Design, Development, and Evaluation of a Patient-Centered Health Dialog System to Support Inguinal Hernia Surgery Patient Information-Seeking

Ross James Lordon

A dissertation

submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

William Lober, Co-Chair

Uba Backonja, Co-Chair

Heather Evans

Sean Munson

Program Authorized to Offer Degree: Biomedical Informatics & Medical Education

University of Washington

Abstract

Design, Development, and Evaluation a Patient-Centered Health Dialog System to Support Inguinal Hernia Surgery Patient Information-Seeking

Ross James Lordon

Chairs of the Supervisory Committee:

William Lober and Uba Backonja

Biomedical Informatics and Medical Education

Surgery patients engage in health information-seeking activities to better understand their health conditions. An example of this activity is patients collecting data outside of the hospital to track their surgery recovery. Patients can also seek health information from resources such as clinicians, patient education materials, multimedia, friends or family members, and websites to answer their questions. However, surgery patients could encounter barriers when trying to make sense of their collected data or engaging in health information-seeking. For example, clinicians have limited availability to help make sense of the collected data or answer patient questions. Additionally, surgery patients may have low health literacy levels or have difficulties recalling their discharge teaching.

In this dissertation, I will first describe how patients collecting data outside of the hospital has an impact on patient-clinician relationships and patient efforts to make sense of the data through information-seeking. Second, I will detail an innovative approach to aid inguinal hernia surgery patient information-seeking activities - a prototype health dialog system, called “Hernia Coach,” built as a Google Assistant “Action.” To develop Hernia Coach, I investigated the information needs of inguinal hernia surgery patients by engaging patients and clinicians in participatory design sessions. Then I designed and developed Hernia Coach. Finally, I evaluated Hernia Coach to identify the system's ability to aid patient health information-seeking by recruiting design experts to perform a heuristic evaluation, engaging patients in user testing sessions, and evaluating the Hernia Coach’s ability to provide relevant health information to participant queries.

This research contributes to an improved understanding of how patients collecting data to address health concerns affect patient-clinician relationships, expands on inguinal hernia surgery patient informational needs research, and demonstrates how health dialog systems can act as a tool to aid patients with their health information-seeking activities. My work also has implications beyond inguinal hernia surgery. When patients engage in health information-seeking online, the onus is on them to identify material relevant to their health condition that also contains accurate information. Health dialog systems have the potential to act as patient-centered health information repositories and allow patients to easily ask questions and quickly receive reliable answers from a reputable resource. Additionally, my work provides design recommendations for researchers creating these systems and identifies opportunities to improve the evaluation of patient-centered health dialog systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I first would like to acknowledge my reading committee members, William Lober, Uba Backonja, Heather Evans, and Sean Munson. Your mentorship and guidance over the years was instrumental to the success of this work. In addition to my reading committee members, I also acknowledge the mentorship and time invested by my Graduate School Representatives, Lingtak-Neander Chan (current) and Gary Hsieh (former). I would also like to acknowledge my team at the University of Washington's Graduate School, John Drew, Jen Lehner, Steven Allen, and Doug Pardo. Jen, thanks for giving me the opportunity to work and grow my technical skill set while pursuing my PhD. Finally, I would also like to note my appreciation for the NLM training grant (T15LM00744), which provided me with funding for the initial years of my PhD pursuits.

I also would like to acknowledge the wonderful group of people I have worked alongside with during my PhD career. To the many members of the mPOWEr research group, thanks for the opportunity to learn and grow as a researcher during the first few years in my program. To my team at Microsoft's Artificial Intelligence and Research Group, thank you for the opportunity to apply my research skill set to a new and innovative problem. I would also like to specifically thank the wonderful BHI postdocs and students I had the pleasure of working with - Amanda Hall, Shefali Haldar, Jimmy Phuong, Sean Mikles, Nick Robison, Rebecca Hazen, Graham Kim, Laura Kneale, Abdul Alshammari, Ahmad Aljadaan, Lucy Wang, Maher Khelifi, Adam Nishimura, Justin McReynolds, Marea Cobb, Patrick Sanger, Amanda Lazar, Nikhil Gopal, Maggie Dorr, Ryan James, William Kearns, Dae Lee, Piotr Mankowski, and Abhishek Pratap.

DEDICATION

This work is dedicated to my family.

To my parents and sister Claire, your encouragement, empathy, and loving support was and is fundamental to me achieving success both in academics and life.

To my loving partner Anya, you lifted my spirits when I was down, you reminded me there is more to life than work, and you were always ready to head out on an adventure with me when I needed to seek solace and rejuvenation in the great outdoors.

Thank you.

CHAPTER 1. DISSERTATION INTRODUCTION

1.1 Introduction

Surgery patients engage in health information-seeking activities to better understand their health conditions. An example of this activity is patients collecting data outside of the hospital to track their surgery recovery. Patients can also seek health information from resources such as clinicians, patient education materials, multimedia, friends or family members, and online resources to answer their questions. However, surgery patients could encounter barriers when trying to make sense of their collected data or engaging in health information-seeking. For example, clinicians have limited availability to help make sense of the collected data or answer patient questions. Additionally, surgery patients may have low health literacy levels or have difficulties recalling their discharge teaching.

Inguinal hernia repair is a common type of surgery. Patients undergoing inguinal hernia surgery face many challenges throughout their surgery journeys, which are the times before, during, and after surgery. Their journeys consist of multiple clinic appointments, coordinating logistics with friends and family members, arranging time off of work, transitions into and out of the hospital on the day of surgery, recovery after surgery, potential hernia or surgery complications, and health information-seeking activities.

Dialog systems (DSs) and health dialog systems (HDSs) have the potential to act as an additional information resource to support inguinal hernia surgery patient information-seeking. DSs (i.e., virtual assistants, conversational agents, chatbots) are computer systems designed to converse with users using natural language. Users can interact with DSs to accomplish tasks or answer general questions. For example, a surgery patient recovering at home could use a DS to play music or find out additional information about prescription medications. HDSs are a form of DSs that have been created to support patient health information-seeking, promote behavior change, and triage patient concerns. Over the past decade, there has been a large scale adoption of commercial DSs, such as Amazon’s Alexa, Apple’s Siri, Google’s Google Assistant, Microsoft’s Cortana, and Samsung’s Bixby. Developers also now have the ability to extend the functionality of these DSs through the creation of “Skills” or “Actions.” “Skills” and “Actions” are analogous to creating “apps” for smartphones. The ability to create DS “Skills” or “Actions” suggests it is possible to create inguinal hernia focused patient-centered HDS using to aid health information-seeking.

Yet, little to no research has been conducted to identify the scope of patients collecting data to their better understand their health; and how these patients are engaging in health information-seeking with their clinicians to make sense of the collected data. Additionally, the information needs of inguinal hernia surgery patients are underexplored. Furthermore, there are few precedents of patient-centered HDSs, built as DS “Skills” or “Actions”, to aid patient health information-seeking.

