE visit 3)

There are also times when a change is not made right away despite an issue but is postponed or is anticipated in the future based on expected future events. For example, in the following case, the CDE notes a pattern of high numbers after dinner but does not make those changes yet. She suggests waiting because school was starting in the next few days and would mean drastic changes to the patient's routine,

> The educator points to the post dinner highs and asks if the patient has any snacks after dinner. The caregiver mentions that the snacks are covered by insulin and are mostly without carbohydrates. The CDE asks the caregiver to keep an eye on post meal numbers. She does

> not make any changes and asks them to wait until after school starts. The CDE mentions she might change the carb ratio for lunch in the future. (P04’s CDE visit 4)

Lastly, in some cases, the data was not enough to clearly understand the nature of an episode of suboptimal management, determine its cause, or explain it. In such cases, providers suggested needing more data and instructed the patient to gather more evidence.

> The CDE suggested the patient to restrict eating after 8 pm and to track blood sugars closely beyond that time “*We need to figure out a way to prevent [lows] and I’m not sure if it’s your basal or your food. The only way to tell if it’s your basal is if you don’t eat after 8 pm for a couple nights and watch what your blood sugar does. But you cannot eat anything after 8. . . When do you want me to peek at your numbers again then?*” (CDE in P07’s CDE visit 07)

Challenge: Prioritizing between episodes

When multiple episodes of suboptimal management are observed and they all seem to be problematic, providers and patients need to choose which episode to consider for an intervention. This is particularly challenging when making changes for an active episode of suboptimal management could result in another problematic episode. In such cases, the extent of dosage change needs to be carefully calibrated to avoid creating more problems. The following vignette describes a case when the patient and the clinician decided against increasing insulin dose for an episode of high BG because a higher dose had been causing hypoglycemic events previously.

> The caregiver and the educator are discussing how the patient’s morning insulin dose does not seem to be bringing down the high numbers after waking up. The caregiver notes, “*He doesn’t go down when he gets up.*” After asking about when the patient takes insulin, how much he takes, and what kinds of food he eats in the morning, the educator thinks about the long-acting insulin being a potential cause, stating, “*The only thing it could be is his lantus [long-acting insulin] dose. Because it’s the only insulin that’s working in his body. But if we increase the insulin, increase the lantus, then he’s gonna be too low over night like we were having him, remember?*”

They are not sure about increasing the dose because a higher dose was causing low BG numbers and they had lowered the dose only recently avoid the low BG values. They did not make any changes in this visit. (P07’s CDE visit 6)

7.5 FRAMEWORK FOR EPISODE-DRIVEN SENSEMAKING FOR DATA ANALYSIS, INTERPRETATION, AND DECISION-MAKING

Based on observations of 71 data review sessions between Type 1 diabetes patients, caregivers and clinicians, this study describes the process involved in analysis, interpretation, and decision-making with multidimensional diabetes data. I show that using data to assess and inform the management of Type 1 diabetes is a nuanced multi-step process. For instance, not every pattern of abnormal blood glucose depicted in the data requires attention or an intervention. The need for an intervention is decided by careful assessment of multiple types of data, including device data and contextual data elicited from patients. Prior models of chronic condition management have described reflection and sensemaking activities for informing disease management. However, they do so in terms of high-level activities. For example, Mamykina et al.’s sensemaking framework for chronic condition management outlines three steps—perception, inference, and action [122]. Similarly, Hill-Briggs’ problem-solving model for diabetes management describes the different components of problem-solving that include the problem-solving process, the problem-solving orientation, disease specific knowledge, and transfer of past experiences[74]. My prior research noted that these frameworks are limited in the extent to which they represent the use of patient-generated data and called for better understanding how sensemaking and problem-solving happen through periodic and retrospective review of data in practice[144]. This work adds to the sensemaking and problem-solving frameworks by describing the different analytical activities through patients and clinicians make sense of the data for identifying potential problems, their causes, and their solutions.

Based on the results of this study, I will now synthesize the different analytical components involved in assessing self-management with multidimensional data into a descriptive framework that captures the analytic practices that I observed. In this framework, I propose that the analysis of multidimensional patient data for assessing and informing self-management of T1D involves the following four stages—detecting episodes of suboptimal outcomes or behaviors, elaborating episodes through narratives, classifying episodes to assess the extent to which an episode is problematic and needs an action, and generating recommendations for care actions in response to episodes that are potentially problematic (Figure 23). Below, I describe these components and present opportunities for informatics tools to support them using a combination of computational techniques and human-computer interaction techniques for user involvement.

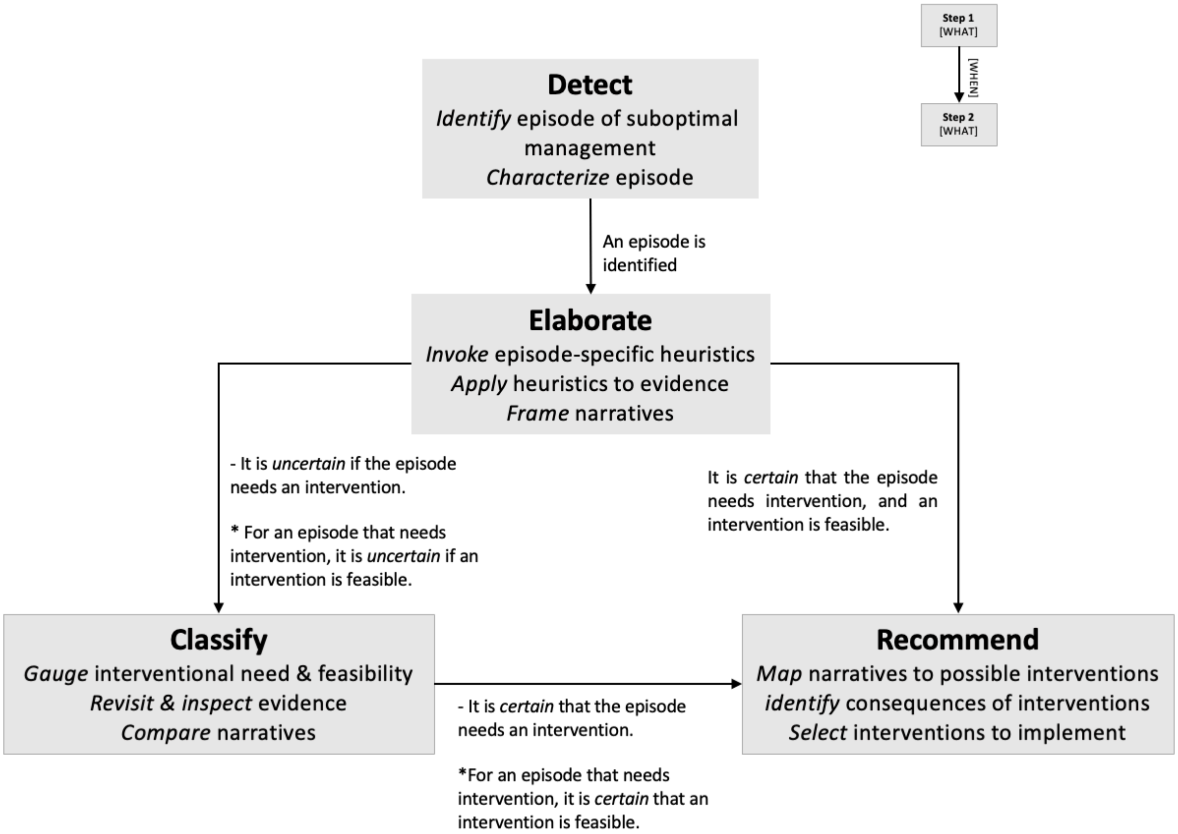


Figure 23: Framework for episode-driven sensemaking for data analysis, interpretation, and decision-making.

7.5.1 Detecting Episodes

The first step in assessing self-management involves identifying episodes of suboptimal clinical outcomes and suboptimal patient or system behaviors. In the findings, I noted that this involved identifying patterns in blood glucose readings, as well as carb-tracking and insulin intake behaviors over different time periods. It also involved identifying one-time unusual events in one or more data streams. Episodes served two functions in the analytic process. First, they offered a representation of a patient's lived experiences—a basis for identifying opportunities to improve management. Second, they guided the analysis of multidimensional data by helping direct attention to relevant data streams and domain knowledge (e.g., decision-making heuristics).

The identification of episodes from data is not straightforward. Patients and clinicians could have different understanding of an episode. These differences in understanding could be attributed, at least in part, to the limited support offered by existing tools. While pattern identification is well supported in CGM displays, it is not as well supported in insulin pump displays (Table 13). Finding patterns in the insulin pump data using the Logbook view (see Table 13 #3) involved visual analysis and inference (e.g., counting specific events in the Logbook view—red and green colored cells, visual chunking of data from specific time periods). This could potentially result in patients and clinicians having a different understanding of what is a pattern and how to observe it in the data, which I noted in this study and my prior research [144].

Supporting Episode Detection

Informatics tools can improve episode detection by using pattern recognition, anomaly detection, and information visualization techniques[59,174,180]. One key consideration in creating algorithms or user-facing displays of data for episode detection is for systems to understand an episode both phenomenologically and computationally and help users translate between the two. As a phenomenon, an episode depicts lived experiences of patients that may or may not be flagged by an algorithm, resulting in different understandings between patients and clinicians. What might be detected by an algorithm may not be concerning for patients and clinicians[123]. Such inconsistencies require the need to translate between the phenomenological and computational representations, which can be achieved by having patients and clinicians augment automated episode detection through interpretation and discourse [53,76]. For example, user feedback mechanisms could be used to enable patients and clinicians in assigning labels to computationally detected episodes, including labels for relevance, severity, accuracy, and representativeness[13].

In this study, episodes of suboptimal behaviors were identified in addition to episodes of suboptimal outcomes. The term "episode" in standard chronic care typically refers to a spell of illness where symptoms flare up and are experienced by patients. In addition to noting such episodes, my study participants, especially clinicians, noted one time or recurring *patient behaviors* that deviate from the prescribed standards for self-care and that could eventually result in or had already resulted in a symptomatic episode. Hence, I use an expanded notion of "episodes" prevalent in the domain of chronic conditions to include episodes of suboptimal patient behaviors, in addition to symptomatic outcomes. While much research has gone into computationally detecting different types of events from sensor data, existing computational representations largely focus on outcome data[180,181,185]. This study suggests the opportunity to create computational representations of episodes that can detect behavior-related episodes from different types of data in addition to outcome-related episodes; and that can help connect the two types of episodes. Approaches to detecting patterns in adherence or non-adherence to specific behaviors, such as those developed for detecting mental health states, and medication adherence/non-adherence, could be adapted to detect diabetes-specific behaviors[15,168,170].

7.5.2 Elaborating Episodes

Once an episode is identified from the data, it needs to be explained in terms of the different factors/ other data streams associated with it. In this study, I noted that deriving explanations involved building narratives guided by clinical heuristics that map a specific type of outcome to associated patient behaviors, and to the treatment or intervention that would be suitable. Depending on the type of episode detected, participants sought different data subsets (such as days with certain characteristics) that represented an episode in relation to potential causes and consequences. Elaboration was challenged when there was insufficient data to explain the episode and when multiple factors seemed to be involved. With the current tools, patients and clinicians must spend time and effort in visually extracting relevant events and data, for example using the logbook format display shown in Table 3, to identify subsets of data (e.g., episodes that fit certain criteria, episode-specific cause-effect mappings) that provide evidence to explain the episode.

Supporting elaboration

Elaboration can be supported by informatics tools through varying degrees of automation. One of the simpler approaches to support elaboration would be to enable episode-driven filtering and navigation of data, which would go beyond the general filtering and navigation techniques highlighted by prior work[179]. Episode-driven data filtering and navigation capabilities in tools could help users identify subsets of data most relevant to an episode. To implement such a feature, tools could employ decision-making heuristics to encode the data [68,161]. Based on the findings, these heuristics would connect episodes to relevant events/factors/data, potential corrective actions, and consequences to detect subsets of data associated with an episode (e.g., connecting a particular factor to its effect on an episode in a particular time window).

More sophisticated approaches that could be valuable include automated narrative generation, especially approaches that can meaningfully extract and combine different units of information or insights from multivariate data to provide recommendations and explain those recommendations to support actionability (e.g., episode-cause-recommendation). For example, approaches in automated journalism include extraction of atomic or semantic units of news/events/information that can be combined to create narratives [24,25,84]. In the context of health data, similar automated generation of narrative components could be applied, using the episode, its characteristics, the episode context, its interpretation, and episode-specific recommendations as the semantic units. Such an approach would need to rely on a clinical knowledge base (e.g., containing different types of episodes, typical causes, type of data required to interpret episodes, possible interpretations, and general corrective actions/decisions) to provide a foundation for meaningfully identifying and connecting different units of information into a narrative that fits the data[19,42].

7.5.3 Classifying Episodes

Depending on the clarity achieved during the elaboration step, classification seeks to understand if an intervention is needed and is feasible in response to an episode of suboptimal management. In my observations, classification involved closely considering different characteristics of the episode, such as its severity/strength and frequency, along with the confidence with which it could be explained. This process resulted in identifying episodes for which an intervention was absolutely required and feasible, an intervention was absolutely required but not entirely feasible, an intervention may or may not be required and is feasible, and an intervention is not required at all. The lack of a quantified understanding of the different episode characteristics and its association with different factors challenged classification, especially when there was mixed evidence in support of a causal explanation for an episode. Quantification of episode characteristics and its association with factors is not available in the existing tools and therefore must be inferred by participants in the analytic process.

Supporting episode characterization for classification and shared understanding

Classification algorithms could be employed to identify different episode categories described above by establishing and labeling episode characteristics, such as how severe an episode is, how long the episode has existed, and factors that might cause an episode. However, a bottleneck for classification algorithms, and machine learning tasks in general, is the difficulty to label dataset because of which methods to automatically generate labels for training data, also called weak supervision, have been developed [146,173]. One weak supervision approach employs user-created domain heuristics as functions for labeling subsets of data. Models trained on data labeled through weak supervision have matched the performance of fully supervised models (e.g., pairwise comparison[161]). A heuristic-based weak supervision approach could be suitable here to automate the generation of training labels and train classifiers for identifying episodes[161,173]. Implementing this approach would involve having clinicians express and rank the heuristics used to understand, assess, and compare episodes for an intervention (e.g., when comparing two episodes to select one that needs intervention, heuristics such as choose the episode with larger variability, choose the episode with a higher number of contributing days, choose the episode whose occurrence has the strongest correlation with bolus timing relative to meals, or choose the episode that is the longest in duration). Such heuristics can be programmed to create training data, which can then be used to train models that can automate the classification of episodes to understand the need and feasibility of an interventional response.

In addition to automating data labeling and episode identification and classification, information on different episode properties/labels could also be visualized to enable a comparative understanding of different episodes based on their features, such as whether one episode is more severe or more evident than another. This would support informed prioritization and classification, especially in the challenging situations when there is inconsistent evidence in support of an episode or when there are multiple episodes of potential interest. More precise characterization could also support a shared understanding of the episodes among clinicians, patients, and caregivers. When defining episode properties, it would be important to identify characteristics that are inclusive of both patients' and clinicians' needs and preferences as they may have different perceptions about what counts as a problem or an episode and different tolerance for deviations from the normal [144].

7.5.4 Recommending Interventions for Episodes

Once it is established that an episode requires an intervention that is feasible and likely to prevent future episodes, I saw clinicians recommend and discuss one or multiple corrective interventions along with their potential consequences. Different types of recommendations involved one-time changes, situational changes, and anticipated changes to regimen or behaviors, and recommendations to gather more data. When multiple episodes were observed in the data and required intervention, participants faced challenges in prioritizing the episode that should be acted upon.

Supporting recommendation generation and presentation

T1D automated insulin delivery systems and decision support systems can improve glycemic outcomes by showing predicted blood glucose values and offering recommendations for insulin dosage changes. However, their real-world use by patients is not entirely understood[172]. Like intelligent clinical decisions-support systems for experts, they may present issues such as limited transparency into system recommendations, which may limit users in understanding or implementing those recommendations[123]. Presenting recommendations alone may not suffice to support human judgment and validation. Based on the episode-driven framework, tools could support episode-specific recommendations with evidence or examples from the data (e.g., as has been done in context-based systems to explain inferred contextual states[14], and in ML systems to explain system decisions[23]), which may include presenting representative data to explain why a particular recommendation was generated. Presenting examples has shown to aid user understanding of system generated information[23]. Another potential technique that I have discussed above is to use narrative visualizations to present a sequence of displays to logically explain a recommendation step-by-step. In addition, in case of multiple episodes that result in generation of multiple recommendations, visualization techniques to compare episode-recommendation combinations could be employed[1,44].

7.6 CONNECTIONS WITH OTHER FRAMEWORKS

The episode-driven sensemaking framework extends prior frameworks of sensemaking[122] and problem-solving[74] that describe the management of chronic conditions. Hill-Briggs' problem-solving model focuses on identifying problems or barriers in self-management and describes four aspects of successful chronic disease management: a) problem-solving process, b) problem-solving orientation, c) disease specific knowledge, and d) transfer of past experiences[74]. For diabetes management the problem-solving process involves engagement with one's data to identify problems that need an intervention. However, the problem-solving model does not seek to account for the data-based activities involved in the problem-solving process for problem identification and resolution. This study thus complements the problem-solving model by unpacking the different analytical components involved in assessing self-management to identify noteworthy episodes, establish whether those episodes are problematic and require an intervention, and establish the potential care action in response to the problematic episodes. This study also shows that problem-solving is not as straightforward as applying clinical guidelines to data. For example, episodes of suboptimal management flagged by systems are not always considered problematic enough to be acted upon. To support problem-solving, Mamykina et al. developed MoDD, a web-based tool that engaged patients in problem-solving using their glycemic data. Based on the glycemic pattern chosen by patients, MoDD directed educational content to patients to show common behavioral triggers of a glycemic pattern and potential solutions to fix the pattern. The concept of episode-driven sensemaking is similar in that it uses glycemic patterns to guide analytic focus. The episode-driven sensemaking framework builds upon pattern-based problem-solving and education by accounting for the nuanced analytical activities involved in problem-solving with multiple streams of data.

My framework also extends Mamykina et al.'s sensemaking framework that describes diabetes management in terms of three activities (perception, inference and action) and two modes of self-management in which these activities are performed—habitual and sensemaking [122]. The habitual mode of management is in effect when a person's experiences follow their existing mental models of disease management. However, when experiences do not fit existing mental models, sensemaking is required to adjust old mental models or create new mental models. While sensemaking focuses on generating explanatory models for breakdowns for unusual experiences during self-care, the model does not seek to clarify how those explanatory models are generated using data. For the habitual mode, the model does not describe how existing models are invoked based on data, or how these models are translated to different types of possible care recommendations. The episode-driven sensemaking framework contributes an understanding of how multidimensional data along with domain knowledge is used for analytic activities related to habitual and sensemaking mode. Given the granular focus of this work on the analytical practices that support data-driven management of chronic conditions using patient-generated data, this work contributes to both frameworks by offering concrete suggestions for how informatics tools could better support both problem-solving and sensemaking.

7.7 DISCUSSION

Despite growing evidence showing the availability and utility of patient-generated data, its potential remains to be harnessed because of the lack of adequate informatics tools. Limited accounts of analytical activities and processes through which multidimensional data is translated to decisions has challenged the development of tools[100]. To understand the use of patient-generated data, prior work has primarily focused on establishing information needs of data consumers (what data or insights users want), but limited work has focused on understanding the process of reflection and insight generation[10]. For example, MacLeod et al. identified different types of questions related to episodes and triggers and common information needs of patients in managing diverse types of chronic conditions but do not provide an account of how patients or systems can use patient-generated data to answer these questions[114]. Similarly, frameworks of chronic disease management, such as sensemaking and problem-solving, provide an account of high-level activities involved in sensemaking or problem-solving but fall short of describing how data is used to realize these activities. Better developing technology to support users requires understanding reflection processes for insight generation, particularly when multiple data streams are involved in understanding and managing health[10,118].

Through this study, I address the abovementioned gaps by investigating 71 data review sessions between T1D patients/caregivers and their clinicians held remotely on a biweekly basis for regular assessment of T1D self-management. I provide a detailed account of analytical activities and challenges involved in the use of multidimensional patient-generated data to regularly assess and inform the management of T1D. The analytical activities described here are coherent with the high-level activities described in Mamykina et al.'s sensemaking framework, such as using data to invoke existing models or create explanatory models of the disease[122]. However, I also studied *how* data is used to invoke or create explanatory models. Prior work also falls short of accounting for how explanatory models map to recommendations for care. This study shows that heuristics governing the relationship of outcomes, causes, and potential actions drive the generation of recommendations. User agency and judgment play a key role in the application of these heuristics to data for insight/recommendation generation.

Several computational techniques are emerging that promise to improve insight generation from patient-generated data. Prior work on automated insight generation has noted that the outputs of systems could be misrepresentative of patients' lived experiences and could also be misinterpreted by patients[11,85,86,119,120](e.g., oversimplification of correlations), suggesting the need to better support users in the interpretation and application of these insights. Similar observations have been noted in the study of intelligent clinical decision-support systems. Notable findings from the use of such systems include their inability to work with real-world data in the wild, their "black box" nature which undermines comprehensibility and trust, and the clinicians' need to better understand system behavior and decision-making [123,186]. In addition to such issues, in a real-world deployment, computational systems might also face challenges that I described in the findings, such as providing recommendations with inconsistent evidence, and prioritizing recommendations when multiple problematic episodes and corresponding recommendations are present. Taken together, these issues suggest that in addition to devising techniques for automating insight generation, there is a need to engage users in the process of insight generation and support intelligibility of system-generated insights. What remains to be explored is how systems can combine computational techniques with user-facing interactive components such that user agency, cognition, and judgment are not replaced but facilitated and empowered[118]. This calls for an interdisciplinary and human-centered approach to understand user practices with existing tools, the design space and limitations of existing tools, and the opportunities for applying computational and human-computer interaction techniques.

7.8 LIMITATIONS

This study reflects the practices of providers and patients from a single facility in the context of Type 1 diabetes. While these practices may differ from one clinic facility to another and from one condition to another, the framework for episode-driven review of data proposed in this paper could be applied to the assessment of health data for other chronic conditions, as many of these conditions are marked by symptomatic episodes that offer opportunities for patients and clinicians to improve care. Future work will need to be done to extend these findings and generalize to different clinics, providers, and chronic conditions.

7.9 CONCLUSION

In this chapter, I provide a framework for episode-driven sensemaking with data that involved four stages - episode detection, characterization, classification, and recommendation generation. I identify the data-related challenges encountered in these different stages. I describe ways in which this framework provides a structure for technology design to making personal health data actionable. The episode-driven framework provides a novel perspective on understanding data practices and a basis for designing technology to facilitate engagement with data. It opens opportunities for future research - automating episode detection and characterization to rank them for prioritization, designing tools to filter and package data relevant in the context of an episode, and designing interfaces to present data to support different phases of episode-driven data review.