To address these gaps, in Aim 1, I conducted a systematic review of the literature to identify the effect of patients collecting data outside of the hospital on patient-clinician relationships within surgery and primary care and how patients engage in health information-seeking to understand their collected data. Then, in Aim 2, I performed a needs assessment using participatory design to further explore the information needs of inguinal hernia surgery patients. I also identify how an HDS could support these patients' health information-seeking activities throughout their surgery journeys. Finally, in Aim 3, I designed, developed, and evaluated a prototype Google Assistant “Action” to aid inguinal hernia surgery patient health information-seeking.

1.2 Dissertation Outline

The subsequent sections of Chapter 1 are comprised of the dissertation outline, the dissertation setting, background and significance, and related work to this dissertation research. Chapter 2 concerns Aim 1, Chapter 3 concerns Aim 2, Chapter 4 concerns Aim 3, which are all detailed below. Chapter 5 discusses the conclusions, limitations, future work, and contributions of this dissertation research. I have included a flowchart in Figure 1 depicting the flow of this dissertation research and how each of the research activities are interconnected

1.2.1 Aim 1: To identify the effect of patients collecting health data outside of the hospital on patient-clinician relationships and how patients engage in health information-seeking to understand their collected data.

In Aim 1 (Chapter 2), I conducted a systematic review of the literature to identify the effect of PGHD and PROs on patient-clinician relationships within surgery and primary care and how patients engage in health information-seeking to understand their collected data. We specifically patient-generated health data (PGHD) and patient-reported outcomes (PROs). Research indicates some patients collecting PGHD or PROs desire their clinicians to regularly review and help make sense of the recorded data. However, clinicians expressed this practice is not feasible for all patients, in all circumstances, which could result in conflicting views regarding the role of PGHD and PROs used in clinical care. The scope of how these patient health information-seeking activities impact clinician relationships is underexplored.

To address this gap, I queried six research databases to identify literature describing or documenting the effects of PGHD or PROs on patient-clinician relationships in surgery and primary care. I then synthesized the identified research articles. Overall, this work demonstrated PGHD and PROs supported patient-clinician health communication. Yet, the technology used to collect data had varied abilities to support patient-clinician collaboration. Furthermore, some patients were unable to independently make sense of their data, sought interpretation assistance from their clinicians, and some clinicians were unable to aid with data interpretation due to barriers such as clinic appointment time constraints.

The final finding was the major motivation for the subsequent work in this dissertation, which was to identify, develop, and evaluate a tool to aid independent surgery patient health information-seeking.

1.2.2 Aim 2: To identify how health dialog systems can support inguinal hernia surgery patient health information-seeking.

In Aim 2 (Chapter 3), I identify how an HDS could support inguinal hernia surgery patient health-information seeking. Little to no research has been conducted to evaluate inguinal hernia surgery patient information needs. Additionally, there are few examples of surgery patient-centered HDSs.

To address this gap, I engaged patients who had previously undergone inguinal hernia surgery and clinicians who care for this type of patient as active design process members in participatory design sessions. The participatory design sessions primarily consisted of brainstorming activities, identification of hernia surgery questions, and exit interviews. Through this work I identified: the patient participant's surgery journeys from diagnosis through recovery, the questions patients ask during surgery journeys, and which patient questions are best suited for an HDS to answer.

The findings from the participatory design sessions provided a foundation for the design, development, and evaluation of an HDS prototype in Aim 3 (Chapter 4) to support inguinal hernia surgery patient health-information seeking.

1.2.3 Aim 3: To design, develop, and evaluate a health dialog system prototype to aid inguinal hernia surgery health patient-information seeking.

While the ability to create DS "Skills" or "Actions" had been possible for a few years, there was little precedent in the literature describing the advantages or disadvantages of using certain commercial DS “Skills or “Actions” for building HDSs. Additionally, we identified a number of examples illustrating the process of building an HDS, albeit not specifically for surgery. Finally, we identified examples evaluating DSs to aid patient health information-seeking, but not HDSs built as “Skills” or “Actions.”

To address this gap, In Aim 3 (Chapter 4), I used the insights generated from Aim 2 (Chapter 3) to design, develop, and evaluate an HDS prototype called “Hernia Coach.” Hernia Coach was built as a Google Assistant “Action” to aid inguinal hernia surgery patient-information seeking. In this study, I first evaluated the top five most widely adopted smartphone-based DSs to identify the best-suited DS for building Hernia Coach as a “Skill” or “Action”. Then I generated Hernia Coach’s health content to answer inguinal hernia patient questions. Subsequently, I created personas to focus Hernia Coach's development efforts on target users and make the scenarios I created and used in the evaluation of Hernia Coach more effective. Then I developed Hernia Coach itself. Finally, I evaluated Hernia Coach using a three-step approach. First, I recruited design experts to conduct a heuristic evaluation of Hernia Coach. Using the findings from the heuristic evaluation I improved Hernia Coach’s design before the next evaluation step. Second, I recruited previous inguinal hernia surgery patients to engage with Hernia Coach in usability testing sessions. Third, I assessed Hernia Coach’s query responses from the heuristic evaluation and patient usability testing sessions for accuracy.

The findings from Hernia Coach's design, development, and evaluation demonstrated HDS "Skills" or "Actions" built using commercial DSs have the potential to provide an innovative platform to facilitate surgery patient health information-seeking. Furthermore, I generated design recommendations for future HDS "Skill" or "Action" design and development research.

1.3 Dissertation Setting

To achieve the three aims previously described, I conducted three studies - a systematic review study, a needs assessment study, and an evaluation study - which are presented in the subsequent chapters of this dissertation. Each of these studies were conducted in the Pacific Northwest Region of the United States. All study procedures involving human participants were approved by the University of Washington's Human Subjects Division (UW HSD). Participant eligibility criteria are described in each dissertation chapter. Participants who enrolled in the studies in-person provided documented informed consent. Remote participants were provided with copies of consent documents and encouraged to ask questions prior to participating in the research study. The UW HSD granted a documented consent waiver for participants remotely engaging in the research studies.

1.4 Background and Significance

I will first describe surgery patient information needs (1.4.1). Second, I will give an overview of the barriers patients currently face when engaging health information-seeking (1.4.2). Third, I will describe why I selected inguinal hernias as the domain focus for Hernia Coach (1.4.3). Fourth, I will illustrate the growth of mobile health applications, DSs, and the scope of these technologies health-related functionalities (1.4.4).

1.4.1 Surgery Patient Information Needs

Surgery patients have information needs prior to surgery [1–8], during the day of surgery [9], and after undergoing surgery [10–14]. Prior to surgery, patients wanted a better understanding of the operation itself [1,3,6,7]. Specifically, patients wanted to better understand common and rare complications [6], pain management, permitted activities [4], and recovery [1,4,7]. In one study, researchers identified head and neck cancer patients were not receiving adequate information pre-surgery cornering outcomes such as altered appearances, functional difficulties, lifestyle changes, and the significance of their condition [2]. Setting expectations about the surgery process and outcomes can improve patient satisfaction [6–8]. In particular, enhanced recovery after surgery (ERAS) is a framework that provides information and sets expectations prior to surgery. This framework gave pre-surgery colorectal patients the ability to ask questions, provided reassurance, facilitated identifying gaps in care, improved understanding, and created trust with their clinicians [7].

A systematic review by Suhonen and Leino-Kilpi (2006) identified the information needs of surgery patients during the day of surgery. Patients wanted information regarding the surgery itself such as outcomes, consequences, risks, anesthesia, and events in the hospital. Nurses also identified psychosocial support as an important component of patient information needs on the day of surgery [9].

After surgery patient information needs primarily concerned self-care at home [10,13,14], identifying concerning signs and symptoms requiring medical care [10,14], and activity levels or activity restrictions [10,11,14], pain management [10,11], and wound care [10,11]. Other information needs concerned wound care, quality of life [10], and physical and psychological needs [12]. I identified only one research article with a focus on inguinal hernia surgery patients [11]. This study focused exclusively on the postoperative period and identified pain management, wound infections, return to intimacy, exercise, driving, and stretching to be the most important information needs for inguinal hernia surgery patients.

1.4.2 Patient Health Information-Seeking Barriers

Patients engaging in health information-seeking may be affected by a number of barriers such as their ability to understand health information, the methods by which they are provided health information, or limited clinician availability to answer questions. Health literacy is the degree to which individuals have the capacity to obtain, process, and understand basic health information needed to make appropriate health decisions [15]. In 2004, the Institute of Medicine reported 90 million Americans have difficulty understanding and acting upon health information [16]. Low health literacy negatively impacts a patient’s ability to understand their personal health, disease, medications, treatment [2,17–20], and written health information [21,22]. Hospitalization, emergency care, and mortality rates have been associated with low health literacy [22,23]. For surgery patients, in particular, low health literacy was associated with poor medical information comprehension and non-adherence to medications [24].

One key factor contributing to low health literacy levels is the ability of patients to read and comprehend written information. Researchers have identified the average U.S. resident reads at an 8th-grade reading level [25] and the average Medicare beneficiary reads at a 5th-grade reading level [26]. Patient low health literacy levels and reading abilities have driven the creation of policies pertaining to patient-education materials. For example, the Joint Commission states patient education materials, such as leaflets and brochures, should be written at a 5th-grade reading level [27]. Patient education materials are provided to patients to aid with health information recall and health condition understanding [28]. Patients have identified education materials as valuable sources of information [29]. Yet, patient education materials have been shown to have high levels of readability, which impacts these documents' ability to improve health understanding [30–34] and patient comprehension prior to surgery [35].

Another common tool to provide surgery patients health information is verbal communication with clinicians. Yet, the verbal exchange of health information makes it difficult for surgery patients to recall the information provided by clinicians [2,36,37]. Surgery patients may seek out additional information from online sources or support groups to aid their recall [2]. Multimedia, such as interactive computer programs and websites, have been shown to improve patient surgery patient comprehension [35,38].

Additionally, patients collecting PGHD may be unable to make sense of their collected data and seek out interpretation assistance from their clinicians. Sanger et al. (2016)

identified surgery patients collecting postoperative wound infection data with an mHealth app, mPOWEr, expect clinicians to review the data and make an assessment. However, clinicians expressed concerns that routine mPOWEr PHGD review would not be practical given current work obligations. The clinicians expressed a preference to only routinely review high-risk patient mPOWEr PGHD [39]. We encountered instances of this occurring while conducting mPOWEr pilot implementation work, which resulted in patient frustrations because their clinicians were not actively reviewing their mPOWEr PGHD and communicating with them about the data.

1.4.3 Inguinal Hernias

We chose to create Hernia Coach to support inguinal hernia surgery patient information-seeking for a number of reasons. Inguinal hernias are a common health condition [40] with an estimated 500,000 - 600,000 people in the United States [41,42] and twenty million people worldwide [43] undergoing surgery to have their inguinal hernia repaired every year. Inguinal hernia surgery patients typically are male (96% of surgeries), middle-aged or older (mean 54 years old) [44–46], and undergo outpatient surgery to repair the hernia [47–49]. Outpatient surgery patients are generally discharged home from the hospital the same day as the surgery. I identified I would be able to focus my efforts while working towards achieving significant impact when creating Hernia Coach due to the high prevalence of inguinal hernias, homogenous demographics, and the patient population’s outpatient hospital experiences.

1.4.4 Smartphone and Dialog System Adoption

Smartphone Adoption

Smartphone adoption has been steadily increasing over the past 14 years. In 2006 sixty-six percent of U.S. adults owned a cell phone [50] but only two percent of all mobile devices were smartphones [51]. In February 2019 the number of U.S. adults who own a smartphone rose to ninety-six percent and eighty-one percent of those devices were smartphones [50]. The level of adoption for persons aged 49 or younger in 2019 is even greater, where ninety-nine percent had a cell phone and ninety-two percent of those cellphones were smartphones [50].

Smartphone owners are using their devices to better understand and improve their health. In 2017, the most recent data available, there were 325,000 mobile health applications (mHealth apps) available for download in major smartphone app stores [52]. In 2018 sixty-four percent of teens and young adults reported using mHealth apps [53] and in 2015 more than half of smartphone owners had downloaded an mHealth app to track physical activity levels or food intake PGHD [54].

Yet, while smartphone owners are using mHealth applications there have been concerns noted in the literature. The majority of mHealth apps and health content incorporated into the mHealth apps are not developed with the input of medical experts or stakeholders [55–58]. Furthermore, most mHealth apps do not adhere to medical evidence guidelines [59].

Dialog System Adoption

DS adoption has also been steadily increasing in the past two years. In 2016, only one percent of U.S. adults had access to a smart speaker providing access to a DS. In 2019 the number of U.S. with access to DSs through a smart speaker rose to twenty-six percent [60,61]. The adoption of DSs has not been limited to just smart speakers, U.S. adults are also accessing DSs on their smartphones. In 2018 sixty-six percent of U.S. adults reported using a smartphone DS and twenty-five percent use their smartphone DS daily [62]. The most common DS use case is information-seeking [61–63] and in a recent survey, seventy-one percent of physicians stated a DS could provide health information to patients [64].

DS users are already engaging in health-information seeking with DSs. A recent systematic review identified 309 health and fitness Alexa “Skills” and Google Assistant “Actions” [65]. The identified “Skills” and “Actions” covered 24 distinct health domains including health education. Amazon, the organization behind Alexa, has also been investing in extending Alexa’s functionality into the health domain. In spring 2019 Amazon announced a Health Insurance Portability and Accountability Act (HIPAA) compliant invite-only “Skills” kit for developers [66]. Amazon also announced some early HIPAA compliant skills created with partnering health organizations including Boston Children’s Hospital [67,68], Providence St. Joseph [69], and Cedars-Sinai [70]. In addition to these “Skills”, others have been created to National Health Service patients in the United Kingdom [71], by WebMD [72], and the Mayo Clinic [73].

Similarly to mHealth apps, researchers have identified notable concerns when using DSs for health information-seeking [74–76], which I will describe in greater detail in the following related work section.

1.5 Related Work: Dialog Systems to Support Patient Health Information-Seeking

In this section, I will describe related work about HDSs, commercial DSs, and surgery specific DS "Skills" or "Actions" to support patient information-seeking.