Balancing automation and user agency in generating insights is important in making multidimensional data actionable, particularly when decisions need to be made from smaller data sets with missing data or data of suboptimal quality, requiring user judgment. Combining automated insights with visual evidence can help address issues reported in prior work on the user experience of automated insight generation, which include the automated insight not representative of lived experiences, the insight not being actionable enough, and the insight being misinterpreted by the user. An episode-driven approach to system design can enable balancing automation and user engagement functions in a tool. Different stages of the episode-driven framework lend themselves to combining computational and interactional techniques in nuanced ways.

Episode-driven approach to design tools also provides a basis to direct users towards predefined narratives (e.g., cause-effect, behavior-consequence) by combining relevant slices of data and presenting them in an appropriate sequence using the four analytic phases as a basis (i.e., detection, elaboration, classification, resolution). For example, the detection phase can involve presenting an automatically detected episode while the elaboration phase can involve presentation of actual evidence to contextualize and explain the episode with more information. Given the lack of explanation-based interfaces studied in prior work, the episode-driven framework can guide future research to design and evaluate explanatory data interfaces for the presentation of narratives.

Considering the episode-driven framework as a basis for tool design can also help bridge the data practices of patients and clinicians. Episode-driven data practices as identified in this study resemble the analytical activities conducted by patients and reported in Chapter 5. For example, patients identified trends in the data and explained those trends in terms of the available data. In a similar direction, collaborative data analysis happened in the context of episodes. There were a few differences in how patients and clinicians engaged with data, which the episode-driven sensemaking framework could help bridge. One key difference between patients and clinicians in interpreting the data was regarding the explanatory frames used. The frames that patients employed were primarily sought or developed from the data. The frames employed by clinicians in the collaborative data reviews were clinical heuristics, templates that mapped outcome-cause-corrective actions. These heuristics did not emerge from the data through exploration but rather directed what data was important to look at (i.e., data that served as evidence matching the heuristics). The heuristics simplified data interpretation and decision-making as they helped improve analytical focus. Another difference between patients and clinicians was regarding the determination of whether a decision should be made from the data. Patients were unable to make sense of the data when they could not find a trend or when they identified counterintuitive findings. In the collaborative context, the extent to which the data fit one of the explanatory frames helped assess when the data could be used for making decisions and when not. Lastly, patients did not engage in assessing the data sufficiency and the regimen quality, which was done in the collaborative reviews by the clinicians. The episode-driven framework can help bridge these differences as it enables the incorporation of clinical and contextual knowledge required to interpret the episodes (explanatory frames, data sufficiency and regimen quality) in the design of data review tools.

Based on the above-described potential of the episode-driven sensemaking framework in guiding system design, in the next chapter, I present a prototype informed by this framework and describe the results from a qualitative evaluation of the prototype with patients. This prototype complements the interfaces I studied in chapter 5, which had similar limitations as reported in prior work (e.g., counterintuitive insights, limited actionability).

Chapter 8 Designing Episode-Driven Data Narratives to Support Sensemaking and Decision-Making with Multidimensional Health Data

In chapter 6, I synthesized the episode-driven sensemaking framework and established its potential as a basis for designing tools. Additionally, I provide guidelines for creating tools to facilitate episode-driven sensemaking. In chapter 3, I established the promise of narrative visualizations and visual data stories as techniques to simplify the engagement with multidimensional health data for patients, the non-expert consumers of data. In this chapter, I bring these learnings together to establish the potential of narrative visualization techniques for implementing a tool for episode-driven sensemaking. I draw from the findings and implications of Chapter 6 to articulate the requirements for creating *episode-driven narratives* for sensemaking and decision-making with data. By designing a set of data interfaces informed by the episode-driven sensemaking framework and by evaluating them with patients, I show that data-driven storytelling is a promising technique to design for episode-driven presentation of multidimensional data. This technique can help systematically organize and present evidence in an ordered sequence and guide user attention to different types of information to facilitate structured review of data. I describe the results of a study I conducted to compare patients' experiences of sensemaking and decision-making with diabetes data using the commercially available platforms and the episode-driven narratives. For this study, I led the prototype design, data collection and analysis, and writing.

8.1 SUMMARY

Background: Given the limitations of exploratory data interfaces studied in prior work (misinterpretation, misrepresentation, complexity), and the limited data interpretation skills of non-expert consumers, there is a need to move beyond exploratory interfaces and investigate techniques to better communicate personal health data to patients.

Objective: The objectives of this study are to – 1) establish the feasibility of the episode-driven sensemaking framework and narrative visualization techniques to design episode-driven

data narratives from multidimensional data, and 2) compare patients' experiences of reviewing data using the commercially available data review platforms and the episode-driven narratives.

Methods: The study involved two data review sessions for each participant. For each session, participants reviewed data from continuous glucose monitors and insulin pumps while thinking aloud and completing a questionnaire based on the data. After the review, participants took a cognitive load assessment. In the first session, participants reviewed the data using the commercially available PDF reports. In the second session, participants reviewed the data using the episode-driven narratives. In the first session, participants also took standard assessments on diabetes numeracy and graph literacy. In both the sessions, participants were interviewed using a semi-structured interview protocol to understand their overall experience and challenges of reviewing data through two different platforms.

Results: 14 participants enrolled in the study but only 6 participants completed the two sessions. The episode-driven narratives resulted in a higher average score on the questionnaires and a low average task load than the commercial PDF reports. Episode-driven narratives enabled decisional agency of participants in different ways. Overall, they improved data interpretation, data navigation, and the ability to identify self-care actions from the data as compared to the commercial PDF reports. Findings suggest two design implications – designing to enable user agency and designing different types of explanations.

8.2 MOTIVATION

One key limitation of personal informatics systems for general wellbeing and chronic illness management is the lack of support for enabling the use of multidimensional health data. Particularly, interfaces designed to help make sense of multiple data streams are primarily exploratory, with limited support for translating data-driven insights to care actions in the context of personal health. Less exploratory interfaces include the presentation of correlations or conclusive insights using natural language summaries and graphs but are known to result in misinterpretation, perceived misrepresentation of lived experiences, and poor actionability [11,64,85]. With regards to design, they are limited in how they combine insights with actual data to explain the insights to the users. Additionally, while systems are built with a general understanding of information needs identified by prior work (e.g., correlations, distribution) [34], they are designed with a limited understanding of what reflection entails or the nuanced ways in which it could unfold for generating insights and translating them to actions [31]. As a result of these limitations, systems do not entirely align with user needs or expectations and result in data interfaces that are not actionable enough [64]. Lastly, despite many of these systems relying on different types of visualizations to present multidimensional personal data, there is a lack of guidelines for designing personal data visualizations, especially for use by non-experts, such as patients [57,132]. These research gaps present a need to better understand analytical tasks and workflows involved in reflection on personal health data, scaffold those workflows through data interfaces by moving beyond exploratory interfaces, and support translation of data to actionable information.

To address these gaps, I presented several investigations in the previous chapters that, taken together, contribute an improved empirical understanding of individual and collaborative data practices of patients and clinicians and provide implications for designing tools to support these data practices. Specifically, I provide a framework of analytical stages involved in data-based sensemaking (which I call episode-driven sensemaking - Figure 23); provide an understanding of different types of assessments performed using health data to inform the care of Type 1 diabetes (assessment of data quality, clinical outcomes, patient behaviors and regimen quality); demonstrate the challenges of using data to make care decisions; and point to implications for designing tools. In this chapter, I implement design implications from my previous investigations into episode-driven data interfaces and evaluate them with patients.

8.2.1 Chapter Outline

The outline for the remaining chapter is as follows. I first review relevant prior work to demonstrate the need for explanatory approaches to support engagement with health data and the potential of using visual data narratives as a basis for engagement. I then present the design rationale for creating a narrative-based tool to facilitate episode-driven sensemaking with data and present a prototype. Next, I describe the research questions and the methodology for an exploratory evaluation of the prototype. This is followed by the findings from the study to compare commercial data review platforms with episode-driven narratives. I conclude the chapter by discussing the design implications for tools to support episode-driven sensemaking with patient-generated data.

8.3 RELATED WORK

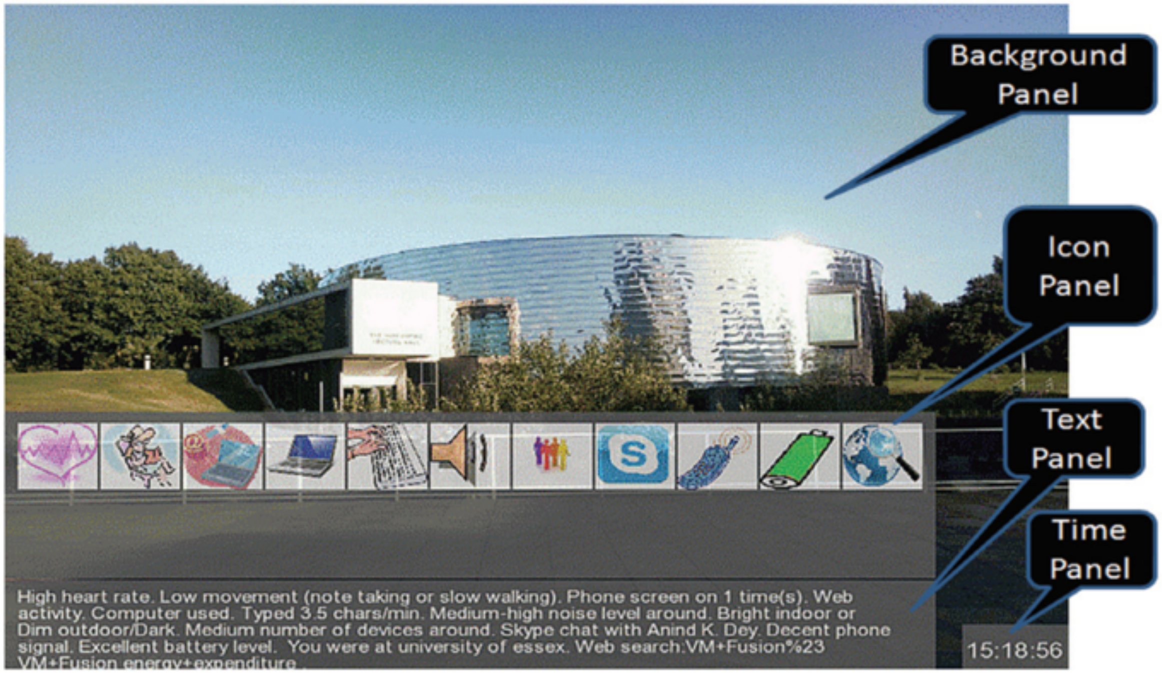
In this section, I point to the potential of data narratives/stories as a promising technique to present data. I do so by drawing upon prior studies using similar techniques in different contexts – personal informatics and health informatics. Particularly, I review a tool for story generation from lifelogs, a tool for story authoring from personal data, a tool for storytelling from multimodal data, and a study that demonstrates the value of communicating data to patients in the form of stories.

8.3.1 The Promise of Stories in Personal Informatics Systems to Present Data

Summary: I describe two studies below that establish the usefulness of stories of personal data (correlating different types of information, presenting data to others to seek social support). However, these systems are considerably exploratory in nature. While they enable users to create and explore different types of stories from their data, they do not assess how the audience (those reading the stories) interprets these stories and the associated data. Using stories to present data is promising and there is an opportunity to explore explanatory interfaces, such as visual data stories, for engagement with personal data.

Correlating lifestyle data through stories

Pavel et al. contributed an automated story creation framework to enable the presentation of multidimensional data. The framework extracted events from wearable sensors, mobile phones, and desktop machines and the context for those events. It determined the value of different types of contexts (social, environmental, mental, emotional, spatial, etc.) for a particular event, mapped data about contexts to semantic labels, and mapped numbers to media for presentation. The stories were generated based on meaningful events identified by the user or by the system. The platform allowed for users to customize the stories (Figure 24) by flagging meaningful events through annotations. For different meaningful events, the system then presented an automatically generated multimedia story to compile the multimedia data associated with the event and present to users as shown in Figure 3. Users found different types of contexts important for different types of events. For instance, for a meeting event, stress, noise, and room temperature were important. Users found that the data was easier to understand through stories as compared to a traditional diary. Stories helped get the essence of the data quickly. This work establishes that stories provide a promising approach to help people engage with multivariate data. Additionally, they help present evidence most relevant for the story. However, stories have not been employed to present multivariate data about events experienced by people in the context of a chronic condition.



High heart rate. Low movement (note taking or slow walking). Phone screen on 1 time(s). Web activity. Computer used. Typed 3.5 chars/min. Medium-high noise level around. Bright indoor or Dim outdoor/Dark. Medium number of devices around. Skype chat with Anind K. Dey. Decent phone signal. Excellent battery level. You were at university of essex. Web search:VM+Fusion%23 VM+Fusion energy+expenditure .

Figure 24: A story event displayed as a multi-media object to represent different types of data in Pavel et al.'s system. Different types of icons in the icon panel show the types of data that the system presents for a meaningful event it identified. Text description at the bottom presents data values for users to infer correlations in the different types of data.

Telling stories from personal data to seek social support

Epstein et al. studied Yarn [54], a mobile application to support authoring stories from personal data related to DIY projects and marathon preparation. Yarn offered several templates to guide people in authoring stories of accomplishment using images, numbers, text, and time. Through pre-specified templates, users could create several moments related to a larger story, such as moments related to different practice runs when training for a marathon, which were then presented in a feed and ranked based on importance (importance was determined through several factors, such as the length of the run). This feed could be viewed by friends to understand the user’s story and progress on a project, which could then become a basis for conversations and social support. Yarn was a story authoring tool and did not involve studying data interpretation through stories, which my study investigates.

8.3.2 The Promise of Stories in Health Informatics Systems

Summary: The two studies described next exemplify the use of data-based stories/narratives in two different contexts within healthcare – one to support collocated teamwork, reconstruct sequence of events, and create common ground; and another to support communication with the patients. Taken together, these studies establish the potential of using narrative techniques to make sense of multimodal data and to communicate the data to non-experts. Additionally, they point to the lack of guidelines for designing storytelling visualizations from health data.

Storytelling from Multimodal Data to Aid Collocated Teamwork

Given the lack of explanatory techniques to promote the use of data by lay consumers of data, Maldonado et al. used the storytelling approach to create visualizations from multimodal data to aid collocated teamwork in a nursing classroom [124]. Different types of sensor data, such as team member positions, arousal peaks through skin conductance, were combined and visualized to show students' activity to the teachers and to students themselves. For example, visualizing sensor data on student positions and timeliness of actions helped teachers identify student mistakes and provide corrective feedback. The prototype consisted of a timeline of actions of different nurses as a background, which was then enhanced with highlights and annotations to draw attention to mistakes or wrong actions and to validate correct actions in different situations (Figure 25). A rule-based algorithm was used to highlight different sections of the timeline and add appropriate annotations. The layered story-telling approach using a common event timeline layer with additional layers highlighting different sections of the timeline helped students reconstruct the sequence of events and understand their performance in that context. Martinez-Maldonado et al.'s work highlights the feasibility of using a storytelling approach to enable sensemaking with multimodal data. Particularly, bringing attention to specific information elements in the context of a larger narrative (the timeline here) and an overarching objective (that determined what gets highlighted) can simplify sensemaking with different types of data. In a similar direction, the design of my prototypes involves highlighting different types of data through episode-driven data narratives. It explores the role of narrative visualizations in the context of chronic disease management.

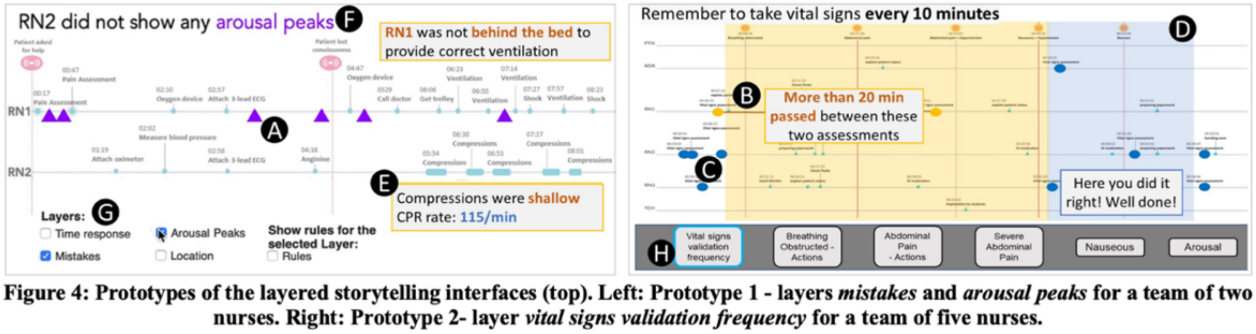


Figure 4: Prototypes of the layered storytelling interfaces (top). Left: Prototype 1 - layers *mistakes* and *arousal peaks* for a team of two nurses. Right: Prototype 2- layer *vital signs validation frequency* for a team of five nurses.

Figure 25: Storytelling with sensor data from nursing team work on a trauma patient (Maldonado et al.). A timeline of events is annotated and highlighted to surface different types of information, such as mistakes and achievements of students, to help educators provide feedback.

Using Verbal Stories Created from Patient Data for Tailored Communication with Patients

Storytelling from patient data has been established as a potential approach for tailored communication with patients. Hougaard and Knoche studied the use of storytelling by therapists working with cognitively impaired patients to explain to them their data on game playing performance [77]. Through timeline and heatmap visualizations, therapists were able to tailor stories using cognitive training data to fit the level of awareness a patient had. The objective of creating narratives for patients with different levels of awareness was to focus on different insights important for different patients and shift their views on the data based on their awareness. For example, for patients with lack of self-awareness rejecting the existence of a performance deficit, the aim of storytelling was to bring attention to the deficit and create awareness. For patients with emergent awareness, the aim of the narrative was to provide strategies to overcome problems related to the deficit. Given the value of stories, this study points to the need to design storytelling visualizations tailored to the needs of patients (e.g., level of awareness for patients with cognitive impairments). It also points to the lack of guidance on how to design visualizations to support storytelling in the context of health data.

8.3.3 Limitations of Patient-Facing Health Data Visualizations and Opportunity for Presenting Multidimensional Health Data Through Stories

There is an increasing focus on improving patient-engagement with consumer-generated health data through visualizations [3], more so for patients with chronic conditions [113]. Within this context, researchers have called for improving the design of visualizations to accommodate audiences with different skills by going beyond traditional and simple graphical presentations of data to information-rich graphics or infographics [78]. Contrary to the practices of using simple visualizations for patients, researchers have called for designing patient-facing infographics that contextualize the data presented and allow for extracting meaning when literally interpreted [4] to make the designs independent of literacy skills.

While the increasing emphasis on designing information rich presentations or infographics (as opposed to simple graphs) to engage patients with diverse data literacy skills is a promising change, more research is needed to go beyond presenting single variables, that is presenting multiple streams of data for easy comprehension and interpretation. Towards supporting patients in interpreting multiple streams of data, data visualization techniques, such as data-driven storytelling [98], and different types of guidance mechanisms (e.g., orienting, directing, prescribing) used in visual analytics [26], can help create simpler interfaces to guide lay users in navigating their data and can reduce the burden of translating data to actionable information. To help users generate value from multidimensional data, visualizations may need to serve both communicative and analytic functions. The communicative function of visualizations is to make it easier for the users to get the point. The analytic function of the visualizations is to help exercise their agency in identifying and understanding associations across different data streams and make decisions based on these associations. Data-driven storytelling and narrative visualizations can enable communicative as well as controlled analytic functions for users to understand data and infer insights. These techniques have also emerged as powerful for engaging non-expert consumers of data [28]. To the best of my knowledge, narrative visualizations have not been explored for presenting multidimensional patient-generated data. My study qualitatively evaluates episode-driven visual narratives based on diabetes data with patients as users.

8.4 PROTOTYPE DESIGN FOR EPISODE-DRIVEN DATA REVIEW

In this section, I describe the high-level requirements for an episode-driven data review tool embodying the episode-driven sensemaking framework. I then present a prototype showing an episode-driven narrative. I describe the design choices and the rationale for visualizing the different prototype components and for connecting the components together. Lastly, I establish the feasibility of algorithmic identification of the different prototype components.

8.4.1 High-Level Requirements for An Episode-Driven Data Review Tool

Below, I summarize the different analytical activities involved in episode-driven sensemaking (Chapter 6). I then describe how they map to different system features for an episode-driven data review tool. In making sense of multivariate data, patients and clinicians performed the following activities:

- identify patterns or atypical events in the data that denote suboptimal management and potential problems – I call these patterns and atypical events **episodes**. An episode represents a single event or a collection of events that have certain characteristics, such as violating a threshold, or satisfying a rule (e.g., occurring at the same time, different times, or one after the other). Episodes represent phases from the past where management was not optimal and could require certain self-care actions to avoid their recurrence. *A tool for episode-driven review needs to detect and present episodes*.
- elaborate the episode to identify a cause – Elaboration involves the use of **clinical and contextual information** to frame an explanation for the episode. This information could include patient behavior as represented by the data or as reported by the patient (food intake, insulin intake, activity, manual interventions, data recording), regimen quality (insulin dose settings in the pump), events surrounding the episode, interpretive frames (templates that map outcome-cause-action or behavior-consequence-action), and patient's information from the medical records. *A tool for episode-driven data review needs to present contextual information, that is the context in which an episode occurred as depicted by the data, and clinical information, such as the interpretive frames used to render data meaningful.*
- classify an episode as a non-problem or a problem – This involves understanding the extent to which an episode represents a problem and whether it can be fixed. Here, information such as episode strength, its likelihood of recuring, the extent to which a cause was established, the episode frequency, and overall glycemic performance, is considered to assess if an episode requires an action. Providing the abovementioned **episode characteristics** is important to help distinguish between problems and non-problems. *A tool for episode-driven sensemaking needs to present episode characteristics to enable classification*.

- identify a solution for a problematic episode by using **interpretive frames** that map outcomes to a cause and an action or behavior to outcomes and an action – For example, daytime high BGs are meal related and are caused due to insufficient carb ratios at meals. Changing carb ratio can fix the high numbers during the day. *A tool for episode-driven sensemaking would need to present the interpretive frame and the evidence on the outcome, cause, and action represented by the interpretive frame.*

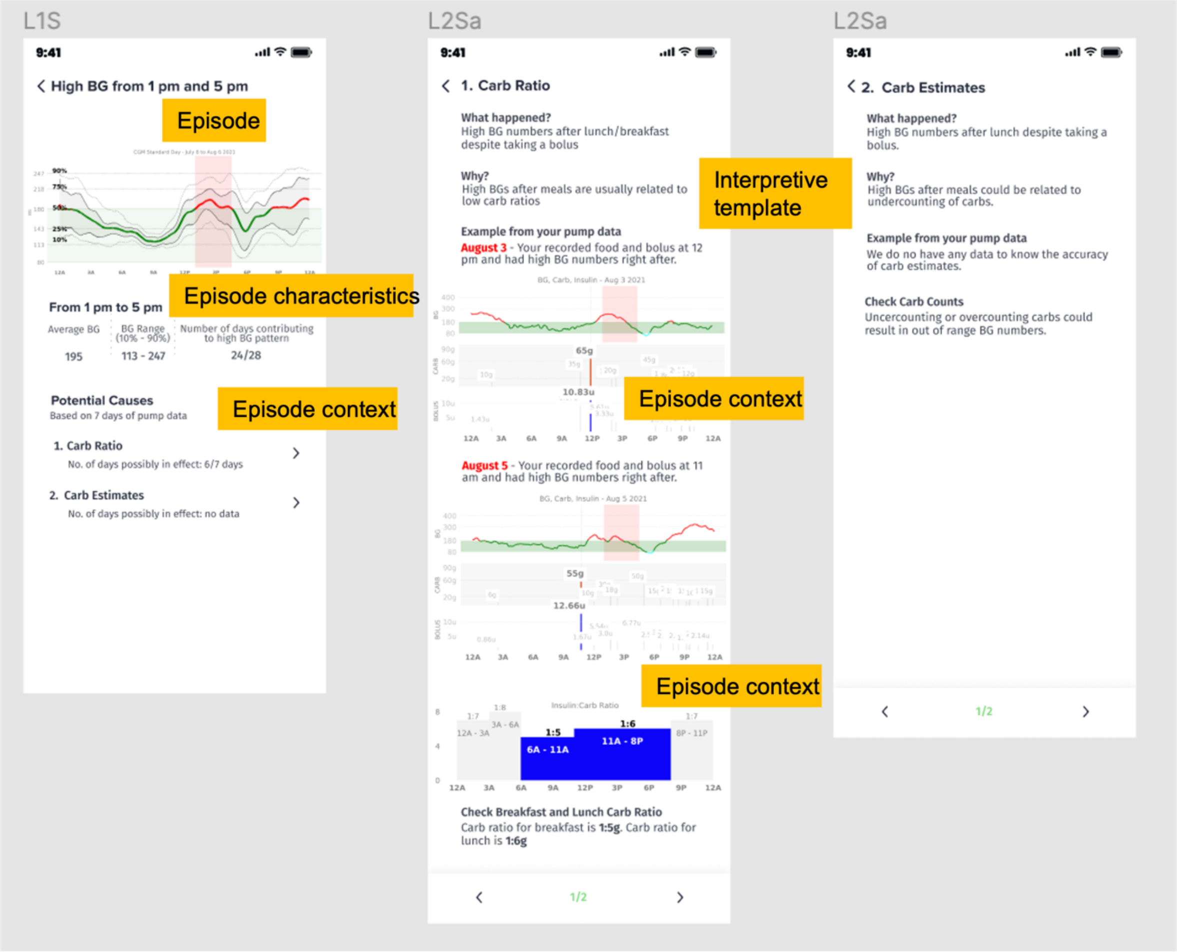
Following the description above, an episode-driven data review tool must include the following components - episode, episode characteristics, episode context, and interpretive frames. In essence, an episode-driven tool must package these components and present them for sensemaking. More specifically, it needs to do the following:

1. **Identify** and **visualize** the following information - an episode, its characteristics, the episode context to identify potential causes, and an interpretive frame to help understand the episode and its context.
2. **Organize** different types of information (episode-characteristics-context-interpretive frame) into a meaningful sequence or a narrative to explain what happened and why, and to help orient users towards different options for self-care actions for an episode.
3. **Communicate** the different types of information by visually connecting the episode, its characteristics, the context, and the interpretive frame and guide user attention to different information and the underlying data.

In the next section, I present the design of interfaces that I created using patient data to visualize and present an episode, its characteristics, the context and the interpretive frame.

8.4.2 Interfaces for Episode-Driven Data Review

Figure 26 presents prototype interfaces, an episode-driven narrative, showing different types of information – an episode, the episode characteristics, the episode context, and the interpretive frame/template. Each of these can be mapped to a specific component in the prototype. Another prototype example can be seen here.



High BG from 1 pm and 5 pm

From 1 pm to 5 pm

Average BG	BG Range (10% - 90%)	Number of days contributing to high BG pattern
195	113 - 247	24/28

Potential Causes
Based on 7 days of pump data

1. **Carb Ratio**
   No. of days possibly in effect: 6/7 days
2. **Carb Estimates**
   No. of days possibly in effect: no data

1. Carb Ratio

What happened?
High BG numbers after lunch/breakfast despite taking a bolus

Why?
High BGs after meals are usually related to low carb ratios

Example from your pump data
August 3 - Your recorded food and bolus at 12 pm and had high BG numbers right after.

August 5 - Your recorded food and bolus at 11 am and had high BG numbers right after.

Check Breakfast and Lunch Carb Ratio
Carb ratio for breakfast is **1:5g**. Carb ratio for lunch is **1:6g**

2. Carb Estimates

What happened?
High BG numbers after lunch despite taking a bolus.

Why?
High BGs after meals could be related to undercounting of carbs.

Example from your pump data
We do no have any data to know the accuracy of carb estimates.

Check Carb Counts
Uncercounting or overcounting carbs could result in out of range BG numbers.

Figure 26: Prototype interfaces with different components of the episode-driven data review tool highlighted.

8.4.3 Design Choices and Rationale for Visualizing Individual Interface Components

For each of the components shown in the interfaces above, the data presented, and the presentation formats are listed in Table 15. The format choices for the different data visualizations were guided by the review of existing visualizations (Figure 27) that I presented in chapter 6 and following general guidelines from prior work.

Element	Data	Values and Format	Visualizations and their placement
Time presentation	Time shown in 12-hour format with 3-hour intervals	Horizontal labelled axis. Labels - 12A, 3A, 6A, 9A, 12P, 3P, 6P, 9P, 12A	The most common presentation in the existing visualizations is midnight to midnight with 1-hour

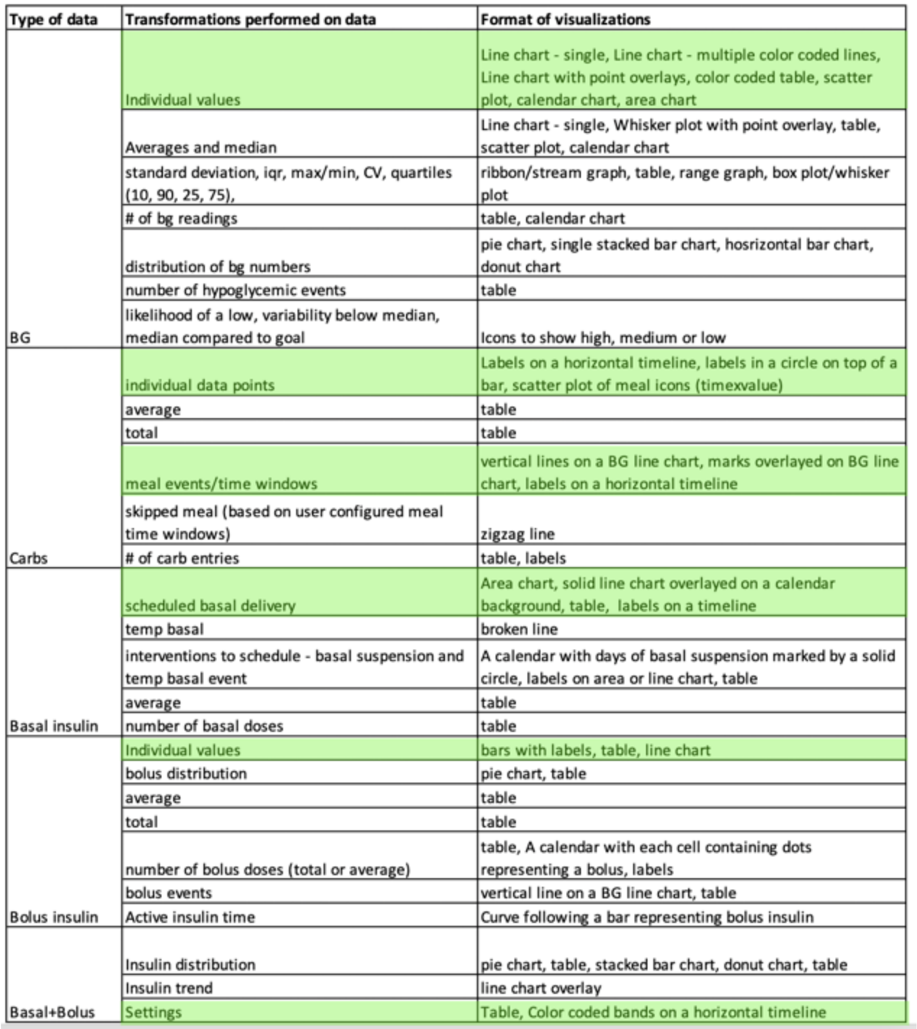
			intervals. I am choosing 3-hour intervals for a mobile interface
Episodes Episodes in this context are glycemic patterns. These patterns denote recurring high or low BG numbers during a certain time window of the day.	Data used were BG values from the CGM.	A reference band to denote normal BG range. Median BG values shown via a line chart with color coded segments that highlight above or below range segments. 10th%, 25th%, 75th % and 90th % lines presented as dashed lines. Vertical bar highlighting the segments of the median BG curve where it goes above range and the 5th % curve where it goes below range. 5th % curve was used to show the low BG patterns. 50th % curve was used to show high BG patterns.	One of the presentations of BG values in the Dexcom Clarity report is the Ambulatory Glucose Profile. I am choosing to base my visualization of BG curves on this presentation.
Episode characteristics	Data used were BG values from the CGM.	Text labels placed below the BG pattern visualization to show the following:	I chose to use text labels instead of visually depicting these because of the limited screen space on a mobile platform.

		Number of days a particular BG trend was present 10th % - 90th % BG range The average BG during the time duration of the segment where the curve is either below or above range.	
Episode context	BG	Median BG value curve with color coded segments that are above or below range A reference band to denote normal BG range Vertical bar highlighting the segments of the median BG curve where it goes above range and the 5th % curve where it goes below range.	I chose to display BG, carb, and insulin values for single days in a stacked time series following what is done by existing platforms to show multivariate data.
	Carb	Vertical lines displayed across a timeline; values shown through text labels The carb values relevant to the episode were highlighted using a bright color and the irrelevant	

		values were muted by coloring them grey.	
	Bolus	Vertical lines displayed across a timeline; values shown through text labels The bolus values relevant to the episode were highlighted using a bright color and the irrelevant values were muted by coloring them grey.	
	Basal and carb ratio settings	Vertical bars displayed across a timeline; values shown through text labels The basal settings relevant to the episode were highlighted using a bright color and the irrelevant values were muted by coloring them grey.	These were placed below the stacked time series to establish associations between the time duration of the settings and the time duration of the episode.
	Insulin distribution	Pie chart with labels for different types of insulin (basal, bolus, auto-corrections) and the proportion values	
Interpretive frames	Clinical or patient-specific contextual knowledge extracted from	Text	These were placed before the data displays

	expert-labelled cases – e.g., high after meals implies issues with carb ratio, etc.		

Table 15: The design choices and rationale for the individual components of an episode-driven data review tool for Type 1 diabetes.



Type of data	Transformations performed on data	Format of visualizations
BG	Individual values	Line chart - single, Line chart - multiple color coded lines, Line chart with point overlays, color coded table, scatter plot, calendar chart, area chart
	Averages and median	Line chart - single, Whisker plot with point overlay, table, scatter plot, calendar chart
	standard deviation, iqr, max/min, CV, quartiles (10, 90, 25, 75),	ribbon/stream graph, table, range graph, box plot/whisker plot
	# of bg readings	table, calendar chart
	distribution of bg numbers	pie chart, single stacked bar chart, hosrizontal bar chart, donut chart
	number of hypoglycemic events	table
	likelihood of a low, variability below median, median compared to goal	Icons to show high, medium or low
Carbs	individual data points	Labels on a horizontal timeline, labels in a circle on top of a bar, scatter plot of meal icons (timexvalue)
	average	table
	total	table
	meal events/time windows	vertical lines on a BG line chart, marks overlayed on BG line chart, labels on a horizontal timeline
	skipped meal (based on user configured meal time windows)	zigzag line
	# of carb entries	table, labels
Basal insulin	scheduled basal delivery	Area chart, solid line chart overlayed on a calendar background, table, labels on a timeline
	temp basal	broken line
	interventions to schedule - basal suspension and temp basal event	A calendar with days of basal suspension marked by a solid circle, labels on area or line chart, table
	average	table
	number of basal doses	table
Bolus insulin	Individual values	bars with labels, table, line chart
	bolus distribution	pie chart, table
	average	table
	total	table
	number of bolus doses (total or average)	table, A calendar with each cell containing dots representing a bolus, labels
	bolus events	vertical line on a BG line chart, table
	Active insulin time	Curve following a bar representing bolus insulin
Basal+Bolus	Insulin distribution	pie chart, table, stacked bar chart, donut chart, table
	Insulin trend	line chart overlay
	Settings	Table, Color coded bands on a horizontal timeline

Figure 27: The display formats for different types of data that emerged from the review of commercially available visualizations (Chapter 5)

The design rationale for visualizing different components (episode, episode characteristics, etc.) are as follows:

- Formats for different components were selected depending on the data or the insight to be presented [142]. The review of existing tools from chapter 6 (Figure 27), literature on review of different patient-facing visualization formats [2,171], and evidence on effective task-visualization mappings from prior work [17] informed the choices here.

- Formats for components were selected to maintain consistency within components to the extent that was possible [142,143,175].

- Because prior work has shown that patients prefer complete information using simple designs [51], the data interface aimed to balance information thoroughness and simplicity of presentation. Simple displays were favored over relatively more complex displays.

8.4.4 Design Choices and Rationale for Connecting Different Interface Components

To organize and communicate different components, I sought guidance from research on narrative visualizations creatie visual narratives by bringing the different components together (Hullman et al., 2013; Segel & Heer, 2010; Stolper et al., 2016). Additionally, prior work on designing multiple visualization systems provides guidelines to design for consistency and connectedness of different visualizations (Wang Baldonado et al., 2000). I now describe my design rationale for connecting individual visualization components to create a narrative and communicate it.

- Visualizations corresponding to different types of data on episode, context and interpretive frame needed to be sequenced. Visualizations can be sequenced to surface different types of associations - question-answer association, chronological association, causal association, granularity related association (overview to details), comparison invoking association, or spatial association (Hullman et al., 2013). These associations need to be self-evident and perceivable through annotations or visual cues (Wang Baldonado et al., 2000). Given the context of problem-solving, I used causal associations – problem/episode-cause-evidence-suggestions, to organize the visualizations.

- To maintain user’s orientation during screen transitions, I ensured that the different screens share visual context to the extent possible to make the connections across the screens evident (Hullman et al., 2013). For example, different types of data were presented using the same time scale. The episode was presented in the same format across all screens.

- To make data easy to understand and to limit reliance on the user's data literacy skills to interpret the data, I supplemented the visualizations with simple text annotations to communicate gist information to the patients (Desai et al., 2018). The interfaces were designed to yield an accurate meaning when interpreted literally through annotations (Arcia et al., 2016). That is, reading the annotations alone could enable a user to understand the information.

On a more detailed level, one of the foundational works on narrative visualizations [156] helped me identify and implement design dimensions using concrete design features. Table 16 lists these design dimensions, their definitions, and how they were implemented in the prototype.

Design Dimension	**Definition**	**Implementation Choice**
Genre (type of visual narrative)	Taxonomy of types of visual narratives – magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show and film/video/animation	a mix of slide show, chart with text descriptions, and partitioned poster
Visual structuring	"mechanisms that communicate the overall structure of the narrative to the viewer and allow him to identify his position within the larger organization of the visualization" [156].	carousel cues with dots or lines
Highlighting	"visual mechanisms that help direct the viewer's attention to particular elements in the display" [156].	color coding and text descriptions to guide attention – specific sections of the timeline highlighted; specific data points highlighted

Transition guidance	"techniques for moving within or between visual scenes without disorienting the viewer" [156].	shared context across screens – same timeline, a time window of interests highlighted across screens
Ordering	"ways of arranging the path viewers take through the visualization" [156].	Linear and hierarchical
Interactivity	"ways a user can manipulate the visualization (filtering, selecting, searching, navigating), and also how the user learns those methods (explicit instruction, tacit tutorial, initial configuration)" [156].	left and right tap in the carousel with current position highlighted
Messaging	"ways a visualization communicates observations and commentary to the viewer" [156].	Text descriptions/annotations and labels

Table 16: Design choices made for narrative creation following design dimensions laid out by Segel and Heer

8.4.5 Potential for Algorithmic Story Generation – Episode-Driven Identification and Display of Relevant Data

I created visualizations from patient data using python libraries and then arranged them into an interactive mobile prototype using Figma. To build a system for episode-driven narrative generation, one approach is to algorithmically identify different components – episode, episode characteristics, episode context, and interpretive frames to further automate this process as indicated by prior work. While this is challenging, particularly for identifying potential causes, it is feasible. For example, combining machine learning to predict glycemic outcomes with a template-based expert system has been used to generate nutrition recommendations [64]. Additionally, algorithmic techniques to identify causal moderators in time series data can be used to identify potential factors influencing glycemic patterns [191,192]. These prior works establish the feasibility of algorithmically identifying different components.

Another approach to identify episode-driven narratives is to use episode-specific templates developed in collaboration with clinicians as the basis. For instance, for a system that utilizes machine learning for episode detection, given the episode detected, episode-specific templates can enable the selection and visualization of episode-specific information. For instance, Table 17 presents a list of episodes that can be mapped to different potential causes or outcomes and potential actions. These can be expanded to generate detailed templates, such as that shown in Table 18, which presents a template for an episode of high blood glucose in the afternoon. Given an episode, the template can be used to specify potential causes, events of interest, insights to show, annotations, display format for the insights, and the potential solutions for the episode.

Disease-specific phenomenon or behavior (episodes)	**Interpretive frames**	
	Potential causes or outcomes	**Potential action**
High BG during the day	Carb ratio for one or multiple meals is not sufficient	Increase the insulin to carb ratio for meals that pose a problem
High BG overnight or early morning before eating	Basal insulin overnight is not sufficient	Increase the basal insulin overnight
	Eating uncovered food at bedtime	Cover bedtime food with insulin or avoid bedtime food
High BG after giving correction insulin	Correction factor is not sufficient	Increase the insulin taken for every unit difference in BG number from the target
Repeated administration of correction insulin	Basal insulin is not sufficient	Increase basal insulin in time windows when correction is needed
High BG throughout the day	Basal insulin is not sufficient	Increase basal insulin for the day
	Food bolus is being missed	Improve adherence to boluses
Low BG during a specific time of the day	Overcorrection with insulin	Avoid correction boluses too often
	Physical activity could cause lows	Take uncovered carbs before physical activity
Missing carb entries	Less bolus insulin than expected and potentially high BGs	Enter carb data in pump
Missing food boluses	Less insulin than expected and potentially high BGs	Take bolus insulin for food

Missing BG entries	Less insulin than expected and potentially high BGs	Enter BG data in pump

Table 17: Potential combinations of outcomes-causal behaviors and behaviors-subsequent outcomes mapped to potential corrective actions.

Episode	Potential Cause	Data/Event of Interest	Insight to show	Annotations	Format of display	Potential solution
High BG after lunch (2PM-6PM)	Carb ratio for one or multiple meals is not sufficient	BG numbers from t1 to t2	High BG pattern from t1-t2 (more than 50% of BG numbers are high)	You have a <episode type> BG pattern from <t1> to <t2>	Line chart	Increase carb ratio at lunch
		Carb and insulin at t3, which is right before (t1 minus 2 hours) or during the time window t1-t2.	Nature of BG numbers at t3.	On <date 1>, you ate at <t3>. After <t4-t3> hours, your BG number changed from <BG at t3> to <BG at t4> despite taking a bolus.	Line chart overlay with green marks showing normal BGs and red marks showing high BGs	
		BG number after the above-selected carb and insulin events at t4	Nature of BG numbers at t4.		Line chart overlay with red or green marks	
		Carb ratio during the lunch time window of t1-t2	Description of carb ratio	Your carb ratio during <t1> and <t2> is <value of carb ratio>	Bars on a timeline	

	Missing lunch bolus	Carb only entries at t3, which is before or during the time window with incidence of high BG numbers.	Carbs were consumed without insulin	On <date 1>, you ate at <t3>. You did not take insulin at t3. After <t4-t3> hours, your BG number changed from <BG at t3> to <BG at t4>.	Events on a timeline using circles to mark the events	Cover carbs

Table 18: Example of a template specifying different types of information for an episode.

8.5 RESEARCH QUESTIONS

By qualitatively evaluating a prototype (i.e., episode-driven narratives) embodying the episode-driven sensemaking framework, I aim to address the following research questions:

- What is the feasibility of the episode-driven sensemaking approach in facilitating structured engagement and decision-making with multiple streams of data?
    a. How does a prototype embodying the episode-driven sensemaking framework compare with the commercially available data platforms?
- What is the feasibility of episode-driven narratives in facilitating review of multidimensional health data for identifying self-care actions?
- How can we design mobile data interfaces to help patients navigate different types of data, identify problems and potential self-care actions to address those problems?
    a. In what ways narratives or data-driven stories enable or limit interaction with multidimensional data?
    b. How can we design for the presentation of data narratives from complex multivariate health data to aid decision-making for self-care?

The answers to these questions are important for generating design insights to guide the development of patient-centric tools supporting data-driven management of chronic health conditions. Managing a health condition based on data is characterized by regular data interaction for assessment, sensemaking, and decision-making. The episode-driven sensemaking framework provides a basis to facilitate and investigate interaction with health data to begin to answer the above research questions.

8.6 METHODOLOGY

The objective of the study was to understand how T1D patients interact with data from continuous glucose monitor (CGM) and insulin pump (data on carbs, different types of insulin) as presented through two platforms - the commercially available data review platforms and the episode-driven narratives embodying the episode-driven sensemaking framework. Specifically, I wanted to understand how patients come up with plans for changing their behaviors or insulin doses based on the glycemic patterns and their potential causes observed in the data.

The study involved the following phases: recruitment and enrollment, prototype development, commercial data platform review and evaluation session, and research prototype review and evaluation session

8.6.1 Recruitment and Enrollment

The study information was made available to patients on the website of the University of Michigan pediatric endocrinology clinic and on the health research platform of the University of Michigan (https://umhealthresearch.org). Patients who expressed interest through either of the two platforms were contacted to assess eligibility. Eligible patients were provided with the informed consent form. Participants who signed the form were considered enrolled and were contacted for the next steps. Enrolled patients also filled a baseline survey to ensure eligibility and to gather demographic and general information about their diabetes management (Appendix C). The inclusion and exclusion criteria for recruitment were as follows:

Inclusion criteria

- Years since diagnosis > 2
- Uses a continuous glucose monitor and an insulin pump

- Has a Dexcom Clarity account and an account on another platform to view insulin pump data.
- Had downloaded data during their last clinic visit and had it reviewed by a certified diabetes educator or an endocrinologist
- Patient and the parent/guardian are willing to complete all study activities and are willing to provide consent for the use of their data to develop prototypes
- Has a laptop and a mobile phone with internet connection and a camera on both the devices
- Patient or parent/guardian is willing to download and send diabetes data when required by the study activities
- Patient and caregiver can speak English

Exclusion criteria

- Patients who do not have a Dexcom Clarity account or an account for viewing the insulin pump data.
- Patients and parent/guardian who do not speak English.