1.5.1 Health Dialog Systems

Within the last decade, HDSs have been developed to support patient-health information seeking. Goldenthal et al. (2019) created an HDS to support patient information-seeking after ureteroscopy. Ureteroscopy is a common outpatient procedure to treat kidney stones [77]. Vaira et al. (2018) created an HDS, called "MamaBot," to provide pregnant women, mothers, and families with young children information about hospitals, pharmacies, nutrition, emergency management, and pregnancy or child growth issues [78]. Crutzen et al. (2010) created an HDS, called "Bzz," to answer adolescent's questions about sex, drugs, and alcohol [79].

Descriptions of these HDS's development and evaluation efforts and were varied. The ureteroscopy HDS health content generation methods were described but information about the underlying technology powering the HDS was very limited. This HDS's evaluation efforts were also limited in scope. Seven of the twenty patients enrolled in the study used the HDS and participated in a semi-structured follow-up interview. Only two of the patients' perceived HDS benefits are described in the article. Mamabot's development efforts were described in detail. However, we could not identify literature evaluating or describing Mamabot being used by patients. Descriptions of Bzz's health content generation methods and development efforts were very limited. However, the researchers conducted a pilot implementation of Bzz with 929 participants. The participants rated Bzz with low ease of use, highly reliable information, and moderate information quality. The participants also preferred Bzz more than phone hotlines and search engines for health information-seeking. This research demonstrates HDSs to support patient health information-seeking needs to be more rigorously reported, evaluated, and investigated further.

1.5.2 Commercial Dialog Systems

Researchers have also assessed commercial DSs themselves, not "Skills" or "Actions," to aid patient information-seeking. Miner et al. (2016) queried Siri, Google Now (Google Assistant precursor), Cortana, and S Voice (Samsung Bixby precursor) with statements and questions regarding mental health, interpersonal violence, and physical health to identify if the DSs were capable of identifying a crisis situation, responding with respectful language, and referred the user to medical services. The results were varied. Siri and Google Now were able to recognize suicide ideation and refer users to a suicide prevention hotline. Siri recognized statements concerning depression but did not refer users to medical resources. Cortana identified queries concerning rape and referred users to a sexual assault hotline. Siri identified physical health concerns and referred users to nearby hospitals [74].

Bickmore et al. (2018) recruited participants to query Alexa and the Google Assistant for medical, medication, and emergency situation information. The researchers identified 44 instances that potentially could have resulted in user harm. Thirty percent of these instances were attributed to incorrect information provided the DS, 46% to user error, and 25% consisting of both incorrect DS information and user error. In 21% of instances, the user-provided partial information in their query, which yielded an incorrect query response. In 16% of instances, the user-provided complete information in their query but the DS responded with a partially valid query response. Overall, Siri was identified to have caused the most potential harm and potential death, the Google Assistant was second, and Alexa was third [75].

Palancia et al. (2019) identified how effectively Alexa, the Google Assistant, and Siri identified the top 50 prescribed medication names. The Google Assistant had the highest accuracy rates, followed by Siri, and then Alexa. The researchers also found foreign accents affected Alexa’s and Siri’s medication name understanding [76].

Cho (2019) found when DS users engaged in non-sensitive health information-seeking using the Google Assistant, voice-based interaction modalities enhanced the social presence of the DS and users reported fewer privacy concerns. When users engaged in sensitive health-information seeking with high privacy concerns, the social presence of the Google Assistant did not differ based on voice-based or text-based interaction modalities. Cho also found the hardware used to interact with the DS (e.g., smart speaker, smartphone) did not have an impact on user attitudes when engaging in health information-seeking [80].

This research indicates commercial DSs are poorly equipped to support patient-information seeking. These technologies have been developed to support accomplishing tasks, solving problems, or answering general questions. Supporting patient health information-seeking has not been a focus for the organizations developing these commercial DSs. The creation of health domain-specific DS “Skills” or “Actions” has the potential to mitigate the poor health information provided by DSs and better support patient health information-seeking.

1.5.3 Surgery Focused Dialog Systems “Skills” and “Actions”

I identified two surgery patient-centered HDS “Skills” or “Actions” available through DSs. The first HDS I identified was “Hospital for Special Surgery,” which was available as both an Alexa “Skill” and Google Assistant “Action.” This HDS had limited functionality focused on connecting patients with resources at the New York-based hospital [23]. Additionally, the HDS only had one health information component about back injury pain, which then directed users to contact the hospital for an appointment. The second HDS I identified was “My Children’s Enhanced Recovery After Surgery” [67,68], which was an early HIPAA compliant Alexa “Skill.” This “Skill” was created by Boston’s Children Hospital to support parents and caregivers of children after cardiac surgery. Specifically, the “Skill” facilitated parents and caregivers providing updates to their child’s healthcare team concerning their child’s recovery progress after cardiac surgery.

Parents and caregivers could also obtain information about post-surgery appointments using the “Skill”.

Of the two identified surgery “Skills,” one primarily supported connecting users with hospital resources and provided very limited health information. The other focused on collecting pediatric surgery recovery outcomes. This related work demonstrates the potential of DS to support health information-seeking and HDS built as “Skills” or “Actions.” However, this research leaves many questions unanswered. Little is known about which DS is best suited to build HDSs, the development process required to build HDSs as “Skills” or “Actions”, and if HDS “Skills” or “Actions” can effectively support patient health information-seeking.

1.6 References

1 Hörmann S, Doering S, Richter R, *et al.* Patients’ need for information before surgery. *Psychother Psychosom Med Psychol* 2001;**51**:56–61.
2 Newell R, Ziegler LE, Stafford N, *et al.* The information needs of head and neck cancer patients prior to surgery. *Ann R Coll Surg Engl* 2004;**86**:407–10.
3 McGaughey J, Harrisson S. Understanding the pre-operative information needs of patients and their relatives in intensive care units. *Intensive Crit Care Nurs* 1994;**10**:186–94. doi:10.1016/0964-3397(94)90019-1
4 Sjöling M, Norbergh K-G, Malker H, *et al.* What information do patients waiting for and undergoing arthroplastic surgery want? Their side of the story. *J Orthop Nurs* 2006;**10**:5–14. doi:10.1016/j.joon.2005.09.014
5 De GJ, Abbotts S, Collins B, *et al.* Same day surgery: results of a patient satisfaction survey. *J Qual Clin Pract* 1996;**16**:165–73.
6 Ivarsson B, Larsson S, Lührs C, *et al.* Extended written pre-operative information about possible complications at cardiac surgery—do the patients want to know? *Eur J Cardiothorac Surg* 2005;**28**:407–14. doi:10.1016/j.ejcts.2005.05.006
7 Aasa A, Hovbäck M, Berterö CM. The importance of preoperative information for patient participation in colorectal surgery care. *J Clin Nurs* 2013;**22**:1604–12. doi:10.1111/jocn.12110
8 Pager CK. Randomised controlled trial of preoperative information to improve satisfaction with cataract surgery. *Br J Ophthalmol* 2005;**89**:10–3. doi:10.1136/bjo.2004.048637