8.6.2 Prototype Development

Prototype development involved patient data collection, data annotation by a clinician, patient data analysis (to identify episodes, episode characteristics, episode context, and interpretive frame), prototype development, and questionnaire development.

Patient data collection

Data from enrolled patients was obtained from their medical devices (CGM and insulin pump) and their electronic health records. These data included raw data from the CGM and insulin pump over 90 days from the date of their last clinic visit, the PDF reports downloaded from the commercial data platforms and reviewed in the last clinic visit, and the corresponding clinical notes.

Data annotation by a clinician

The PDF reports from different patients was reviewed by an endocrinologist (Dr. Lee) to obtain an interpretation, that is in addition to the interpretation documented in the clinical notes. The PDF reports do not provide any interpretation and only provide data through different types of visualizations. Providers review the data and record their interpretation in their notes. As a part of their interpretation, they note the glycemic patterns, their potential causes, and the changes that were recommended.

I chose to obtain data from the clinic visits as it provided an additional interpretation done together by patients and clinicians in the form of clinicians' notes in the EMR. These notes provided a source for the ground truth for the insights contained in the data. The PDF reports were used as a comparison platform. The raw data was used to reproduce information in the PDF reports and clinicians' notes and present them in the prototypes.

Patient data analysis for episode module creation

The interpreted reports and the raw data from diabetes devices were used to create different cases for presentation through a mobile prototype. I refer to these cases as *episode modules*. Specifically, for each glycemic pattern or episode that was noted by the experts, creation of an episode module involved identifying and documenting the following information:

- episode - the disease-specific phenomenon or behavior (e.g., a pattern of low blood sugar overnight, pattern of insulin administration overnight),
- episode characteristics - the severity and/or the strength of the phenomenon or behavior (e.g., number of days it happened),
- episode context - the events surrounding the episode linked to the potential causes of an outcome (e.g., low BG caused by physical activity in the evening) or outcomes of a behavior (missing lunch insulin caused high BG at dinner),
- interpretive frame - the potential causes of the episode mapped to data from the devices and the actions that could be taken (e.g., overnight high BGs are typically due to low basal rates and can be fixed by increasing overnight basal insulin)

The identification of the episode modules was done manually using clinicians' interpretations of the data. Text descriptions to accompany different components of the episode modules were created manually. Different components of the episode modules were then visualized using the raw data from the CGM and the insulin pump and using python libraries. These visualizations were then brought together in a mobile prototype on Figma.

In total, PDF reports from 9 patients (P01, P03, P04, P05, P07, P08, P09, P10, P12) were interpreted by Dr. Lee, in addition to the interpretation available from the electronic medical records. From these, a total of 7 episode modules were identified for creating the prototypes. These modules involved data from five patients (P01, P04, P05, P07, P09). The episode modules were selected for prototype development to ensure diversity of the disease-specific phenomenon, potential factors, and potential solutions being represented. Table 19 presents the different episode modules.

Episode Type	**Episode Time Window**	**Time period – CGM Data**	**Time period – Pump data**	**Factors**	**Patient ID**
High BG [afternoon]	12 pm – 2:45 pm	May 25 – May 31	May 25 – May 31	Morning carb ratio	P01
High BG [night]	8 pm – 12 am	July 8 – Aug 6	Aug 02 – Aug 8	Evening carb ratio Frequent meals	P05
High BG [afternoon]	1 pm – 4:30 pm	July 8 – Aug 6	Aug 02 – Aug 8	Morning/Afternoon carb ratio	P05
High BG [evening + night]	7 pm – 12 am	May 18 – May 31	May 18 – May 31	Evening carb ratio Bolus timing Missed bolus Basal insulin/ Basal bolus distribution	P07

Low BG [afternoon + evening]	2 pm – 7 pm	May 18 – May 31	May 18 – May 31	Bolus timing Auto corrections Morning/Afternoon carb ratio	P07
Low BG [evening + night]	7 pm – 11 pm	June 6 – June 19	June 6 – June 19	Evening carb ratio Insulin intake without BG Physical activity	P04
High BG [overnight + morning]	12 am – 9 am	Aug 17 – Aug 30	Aug 17 – Aug 30	Basal insulin/ Basal bolus distribution	P09

Table 19: Summary of episode modules used to create the prototypes.

Prototype development

The raw data from CGMs and insulin pumps corresponding to different episode modules was used to create visualizations using python libraries. The format of these visualizations was similar to some of the visualizations present in the commercially available PDF reports to avoid introducing new formats that the patients might not have seen before. The prototype visualizations aimed to reproduce the information in the PDF reports to the extent that was possible. The prototype visualizations were then put together in an interactive mobile prototype built on Figma to embody the episode-driven sensemaking framework following the design choices described in section 8.4. The information in the prototype visualizations had minor differences from the information in the PDF reports. For instance, some of the elements that are only visually presented in the PDF reports (10th % - 90th % range) were quantified and displayed as numbers in the prototype visualizations. Similarly, the time duration of the glycemic patterns had small differences from the PDF reports because the prototype visualizations used the median instead of the average as the criteria for visually representing high BG patterns. This was based on the information from clinicians that 50% or more values being high constitutes a high BG pattern. For low BG patterns, the 5th percentile curve was used.

Questionnaire development

In order to assess patient's comprehension of the data as seen through different platforms, and to assess decision-making based on the data, questionnaires were created corresponding to each episode module. For each episode module, there were two versions of the questionnaire, one corresponding to the PDF reports and one corresponding to the prototypes. Given the slight differences in the information across the two platforms, as noted above, I could not use the same questionnaire. The questionnaires asked about the following: patterns observed, potential causes, potential care actions, the most evident cause, and the most likely care action. Appendix D shows the specific questions asked.

Episode modules were assigned to participants to balance the number of participants across the available modules. The aim was that each episode module should be seen by the same number of participants. Additionally, I manually assigned episode modules to ensure that each participant sees variations across the time duration of episode, the type of episode, and the number of factors associated with the episode.

Participants did not see their own data in any of the sessions. They reviewed other participants' data. The reason behind this choice was to ensure that all the participants only see data that is unfamiliar. Not all participants' data contributed an episode module for prototype development, which meant that not all participants could be shown an episode module extracted from their data. Showing them data from other participants ensured that every participant was unfamiliar with the data that they are shown. Across the two sessions, participants saw the same data/episode module to enable comparisons. In the session to review the prototypes, participants saw additional episode modules to enable the in-depth assessment of the prototypes.

8.6.3 Commercial Data Review Platforms: Review and Evaluation Session

After the creation of the prototypes, participants were invited for the first session to review the PDF reports from the commercial data review platforms. In this session, participants began by taking two standard assessments – the diabetes numeracy test [80] and the graph literacy assessment [65]. After these, participants had the option to take a 5-minute break, but all the participants wanted to continue without a break. Next, participants reviewed the PDF report (CGM and insulin pump report) focusing on the assigned episode module and completed the questionnaire corresponding to that episode module. While reviewing the report and answering questions, participants were instructed to think aloud. After completing the questionnaire, participants were asked to take the cognitive load assessment [71] while also thinking aloud. The session concluded with a semi-structured interview to ask participants about their overall experience reviewing data and to probe them on specific aspects observed during the data review session (instances when they seemed to struggle or when they seemed surprised). I took notes during the think aloud session. I also asked participants about their general practices of reviewing diabetes data for self-care.

8.6.4 Episode-Driven Narrative Prototypes: Review and Evaluation Session

In the prototype review session, participants began with a simple episode module (one episode mapped to one potential cause). This episode module was presented to all the participants as a tutorial. Before they began, they were informed about ways to interact with the prototype and select the size of the display appropriate for them. Like the previous session, they were asked to think aloud while reviewing the prototype and completing the questionnaire. Once they completed the questionnaire for the first episode module, they took the cognitive load assessment. After that, participants engaged in a short open-ended interview to share their experiences of the prototype. Next, participants were shown at least one or at the most two more episode-modules depending on the time it took them to finish the activities related to each module. Overall, I did not want the session to exceed 90 minutes, as outlined in the informed consent. The next episode module that each participant saw was the same as the one they worked on in the previous session with the PDF reports. This enabled comparison of participants' experiences and scores on the questionnaires across the two sessions. Two participants saw two modules and the remaining four saw one module (in addition to the tutorial episode module). For each module, participants completed a questionnaire and the cognitive load assessment, and engaged in a short debrief to talk about their overall experiences. At the end, I conducted a semi-structured interview to ask participants more comprehensively about their experience reviewing the data, to understand how they compare the experience in this session with the previous session of using the PDF reports, and to probe on specific aspects of the data review session. To ensure that participants have little memory of the data from the previous session, the prototype review session was conducted two weeks after the first session except for P04 who had time constraints.

All the sessions were conducted on Zoom. Participants used their laptop for both the sessions. All the sessions were video recorded and were transcribed using the live transcription feature of Zoom. The semi-structured interviews in both the sessions focused on understanding to what extent and how participants engaged in the different analytical phases of the episode-driven sensemaking framework (episode detection, episode elaboration, episode classification). Additionally, participants were asked about positive and negative aspects of both the platforms.

8.6.5 Data Analysis

For each participant, data was analyzed in multiple steps. I first scored their questionnaire responses for the episode module that was common across the two sessions for each participant. Next, I looked at the videos to make detailed notes about their responses to each question on the questionnaire. I noted their thoughts and identified why they selected certain answers and which questions seemed particularly challenging. The responses related to the same episode module were compared across the two sessions to identify differences and similarities. Lastly, I analyzed the transcript for each session guided by the episode-driven framework to identify ways in which participants engaged with the analytical phases described by the framework. I performed descriptive coding to identify instances of episode detection, episode elaboration, episode classification, and episode-specific recommendation generation (wherein an episode represented a glycemic pattern present in the data). The data specific to each instance was then analyzed and coded across participants using in-vivo coding to label the types of information participants used for specific activities, the challenges they faced, the suggestions they provided for improving the design, and the differences in their ways of interpreting the data from the previous session. Additionally, I coded participants' descriptions of the helpful and unhelpful aspects of the different platforms using in-vivo coding.

8.7 FINDINGS

In this section, I first present details about the participants who completed the study followed by their scores on standard assessment of diabetes numeracy and graph literacy. I then present their scores on the questionnaires and the task load scores from the two sessions.

8.7.1 Participants

14 people enrolled in the study from which 6 went ahead with the study and completed both the sessions. Among the 6 participants there were 5 adult patients and 1 caregiver of a pediatric patient. All the 6 participants were White. 3 participants were employed full-time and 3 were not employed. Table 20 presents the details about each participant.

ID	**Age**	**Gender**	**Education**	**Years with T1D**	**CGM**	**Insulin Pump**	**Data Review Platform**
P01	45	F	Bachelor’s degree	40	Dexcom G6	Tandem t:slim	Dexcom Clarity,t:connect
P04	19	M	Some college (student)	7	Dexcom G6	Tandem t:slim	Dexcom Clarity,Dexcom Share
P05	21	F	Some college (student)	14	Dexcom G6	Tandem t:slim	Medtronic Carelink/Connect,Dexcom Clarity
P09	49	M	Master’s, Professional, or Doctorate degree	3	Dexcom G5	Tandem t:slim	Dexcom Clarity
P10	39	M	Bachelor’s degree	37	Dexcom G5	Tandem t:slim	Dexcom Clarity,Dexcom Share
P13	18	M	Some college (student)	5	Dexcom G6	Tandem t:slim	Dexcom Clarity,Dexcom Share

Table 20: Participants’ details

8.7.2 Scores on Standard Assessments, Questionnaires, and Task Load Assessment

Table 20, 21 and 22 present the different scores. The average scores obtained on the questionnaires completed with the prototypes was higher than the commercial PDF reports (Table

22). The average task load of completing the questionnaires with the prototype was lower than completing the questionnaires with the commercial PDF reports (Table 23). P04 and P13 had relatively higher diabetes numeracy and graph literacy (the highest that was scored, see Table 21) and obtained full score on the questionnaires with both the platforms. They both had a lower task load score with the prototype. The scores presented in the tables correspond to the same data set reviewed through the two different platforms about 2 weeks apart except for P01.

Participant	**Diabetes Numeracy Test**	**Graph Literacy Assessment**
P01	73	85
P04	87	92
P05	87	77
P09	80	62
P10	67	77
P13	87	92

Table 21: Diabetes numeracy and graph literacy scores.

Participant	**Score based on commercial PDF Reports**	**Score based on episode-driven narratives**
P01	60	90
P04	100	100
P05	83	100
P09	56	67
P10	21	82
P13	100	100
Mean score	**70**	**90**

Table 22: Participants' scores for questionnaires completed with the commercial PDF reports and the episode-driven narratives. The scores correspond to the same data set reviewed through the two different platforms about 2 weeks apart except for P01.

Participant	**Task load with commercial PDF Report**	**Task load with episode-driven narratives**
P01	79	49
P04	49	25
P05	53	36
P09	33	28
P10	81	42
P13	57	41
Mean score	**59**	**37**

Table 23: Participants' scores for task load across commercial PDF reports and episode-driven narratives. The scores correspond to the same data set reviewed through the two different platforms about 2 weeks apart except for P01

Next, I describe the qualitative findings from the analysis of participants' experiences with the two platforms. I first describe how the episode-driven narratives performed as compared to the commercial PDF reports. I then report the rest of the findings in themes that correspond to the different components of the episode-driven sensemaking framework. These include episode detection, episode elaboration, episode classification, and episode-specific recommendation generation. Lastly, I describe four dimensions in which the information and decisional needs of the participants varied.

8.7.3 Episode-Driven Narratives Compared to Commercial Data Review Platforms

Episode-driven narratives enabled better navigation of data, better interpretation of glycemic patterns, and improved comfort in making plans for self-care

Episode-driven narratives enabled better interpretation of blood glucose data by combining data from insulin pumps and continuous glucose monitors. P01 and P04 compared the prototypes with Dexcom Clarity's pattern analysis (Table 23, row 3), which they noted does not provide insulin pump data required to interpret the patterns. P04 mentioned, "*Dexcom clarity can analyze patterns for like weeks or months I think, but it's less detailed than this. It doesn't take into account your pump boluses [insulin] and things like that*."

When talking about the commercial reports, P01 mentioned,

> *"[They need to] do better at helping you easily understand things about your data, things like what patterns are going on. And what would be done to fix those patterns. When you run the clarity report, it says, you know pattern detected, you know we detected you are having highs between 9 pm and midnight, whereas I don't really see that in the tandem [insulin pump] application."*

The existing reports do not support the interpretation of patterns using the insulin pump data because the patterns and the data from the insulin pumps are visualized separately. The patterns presented by the CGM report may not be presented by the insulin pump report even though the two platforms share information.

Episode-driven narratives made it easier to navigate the different types of information compared to the commercial reports. With the commercial reports, P05 noted that it was challenging to go through the data without any guidance and did not like the amount of information shown in the pump reports, "*It can be like very overwhelming to have all of this data without, like a very clear way to go through it, unless you're like pretty familiar with going through you know, different data graphs and stuff like that… I don't like how the thing with the pump report is just like I feel like they're not that easy to read, and they're kind of cluttered, and I think it should be better by now to be honest with you.*" With episode-driven narratives, she liked that it listed out the issues which could guide people, "*it's like listing out the issues that might be really helpful for someone who's kind of like feeling really lost about what to do*". She also described that the episode-driven narratives reduced the mental demand involved in figuring out the patterns and the causes, "*the mental demand was really low like I just think that by pointing out all the reasonings behind the patterns, it did like pretty much most of the mental work for me. Normally like a lot of the mental demand for me is like figuring out these patterns and, like the causes of these patterns myself.*"

Lastly, participants felt more confident and comfortable about coming up with plans for action with the episode-driven narratives. For P04, having the analysis and suggestions supported with examples from the data made him more confident about the changes he thought are required, *"[I'm] extremely confident because I know there's a lot of data backing it up.*" P10 mentioned being more willing to make changes based on the episode-driven narratives as compared to the

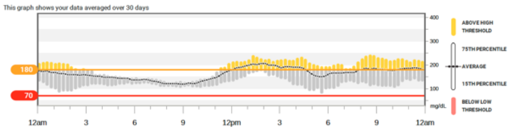
PDF reports, "*These [episode-driven narrative interfaces] are a lot easier to understand. I'm a fan of these. I would probably be more susceptible to making these changes myself from these. I felt a little more comfortable with this data, making the decisions because there wasn't so much and it was kind of bite sized data, really tangible. Everything was kind of at one screen I didn't have to cycle through report.*"

In the subsequent sections, I present the findings related to the different phases of the episode-driven sensemaking framework (episode detection, episode elaboration, episode classification, episode-specific recommendation generation).

8.7.4 Episode Detection

Participants missed to note episodes (blood sugar patterns) when using the commercial PDF reports, particularly those related to low blood glucose

When reviewing the commercial reports, participants often did not note/record all the patterns that were present. Different participants used different displays from the Dexcom Clarity report (CGM data report) to identify the patterns. These include the standard day overview, the daily overlays, and the pattern analysis section (Table 24).

Dislpay	**Description**
	The standard day overview of blood glucose numbers aggregated over several days. The black line trace shows averages. Each bar shows the 25th – 75th percentile range for 15-minutes time windows over 24 hours. Section of the bars above threshold are yellow.

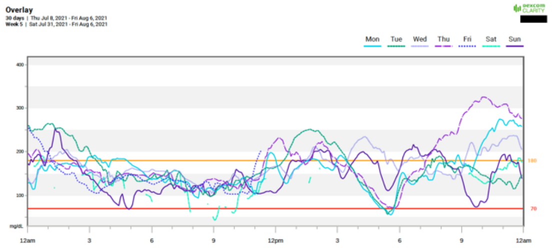
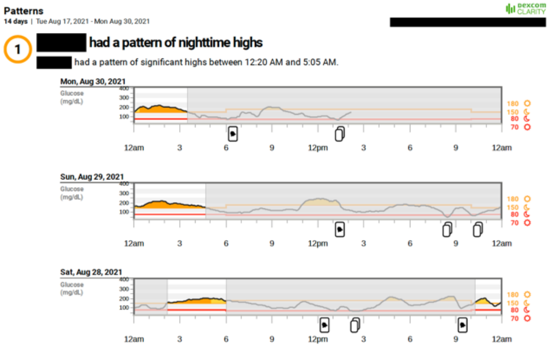
 Overlay 30 days | Thu Jul 8, 2021 - Fri Aug 6, 2021 Week 5 | Sat Jul 31, 2021 - Fri Aug 6, 2021 [redacted] 	The overlay view of blood glucose numbers showing blood glucose curves of individual days in different colors.
 Patterns 14 days | Tue Aug 17, 2021 - Mon Aug 30, 2021 [redacted] 1 [redacted] had a pattern of nighttime highs [redacted] had a pattern of significant highs between 12:20 AM and 5:05 AM. 	The pattern analysis section of Dexcom Clarity showing a nighttime high pattern.

Table 24: Displays used for identifying patterns in blood glucose data.

P01 used the overlays and identified only one pattern among the two that were present in the data. She found it easier to look at the daily overlays to identify patterns as compared to the other displays in the CGM and the insulin pump PDF reports. With the prototype, she was able to note both the patterns.

P10 used the pattern analysis section of the Dexcom Clarity report and noted only the most severe pattern that was presented. He did not record the other less severe patterns. With the episode-driven narratives, he recorded all the patterns that were presented.

The low BG trends displayed in the standard day overview were not noted or considered by several participants (P05, P09, P10, P13) despite the color coding that is used to denote low patterns and despite the general understanding that low blood glucose numbers need to be handled with more urgency than the high blood glucose numbers.

There are several potential reasons participants did not observe or disregarded the episodes of low blood glucose. First, when looking for BG patterns in the commercial PDF reports, participants used the average blood glucose curve. The average curve going above the high threshold was labelled by participants as a high BG pattern. Going by this rule, a low BG pattern would mean the average curve going below the low threshold. However, the average curve cannot be used to identify patterns of low BG because low BGs are a lot less frequent than high BGs. Hence, about 5% to 10% of BG numbers being below the lower threshold is considered a low BG pattern and worth investigation. Since participants seemed to not have as clear a criterion to differentiate the two types of patterns (high and low) and label low BG patterns, they perhaps failed to observe the low BG patterns. Second, regions of high BG are presented in yellow color and regions of low BG are presented in red color in the Dexcom Clarity standard day overview display. High BG numbers being more frequent than low BG numbers, the display typically had a lot more yellow regions than red regions. Consequently, participants eyes were primarily drawn to the relatively more dominant regions of yellow color that showed high BG values. There is a need to better emphasize low BG patterns given their infrequent prevalence and serious implications.

8.7.5 Episode Elaboration

Elaboration to identify potential factors and care actions using commercial reports involved more guesswork as compared to the episode-driven narratives prototypes

When reviewing data using the commercial PDF reports to identify potential causes of the patterns, participants made guesses based on their general knowledge about all the different factors that impact diabetes without necessarily looking at or considering the insulin pump data that was available. They tended to think very broadly about every possibility perhaps because of the multiple factors that influence a condition like diabetes and the difficulty of finding specific evidence from the reports. For instance, when asked to select potential causes and action about a particular episode, P05 did not look at the pump data at all. She only looked at the CGM curve and hypothesized about what could be happening. She went through the options one by one and selected every possible cause without considering any evidence.

> P05 with the PDF reports: "*I'm just gonna go through one by one and click on the ones that I think are correct*".

With the prototype, she did not make as many guesses and selected a fewer number of responses. Having a narrative structure to navigate the data, having fewer days of data to look at in the form of episode-specific examples, and having the data features visually highlighted and described in text improved P05's consumption of the data.

> P05 with the prototype: *"[reads the answer choice] not taking the meal bolus. Well, wait let's see actually. OK, so the six of the seven days are contributing to the blood sugar pattern. I think that I'm actually not going to include not taking a meal bolus, because it says in here they recorded the food and bolus at 11:45 am.*"

Similarly, P01 engaged in guesswork when asked about potential causes without looking at the data. She selected every possible cause from the list of presented options. Specifically, she made assumptions about missed boluses and carbs without looking at the pump data, perhaps because data related to patients' behaviors is spread across several displays of individual days making it harder to visually scan for evidence that matters. Later, she had trouble finding meal information form the pump reports. With the episode-driven narratives, having looked at an example day that showed pump and CGM data representative of the issue being narrated, she was sure about a few types of information, such as the patient took a bolus and recorded carbs. She did not select those as the potential causes, "*looks like for the rest of the time, the person is in range, except for that spike [looking at the BG pattern highlighted]. [Clicked to open the detailed information about the factors listed for the BG pattern] entered in a bolus, entered in their meals, so they didn't skip a meal didn't skip a bolus there. They are entering in carbs, and they are doing insulin. I would say that their meal carb ratio is off for breakfast."*

Elaboration with commercial reports involved mapping data features that represent potential problems and causes

Participants who scored well with the commercial reports (P04, P13, P05) were able to better map data features to lived experiences and hence, had a better understanding of which data features represent which type of issue or what the data should or should not look like for a specific issue. For example, P05 looked at the rise in the BG curve before an insulin intake event to infer that the patient might be blousing [taking food insulin] late, "*it still looks like they might be bolusing late because it's already spiking before they get insulin sometimes."* In the same context of the timing of the food insulin boluses, P13 described how taking the insulin sooner would not have caused the BG peaks that were present, "*So if they injected [insulin] like slightly earlier then you wouldn't see as much of these peaks up here.*"

For the issue of carb ratio insufficiency, participants looked for the BG curve in relation to a meal event. If the BG curve crossed the normal range after a meal and stayed up for a while, it was an indication that the carb ratio might be insufficient. Similarly, a failed pump site would show above range BG curve for a longer time duration unless the patient realized it and changed the pump site. While the standard error was hard for the majority of the participants to relate to, P09, who was a diabetes researcher found standard error helpful in knowing if the pattern was a systematic problem, "*The standard error for me is critical because it tells me whether it's a systematic problem or once in a blue moon issue that needs to be addressed in a different way. So like a larger variability would mean that it's less of a systematic problem*".

Not all participants were able to identify and leverage particular features of the data depicting the presence or absence of an issue.

Elaboration with commercial reports involved back and forth based on users' ability to find meaningful evidence (or data features) and the ease with which the evidence could be found

In most of the cases, more than one factor was involved in affecting the blood glucose numbers. When reviewing data to identify different types of factors and to select the few most compelling factors, participants went back and forth on what could be the potential causes and interventions for a particular pattern. P13 and P09 changed their minds several times as they scrolled through the reports while selectively attending to different types of evidence. For instance, P09 was analyzing the BG curve in relation to the meal and insulin intake events. When investigating the potential causes, he did not think that the insulin timing was an issue based on the evidence that came to his attention on first look. Later, when thinking of possible changes to make, a second look at the same evidence made him change his mind to include improvement in bolus timing as a behavior change recommendation.

> P09, when thinking of the causes, pointed to specific days to establish that bolus behavior is appropriate, *"How does it spike [opens the CGM report standard day overview to check the CGM curve]. So, they have it around 8 [goes back to insulin pump report], not always [talking about mealtimes for May 18, 19, 20, 21, 22, 23]. The question here is do they not take bolus before their meals? I don't think so because here [looking at May 21] for instance I see carbs and it is not spiking immediately after. Here I see carbs, it does spike a*

> *little but that is a one-time thing. Here [looking at May 23] I see carbs but they are not spiking after, so I suspect that they do pre-bolus.*"

> Later in the process when thinking of the solutions, P09 pointed out the potential to improve bolus timing, "*I am looking at the curves to see how long.. I mean whether it goes down too much before spiking up and I don't see that trend. I see that when they take carbs they never go much lower so they could pre-bolus more [looking at May 19 and May 22]"*

Different ways in which the BG curve behaved after a meal and insulin intake event only came to P09's attention after spending enough time with the reports and scanning the data from all the individual days twice. Users who do not spend as much time and do not know what data features to look at would not be able to come to similar conclusions. This also implies that the ease with which a particular type of evidence or a data feature comes to users' attention would influence the decisions they make.

Elaboration with episode-driven narratives involved utilizing episode-specific evidence and episode-specific domain knowledge to make informed choices from different factors

When understanding the list of factors potentially causing a pattern, the narratives presented to explain the factors through examples from data contextualized with domain knowledge enabled participants to make choices in nuanced ways. Participants chose which factor or combination of factors to consider in their decisions for care actions targeting a particular pattern based on different criteria. These included the number of days a factor was influential, the ease with which a factor could be worked upon, a factor for which there was evidence on what worked, and the standard deviation of the time duration in which a pattern exists.

P05 and P10 choose to work on the factor that was impacting outcomes on the greatest number of days. P05 mentioned,

> *"Even though evening carb ratio is listed first [in the list of factors] I'm going to pick basal insulin for 14 of 14 days, because that seems to have the highest impact.. on like every day. It's not just like occasional user error or something. I mean I guess everybody misses boluses [looking at missed bolus, another factor that was in the list] sometimes but I mean I*

guess if it's three of every 14 days that's not really ideal, but you know it happens so. . . I'm gonna [going to] say basal insulin just because it seems to be impacting every day, and also, it seems really far off."

She also considered the extent to which the factor deviated from the guideline. For instance, in this, case the basal insulin was only 25% of the total insulin [basal and bolus insulin] whereas according to the guidelines basal insulin should be 40% of the total insulin intake.

P01 chose to work on a factor that would have been easier for her to work with. She saw two factors – carb ratio and frequent meals. She chose to change the carb ratio instead of avoiding frequent meals because she had a condition which required her to divide her meals into smaller portions instead of eating it all at once, "*so I had ran into a problem where I couldn't bolus before my meals, because I was never quite sure how much I would be able to eat, so it was kind of like you bolus for half your meal and then, if you ended up finishing it, then you bolus for the other half*". Since there was only one carb ratio setting to change, that seemed easier than avoiding frequent meals, "*I might probably do the carb ratio, because it would seem like it would be the easier of the two. Because we have one carb ratio during that time so.*"

P04 and P13 chose the factor for which there was evidence about what worked. Between carb ratio and frequent meals, P04 chose to correct frequent eating behavior because the explanation for frequent eating showed a day where less frequent eating was associated with better BG numbers. Instead of relying on overall measure of the number of days a factor may have influenced outcomes, he chose to select a factor that had more concrete information on what worked, "*I was kind of more convinced by the fact that this had like an example of where you did what you would do to fix it and it worked. this was like concrete like a guaranteed way to fix it. Because you did the solution once and it worked.*"

For P09, the choice of factors was dependent the standard deviation during the time window in which the pattern exists. This information helped him decide whether the causal factor could be something systemic, like the insulin doses or could be something more variable, like the patient's behavior, "*if the standard error is tight, I know that a slight correction in a program [insulin dose setting programmed in the pump] will be enough to fix the issue*".

Unlike the PDF reports where participants had to identify potential causes, episode-driven narratives provided a list of potential causes with explanations for each cause. These explanations included examples of data and domain knowledge required to interpret the data. Participants found these helpful and relied on them in nuanced ways to make choices. These explanations provided an opportunity for participants to exercise their agency in making decisions.

8.7.6 Episode Classification

Classification with PDF reports relied on assumptions, did not involve engagement with the available evidence, and was done upfront

When looking at the PDF reports, 4 (P05, P09, P10, P13) participants registered the several glycemic patterns that were there but did not consider them noteworthy enough to select them as an answer or to analyze them further. They carried out pattern detection and classification simultaneously. In selectively paying attention to one pattern or the other, all these participants particularly deprioritized the low BG pattern in the data upfront in favor of the high BG pattern without looking at the detailed evidence, such as the potential causes of the low BG patterns or their frequency. This was particularly surprising as clinicians often prioritize low BG patterns for interventions. For P09, high BG patterns were more easily fixable than a low BG pattern because typically a low BG pattern comes with larger variability (implying a greater number of possible causes). He noted the pattern with the least variability and skipped the remaining two patterns that had higher variability when looking at the PDF report. For him, a pattern with lesser variability was more likely due to a systematic issue than the one with relatively higher variability, "*There is quite some standard error. You see the values on average are going higher and even with the standard error they are systematically high here (noting the pattern that seemed more severe). So, you know something needs to be done to address this here. I am less confident when you are just above average and with standard error you are going way down as well (noting the other pattern in blood glucose). It would be more risky to target this time frame with something specific than it would be to target this one. It is bad and it is systematically bad almost a 100% of the time bad when you come to this part of the curve"*

P05 only noted patterns of highs from the Dexcom standard day overview and did not think of the low BG patterns to be noteworthy. She hypothesized that the high BG pattern could be driving the low BG pattern because of an overcorrection of the highs. She did not go further to check on how many days were the lows happening as she thought that fixing the highs might fix the lows, "*It [the low at 6 pm] seems to be coming after you're like coming down from a high blood sugar, which makes me think it has something to do with like, maybe an over correction, or like over-bolus of some kind. I guess it seems like it's I don't know maybe it is counterintuitive, but like if you have better control over your blood sugars during like eating times like those lows, will kind of go away.*"

Participants' decisions were primarily based on the CGM curve and did not involve looking at more detailed information from the days when low BGs occurred. While low BGs receive immediate attention in real time when they are happening, they do not receive as much attention in retrospective review of data.

Classification with episode-driven narratives involved prioritizing one or multiple episodes for intervention based on the measures, factors, and evidence specific to the episodes

When looking at multiple patterns from the same data, participants did not dismiss any of the patterns upfront as done with the PDF reports. They chose one or more patterns for an intervention based on the different types of information presented. These included the measures associated with the patterns (average, time duration, range of BG numbers, and number of days of occurrence) and the factors that were listed as potentially causing the pattern. In addition to the information explicitly presented, participants also based their choices on the understanding of the risks associated with not fixing a pattern and the benefits of fixing a pattern.

P04 and P13 looked at the different measures associated with the patterns when selecting which pattern to give more attention to. P04 mentioned the information he was considering, "*I'd probably be looking at like the average BG so like, this one's a 195 this one's 185 so I'd probably say this one would be more severe if the average is higher. I would also look at the timeframe because being high for longer is worse – you want to come back down."* Based on the measures observed, P04 wanted to fix both as the two patterns had similar averages, time durations, and range, "*It would be hard to quantify specifically. I would probably try to do both at the same time since they're separated timewise like there's a gap in between. So, they're not necessarily related.*

I think I'd be comfortable changing the settings for the like eight to 12 and then one to four at the same time."

Given that both the patterns seemed similar numerically, for the same dataset, P13 chose to fix only one pattern based on his understanding of what would have the greatest benefit. Specifically, he chose to fix the nighttime pattern because that could result in having a better start to the next day, "*They're both pretty like similar. The length is first of all, quite similar like being highest they're both four hours long. And also, like their peaks. but I would go with the nighttime one just because I know that if you're high when you go to sleep, you're going to be high the entire night until you wake up and correct. Control IQ will definitely help with that, but it only can help so much."* Similarly, P10 chose to fix the overnight pattern over the daytime pattern because of the risks of having hyperglycemia while asleep, "*you're sleeping and that's kinda scary and then that's going to set up the rest of your day to kind of flow better."*

Instead of only considering the measures associated with the BG pattern (the is, the episode characteristics), P01 chose to look at the factors to see which pattern to fix. She chose to fix the pattern that had a factor with the most evidence, that is the episode that was linked to a factor that seemed influential on all days of the week versus the one that was linked to a factor that was influential on fewer days of the week, "*I like underneath the graphs it says, the number of days [referring to the number of days mentioned for each of the factors]. You know the first one [factors for the first pattern, 8pm to 12 am] says, seven out of seven days, so it [the issue with the factor related to 8pm-12am pattern] happened every day that week, and the second one [factors for the second pattern, 1 pm to 5 pm] says it [the issue with the factor related to 1pm-5am pattern] happened six out of seven days, so I would start with the 8pm to 12am because it [the issue with the factor related to 8pm-12am pattern] happened every day. Granted the other one happened almost every day, too, but um. I would start with the one that it was happening every day and try to fix that first then move down to one that maybe you know wasn't happening as frequently, even though it's just a day difference."*

Lastly, P09 was inclined to fix a pattern that could be corrected by changing the insulin doses over changing behaviors. For him, a smaller standard error associated with the pattern was an indication that the cause underlying the pattern was more systematic, such as problems with the insulin dose and not behaviors, "*So you see here [pointing to a segment of the CGM curve], the* *standard error is super tight. Whether it's good or bad I know that it's something that can be addressed by changing I don't know the basal or something like that but not behavior or not type of food, or anything like that. Not something that would induce variability. But here [pointing to another segment of the CGM curve] the standard error is a massive which tells me it could be that people are not very good at measuring carbs. It could be that they change the pre-bolusing in weird ways or something else, but it's not something in the calculations [insulin dose].”* He deprioritized pattern with higher variability.

8.7.7 Episode-specific Recommendation Generation

Episode-driven narratives improved participants’ ability to translate data to self-care actions

P10 and P13 were unable to identify self-care actions using the PDF reports but were able to come up with an action plan when reviewing the data with the prototypes.

> Changes suggested based on the PDF reports by P10: “*N/A”*
> Changes suggested based on the episode-driven narratives by P10: “*Increase basal insulin gradually and see if changes improve over several days. Also, take dinner meals earlier and take insulin 20 minutes before dinner.”*
>
> Changes suggested based on the PDF reports by P13: “*Nothing”*
> Changes suggested based on the episode-driven narratives by P13: “*Change it [dinner carb ratio] from 1:9.2 to 1:8.5”*

With the PDF reports, different types of evidence came to participants’ attention piecemeal based on their ability to spot meaningful evidence, as described in another theme of findings. Without additional information on how to interpret different types of evidence in relation to each other or how to prioritize them, participants suggested multiple changes despite having a general preference to make only one change at a time. For instance, P04 suggested multiple changes with the PDF reports as compared to episode-driven narratives.

> Changes suggested based on the PDF reports by P04: “*I would eat all of the nighttime snacks as one. I would wait 15 minutes after my bolus to eat. I would up the correction*

> *factor from 7-12 to 1:20 to start with. I would also increase the carb ratio in this time frame to 1:6 to start.*"

With episode-driven narratives, P04 was able to come up with a more focused and methodical plan for making changes to his care.

> Changes suggested based on the episode-driven narratives by P04: "*I would switch to only having one meal in this time frame to start with, and see if this fixes the issue. If the issue persists, I would then change the carb ratio during this time period as necessary*."

A potential reason for these findings could be that the commercial PDF reports did not help participants compartmentalize different issues and focus on the few most important issues to work on. This may result in patients getting overwhelmed with the complexity of all the changes they think they might need to make, eventually not make any changes.

8.7.8 Information Presentation Tradeoffs

Participants had diverse information and decisional needs along four dimensions, which I describe next.

Presenting all versus a fewer number of days from the time period under analysis

While most of the participants felt that having just one day of data as a representative example for an issue is enough and reasonable, two participants felt the need for more data. P04 specifically liked having to look at a fewer number of days instead of multiple days, "*I think having just a few days is better. I don't know, at least for me like looking at a bunch of days it's like, you know, I don't remember what I did that day. Maybe it was. I don't know, like your day to day can be different. What really matters is the average of those days or issues that persist across all days despite all those days being different*".

P09 and P13 on the other hand preferred to see more information like what was shown in the commercial reports (details about all the days). P09 was generally used to looking at multiple days. P13 wanted more information (a greater number of days showing a particular issue) because of his concern about the system misinterpreting the data and his need to cross check that he did not miss any other factor, "*As good as a computer gets as of right now, they are still not comparable to a* *human, especially with something like type one diabetes, which is so imprecise and imperfect science like there's no way to know exactly what the best change is but it can definitely hint towards it or bring you in the right direction which is why these potential causes are great. But I would want to be able to see different occurrences of this [the issue] happening, as opposed to just this one random day. The computer could have messed up somewhere or something or could be misinterpreting it.*"

Showing what is not going well and what is going well

In addition to knowing the potential causes of a pattern that were listed in the prototype, participants also wanted to know information on other factors, particularly those that were not considered related to the pattern by the system and those related to patient's compliance to self-care behaviors, such as recording data in the insulin pump.

P13, who tended to think very comprehensively about all possible causes, struggled to believe that the factors not listed under potential causes were not an issue and later indicated that he would want more information, "*It doesn't say that this [not recording certain data in the insulin pump] is an issue, but just from looking at this [the presented data], you can't tell if they're putting it in [on all days], which is why it's interesting that it doesn't say that that.*" Here, P13 could not tell from looking at the examples presented if the patient was consistently recording information in the pump. Because the patient was recording data, this was not listed as a potential cause for the pattern but P13 wanted to check for himself. In a similar direction, P01 and P09 wanted to see aggregate information about general insulin and meal administration behaviors to rule out the possibility of behavior related causes before thinking about insulin. P01 wanted to see it upfront on the first screen regardless of the association of that data with the glycemic patterns, "*The only thing that I would really like is, this particular graph is only showing what your CGM is showing. It's not showing you know you've got to click a little farther into it [to see data on meal and insulin recording]. It's not showing when you had carbs when you had a bolus and to me I just like that, on the very first thing [screen] to show that this is where you're entering carbs and bolusing, and this is when it goes up. Personally, I'd rather have on the beginning screen. You know if I had circles of this is where my entered my carbs and a diamond shape of this is, you know when the bolus was delivered, even though they would be layered over each other you'd at least have some idea of when you entered them.*"

Where possible, a potential factor was explained by presenting a good day and a bad day in relation to the factor. Multiple participants appreciated this presentation. The good day provided evidence on what worked well in relation to a factor and the bad day provided evidence on what did not work well, "*Okay, this is kind of cool so shows, like the bad versus the good.*" Having the evidence about what worked helped them better identify the action that could fix an issue. Further emphasizing the value of comparing good and bad days, P09 suggested the need to compare good days and bad days for specific issues, such as patterns of low blood glucose. Understanding potential causes of low BG patterns required comparing days with and without lows. Because lows are less frequent, they need to be compared with other days to identify the key differences that may have caused the lows, "*here you cannot base your analysis on the number of days that are affected for a low if the standard error is super wide. I would worry about looking at the days where it went low and see what was different compared to the days where it didn't go low.*"

Showing factors with data versus without data: De-emphasize factors without data

A few interfaces showed prescriptive information about general factors about which data was not available, such as growth and the quality of carb estimation. Detailed information about such factors did not contain any data (because there wasn't any data available on them) but only provided general information about how that factor could affect blood glucose. P10 and P13 did not like to see such factors. For them, it took away attention and time without any value in return. They suggested to keep another section on the screen where you could put secondary factors of such kind to only let the user know briefly. P10 mentioned, "*it's just waste space it's more for me to read it, you don't want I'm busy. It doesn't tell me anything.*" When asked about how he would like that information, he said, *"kind of have a call out just to kind of get your brain lubricated to kind of think about what else could it be, or you know, maybe, have you thought about or have you considered or talk to your doctor about you know kind of those types of things*".

Suggesting one change at a time versus a combination of changes

The majority of the participants noted that they preferred to make only one change at a time. However, after having looked at the data, they felt comfortable making more than one change. P10 and P13 looked at patterns with two potential causes that seemed equally influential. Instead of choosing one factor over the other, they said they could change both at the same time (one was behavior related and the other was insulin related). For instance, P13 described the two changes as follows, "*for this one definitely like that this is a behavioral thing you just don't inject as much, and then you can up your carb ratio little bit. And then just see how that helps and then like this is a relatively simple thing that you can do in one day.*" P10, who typically preferred to make only one change at a time and was not comfortable making changes with the commercial reports, reported feeling comfortable making two changes with the porotypes, "*they both have equal weight to me and they're both convincing. And I think combined if you're having less meals and giving the correct carb ratio that's a perfect thing there. It almost needs to say you know combine these to fix your carb ratios and [inaudible] frequent meals and, and this will fix your issue*."

Because multiple factors were listed as the potential causes of a BG pattern, P09 wanted guiding information to know how much effort to put into what factor when thinking of care actions, "*I would like it if the user interface could rank the potential causes by likelihood. Well, so the most likely culprit is the bolus timing. Then the second most likely culprit is the carb ratio, and potentially the third most likely culprit is the auto correction, so that people know Well, let's work on that but if it's not enough I know I need to make adjustment to let's say the carb ratio, eventually to the auto correction.*" He mentioned that without this information, it would be a challenge to deal with a list of causes, "*the problem is they give you multiple options and that's always a problem. Here they give you 3 possibilities. So, people have to go through them. Read them forever and decide [what] to tackle first. That's hard*!" While the research prototypes took lesser time to review than the PDF reports, P09's suggestion indicates the need to further improve the navigation of different options users have in the form of different potential causes.

8.8 DISCUSSION

In this study, I compared how T1D patients make sense of their data from continuous glucose monitors (CGMs) and insulin pumps using the commercially available data review platforms and using a research prototype (episode-driven narratives) informed by the episode-driven sensemaking framework synthesized in Chapter 6. The study found that the episode-driven narratives enabled better navigation of data, better interpretation of glycemic patterns, and improved comfort and confidence in identifying self-care actions as compared to the commercially available tools to review diabetes data. Additionally, the episode-driven narratives imposed a lower cognitive burden as compared to the PDF reports. The study also shows the ways in which episode-driven narratives enabled user agency through data interaction and comprehension for sensemaking and decision-making. Specifically, interacting with different levels of insights (glycemic patterns, potential causes of patterns, explanation and example for each cause) along with insight-specific evidence and domain knowledge enabled participants to exercise their agency in nuanced ways to utilize data for making care decisions. Through their interaction with the data, the study found dimensions along which participants' information and decisional needs varied. This study contributes a novel prototype embodying the episode-driven sensemaking framework in the form of episode-driven narratives. Additionally, it contributes an empirical understanding of how episode-driven narratives improve patients' comprehension of data, their understanding of potential self-care actions, and their experience of reviewing data.

While patient-generated data can support and improve the ways in which chronic conditions are managed, the majority of the research on interfaces for patient-generated data has focused on clinicians [94,155,189]. The limited work on patient-facing systems primarily focuses on exploratory systems or on providing recommendations from patient-generated data. This presents an opportunity to explore systems that help patients interact with their data in a lightweight manner and make decisions using insights that the system presents [64,93]. For diabetes management, HCI researchers have only begun exploring the presentation of system-generated insights to patients. One focus of this research has been to present educational content to users based on their data. For instance, Mamykina et al. studied MoDD, a web-based tool that identifies glycemic patterns and shows educational content relevant to that pattern. Episode-driven narratives are similar in that they use episodes as the basis for data presentation. They differ in the content presented. While MoDD focused on educational content, episode-driven narratives focus on presenting data slices relevant to interpret a glycemic pattern. Additionally, based on the concept of episode-driven sensemaking, the episodes need not only be glycemic patterns as was the case with MoDD, but episodes can also be behavioral patterns. Lastly, MoDD was a web-based application whereas episode-driven narratives are mobile interfaces.

Another focus of HCI research in diabetes management has been to show predictions of blood glucose numbers and dietary recommendations in the context of predictions [51,52]. There is limited work on designing data-driven sensemaking and decision-support interfaces for the presentation of multidimensional health data. This study contributes a novel prototype that presents multidimensional device data using narrative data interfaces to enable patients in better understanding their data, in identifying potential self-care actions based on the insights shown, in exploring their own decision-making needs, and in exercising their agency.

As described in Chapter 3, presenting multidimensional data or insights from them in personal informatics systems remains a challenge. Primary techniques employed by prior work include context-enhanced presentation of primary data (such as step count for physical activity, blood sugar values for diabetes), visual cuts of different types of data, natural language summaries to present correlations among different types of data, visual analytics systems to enable data exploration, and conversational agents to prompt reflection on specific data streams. There has been limited exploration of storytelling as an explanatory technique for the presentation and communication of multidimensional personal (health) data. By designing and studying episode-driven narratives, this work shows how data-driven narratives can be designed and how they can improve patients' experience of interacting with multidimensional health data.

As also described in Chapter 3, systems supporting the self-management of chronic conditions have evolved from promoting free form reflection to presenting deterministic insights by engaging patients in systematic data collection. These insights range from descriptive correlations with information on statistical significance, template-based recommendations, and predictions of specific measures. Notable patient experiences of using such systems include misrepresentation of lived experiences or user understanding of a situation, misinterpretation of insights, and difficulty of translating insights to actions (e.g., which recommendation to choose from the many that the system provides, whether to ignore associations that are not statistically significant, and deriving implications from obvious insights) [39,64,86,169]. This work shows that providing different types of explanations that include episode-specific evidence and domain knowledge can improve users' understanding of the potential self-care actions related to specific insights and result in fewer assumptions and misinterpretations.