9 Suhonen R, Leino-Kilpi H. Adult surgical patients and the information provided to them by nurses: A literature review. *Patient Educ Couns* 2006;**61**:5–15. doi:10.1016/j.pec.2005.02.012
10 Pieper BA, Sieggreen MY, Freeland BS, *et al.* Discharge information needs of patients after surgery. *J Wound Ostomy Cont Nurs Off Publ Wound Ostomy Cont Nurses Soc* 2006;**33**:281–9; quiz 290–1.
11 Bradshaw C, Pritchett C, Bryce C, *et al.* Information needs of general day surgery patients. *Ambul Surg* 1999;**7**:39–44. doi:10.1016/S0966-6532(98)00048-1
12 Yiu HYM, Chien W-T, Lui MH-L, *et al.* Information needs of Chinese surgical patients on discharge:a comparison of patients' and nurses' perceptions. *J Adv Nurs* 2011;**67**:1041–52. doi:10.1111/j.1365-2648.2010.05528.x
13 Galloway S, Graydon J. Uncertainty, symptom distress, and information needs after surgery for cancer of the colon. https://ovidsp.tx.ovid.com/sp-3.33.0b/ovidweb.cgi?QS2=434f4e1a73d37e8c163b3d0d52d889b103368af7d714733736581ed5dcdf04c2b1ee3a51911cf2b8271926549a89548cd66fc62a00111c8e98fa78fbc75c179ac98286330e78d93642e54a05be9b20792c53e86fe807fe342b7ae67c4705f330400637dd5869e0c43c24b4035638ddfadfac258e845bbd162bb5ccbcdc89f340c993bae02e59cdedf1005ba47857bc4421d073d362ab029345a3891a9c304bbc153f511d009633d65cc0f5875b9951dbe54569726a8baa52833305af25f321614252ba1fbdd82c63454b28e0adc0ef340f877469c24b649315452d27bbc3a6a0dd4eb02a905dfe503a5a689c50488f5da6eeb560b6bc7cb26d32808f7ba82625a89652af7c679153 (accessed 6 Jun 2019).
14 Henderson A, Chien WT. Information needs of Hong Kong Chinese patients undergoing surgery. *J Clin Nurs* 2004;**13**:960–6.
15 National Center for Health Statistics (U.S.), editor. *Healthy People 2010: final review*. Hyattsville, MD: : U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics 2012.
16 The Institute of Medicine. *Health Literacy: A Prescription to End Confusion*. National Academies Press 2004. doi:10.17226/10883
17 Gerber BS, Pagcatipunan M, Smith EV, *et al.* The assessment of diabetes knowledge and self-efficacy in a diverse population using Rasch measurement. *J Appl Meas* 2006;**7**:55–73.
18 Guerra CE, Dominguez F, Shea JA. Literacy and knowledge, attitudes, and behavior about colorectal cancer screening. *J Health Commun* 2005;**10**:651–63. doi:10.1080/10810730500267720
19 Kalichman SC, Benotsch E, Suarez T, *et al.* Health literacy and health-related knowledge among persons living with HIV/AIDS. *Am J Prev Med* 2000;**18**:325–31.
20 Paasche-Orlow MK, Riekert KA, Bilderback A, *et al.* Tailored education may reduce health literacy disparities in asthma self-management. *Am J Respir Crit Care Med* 2005;**172**:980–6. doi:10.1164/rccm.200409-1291OC
21 Weiss BD, Reed RL, Kligman EW. Literacy skills and communication methods of low-income older persons. *Patient Educ Couns* 1995;**25**:109–19.
22 Berkman ND, Sheridan SL, Donahue KE, *et al.* Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;**155**:97–107. doi:10.7326/0003-4819-155-2-201107190-00005
23 Baker DW, Gazmararian JA, Williams MV, *et al.* Functional health literacy and

the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health* 2002;**92**:1278–83. doi:10.2105/ajph.92.8.1278
24 De Oliveira GS, McCarthy RJ, Wolf MS, *et al.* The impact of health literacy in the care of surgical patients: a qualitative systematic review. *BMC Surg* 2015;**15**:86. doi:10.1186/s12893-015-0073-6
25 Davis TC, Wolf MS. Health literacy: implications for family medicine. *Fam Med* 2004;**36**:595–8.
26 United States Government Accountability Office Report to Congressional Requesters. Medicare: Communications to Beneficiaries on the Prescription Drug Benefit Could Be Improved. http://www.gao.gov/new.items/d06654.pdf. (accessed 18 Jun 2019).
27 The Joint Commission: Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care: A Roadmap for Hospitals. Oakbrook Terrace, IL: : The Joint Commission 2010.
28 Cornet S, Neal W, Ordelt K. In Plain Words: Creating Easy-to-Read Handouts Writer's Tool Box. The Ohio State University
29 Moerenhout T, Borgermans L, Schol S, *et al.* Patient health information materials in waiting rooms of family physicians: do patients care? *Patient Prefer Adherence* 2013;**7**:489–97. doi:10.2147/PPA.S45777
30 Agarwal N, Hansberry DR, Sabourin V, *et al.* A Comparative Analysis of the Quality of Patient Education Materials From Medical Specialties. *JAMA Intern Med* 2013;**173**:1257–9. doi:10.1001/jamainternmed.2013.6060
31 Walsh TM, Volsko TA. Readability Assessment of Internet-Based Consumer Health Information. *Respir Care* 2008;**53**:1310–5.
32 Johansson K, Nuutila L, Virtanen H, *et al.* Preoperative education for orthopaedic patients: systematic review. *J Adv Nurs* 2005;**50**:212–23. doi:10.1111/j.1365-2648.2005.03381.x
33 Hansberry DR, Agarwal N, Shah R, *et al.* Analysis of the readability of patient education materials from surgical subspecialties. *The Laryngoscope* 2014;**124**:405–12. doi:10.1002/lary.24261
34 Protheroe J, Estacio EV, Saidy-Khan S. Patient information materials in general practices and promotion of health literacy: an observational study of their effectiveness. *Br J Gen Pract* 2015;**65**:e192–7. doi:10.3399/bjgp15X684013
35 Mulsow JJW, Feeley TM, Tierney S. Beyond consent—improving understanding in surgical patients. *Am J Surg* 2012;**203**:112–20. doi:10.1016/j.amjsurg.2010.12.010
36 Pollack AH, Backonja U, Miller AD, *et al.* Closing the Gap: Supporting Patients' Transition to Self-Management after Hospitalization. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. Santa Clara, California, USA: : ACM Press 2016. 5324–36. doi:10.1145/2858036.2858240
37 Kessels RPC. Patients' memory for medical information. *J R Soc Med* 2003;**96**:219–22. doi:10.1258/jrsm.96.5.219
38 Özkan S, Mellema JJ, Nazzal A, *et al.* Online Health Information Seeking in Hand and Upper Extremity Surgery. *J Hand Surg* 2016;**41**:e469–75. doi:10.1016/j.jhsa.2016.09.006
39 Sanger P, Hartzler A, Armstrong C, *et al.* A patient-centered system in a