In what follows, I present design implications for tools to support sensemaking and decision-making with multidimensional device data drawing on the themes described in the findings.

8.8.1 Enable Choices to Enable Agency While Limiting Assumptions

Glycemic outcomes in Type 1 diabetes are influenced by several factors at the same time. Consequently, when explaining a glycemic pattern, the research prototype presented multiple factors and suggestions related to each of those factors. While the existence of numerous possibilities and detailed explanation of each could challenge making choices, I saw that participants used it as an opportunity to better align their choices for self-care actions with their needs and preferences. For instance, participants did not necessarily choose to work on the factor that was most frequently involved. They often chose the factor that was the easiest for them to work on or what seemed to have the most benefit. Additionally, despite their general preference to make one change at a time, looking at the episode-driven narratives made them comfortable thinking of two changes at a time, something that they might not have done based on the commercial data review platform.

On one hand providing multiple potential factors to explain an episode enabled choices, on the other hand excluding factors that did not seem to be relevant for an episode resulted in skepticism. In the context of the PDF reports, not being able to easily find information about a factor resulted in assumptions. It may happen that excluding certain factors from the explanation of episodes in the episode-driven narratives might also result in users making assumptions.

Taken together, these experiences suggest that scaffolding data interaction through a narrative structure and presenting multiple options supported by relevant data, as enabled through the episode-driven narratives, can eventually help participants make decisions that they otherwise would not have been comfortable making. Additionally, these experiences suggest that tools need not only present one absolute cause or one directive recommendation for a problem but should also help users navigate the several reasonable and feasible causes or solutions that the data indicates. Depending on users' preference (e.g., the effort they want to put, the risk they want to take), this could mean helping the user select one cause to work on or helping them identify a combination of causes. Additionally, the inclusion and exclusion of specific information from an explanation must consider the scope for making assumptions in a decision-making scenario. In some situations, enabling assumptions might lead to misinterpretations. Lastly, the designers of decision-support systems need to consider the tradeoffs involved in presenting one possibility (e.g., one recommendation) and presenting multiple possibilities. The former restricts the choices users have resulting in limited scope for accommodating user preferences and could result in assumptions about the information not made explicit [31]. The latter enables users to exercise their agency to make choices, which can provide more value but would also require providing users with extra support to navigate these choices, particularly the newly diagnosed users who may not have established strong preferences about disease management.

8.8.2 Support Different Types of Explanations with Different Benefits

Explaining factors related as well as unrelated to an insight

Half the participants liked to see fewer number of factors, that is only the problematic ones, being presented and appreciated having to look at a fewer number of days of data to understand the role of those factors. Three participants, P01, P09 and P13, wanted to see more evidence (how the data relates to every possible factor, a greater number of days, possibly all, for each factor) than what was presented in the prototype to validate the insights presented and overrule the existence of other issues that could have been missed by the system. Participants' experiences suggest that there are variations in the amount of evidence and the types of evidence different users desire to believe system generated insights and to drive them to act. A system showing only the factors/data underlying an insight may not lead a user to believe that the excluded factors are not an issue or are not important to know about. Explaining a particular system-generated insight for such users would not only require the presentation of factors related to that insight, as done by prior work [93] but also those that are not directly involved but were tracked by the user. Having additional explanation about the factors that did not play a role in insight generation and why they were not highlighted by the system can provide an extra layer of transparency, limit assumptions, and improve users' trust.

Explaining with insight-specific examples of data contextualized with interpretation aids

In this work, I presented two levels of explanation in the context of managing a glycemic pattern – providing explanation for a problematic glycemic pattern using a list of potential factors influencing the pattern and explaining each factor in terms of examples from data contextualized with domain knowledge to interpret the data. Participants' choice of glycemic patterns and factors to work on relied heavily on the examples presented in the detailed explanations. Based on the available data, different types of detailed explanations were presented – data from a single day showing an issue, data from a good day and a bad day in relation to a factor, and data to show whether a factor met the guidelines. Each of these explanations had different benefits. For instance, examples of individual days helped show the presence or absence of a problem. Comparisons between good and bad days in relation to a factor enabled identifying not only the problem but also the solution by showing what worked. It also increased participants' confidence in a particular choice. Similarly, displaying a value with information on guidelines or reference values helped determine how far off a factor was from the ideal and the extent of changes that would be required. This suggests that depending on the decisional context, using examples of data contextualized with domain knowledge in detailed explanations may improve how users utilize the information towards making decisions. This observation is in contrast with prior work's findings that providing detailed explanation can negatively affect users' experiences with the system [93]. The differences between this work and prior work can perhaps be attributed to the decisional context. More detailed explanations may be less useful in low-risk decisional situations, such as general sleep or stress management. For high-risk decisional situations, such as decision-making for a chronic condition, this work suggests that detailed explanations are relatively more useful and perhaps necessary to avoid misinterpretation. Additionally, this work shows the benefits of using episode-specific examples of data enhanced with domain knowledge about the condition as explanations, presenting an opportunity to develop computational abilities to identify representative examples from larger datasets for presentation to the user to support the explanation of insights.

8.9 LIMITATIONS

The study involved a small number of participants. All the participants were being seen at the same medical facility. Their experiences with the research prototypes and with T1D management may not be representative of the general T1D population.

8.10 CONCLUSION

In this study, I compared how T1D patients make sense of their data from CGM and insulin pump using the commercially available data review platforms and using a research prototype embodying the episode-driven sensemaking framework. The study contributes a novel prototype embodying the episode-driven sensemaking framework in the form of episode-driven narratives. Additionally, it contributes an empirical understanding of how episode-driven narratives improve patients' comprehension of data, their understanding of potential self-care actions, and their experience of reviewing data. In particular, providing a narrative structure to scaffold data navigation, presenting examples from data contextualized with domain knowledge as explanations, drawing users' attention to specific data features through visual highlights and textual description, and presenting different possibilities as opposed to directive recommendations enabled participants in making choices in more nuanced ways than the commercial data review platforms. These choices relied less on assumptions and more on the available data and participants' preferences. A similar experience would perhaps not be possible when provided with very specific recommendations or when provided with exploratory data interfaces. Tools to support interpretation of multidimensional health data and identification of potential care actions need to support different types of explanations and need to enable choices to respect user agency in complex decisional situations.

Chapter 9 Conclusion and Future Work

My dissertation examines how patient-generated data is used across different settings – individual and collaborative, and how technology can be designed to support these practices towards sensemaking and decision-making from data. I have articulated the analytical activities of patients and clinicians, the analytical challenges, and challenges particular to collaborative interpretation of data. Through this understanding, I have developed a descriptive framework of episode-driven sensemaking with data, which reflects these practices and challenges and provides a basis to design technology. I have reviewed existing tools to identify the limitations of technology in the context of current data practices of patients and clinicians. I have designed and evaluated novel data interfaces for patients that address some of these limitations and account for the episode-driven sensemaking practices. This dissertation establishes the need to better understand and support action-oriented reflection on multidimensional health data. Towards this need, it contributes a descriptive framework anchored in data practices to guide tool design, novel data interfaces to enable the episode-driven presentation of multidimensional health data, and design guidelines for systems to support value creation from patient-generated data.

9.1 IMPLICATIONS

I now summarize the key implications that stem from this dissertation, particularly for designing systems to enable the use of patient-generated data for chronic care management.

Thesis Claim	**Implication**
For data-driven management of chronic conditions, tools need to enable patients and their clinicians in ongoing sensemaking with multidimensional patient-generated data, both individually and collaboratively. One of the shared objectives of sensemaking is problem-solving (identifying problems and	Support effective problem-solving by creating a shared understanding of the data

solutions from data) from multiple streams of data that include clinical and contextual data.	
For informing care through multidimensional data, such tools need to support four types of data-driven assessments - data sufficiency, clinical outcomes, patient behaviors, and regimen quality.	Support different types of assessments from data
Such tools should guide analytic focus by organizing and presenting relevant evidence using a practice-based workflow, an example of which is the episode-driven sensemaking workflow where episodes as phases of suboptimal management guide data analysis and reasoning. Tools supporting episode-driven sensemaking with data should – 1) identify and present an episode along with its characteristics, 2) filter evidence based on the episode identified to surface relevant evidence (contextual information and examples of the episode), and 3) suggest interpretive frames that can help explain the episode with the available evidence.	Guide analytic focus using a scaffold (e.g., a practice-based workflow) to organize and present evidence
A promising technique to implement an episode-driven sensemaking tool for interacting with data is data-driven storytelling. As compared to the commercially available exploratory tools, episode-driven visual data narratives simplify engagement with multidimensional health data; improve comprehension of the data; improve the understanding of potential self-care actions indicated by the data; and impose a lower cognitive burden on patients. Enabling easy and effective sensemaking with data improves patients' experience of data review and their ability to make decisions, which can potentially increase patient engagement with data and can eventually improve self-management and overall control.	Support episode-driven narrative generation and presentation from multidimensional data to simplify reasoning and to support user agency (for several objectives – problem-solving, sensemaking, decision-making)

Table 25: Thesis claims mapped to implications for designing tools to support the use of patient-generated data.

9.1.1 Support Effective Problem-Solving by Creating a Shared Understanding of Data

In chapter 4, I demonstrate that despite problem-solving being a shared and central objective for patients and clinicians in using patient-generated data, there are differences in how the data is perceived that challenge effective problem-solving. Particularly, patients and clinicians differ on the understanding of what is a problem, of what types of problems to focus on, and of the usefulness of the existing data representations. This presents a need for tools to create a shared understanding from the data by automated detection of problems, by providing ways to contextualize problems with additional data, and by creating mutual intelligibility of how sensemaking is performed by different stakeholders. To better understand how sensemaking is performed by different stakeholders, chapters 5, 6, and 7 present a deeper investigation of individual and collaborative sensemaking with patient-generated data.

9.1.2 Support Different Types of Assessments from Data and Help Connect Across Assessments

In chapter 6, the investigation of collaborative sensemaking by patients and clinicians shows four types of assessments that are performed using patient-generated data. These include assessment of data sufficiency, outcomes, patient and system behaviors, and the regimen. When understanding the existing tools that are used, I found that these tools predominantly focus on supporting the analysis of outcomes but provide limited support for analyzing data sufficiency and patient or system behaviors. Additionally, they do not help connect findings from different types of assessments. For example, an outcome might be related to a particular system or patient behavior requiring the need to connect the two types of information, which the existing tools do not help with. These observations present the need for tools to surface insights related to patient and/or system behavior, and the need to help connect different types of assessments performed using data.

9.1.3 Guide Analytic Focus Using a Practice-Based Workflow to Organize and Present Evidence

In addition to understanding the types of assessments that are performed using data, it is equally important to understand the analytical workflows of translating multidimensional data to actionable information, as established in chapter 3. Towards this need, chapters 5 and 7 highlight the analytical activities of patients and clinicians when interpreting data individually and collaboratively. Both patients and clinicians identify episodes from the data, which could include trends and unusual singular events present in outcome or behavior data. This is the detection phase. Once an episode has been identified, they explain the episode using the available data or hypothetical data. I call this the elaboration phase. For an episode that could be successfully explained and that represented a problem, patients and clinicians envisioned potential care actions. This was the episode-specific recommendation generated phase. There were similarities in the analytical activities performed during the individual and collaborative review of data. I synthesize these analytical activities into a descriptive framework called episode-driven sensemaking. I argue that this framework provides a basis to design tools to simplify interaction with multidimensional health data and to bridge the differences in how patients and clinicians understand the data.

9.1.4 Support Episode-Driven Narrative Generation and Presentation of Multidimensional Data

Given the potential of the episode-driven sensemaking framework as a basis for tools design, in chapter 8, I show that a promising technique to implement data review to facilitate episode-driven sensemaking is data-driven narrative visualization. I present prototypes embodying the episode-driven sensemaking framework that uses a narrative structure and I evaluate them with patients. A qualitative evaluation of my prototypes shows that the episode-driven narratives enabled better navigation of data, better interpretation of glycemic patterns, and improved patients' comfort and confidence in coming up with plans for action as compared to the commercially available exploratory interfaces. This study also exemplifies how the episode-driven sensemaking framework can guide tool design. Additionally, the study contributes a novel prototype to facilitate structured engagement with data, that is through episode-driven data narratives. Lastly, it contributes implications for designing tools to support sensemaking and decision-making from multidimensional health data. Episode-driven presentation of different types of information provides a scaffold to enable meaningful interaction with data and enables user agency in making decisions for self-care.

9.2 FUTURE WORK

I am interested in continuing to explore different types of systems for leveraging patient-generated data and the evolving practices of using such systems to create value from data. In this section, I describe the future projects that I want to work on.

9.2.1 Evaluating the Prototype in the Field

While this dissertation concludes at evaluating story-based data interfaces for episode-driven sensemaking, there is an opportunity to design a mobile app that presents a feed of several data stories on an ongoing basis from self-tracked data. I am interested in exploring this opportunity by having patients use the mobile app for day-to-day care of chronic conditions over several weeks to better understand engagement with data and how different types of interactions (short glance-based interactions versus long interactions) can be leveraged to provide value from data towards everyday decision-making.

9.2.2 Studying the Involvement of Clinicians

I am also interested in exploring design opportunities for involving clinicians in the use of my research prototype. Towards this, I would like to conduct a focus group with clinicians to understand how this prototype could be extended to meet their needs for helping patients with their data.

9.2.3 Understanding Data Practices in the Use of Automated Diagnosis/Therapy Systems

Such systems use patient-generated data to make decisions or recommendations for the patients and/or clinicians. Examples of such systems include intelligent clinical decision support tools and automated therapy systems, such as the artificial pancreas. Often, these systems fail at providing the transparency needed to contextualize the recommendations that they offer. Learnings from this dissertation on presenting evidence in the context of an episode can be translated to presenting evidence in the context of automated decisions or recommendations to explain them or make them intelligible to the user. I am interested in identifying and leveraging design opportunities for explanatory interfaces to make intelligent systems in healthcare more comprehensible to their users.

9.2.4 Exploring Visualizations of Behaviors

In chapter 6, I point to the limitations of existing data interfaces. One of these limitations include the lack of visual support for identifying patient and device behaviors from data. Data-driven assessment of patient and device behaviors is critical in contextualizing the outcomes and making decisions. Data interfaces need to be designed to surface behavior related insights, such as how often does the patient exercise, and how often does a system suspend insulin infusion. I am interested in exploring how behaviors can be better visualized and what types of tensions might arise in how patients and clinicians engage with visualizations of patient and/or device behaviors linked to outcomes.

9.2.5 Exploring the design space of data-based explanations

In chapter 8, I show that different ways in which the data was used to explain an insight to the users resulted in different benefits for them. I am interested in further exploring the design space of data-based explanations with the objective to improve users' trust and agency.

9.3 CONCLUSION

The increasing adoption of medical and consumer sensing devices has created a need and an opportunity for technology to help create value from the data produced by these devices. These data in addition to self-reported data from patients provide a basis to understand patients' lived experiences, and to offer care in ways that was not possible without the continuous availability of data. There are several types of systems that offer opportunities for deeper investigation in this direction – exploratory displays, tools for guided data-driven investigation, decision-support tools, recommender systems, and intelligent diagnostic and therapeutic systems. Such systems must continue to be investigated to understand the ways in which they meet or fail to meet user needs and how they could be improved.

As the practices of using patient-generated data evolve, in the process of creating value from the data different stakeholders might require different types of assistance from a system depending on their need and their expertise. Consequently, technology will need to play different roles that would involve varying combinations of system and user agency. In particular, the lay users would require more assistance as compared to the expert users. The focus on understanding how tools can help lay users make sense of complex health data for routine chronic care must continue.

As new tools emerge, different types of engagement with data will surface. Systems for chronic care management would consequently need to support varied types of engagement with the data. For example, users might want different types of guidance for data interpretation ranging from simply orienting the user towards the data to directing the user in performing well-defined analysis to prescribing possible care actions from automated analyses of data. Additionally, engagement may be varied based on impromptu and planned needs for data. Different types of engagement with data and the corresponding role for systems must continue to be studied to reduce the burden of making data actionable.

Appendix A - Interview Protocol for Commercial Platforms

General experience of reviewing data

1. When were the last few times that you downloaded and looked at your data by yourself?
    a. What was the purpose?
    b. What happened as a result of looking at the data?
    c. What platforms (PDF reports, mobile app, etc.) did you use to review your data?
    d. How was your overall experience of reviewing your data?
        i. Did it help? How?
        ii. Was it challenging? What was the challenge? What would help?
2. When were the last few times that you downloaded and looked at your data with your clinician?
    a. What was the purpose?
    b. What happened as a result of looking at the data?
    c. What platforms (PDF reports, mobile app, etc.) did you use to review your data?
    d. How was your overall experience of reviewing your data?
        i. Did it help? How?
        ii. Was it challenging? What was the challenge? What would help?
3. When was the last time you changed your insulin doses on your own?
    a. Can you describe how did you go about it?
    b. Why did you need to change it?
    c. How did you change it? What data did you use to inform the change?
    d. What happened as a result of the change?
    e. Would you have done anything differently?
4. When was the last time you contacted the diabetes educator to make changes to insulin?
    a. Can you describe how did you go about it?
    b. Why did you feel the need to change insulin?
    c. How did you change it? What data did you use to inform the change?
    d. What happened as a result of the change?
    e. Would you have done anything differently?
5. What are some of the challenges of looking at your diabetes data?
    a. What holds you back from engaging with it?
6. What are some of the benefits of looking at your diabetes data?

PDF report review: Questions to probe during or after the data walkthrough

1. Which displays do you usually look at?
    a. Which of these displays do or do not seem familiar?
2. Could you walk us through how you would go about understanding this data and assessing if something needs to change?

    a. What patterns do you see in the data?
    b. What seems to have caused them?
    c. Do these patterns need to be fixed?
        i. Why?
        ii. Why not?
    d. How would you fix them, if you would?
    e. How did you come to this conclusion?
3. Can you explain the patterns you see in the data?

Appendix B – Interview Protocol for Prototypes

Debrief after each task

1. Summarize what did you understand from this data story.
    a. What patterns do you see in the data?
    b. Do these patterns need to be fixed?
        i. Why?
        ii. Why not?
    c. How would you fix them, if you would?
    d. How did you come to this conclusion?

2. Do you find something specifically challenging or burdensome in understanding the displays?

3. Did you need extra information at any point to understand the data?

Final interview after all tasks have been completed

1. How was your overall experience of reviewing diabetes data with these data stories?
    a. What did you like about them?
    b. What did you not like about them?
2. How was this experience different from looking at the PDF reports?
    a. Mobile versus desktop
    b. Cognitive load
    c. Ease of use
3. How would you improve these?
4. Do you need more information in these displays to manage diabetes?
5. Would you like to involve your providers with this tool? Do you have suggestions for how could we involve providers?
6. Given two episodes, could you compare them? (e.g., rank them based on how you might prioritize them for action)
7. Given multiple factors, could you rank them?

Appendix C – Topics for Baseline Questionnaire

Caregiver/Parent demographics

- Age
- Gender
- Race
- Ethnicity
- Education
- Occupation
- Languages spoken
- Have a mobile phone with internet connection?
- Have a laptop?
- Have internet at home?

Child (T1D patient) demographics

- Age
- Gender
- Race
- Ethnicity
- Grade in school
- Employment status
- Members in the household
- Age when diagnosed
- Does child use a phone?

About diabetes management

- Name of insulin pump used
- Name of CGM used
- Name of meter used
- Names of tools used for data review
- Names of diabetes apps used
- How often do you review data outside of clinic visits?
- When was the last time you reviewed data?
- Have you received education about using diabetes data?
- When was your child's last clinic visit?

- Did you download and review your data with the clinicians during your last clinic visit?

Appendix D – Questionnaire for Data Comprehension and Decision-Making

Questions scored – 1, 2, 3, 4, 5, 6 and 9

Questions not scored – 7, 8, 10, 11 and 12

Please answer the questions based on your data.

Q1 What is your time in range between May 25 and May 31?

- ○ 21% (1)
- ○ 2% (2)
- ○ 80% (3)
- ○ 77% (4)
- ○ Not sure (5)

Q2 What BG patterns do you see in the blood glucose (BG) data? Select all that are applicable.

- ☐ I don't see any patterns (1)
- ☐ High BG pattern – 3 am to 5 am (2)
- ☐ Low BG pattern - 3 am to 5 am (3)
- ☐ High BG pattern - 12 pm to 2:45 pm (4)
- ☐ Low BG pattern - 12 pm to 2:45 pm (5)
- ☐ High BG pattern – 10 pm to 11 pm (6)
- ☐ Low BG pattern – 10 pm to 11 pm (7)
- ☐ Not sure (8)

Q3 For the rest of the questions, please refer to the BG pattern that exists around noon.

What time window does the BG pattern occur in?

- ○ 2 pm to 4 pm (1)
- ○ 12 pm to 2:45 pm (2)
- ○ 3 am to 5 am (3)
- ○ 10 pm to 11 pm (4)
- ○ Not sure (5)

Q4 For the time duration of the BG pattern, what is the range of the BG numbers (10%-90% range)?

- ○ 70 - 250 (1)
- ○ 116 - 208 (2)
- ○ 100 - 250 (3)
- ○ 80 - 300 (4)
- ○ Not sure (5)

Q5 What meal(s) could this pattern be associated to? Select all that apply.

☐ Most likely not related to meals (1)

☐ Breakfast (2)

☐ Lunch (3)

☐ Dinner (4)

☐ Not sure (5)

Q6 What are the potential causes of the BG pattern?

Select all that are applicable.

☐ Meal carb ratios are more than needed (1)

☐ Meal carb ratios are lesser than needed (2)

☐ Correction factor is more than needed (3)

☐ Correction factor is lesser than needed (4)

☐ Basal insulin is more than needed (5)

☐ Basal insulin is lesser than needed (6)

☐ Overcounted meal carbs (7)

☐ Undercounted meal carbs (8)

☐ Not recording carbs in the pump (9)

☐ Not taking meal bolus (10)

☐ Not recording BG in the pump (11)

☐ Performing physical activity (12)

☐ Frequent or excessive eating (13)

☐ Taking bolus before meals (14)

☐ Taking bolus after meals (15)

☐ Too much insulin given by the Control IQ technology (16)

☐ Insulin suspension by the pump (17)

☐ Pump site not working (18)

☐ Other (19)

☐ Not sure (20)

Q7 Which cause seems to be the most strongly indicated in the data? In other words, which cause related to the BG pattern do you have the most evidence for? Select one.

○ Meal carb ratios are more than needed (1)

○ Meal carb ratios are lesser than needed (2)

○ Correction factor is more than needed (3)

○ Correction factor is lesser than needed (4)

○ Basal insulin is more than needed (5)

○ Basal insulin is lesser than needed (6)

○ Overcounted meal carbs (7)

○ Undercounted meal carbs (8)

○ Not recording carbs in the pump (9)

○ Not taking meal bolus (10)

○ Not recording BG in the pump (11)

○ Performing physical activity (12)

○ Frequent or excessive eating (13)

○ Taking bolus before meals (14)

○ Taking bolus after meals (15)

○ Too much insulin given by the Control IQ technology (16)

○ Insulin suspension by the pump (17)

○ Pump site not working (18)

○ Other (19)

○ Not sure (20)

Q8 For the BG pattern, choose one of the following:

- ○ Does not need to be corrected (1)
- ○ Needs to be corrected by changing insulin only (2)
- ○ Needs to be corrected by changing behaviors only (3)
- ○ Needs to be corrected by changing both insulin and behaviors (4)
- ○ Not sure (5)

Q9 What all could you do to fix the BG pattern?

Select all that apply.

- ☐ Increase carb ratio at breakfast (1)
- ☐ Decrease carb ratio at breakfast (2)
- ☐ Increase carb ratio at lunch (3)
- ☐ Decrease carb ratio at lunch (4)
- ☐ Increase carb ratio at dinner (5)
- ☐ Decrease carb ratio at dinner (6)
- ☐ Increase correction factor between 12 pm to 2:45 pm (7)
- ☐ Decrease correction factor between 12 pm to 2:45 pm (8)
- ☐ Increase basal insulin between 12 pm to 2:45 pm (9)
- ☐ Decrease basal insulin between 12 pm to 2:45 pm (10)

☐ Take bolus before meals (11)

☐ Take bolus after meals (12)

☐ Eat food less frequently (13)

☐ Eat food more frequently (14)

☐ Change pump site (15)

☐ Nothing needs to be done (16)

☐ Other (17)

☐ Not sure (18)

Q10 What is the one thing that you would do to fix the BG pattern based on the evidence you have?

Select one.

○ Increase carb ratio at breakfast (1)

○ Decrease carb ratio at breakfast (2)

○ Increase carb ratio at lunch (3)

○ Decrease carb ratio at lunch (4)

○ Increase carb ratio at dinner (5)

○ Decrease carb ratio at dinner (6)

○ Increase correction factor between 12 pm to 2:45 pm (7)

○ Decrease correction factor between 12 pm to 2:45 pm (8)

○ Increase basal insulin between 12 pm to 2:45 pm (9)

○ Decrease basal insulin between 12 pm to 2:45 pm (10)

○ Take bolus before meals (11)

○ Take bolus after meals (12)

○ Eat food less frequently (13)

○ Eat food more frequently (14)

○ Change pump site (15)

○ Nothing needs to be done (16)

○ Other (17)

○ Not sure (18)

Q11 Briefly describe what you would do based on the option(s) chosen above. For example, if you chose “increase carb ratio at breakfast”, please state what those changes would be based on the current carb ratio.

____

Q12 How confident are you in making the decision reported in question 11?

○ Not at all confident (1)

○ Slightly confident (2)

○ Somewhat confident (3)

○ Quite confident (4)

○ Extremely confident (5)

Bibliography

1. Danielle Albers, Michael Correll, and Michael Gleicher. 2014. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 551–560. https://doi.org/10.1145/2556288.2557200
2. Jessica S. Ancker, Yalini Senathirajah, Rita Kukafka, and Justin B. Starren. 2006. Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association* 13, 6: 608–618.
3. Adriana Arcia, Jacqueline A. Merrill, and Suzanne Bakken. 2019. Consumer Engagement and Empowerment Through Visualization of Consumer-Generated Health Data. In *Consumer Informatics and Digital Health: Solutions for Health and Health Care*, Margo Edmunds, Christopher Hass and Erin Holve (eds.). Springer International Publishing, Cham, 183–203. https://doi.org/10.1007/978-3-319-96906-0_10
4. Adriana Arcia, Niurka Suero-Tejeda, Michael E Bales, Jacqueline A Merrill, Sunmoo Yoon, Janet Woollen, and Suzanne Bakken. 2016. Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *Journal of the American Medical Informatics Association* 23, 1: 174–183. https://doi.org/10.1093/jamia/ocv079
5. Elizabeth Austin, Jenney R Lee, Dagmar Amtmann, Rich Bloch, Sarah O Lawrence, Debbe McCall, Sean Munson, and Danielle C Lavallee. 2019. Use of patient-generated health data across healthcare settings: implications for health systems. *JAMIA Open* 3, 1: 70–76. https://doi.org/10.1093/jamiaopen/ooz065
6. Uba Backonja, Sarah C Haynes, and Katherine K Kim. 2018. Data Visualizations to Support Health Practitioners' Provision of Personalized Care for Patients With Cancer and Multiple Chronic Conditions: User-Centered Design Study. *JMIR Human Factors* 5, 4. https://doi.org/10.2196/11826
7. Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K. Dey. 2016. Modeling and Understanding Human Routine Behavior. 248–260. https://doi.org/10.1145/2858036.2858557
8. 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 SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 2627–2636. https://doi.org/10.1145/2470654.2481364
9. Tadej Battelino, Thomas Danne, Richard M. Bergenstal, Stephanie A. Amiel, Roy Beck, Torben Biester, Emanuele Bosi, Bruce A. Buckingham, William T. Cefalu, Kelly L. Close, Claudio Cobelli, Eyal Dassau, J. Hans DeVries, Kim C. Donaghue, Klemen Dovc, Francis J. Doyle, Satish Garg, George Grunberger, Simon Heller, Lutz Heinemann, Irl B. Hirsch, Roman Hovorka, Weiping Jia, Olga Kordonouri, Boris Kovatchev, Aaron Kowalski, Lori Laffel, Brian Levine, Alexander Mayorov, Chantal Mathieu, Helen R. Murphy, Revital Nimri, Kirsten Nørgaard, Christopher G. Parkin, Eric Renard, David Rodbard, Banshi

Saboo, Desmond Schatz, Keaton Stoner, Tatsuiko Urakami, Stuart A. Weinzimer, and Moshe Phillip. 2019. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. *Diabetes Care* 42, 8: 1593–1603. https://doi.org/10.2337/dci19-0028

10. Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (DIS ’14), 93–102. https://doi.org/10.1145/2598510.2598598
11. Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Transactions on Computer-Human Interaction* 20, 5: 30:1-30:27. https://doi.org/10.1145/2503823
12. Richard M. Bergenstal, Andrew J. Ahmann, Timothy Bailey, Roy W. Beck, Joan Bissen, Bruce Buckingham, Larry Deeb, Robert H. Dolin, Satish K. Garg, Robin Goland, Irl B. Hirsch, David C. Klonoff, Davida F. Kruger, Glenn Matfin, Roger S. Mazze, Beth A. Olson, Christopher Parkin, Anne Peters, Margaret A. Powers, Henry Rodriguez, Phil Southerland, Ellie S. Strock, William Tamborlane, and David M. Wesley. 2013. Recommendations for Standardizing Glucose Reporting and Analysis to Optimize Clinical Decision Making in Diabetes: The Ambulatory Glucose Profile. *Journal of Diabetes Science and Technology* 7, 2: 562–578. https://doi.org/10.1177/193229681300700234
13. Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Dieter Fellner, and Michael Sedlmair. 2018. Comparing Visual-Interactive Labeling with Active Learning: An Experimental Study. *IEEE Transactions on Visualization and Computer Graphics* 24, 1: 298–308. https://doi.org/10.1109/TVCG.2017.2744818
14. Mustafa Bilgic and Raymond J Mooney. Explaining Recommendations: Satisfaction vs. Promotion. 8.
15. M L Birnbaum, S K Ernala, A F Rizvi, E Arenare, A R Van Meter, M De Choudhury, and J M Kane. 2019. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *NPJ schizophrenia* 5, 1: 17. https://doi.org/10.1038/s41537-019-0085-9
16. M. Brehmer and T. Munzner. 2013. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12: 2376–2385. https://doi.org/10.1109/TVCG.2013.124
17. Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2019. Visualizing Ranges over Time on Mobile Phones: A Task-Based Crowdsourced Evaluation. *IEEE transactions on visualization and computer graphics* 25, 1: 619–629. https://doi.org/10.1109/TVCG.2018.2865234
18. Sue Brown, Dan Raghinaru, Emma Emory, and Boris Kovatchev. 2018. First Look at Control-IQ: A New-Generation Automated Insulin Delivery System. *Diabetes Care* 41, 12: 2634–2636. https://doi.org/10.2337/dc18-1249
19. Marissa Burgermaster, Jung H. Son, Patricia G. Davidson, Arlene M. Smaldone, Gilad Kuperman, Daniel J. Feller, Katherine Gardner Burt, Matthew E. Levine, David J. Albers, Chunhua Weng, and Lena Mamykina. 2020. A new approach to integrating patient-generated data with expert knowledge for personalized goal setting: A pilot study.

International Journal of Medical Informatics 139: 104158. https://doi.org/10.1016/j.ijmedinf.2020.104158

20. Kirsten Butcher and Tamara Sumner. 2011. Self-Directed Learning and the Sensemaking Paradox. *Human-Computer Interaction* 26, 1: 123–159. https://doi.org/10.1080/07370024.2011.556552
21. Joseph A Cafazzo, Mark Casselman, Nathaniel Hamming, Debra K Katzman, and Mark R Palmert. 2012. Design of an mHealth App for the Self-management of Adolescent Type 1 Diabetes: A Pilot Study. *Journal of Medical Internet Research* 14, 3: e70. https://doi.org/10.2196/jmir.2058
22. Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI ’19), 258–262. https://doi.org/10.1145/3301275.3302289
23. Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 258–262. https://doi.org/10.1145/3301275.3302289
24. David Caswell. 2019. Structured Journalism and the Semantic Units of News. *Digital Journalism* 7, 8: 1134–1156. https://doi.org/10.1080/21670811.2019.1651665
25. David Caswell and Konstantin Dörr. 2018. Automated Journalism 2.0: Event-driven narratives. *Journalism Practice* 12, 4: 477–496. https://doi.org/10.1080/17512786.2017.1320773
26. Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jorg Schulz, Marc Streit, and Christian Tominski. 2017. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1: 111–120. https://doi.org/10.1109/TVCG.2016.2598468
27. Philip K. Chan, Matthew V. Mahoney, and Muhammad H. Arshad. 2003. A machine learning approach to anomaly detection. *Department of Computer Sciences, Florida Institute of Technology, Melbourne*. Retrieved April 27, 2017 from http://interfacesymposia.org/I03/intrusiondetection/philip_chan_cs-2003-06.pdf
28. S. Chen, J. Li, G. Andrienko, N. Andrienko, Y. Wang, P. H. Nguyen, and C. Turkay. 2020. Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 26, 7: 2499–2516. https://doi.org/10.1109/TVCG.2018.2889054
29. Yunan Chen. 2011. Health information use in chronic care cycles. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 485–488. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=1958898
30. Zhuo Chen, Kyle Marple, Elmer Salazar, Gopal Gupta, and Lakshman Tamil. 2016. A Physician Advisory System for Chronic Heart Failure Management Based on Knowledge Patterns. *Theory and Practice of Logic Programming* 16, 5–6: 604–618. https://doi.org/10.1017/S1471068416000429
31. Janghee Cho, Tian Xu, Abigail Zimmermann-Niefield, and Stephen Voida. 2022. Reflection in Theory and Reflection in Practice: An Exploration of the Gaps in Reflection Support among Personal Informatics Apps. In *CHI Conference on Human Factors in Computing Systems*, 1–23. https://doi.org/10.1145/3491102.3501991

32. Eun Kyoung Choe, Bongshin Lee, Tariq Osman Andersen, Lauren Wilcox, and Geraldine Fitzpatrick. 2018. Harnessing the Power of Patient-Generated Data. *IEEE Pervasive Computing* 17, 2: 50–56. https://doi.org/10.1109/MPRV.2018.022511243
33. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, 121–132. https://doi.org/10.1145/2750858.2804266
34. Eun Kyoung Choe, Bongshin Lee, and others. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications* 35, 4: 28–37.
35. Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding Self-Reflection: How People Reflect on Personal Data through Visual Data Exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'17). ACM, New York, NY, USA*.
36. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1143–1152. https://doi.org/10.1145/2556288.2557372
37. Chia-Fang Chung, Kristin Dew, Allison M Cole, Jasmine Zia, James A 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 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 768–784. https://doi.org/10.1145/2818048.2819926
38. 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1: 1–27. https://doi.org/10.1145/3314394
39. 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1: 7:1-7:27. https://doi.org/10.1145/3314394
40. Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. 2009. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys* 41, 1: 1–31. https://doi.org/10.1145/1456650.1456652
41. Deborah J Cohen, Sara R Keller, Gillian R Hayes, David A Dorr, Joan S Ash, and Dean F Sittig. 2016. Integrating Patient-Generated Health Data Into Clinical Care Settings or Clinical Decision-Making: Lessons Learned From Project HealthDesign. *JMIR Human Factors* 3, 2: e26. https://doi.org/10.2196/humanfactors.5919
42. Heather J. Cole-Lewis, Arlene M. Smaldone, Patricia R. Davidson, Rita Kukafka, Jonathan N. Tobin, Andrea Cassells, Elizabeth D. Mynatt, George Hripcsak, and Lena Mamykina. 2016. Participatory approach to the development of a knowledge base for problem-solving

in diabetes self-management. *International Journal of Medical Informatics* 85, 1: 96–103. https://doi.org/10.1016/j.ijmedinf.2015.08.003

43. 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 SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 1797–1806. https://doi.org/10.1145/1357054.1357335
44. Michael Correll, Danielle Albers, Steven Franconeri, and Michael Gleicher. 2012. Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1095–1104. Retrieved August 1, 2021 from https://doi.org/10.1145/2207676.2208556
45. Dan Cosley, Elizabeth Churchill, Jodi Forlizzi, and Sean A. Munson. 2017. Introduction to This Special Issue on the Lived Experience of Personal Informatics. *Human–Computer Interaction* 32, 5–6: 197–207. https://doi.org/10.1080/07370024.2017.1324787
46. Thomas Danne, Revital Nimri, Tadej Battelino, Richard M. Bergenstal, Kelly L. Close, J. Hans DeVries, Satish Garg, Lutz Heinemann, Irl Hirsch, Stephanie A. Amiel, Roy Beck, Emanuele Bosi, Bruce Buckingham, Claudio Cobelli, Eyal Dassau, Francis J. Doyle, Simon Heller, Roman Hovorka, Weiping Jia, Tim Jones, Olga Kordonouri, Boris Kovatchev, Aaron Kowalski, Lori Laffel, David Maahs, Helen R. Murphy, Kirsten Nørgaard, Christopher G. Parkin, Eric Renard, Banshi Saboo, Mauro Scharf, William V. Tamborlane, Stuart A. Weinzimer, and Moshe Phillip. 2017. International Consensus on Use of Continuous Glucose Monitoring. *Diabetes Care* 40, 12: 1631–1640. https://doi.org/10.2337/dc17-1600
47. Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. 347–358. https://doi.org/10.1145/2984511.2984534
48. Robin De Croon, Leen Van Houdt, Nyi Nyi Htun, Gregor Štiglic, Vero Vanden Abeele, and Katrien Verbert. 2021. Health Recommender Systems: Systematic Review. *Journal of Medical Internet Research* 23, 6: e18035. https://doi.org/10.2196/18035
49. 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*: 1–11.
50. George Demiris, Sarah J. Iribarren, Katherine Sward, Solim Lee, and Rumei Yang. 2019. Patient generated health data use in clinical practice: A systematic review. *Nursing Outlook* 67, 4: 311–330. https://doi.org/10.1016/j.outlook.2019.04.005
51. 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. 1–13. https://doi.org/10.1145/3173574.3174112
52. Pooja M. Desai, Elliot G. Mitchell, Maria L. Hwang, Matthew E. Levine, David J. Albers, and Lena Mamykina. 2019. Personal Health Oracle: Explorations of Personalized Predictions in Diabetes Self-Management. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300600

53. John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Transactions on Interactive Intelligent Systems* 8, 2: 8:1-8:37. https://doi.org/10.1145/3185517
54. Daniel A. Epstein, Mira Dontcheva, James Fogarty, and Sean A. Munson. 2020. Yarn: Adding Meaning to Shared Personal Data through Structured Storytelling • Graphics Interface. In *Graphics Interface*, 168–182. Retrieved March 24, 2021 from https://graphicsinterface.org/wp-content/uploads/gi2020-18.pdf
55. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. 731–742. https://doi.org/10.1145/2750858.2804250
56. Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*, 667–676. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=2598558
57. Sarah Faisal, Ann Blandford, and Henry WW Potts. 2013. Making sense of personal health information: challenges for information visualization. *Health informatics journal* 19, 3: 198–217.
58. Andrew Farmer, Oliver Gibson, Paul Hayton, Kathryn Bryden, Christina Dudley, Andrew Neil, and Lionel Tarassenko. 2005. A real-time, mobile phone-based telemedicine system to support young adults with type 1 diabetes. *Journal of Innovation in Health Informatics* 13, 3: 171–177.
59. Daniel J Feller, Marissa Burgermaster, Matthew E Levine, Arlene Smaldone, Patricia G Davidson, David J Albers, and Lena Mamykina. 2018. A visual analytics approach for pattern-recognition in patient-generated data. *Journal of the American Medical Informatics Association* 25, 10: 1366–1374. https://doi.org/10.1093/jamia/ocy054
60. Stephanie L. Fitzpatrick, Kristina P. Schumann, and Felicia Hill-Briggs. 2013. Problem solving interventions for diabetes self-management and control: A systematic review of the literature. *Diabetes Research and Clinical Practice* 100, 2: 145–161. https://doi.org/10.1016/j.diabres.2012.12.016
61. Camilla Forsell and Jimmy Johansson. 2010. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces - AVI '10*, 199. https://doi.org/10.1145/1842993.1843029
62. Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2009. UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 1043–1052. https://doi.org/10.1145/1518701.1518861
63. Jon Froehlich, Leah Findlater, Marilyn Ostergren, Solai Ramanathan, Josh Peterson, Inness Wragg, Eric Larson, Fabia Fu, Mazhengmin Bai, Shwetak Patel, and James A. Landay. 2012. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 2367–2376. https://doi.org/10.1145/2207676.2208397
64. Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells, Jonathan N. Tobin, Esteban G. Tabak, David J. Albers, Arlene M. Smaldone, and Lena Mamykina. 2021. From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition

Goal Recommendations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3411764.3445555

65. Mirta Galesic and Rocio Garcia-retamero. 2011. Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, May-Jun.
66. Rocio Garcia-Retamero, Yasmina Okan, and Edward T. Cokely. 2012. Using Visual Aids to Improve Communication of Risks about Health: A Review. *The Scientific World Journal* 2012: 1–10. https://doi.org/10.1100/2012/562637
67. Albert Gatt, François Portet, Ehud Reiter, Jim Hunter, Saad Mahamood, Wendy Moncur, and Somayajulu Sripada. 2009. From data to text in the Neonatal Intensive Care Unit: Using NLG technology for decision support and information management. *AI Communications* 22, 3: 153–186. https://doi.org/10.3233/AIC-2009-0453
68. Gerd Gigerenzer and Wolfgang Gaissmaier. 2011. Heuristic Decision Making. *Annual Review of Psychology* 62, 1: 451–482. https://doi.org/10.1146/annurev-psych-120709-145346
69. R. E. Glasgow, L. Fisher, M. Skaff, J. Mullan, and D. J. Toobert. 2007. Problem Solving and Diabetes Self-Management: Investigation in a large, multiracial sample. *Diabetes Care* 30, 1: 33–37. https://doi.org/10.2337/dc06-1390
70. Garrett Grolemund and Hadley Wickham. 2014. A Cognitive Interpretation of Data Analysis: A Cognitive Interpretation of Data Analysis. *International Statistical Review* 82, 2: 184–204. https://doi.org/10.1111/insr.12028
71. Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9: 904–908. https://doi.org/10.1177/154193120605000909
72. D. A. Hema, S. O. Roper, J. W. Nehring, A. Call, B. L. Mandleco, and T. T. Dyches. 2009. Daily stressors and coping responses of children and adolescents with type 1 diabetes. *Child: Care, Health and Development* 35, 3: 330–339. https://doi.org/10.1111/j.1365-2214.2009.00937.x
73. Jeremy Heyer, Nirmal Kumar Raveendranath, and Khairi Reda. 2020. Pushing the (Visual) Narrative: The Effects of Prior Knowledge Elicitation in Provocative Topics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 1–14. https://doi.org/10.1145/3313831.3376887
74. Felicia Hill-Briggs. 2003. Problem solving in diabetes self-management: a model of chronic illness self-management behavior. *Annals of Behavioral Medicine* 25, 3: 182–193.
75. Anne Holbrook, Lehana Thabane, Karim Keshavjee, Lisa Dolovich, Bob Bernstein, David Chan, Sue Troyan, Gary Foster, Hertzel Gerstein, and COMPETE II Investigators. 2009. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* 181, 1–2: 37–44. https://doi.org/10.1503/cmaj.081272
76. Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 3, 2: 119–131. https://doi.org/10.1007/s40708-016-0042-6
77. Bastian Ilsø Hougaard and Hendrik Knoche. 2019. Telling the Story Right: How Therapists Aid Stroke Patients Interpret Personal Visualized Game Performance Data. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (PervasiveHealth'19), 435–443. https://doi.org/10.1145/3329189.3329239

78. Peter S. Houts, Cecilia C. Doak, Leonard G. Doak, and Matthew J. Loscalzo. 2006. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 61, 2: 173–190. https://doi.org/10.1016/j.pec.2005.05.004
79. Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2015. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3: 420–433. https://doi.org/10.1109/TVCG.2014.2359887
80. Mary Margaret Huizinga, Tom A Elasy, Kenneth A Wallston, Kerri Cavanaugh, Dianne Davis, Rebecca P Gregory, Lynn S Fuchs, Robert Malone, Andrea Cherrington, Darren A DeWalt, John Buse, Michael Pignone, and Russell L Rothman. 2008. Development and validation of the Diabetes Numeracy Test (DNT). *BMC Health Services Research* 8: 96. https://doi.org/10.1186/1472-6963-8-96
81. J. Hullman and N. Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12: 2231–2240. https://doi.org/10.1109/TVCG.2011.255
82. Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. 2013. A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12: 2406–2415. https://doi.org/10.1109/TVCG.2013.119
83. M. Sazzad Hussain, Juchen Li, Louise A. Ellis, Laura Ospina-Pinillos, Tracey A. Davenport, Rafael A. Calvo, and Ian B. Hickie. 2015. Moderator Assistant: A Natural Language Generation-Based Intervention to Support Mental Health via Social Media. *Journal of Technology in Human Services* 33, 4: 304–329. https://doi.org/10.1080/15228835.2015.1105768
84. Rhianne Jones and Bronwyn Jones. 2019. Atomising the News: The (In)Flexibility of Structured Journalism. *Digital Journalism* 7, 8: 1157–1179. https://doi.org/10.1080/21670811.2019.1609372
85. Simon L. Jones. 2015. Exploring correlational information in aggregated quantified self data dashboards. 1075–1080. https://doi.org/10.1145/2800835.2800963
86. Ravi Karkar, Jasmine Zia, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, and Roger Vilardaga. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. 6850–6863. https://doi.org/10.1145/3025453.3025480
87. Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. 2016. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association* 23, 3: 440–448. https://doi.org/10.1093/jamia/ocv150
88. Dmitri S. Katz, Blaine A. Price, Simon Holland, and Nicholas Sheep Dalton. 2018. Data, Data Everywhere, and Still Too Hard to Link: Insights from User Interactions with Diabetes Apps. 1–12. https://doi.org/10.1145/3173574.3174077
89. Dmitri S. Katz, Blaine A. Price, Simon Holland, and Nicholas Sheep Dalton. 2018. Designing for Diabetes Decision Support Systems with Fluid Contextual Reasoning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3174199

90. 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 2012 ACM Conference on Ubiquitous Computing*, 226–234. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=2370253
91. Elizabeth Kaziunas, Mark S. Ackerman, Silvia Lindtner, and Joyce M. Lee. 2017. Caring through Data: Attending to the Social and Emotional Experiences of Health Datafication. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2260–2272. https://doi.org/10.1145/2998181.2998303
92. Saif Sherif Khairat, Aniesha Dukkipati, Heather Alico Lauria, Thomas Bice, Debbie Travers, and Shannon S. Carson. 2018. The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. *JMIR Human Factors* 5, 2: e9328. https://doi.org/10.2196/humanfactors.9328
93. Taewan Kim, Haesoo Kim, Ha Yeon Lee, Hwarang Goh, Shakhboz Abdigapporov, Mingon Jeong, Hyunsung Cho, Kyungsik Han, Youngtae Noh, Sung-Ju Lee, and Hwajung Hong. 2022. Prediction for Retrospection: Integrating Algorithmic Stress Prediction into Personal Informatics Systems for College Students' Mental Health. In *CHI Conference on Human Factors in Computing Systems*, 1–20. https://doi.org/10.1145/3491102.3517701
94. Yoojung Kim, Eunyoung Heo, Hyunjeong Lee, Sookyoung Ji, Jueun Choi, Jeong-Whun Kim, Joongseek Lee, and Sooyoung Yoo. 2017. Prescribing 10,000 Steps Like Aspirin: Designing a Novel Interface for Data-Driven Medical Consultations. 5787–5799. https://doi.org/10.1145/3025453.3025570
95. Gary Klein, Brian Moon, and Robert R. Hoffman. 2006. Making sense of sensemaking 2: A macrocognitive model. *IEEE intelligent systems* 21, 5: 88–92.
96. Gary Klein, Jennifer K. Phillips, Erica L. Rall, and Deborah A. Peluso. 2007. A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 113–155.
97. Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2: 70:1-70:26. https://doi.org/10.1145/3214273
98. Robert Kosara and Jock Mackinlay. 2013. Storytelling: The Next Step for Visualization. *Computer* 46, 5: 44–50. https://doi.org/10.1109/MC.2013.36
99. Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N Kostis. 2012. Scientific Storytelling Using Visualization. *IEEE Computer Graphics and Applications* 32, 1: 12–19. https://doi.org/10.1109/MCG.2012.24
100. A. M. Lai, P.-Y. S. Hsueh, Y. K. Choi, and R. R. Austin. 2017. Present and Future Trends in Consumer Health Informatics and Patient-Generated Health Data. *Yearbook of Medical Informatics* 26, 01: 152–159. https://doi.org/10.15265/IY-2017-016
101. 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 conference on Wireless Health*, 9.