provider-centered world: challenges of incorporating post-discharge wound data into practice. - PubMed - NCBI. *JAMIA* Published Online First: May 2016.https://www.ncbi.nlm.nih.gov/pubmed/26977103
40 Feliciano D, Hawn M, Strand N. Groin Hernia Repair Inguinal and Femoral. American College of Surgeons 2018. https://www.facs.org/~/media/files/education/patient%20ed/groin_hernia.ashx (accessed 31 Jan 2019).
41 Inguinal Hernia Repair Surgery Information from SAGES. SAGES. https://www.sages.org/publications/patient-information/patient-information-for-laparoscopic-inguinal-hernia-repair-from-sages/ (accessed 15 Mar 2018).
42 Cullen KA, Margaret H, Aleksander G. Ambulatory Surgery in the United States, 2006. *CDC - Natl Health Stat Rep* 2009;:28.
43 Bay-Nielsen M, Kehlet H, Strand L, *et al.* Quality assessment of 26,304 herniorrhaphies in Denmark: a prospective nationwide study. *Lancet Lond Engl* 2001;**358**:1124–8. doi:10.1016/S0140-6736(01)06251-1
44 Nordin P, Zetterström H, Gunnarsson U, *et al.* Local, regional, or general anaesthesia in groin hernia repair: multicentre randomised trial. *The Lancet* 2003;**362**:853–8. doi:10.1016/S0140-6736(03)14339-5
45 McCormack K, Scott NW, Go PM, *et al.* Laparoscopic techniques versus open techniques for inguinal hernia repair. *Cochrane Database Syst Rev* 2003;:CD001785. doi:10.1002/14651858.CD001785
46 Beadles CA, Meagher AD, Charles AG. Trends in emergent hernia repair in the United States. *JAMA Surg* 2015;**150**:194–200. doi:10.1001/jamasurg.2014.1242
47 Hewitt DB, Chojnacki K. Groin Hernia Repair by Open Surgery. *JAMA* 2017;**318**:764–764. doi:10.1001/jama.2017.9868
48 Minimally-Invasive Hernia Repair | Hernia Surgery Center | Dartmouth-Hitchcock. https://www.dartmouth-hitchcock.org/hernia/minimally_invasive_hernia_repair.html (accessed 1 Jul 2019).
49 Inguinal hernia repair - How it’s performed. UK NHS. 2017.https://www.nhs.uk/conditions/inguinal-hernia-repair/what-happens/ (accessed 1 Jul 2019).
50 Demographics of Mobile Device Ownership and Adoption in the United States | Pew Research Center. Pew Research Center 2019. https://www.pewinternet.org/fact-sheet/mobile/ (accessed 11 Jul 2019).
51 Lella A. U.S. Smartphone Penetration Surpassed 80 Percent in 2016. ComScore Inc. http://www.comscore.com/Insights/Blog/US-Smartphone-Penetration-Surpassed-80-Percent-in-2016
52 Research2Guidance. mHealth App Economics 2017/2018 Current Status and Future Trends in Mobile Health. Berlin, Germany: 2017. https://research2guidance.com/wp-content/uploads/2017/11/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf
53 Rideout V, Fox S. Digital Health Practices, Social Media Use, and Mental Well-Being Among Teens and Young Adults in the U.S. *Hopelab Well Trust* 2018.
54 Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR MHealth UHealth* 2015;**3**. doi:10.2196/mhealth.4924
55 Cook VE, Ellis AK, Hildebrand KJ. Mobile health applications in clinical practice:

pearls, pitfalls, and key considerations. *Ann Allergy Asthma Immunol* 2016;**117**:143–9. doi:10.1016/j.anai.2016.01.012
56 Wallace LS, Dhingra LK. A systematic review of smartphone applications for chronic pain available for download in the United States. *J Opioid Manag* 2014;**10**:63–8. doi:10.5055/jom.2014.0193
57 Rivera J, McPherson A, Hamilton J, *et al.* Mobile Apps for Weight Management: A Scoping Review. *JMIR MHealth UHealth* 2016;**4**. doi:10.2196/mhealth.5115
58 Lalloo C, Jibb LA, Rivera J, *et al.* “There’s a Pain App for That”: Review of Patient-targeted Smartphone Applications for Pain Management. *Clin J Pain* 2015;**31**:557–63. doi:10.1097/AJP.0000000000000171
59 Subhi Y, Bube SH, Rolskov Bojsen S, *et al.* Expert Involvement and Adherence to Medical Evidence in Medical Mobile Phone Apps: A Systematic Review. *JMIR MHealth UHealth* 2015;**3**:e79. doi:10.2196/mhealth.4169
60 Voicebot.ai. Smart Speaker Consumer Adoption Report - March 2018. Voicebot.ai 2018. https://voicebot.ai/wp-content/uploads/2018/10/voicebot-smart-speaker-consumer-adoption-report.pdf (accessed 17 Jan 2019).
61 Voicebot.ai. Smart Speaker Consumer Adoption Report - March 2019. Voicebot.ai 2019.
62 Voicebot.ai. Voice Assistant Consumer Adoption Report 2018. Voicebot.ai 2018.
63 Microsoft. Voice Report - From answers to action: customer adoption of voice technology and digital assistants. Microsoft 2019.
64 Palanica A, Flaschner P, Thommandram A, *et al.* Physicians’ Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J Med Internet Res* 2019;**21**:e12887. doi:10.2196/12887
65 Chung AE, Griffin AC, Selezneva D, *et al.* Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis. *JMIR MHealth UHealth* 2018;**6**:e174. doi:10.2196/mhealth.9705
66 Alexa HIPAA Compliant Skills. https://developer.amazon.com/alexa-skills-kit/healthcare-skills (accessed 5 Apr 2019).
67 Amazon.com: MyChildren’s Enhanced Recovery After Surgery: Alexa Skills. https://www.amazon.com/MyChildrens-Enhanced-Recovery-After-Surgery/dp/B07QB7PQYW (accessed 3 Jul 2019).
68 My Children’s Enhanced Recovery After Surgery (ERAS) Alexa Skill. Voice.Health. https://www.voice.health/news/my-childrens-enhanced-recovery-after-surgery-eras-alexa-skill (accessed 3 Jul 2019).
69 Providence St. Joseph Health Makes Same-Day Express Care Appointment Scheduling Available On Amazon Alexa. https://www.prnewswire.com/news-releases/providence-st-joseph-health-makes-same-day-express-care-appointment-scheduling-available-on-amazon-alexa-300824573.html (accessed 5 Apr 2019).
70 Cedars-Sinai Taps Alexa for Smart Hospital Room Pilot. Cedars-Sinai Taps Alexa Smart Hosp. Room Pilot. http://www.cedars-sinai.org/newsroom/cedars-sinai-taps-alexa-for-smart-hospital-room-pilot/ (accessed 26 Feb 2019).
71 Schwartz E. Amazon Alexa Offers Health Advice to Brits Straight From the NHS. Voicebot. 2019.https://voicebot.ai/2019/07/11/amazon-alexa-offers-health-advice-to-brits-straight-from-the-nhs/ (accessed 11 Jul 2019).
72 Amazon.com: WebMD: Alexa Skills. https://www.amazon.com/WebMD-Health-

Corp/dp/B01MRM361G (accessed 11 Jul 2019).
73 Amazon.com: Mayo Clinic First Aid: Alexa Skills. https://www.amazon.com/Mayo-Clinic-First-Aid/dp/B0744LJCV2 (accessed 11 Jul 2019).
74 Miner AS, Milstein A, Schueller S, *et al.* Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Intern Med* 2016;**176**:619–25. doi:10.1001/jamainternmed.2016.0400
75 Bickmore TW, Trinh H, Olafsson S, *et al.* Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *J Med Internet Res* 2018;**20**:e11510. doi:10.2196/11510
76 Palanica A, Thommandram A, Lee A, *et al.* Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. *Npj Digit Med* 2019;**2**:55. doi:10.1038/s41746-019-0133-x
77 Goldenthal S, Portney D, Steppe E, *et al.* Assessing the feasibility of a chatbot after ureteroscopy. *mHealth* 2019;**5**.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458345/ (accessed 30 Apr 2019).
78 Vaira L, Bochicchio MA, Conte M, *et al.* MamaBot: A System Based on ML and NLP for Supporting Women and Families During Pregnancy. In: *Proceedings of the 22Nd International Database Engineering & Applications Symposium*. New York, NY, USA: : ACM 2018. 273–277. doi:10.1145/3216122.3216173
79 Crutzen R, Peters G-JY, Portugal SD, *et al.* An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. *J Adolesc Health* 2011;**48**:514–9. doi:10.1016/j.jadohealth.2010.09.002
80 Cho E. Hey Google, Can I Ask You Something in Private? In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2019. 258:1–258:9. doi:10.1145/3290605.3300488

CHAPTER 2. THE EFFECT OF PATIENT DATA COLLECTED OUTSIDE OF HOSPITAL SETTINGS ON PATIENT-CLINICIAN RELATIONSHIPS: A SYSTEMATIC REVIEW

2.1 Introduction

Many people in the United States (U.S.) have access to smartphones and may be interested in health tracking. In 2018, 77% of U.S. adults owned a smartphone [1] and in 2016 at least 325,000 mobile health applications (mHealth apps) were available for download [2]. Sixty-four percent of teens and young adults reported using mHealth apps in 2018 [3]. Additionally, Krebs and Duncan (2015) found that more than half of smartphone owners had downloaded an mHealth app to track a health condition and used the app at least once per day. The most common reasons for downloading an app were to track physical activity levels, record diet intake, or to learn more about exercising [4].

The data collected by patients are examples of patient-generated health data (PGHD) and patient reported outcomes (PROs). PGHD and PROs are "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" [5]. PROs are a form of PGHD captured at discrete intervals in a patient's home, clinic, or hospital [6]. The impetus for collecting these data may be due to a patient's self-motivation to track their health or at the direction of a clinician [5]. Examples of PGHD and PROs include patient health history, symptoms, biometric data, treatment history, health behaviors, satisfaction, and quality of life [5–7]. These data can be passively or actively collected by patients using mHealth apps, wearable devices (e.g., activity trackers), medical devices (e.g. continuous glucose monitoring systems) [8], or validated questionnaires [9] administered using mobile devices [10].

Almost half of patients or caregivers who have collected PGHD report that the practice has changed their approach to maintaining their health [11]. They may ask clinicians new questions, seek second opinions, or reflect differently about how they make healthcare decisions. Unfortunately, patients and clinicians may disagree about how PGHD should be used to address health concerns; these disagreements could negatively impact patient-clinician relationships [12]. For example, patients may expect their healthcare team to review their PGHD and respond within a short time frame about the meaning of those data, which could disrupt clinical workflows. Additionally, clinicians have expressed concerns about workflow impact when using PROs during time limited clinic appointments, such as discussing additional topics with patients [13].

Some research has begun to identify how PGHD and PROs affect patient-clinician relationships. However, previous systematic reviews have focused on: clinician perceptions of PGHD quality [14], PGHD to support diabetes self-management and education [15], the use of PROs in randomized clinical control trials [16,17], strategies to improve PRO data collection [18], and assessing PRO implementations in specific health domains [19–21]. Our objective was to systematically review the literature to identify the effect of PGHD and PROs on patient-clinician relationships within surgery and primary care.

2.2 Methods

2.2.1 Focus and Search Strategy

In this review, we focused on how PGHD and PROs, collected by surgery and primary care patients in everyday life (i.e., outside the clinic), have an effect on patient-clinician relationships. We initially focused on surgery because of the potential tensions between patients and clinicians using PGHD identified in a previous study [12]. Patients and clinicians disagreed about the use of unstructured PGHD, the frequency of recording PGHD, electronic messaging about PGHD, and their goals for using PGHD. In addition to surgery, we included primary care in our search strategy due to the large number of mHealth apps developed to facilitate collecting PGHD for chronic conditions [22–27], which are managed by primary care clinicians [28,29]. We also limited our scope to data collected outside of clinical settings to align with the definition of PGHD and PROs [5].

The research team initially developed our search terms by identifying keywords and MeSH terms associated with the focus this review. Database searches were refined with the assistance of a health sciences librarian. Using a comprehensive search strategy, we queried six research databases that focus on health or information technology domains: MEDLINE, Embase, CINAHL Plus, PsychINFO, IEEE Xplore, and the ACM Digital Library for articles published between January 1st 2006 and October 13th 2017 (the date on which databases were queried). We chose this timeframe for two reasons. First, we wanted to focus on recent PGHD and PRO developments during the transition of mobile devices from predominantly cellular phones to smartphones, which can be used to facilitate PGHD and PRO data collection and sharing with clinicians [8–10]. In 2006, smartphone ownership in the U.S. constituted approximately 2% of all mobile phones [30] and increased to 77% in 2018 [1]. Second, during that time frame the American Recovery and Reinvestment Act (ARRA) was passed in 2009 [31]. A component of ARRA was the Health Information Technology for Economic and Clinical Health Act [32], which required hospitals to accept digital PGHD from a subset of their patients [33]. The search strategy was supplemented by hand searching the citations contained within systematic reviews that were retrieved by the initial search queries. The search strategies used for this review can be found in supplemental table 1.

2.2.2 Eligibility Criteria and Screening

We included all primary research publications describing or documenting the effects of PGHD or PROs on patient-clinician relationships in surgery and primary care when the data were collected outside of clinical settings, whether done at the direction of a clinician, as part of a research study, or initiated by the patient. Publications retrieved by the queries that incorporated other health domains (e.g., gastroenterology) in addition to surgery or primary care were also included. We excluded publications exclusively focusing on patient or clinician satisfaction using PGHD or PRO platforms as they did not address how PGHD and PROs affect the relationships between patients and clinicians.

Retrieved articles were uploaded into Covidence [34], a platform that facilitates abstract screening and full-text eligibility assessment activities for systematic reviews. Duplicates were identified and removed prior to abstract screening using Covidence and manual review. Ross Lordon (RL) and Sean Mikles (SM) independently screened titles and abstracts against the inclusion and exclusion criteria using Covidence. Disagreements were resolved by consensus. Full-text articles were then reviewed and independently assessed against the inclusion and exclusion criteria by RL and SM. Disagreements were resolved with independent assessment by Laura Kneale (LK). Cohen's Kappa was calculated using STATA [35] to assess RL and SM's inter-rater agreement on full-text study eligibility.