102. B. Lee, E. K. Choe, P. Isenberg, K. Marriott, and J. Stasko. 2020. Reaching Broader Audiences With Data Visualization. *IEEE Computer Graphics and Applications* 40, 2: 82–90. https://doi.org/10.1109/MCG.2020.2968244
103. Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5: 84–90. https://doi.org/10.1109/MCG.2015.99
104. 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. 6837–6849. https://doi.org/10.1145/3025453.3026038
105. Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-Ah Kang, and Ji Soo Yi. 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 22, 1: 499–508. https://doi.org/10.1109/TVCG.2015.2467195
106. Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 557–566. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=1753409
107. J. Liang and M. L. Huang. 2010. Highlighting in Information Visualization: A Survey. In *2010 14th International Conference Information Visualisation*, 79–85. https://doi.org/10.1109/IV.2010.21
108. Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Personal and Ubiquitous Computing* 20, 6: 985–1000. https://doi.org/10.1007/s00779-016-0960-6
109. Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. *Why and why not* explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2119–2128. https://doi.org/10.1145/1518701.1519023
110. 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 *UbiComp 2006: Ubiquitous Computing* (Lecture Notes in Computer Science), 261–278. https://doi.org/10.1007/11853565_16
111. Ron Loewe and Joshua Freeman. 2000. Interpreting diabetes mellitus: differences between patient and provider models of disease and their implications for clinical practice. *Culture, medicine and psychiatry* 24, 4: 379–401.
112. Tania Lombrozo. 2009. Explanation and categorization: How "why?" informs "what?" *Cognition* 110, 2: 248–253. https://doi.org/10.1016/j.cognition.2008.10.007
113. Maichou Lor and Uba Backonja. 2020. Visualizations Integrated Into Consumer Health Technologies Support Self-management of Chronic Diseases A Systematic Review. *CIN: Computers, Informatics, Nursing* 38, 3: 120–130. https://doi.org/10.1097/CIN.0000000000000583
114. Haley MacLeod, Anthony Tang, and Sheelagh Carpendale. 2013. Personal informatics in chronic illness management. In *Proceedings of Graphics Interface 2013*, 149–156. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=2532155

115. Haley MacLeod, Anthony Tang, and Sheelagh Carpendale. 2013. Personal informatics in chronic illness management. In *Proceedings of Graphics Interface 2013*, 149–156. Retrieved April 17, 2017 from http://dl.acm.org/citation.cfm?id=2532155
116. Narges Mahyar and Melanie Tory. 2014. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics* 20, 12: 1633–1642.
117. Lena Mamykina, Elizabeth M Heitkemper, Arlene M Smaldone, Rita Kukafka, Heather Cole-Lewis, Patricia G Davidson, Elizabeth D Mynatt, Jonathan N Tobin, Andrea Cassells, Carrie Goodman, and George Hripcsak. 2016. Structured scaffolding for reflection and problem solving in diabetes self-management: qualitative study of mobile diabetes detective. *Journal of the American Medical Informatics Association* 23, 1: 129–136. https://doi.org/10.1093/jamia/ocv169
118. 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: 1–8. https://doi.org/10.1016/j.jbi.2017.09.013
119. Lena Mamykina, Matthew E Levine, Patricia G Davidson, Arlene M Smaldone, Noemie Elhadad, and David J Albers. 2016. Data-driven health management: reasoning about personally generated data in diabetes with information technologies. *Journal of the American Medical Informatics Association* 23, 3: 526–531. https://doi.org/10.1093/jamia/ocv187
120. Lena Mamykina, Elizabeth D. Mynatt, and David R. Kaufman. 2006. Investigating health management practices of individuals with diabetes. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 927–936. Retrieved December 1, 2015 from http://dl.acm.org/citation.cfm?id=1124910
121. 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 SIGCHI Conference on Human Factors in Computing Systems* (CHI ’08), 477–486. https://doi.org/10.1145/1357054.1357131
122. Lena Mamykina, Arlene M. Smaldone, and Suzanne R. Bakken. 2015. Adopting the sensemaking perspective for chronic disease self-management. *Journal of Biomedical Informatics* 56: 406–417. https://doi.org/10.1016/j.jbi.2015.06.006
123. Megh Marathe and Kentaro Toyama. 2021. The Situated, Relational, and Evolving Nature of Epilepsy Diagnosis. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3: 1–18. https://doi.org/10.1145/3432916
124. Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez Nieto, and Simon Buckingham Shum. 2020. From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376148
125. Rene Mayrhofer. 2005. AN ARCHITECTURE FOR CONTEXT PREDICTION. *Schriften der Johannes-Kepler-Universität Linz: Trauner Verlag,*: 8.
126. Hayley McBain, Michael Shipley, and Stanton Newman. 2015. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. *BMC Health Services Research* 15, 1. https://doi.org/10.1186/s12913-015-1221-5

127. 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. 5800–5812. https://doi.org/10.1145/3025453.3025589
128. Marian van der Meulen, Robert H. Logie, Yvonne Freer, Cindy Sykes, Neil McIntosh, and Jim Hunter. 2010. When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. *Applied Cognitive Psychology* 24, 1: 77–89. https://doi.org/10.1002/acp.1545
129. M.B. Miles, A.M. Huberman, and J. Saldana. *Qualitative data analysis: a methods sourcebook*. Thousand Oaks: Sage.
130. Margaret E Morris, Qusai Kathawala, Todd K Leen, Ethan E Gorenstein, Farzin Guilak, Michael Labhard, and William Deleeuw. 2010. Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness. *Journal of Medical Internet Research* 12, 2: e10. https://doi.org/10.2196/jmir.1371
131. Shantanu Nundy, Chen-Yuan E. Lu, Patrick Hogan, Anjuli Mishra, and Monica E. Peek. 2014. Using patient-generated health data from mobile technologies for diabetes self-management support: provider perspectives from an academic medical center. *Journal of diabetes science and technology* 8, 1: 74–82.
132. Francisco Nunes and Geraldine Fitzpatrick. 2015. Self-Care Technologies and Collaboration. *International Journal of Human-Computer Interaction* 31, 12: 869–881. https://doi.org/10.1080/10447318.2015.1067498
133. Francisco Nunes, Nervo Verdezoto, Geraldine Fitzpatrick, Morten Kyng, Erik Grönvall, and Cristiano Storni. 2015. Self-Care Technologies in HCI: Trends, Tensions, and Opportunities. *ACM Transactions on Computer-Human Interaction* 22, 6: 1–45. https://doi.org/10.1145/2803173
134. Tom Owen, Jennifer Pearson, Harold Thimbleby, and George Buchanan. 2015. ConCap: Designing to Empower Individual Reflection on Chronic Conditions using Mobile Apps. In *roceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 105–114. https://doi.org/10.1145/2785830.2785881
135. Sharoda A. Paul and Madhu C. Reddy. 2010. Understanding together: sensemaking in collaborative information seeking. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 321–330. Retrieved April 2, 2015 from http://dl.acm.org/citation.cfm?id=1718976
136. Steffen Pauws, Albert Gatt, Emiel Krahmer, and Ehud Reiter. 2019. Making Effective Use of Healthcare Data Using Data-to-Text Technology. In *Data Science for Healthcare: Methodologies and Applications*, Sergio Consoli, Diego Reforgiato Recupero and Milan Petković (eds.). Springer International Publishing, Cham, 119–145. https://doi.org/10.1007/978-3-030-05249-2_4
137. Dana Pavel, Dirk Trossen, Matthias Holweg, and Vic Callaghan. 2013. Lifestyle stories: Correlating user information through a story-inspired paradigm. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*, 412–415. Retrieved April 17, 2017 from http://ieeexplore.ieee.org/abstract/document/6563980/
138. Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, 2–4. Retrieved April 17, 2017 from

https://www.e-education.psu.edu/geog885/sites/www.e-education.psu.edu.geog885/files/geog885q/file/Lesson_02/Sense_Making_206_Camera_Ready_Paper.pdf

139. Peter Pirolli and Daniel Russell. 2011. Introduction to this Special Issue on Sensemaking. *Human-Computer Interaction* 26, 1: 1–8. https://doi.org/10.1080/07370024.2011.556557
140. Rimma Pivovarov and Noémie Elhadad. 2015. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association* 22, 5: 938–947. https://doi.org/10.1093/jamia/ocv032
141. W.H. Polonsky. 1999. *Diabetes Burnout: What to Do When You Can't Take It Anymore*. American Diabetes Association.
142. Zening Qu and Jessica Hullman. 2016. Evaluating Visualization Sets: Trade-offs Between Local Effectiveness and Global Consistency. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (BELIV '16), 44–52. https://doi.org/10.1145/2993901.2993910
143. Zening Qu and Jessica Hullman. 2018. Keeping Multiple Views Consistent: Constraints,Validations, and Exceptions in Visualization Authoring. 10.
144. 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* 1, CSCW: 1–18. https://doi.org/10.1145/3134723
145. Ramana Rao and Stuart K. Card. 1994. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '94), 318–322. https://doi.org/10.1145/191666.191776
146. Alexander Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. 2017. Data Programming: Creating Large Training Sets, Quickly. *arXiv:1605.07723 [cs, stat]*. Retrieved November 3, 2021 from http://arxiv.org/abs/1605.07723
147. John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1163–1172. https://doi.org/10.1145/2556288.2557039
148. Pavel S. Roshanov, Natasha Fernandes, Jeff M. Wilczynski, Brian J. Hemens, John J. You, Steven M. Handler, Robby Nieuwlaat, Nathan M. Souza, Joseph Beyene, Harriette G. C. Van Spall, Amit X. Garg, and R. Brian Haynes. 2013. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *BMJ* 346: f657. https://doi.org/10.1136/bmj.f657
149. Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 269–276. Retrieved April 27, 2017 from http://dl.acm.org/citation.cfm?id=169209
150. J Saldaña. 2013. The coding manual for qualitative researchers. . Sage, Los Angeles, 260–273.
151. D. Z. Sands and J. S. Wald. 2014. Transforming Health Care Delivery Through Consumer Engagement, Health Data Transparency, and Patient-Generated Health Information: *IMIA Yearbook* 9, 1: 170–176. https://doi.org/10.15265/IY-2014-0017

152. Stefan Särnblad, U. Ekelund, and J. \AAman. 2005. Physical activity and energy intake in adolescent girls with Type 1 diabetes. *Diabetic Medicine* 22, 7: 893–899.
153. Gary Scheiner. 2016. Fine-Tuning Your Basal Insulin Regimen. *Diabetes Self-Management*. Retrieved April 11, 2021 from https://www.diabetesselfmanagement.com/managing-diabetes/treatment-approaches/getting-down-to-basals/
154. J. J. Schensul and M.D. LeCompte. 2013. Chapter 10: Sampling in ethnographic research. In *Essential ethnographic methods: a mixed methods approach*. Lanham: AltaMira, 280–318.
155. Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting Patient-Provider Collaboration to Identify Individual Triggers using Food and Symptom Journals. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, 1726–1739. https://doi.org/10.1145/2998181.2998276
156. Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS* 16, 6: 10.
157. Albert M. Selvin and S. Buckingham Shum. 1999. Repurposing Requirements: Improving Collaborative Sense-Making over the Lifecycle. *International Conference on Product_Focused Software Process Improvement*.
158. Michael Shapiro, Douglas Johnston, Jonathan Wald, and Donald Mon. 2012. Patient-generated health data. *RTI International, April*.
159. Brian K. Smith, Jeana Frost, Meltem Albayrak, and Rajneesh Sudhakar. 2007. Integrating glucometers and digital photography as experience capture tools to enhance patient understanding and communication of diabetes self-management practices. *Personal and Ubiquitous Computing* 11, 4: 273–286. https://doi.org/10.1007/s00779-006-0087-2
160. Aaron Springer and Steve Whittaker. 2020. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Transactions on Interactive Intelligent Systems* 10, 4: 1–32. https://doi.org/10.1145/3374218
161. Ryan Steed and Benjamin Williams. 2020. Heuristic-Based Weak Learning for Automated Decision-Making. *arXiv:2005.02342 [cs]*. Retrieved July 20, 2021 from http://arxiv.org/abs/2005.02342
162. M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. 2011. Context-Preserving Visual Links. *IEEE Transactions on Visualization and Computer Graphics* 17, 12: 2249–2258. https://doi.org/10.1109/TVCG.2011.183
163. Charles D Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories. 14.
164. 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: 165–185.
165. Cristiano Storni. 2014. Design challenges for ubiquitous and personal computing in chronic disease care and patient empowerment: a case study rethinking diabetes self-monitoring. *Personal and Ubiquitous Computing* 18, 5: 1277–1290. https://doi.org/10.1007/s00779-013-0707-6

166. Si Sun and Kaitlin L. Costello. 2018. Designing Decision-Support Technologies for Patient-Generated Data in Type 1 Diabetes. *AMIA Annual Symposium Proceedings* 2017: 1645–1654.
167. P. C. Tang, J. M. Overhage, A. S. Chan, N. L. Brown, B. Aghighi, M. P. Entwistle, S. L. Hui, S. M. Hyde, L. H. Klieman, C. J. Mitchell, A. J. Perkins, L. S. Qureshi, T. A. Waltimyer, L. J. Winters, and C. 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: 526–534. https://doi.org/10.1136/amiajnl-2012-001263
168. Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction* 27, 5: 1–53. https://doi.org/10.1145/3398069
169. Konrad Tollmar, Frank Bentley, and Cristobal Viedma. 2012. Mobile Health Mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, 65–72. Retrieved December 13, 2015 from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6240364
170. Conrad S Tucker, Ishan Behoora, Harriet Black Nembhard, Mechelle Lewis, Nicholas W Sterling, and Xuemei Huang. 2015. Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors. *Computers in biology and medicine* 66: 120–134. https://doi.org/10.1016/j.compbiomed.2015.08.012
171. Meghan Reading Turchioe, Annie Myers, Samuel Isaac, Dawon Baik, Lisa V. Grossman, Jessica S. Ancker, and Ruth Masterson Creber. 2019. A Systematic Review of Patient-Facing Visualizations of Personal Health Data. *Applied Clinical Informatics* 10, 4: 751–770. https://doi.org/10.1055/s-0039-1697592
172. Nichole S. Tyler and Peter G. Jacobs. 2020. Artificial Intelligence in Decision Support Systems for Type 1 Diabetes. *Sensors (Basel, Switzerland)* 20, 11: 3214. https://doi.org/10.3390/s20113214
173. Paroma Varma and Christopher Ré. 2018. Snuba: automating weak supervision to label training data. *Proceedings of the VLDB Endowment* 12, 3: 223–236. https://doi.org/10.14778/3291264.3291268
174. G Vespasiani, A Nicolucci, M Saleh, J Sieber, and S Girolami. Development and validation of a pattern recognition engine for visualization of glycemic patterns. 1.
175. Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. 2000. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces* (AVI ’00), 110–119. https://doi.org/10.1145/345513.345271
176. Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300831
177. Karl E. Weick. 1995. *Sensemaking in organizations*. Sage Publications, Inc.
178. 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 SIGCHI Conference on Human Factors in Computing Systems*, 3066–3078. https://doi.org/10.1145/2858036.2858445

179. 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. 1–13. https://doi.org/10.1145/3173574.3174058
180. Ashenafi Zebene Woldaregay, Eirik Årsand, Taxiarchis Botsis, David Albers, Lena Mamykina, and Gunnar Hartvigsen. 2019. Data-Driven Blood Glucose Pattern Classification and Anomalies Detection: Machine-Learning Applications in Type 1 Diabetes. *Journal of Medical Internet Research* 21, 5: e11030. https://doi.org/10.2196/11030
181. Ashenafi Zebene Woldaregay, Eirik Årsand, Ståle Walderhaug, David Albers, Lena Mamykina, Taxiarchis Botsis, and Gunnar Hartvigsen. 2019. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial Intelligence in Medicine* 98: 109–134. https://doi.org/10.1016/j.artmed.2019.07.007
182. Jenise C. Wong, Aaron B. Neinstein, Matthew Spindler, and Saleh Adi. 2015. A Minority of Patients with Type 1 Diabetes Routinely Downloads and Retrospectively Reviews Device Data. *Diabetes Technology & Therapeutics* 17, 8: 555–562. https://doi.org/10.1089/dia.2014.0413
183. Jamie Wood and Anne Peters. 2017. *The type 1 diabetes self-care manual : a complete guide to type 1 diabetes across the lifespan for people with diabetes, parents, and caregivers.* Retrieved April 11, 2021 from http://main.diabetes.org/dorg/PDFs/living-with-diabetes/T1DSelfCareManual.pdf
184. Pawe\l W. Woźniak, Przemys\law Piotr Kucharski, Maartje M.A. de Graaf, and Jasmin Niess. 2020. Exploring Understandable Algorithms to Suggest Fitness Tracker Goals That Foster Commitment. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. Association for Computing Machinery, New York, NY, USA. Retrieved from https://doi.org/10.1145/3419249.3420131
185. Ke Xu, Shunan Guo, Nan Cao, David Gotz, Aiwen Xu, Huamin Qu, Zhenjie Yao, and Yixin Chen. 2018. ECGLens: Interactive Visual Exploration of Large Scale ECG Data for Arrhythmia Detection. 1–12. https://doi.org/10.1145/3173574.3174237
186. Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 1–11. https://doi.org/10.1145/3290605.3300468
187. Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F. Antaki. 2016. Investigating the Heart Pump Implant Decision Process: Opportunities for Decision Support Tools to Help. *ACM transactions on computer-human interaction : a publication of the Association for Computing Machinery* 2016: 4477–4488. https://doi.org/10.1145/2858036.2858373
188. P. Zhang, D. Schmidt, J. White, and S. Mulvaney. 2018. Towards Precision Behavioral Medicine with IoT: Iterative Design and Optimization of a Self-Management Tool for Type 1 Diabetes. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 64–74. https://doi.org/10.1109/ICHI.2018.00015
189. Y. Zhang, K. Chanana, and C. Dunne. 2019. IDMVis: Temporal Event Sequence Visualization for Type 1 Diabetes Treatment Decision Support. *IEEE Transactions on Visualization and Computer Graphics* 25, 1: 512–522. https://doi.org/10.1109/TVCG.2018.2865076

190. Yixuan Zhang, Andrea Grimes Parker, and Cody Dunne. 2020. *Information Visualization for Diabetes Management: A Literature Review*. OSF Preprints. https://doi.org/10.31219/osf.io/b632g
191. Min Zheng, Jan Claassen, and Samantha Kleinberg. 2018. Automated Identification of Causal Moderators in Time-Series Data. 25.
192. Min Zheng and Samantha Kleinberg. 2020. Using Domain Knowledge to Overcome Latent Variables in Causal Inference from Time Series. 19.
193. Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. *Computer Graphics Forum* 38, 3: 675–685. https://doi.org/10.1111/cgf.13719
194. Haining Zhu, Joanna Colgan, Madhu Reddy, and Eun Kyoung Choe. 2016. Sharing Patient-Generated Data in Clinical Practices: An Interview Study. *Proceedings of the American Medical Informatics Association*. Retrieved April 17, 2017 from https://faculty.ist.psu.edu/choe/download/AMIA-2016-Zhu-PGD.pdf
195. Jasmine Zia, Chia-Fang Chung, Jessica Schroeder, Sean Munson, Julie Kientz, James Fogarty, Elizabeth Bales, Jeanette M Schenk, and Margaret Heitkemper. 2017. The feasibility, usability, and clinical utility of traditional paper food and symptom journals for patients with irritable bowel syndrome. *Neurogastroenterology & Motility* 29, 2: e12935. https://doi.org/10.1111/nmo.12935
196. 2021. Diabetes Education Online - Diabetes Teaching Center at the University of California, San Francisco. Retrieved April 11, 2021 from https://dtc.ucsf.edu/
197. 2021. Insulin Basics | ADA. Retrieved April 11, 2021 from https://www.diabetes.org/healthy-living/medication-treatments/insulin-other-injectables/insulin-basics
198. 2021. Understanding A1C | ADA. Retrieved April 11, 2021 from https://www.diabetes.org/a1c