2.2.3 Data Extraction

The following information was extracted from the included publications: objective, participant demographics, the types of PGHD and PROs collected by patients, and the published results or findings. All text labeled results or findings from the included publications were copied verbatim and uploaded into Dedoose [36].

2.2.4 Synthesis of Findings

To guide our reporting, we followed the Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) statement [37], which informed our thematic synthesis of the publication results and findings [38]. The ENTREQ statement consists of 21 items to promote transparency in qualitative synthesis research and is analogous to the quantitatively focused Preferred reporting items for systematic review and meta-analyses statement. We chose to use thematic synthesis because it is widely used in healthcare systematic reviews [37] and has been previously used in a review focusing on patient perspectives of patient-physician relationships [39].

A hybrid approach using both inductive and deductive methods was used to code the text included for final synthesis [40]. An initial code book was developed based on previous research regarding the potential effects of PGHD on patient-clinician relationships [12]. RL, SM, and LK initially coded a subset of the studies line-by-line independently and met to resolve coding discrepancies. The subsequent studies were all coded by RL and half were coded by SM and LK. SM and LK resolved coding discrepancies with RL.

Once coding was complete, the authors met to develop descriptive themes regarding how PGHD and PROs affect patient-clinician relationships. The ENTREQ statement calls for the comparison method within and across the included studies to be explicitly identified. We chose the Systems Engineering Initiative for Patient Safety (SEIPS) Implementation Model 2.0 [41] as our comparison method. The model was developed to facilitate the comparison of multiple health information system specific socio-technical factors, work processes, and outcomes. RL summarized all of the publication findings according to the descriptive themes organized by the SEIPS implementation model. SM and LK each reviewed half of the publication descriptive themes summaries. Final descriptive theme summaries for each publication were agreed upon by the researchers. Once the descriptive theme summaries were developed, RL, SM, and LK independently drafted analytical themes and met to develop a final set of analytical themes. The final analytical themes reflected the codes and descriptive themes identified during the first two phases of the thematic synthesis.

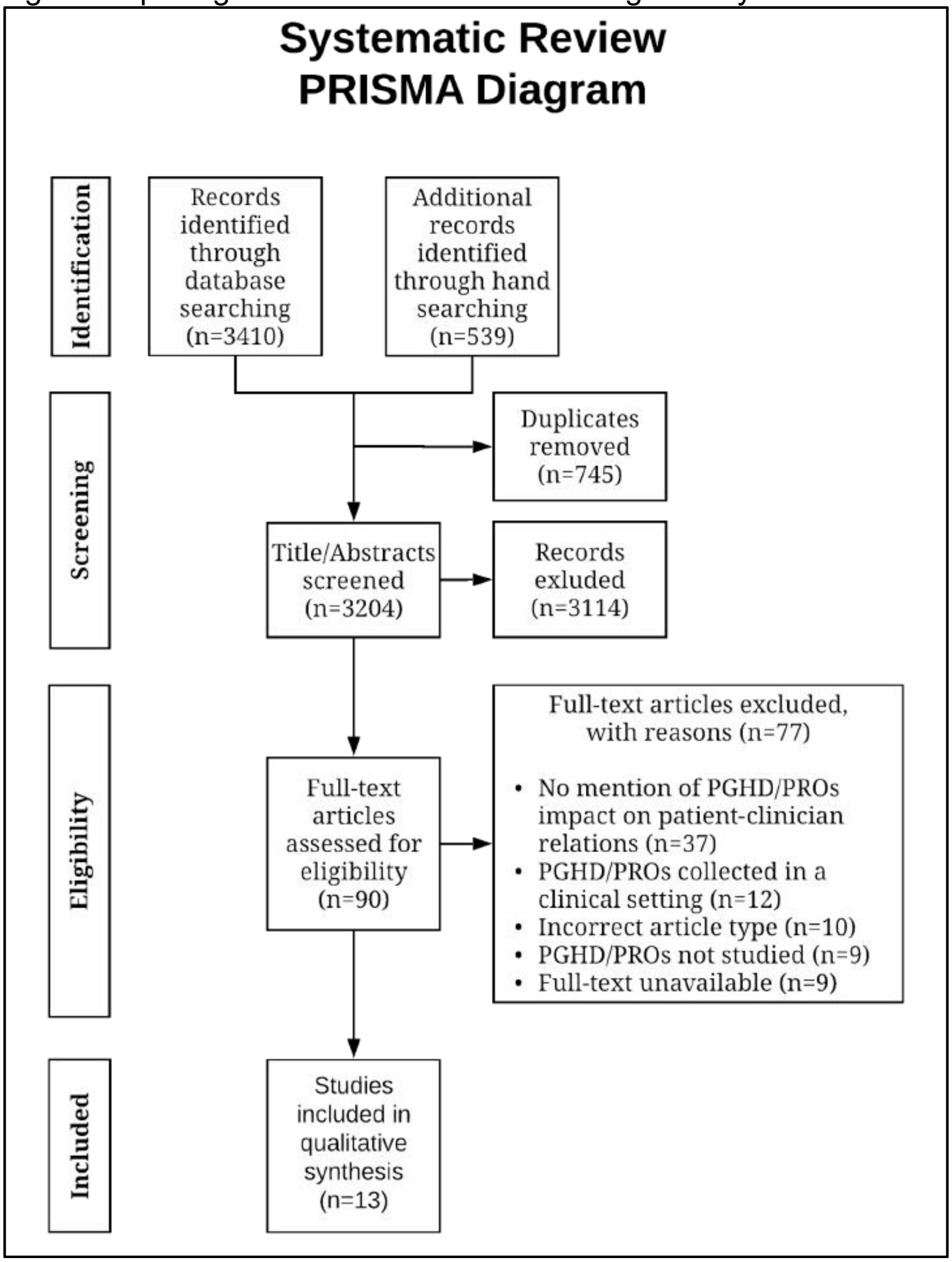
Finally, RL assessed the quality of the included articles using the Mixed Methods Appraisal Tool (MMAT) [42], which is designed to assess the quality of a heterogeneous body of literature utilizing qualitative, quantitative, or mixed methods [43]. MMAT quality scores range from one (one criteria component met) to four (all criteria met). The components assess the studies to determine if criteria, such as the presence of clear research questions or appropriate analysis methods, are incorporated into the included research articles for synthesis.

2.3 Results

2.3.1 Identification and Selection

We identified 3,204 publications. Of those, 3,114 (97.2%) did not meet the inclusion criteria and 90 (2.8%) publications were included in full-text review. Of the 90 full-texts, 77 (85.6%) did not meet the inclusion criteria (see figure 1 for a list of exclusion reasons). Thirteen (14.4%) of the 90 publications were included for final qualitative synthesis [44–56]. None of the included publications were excluded based on MMAT quality ratings. Figure 1 depicts the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRIMSA) diagram [57]. The PRISMA diagram illustrates the number of records identified and included and excluded through the screening process. Inter-rater agreement for the full-text screening process was 0.89, indicating almost perfect agreement [58].

Figure 1: Preferred reporting items for systematic review and meta-analyses (PRISMA) diagram depicting the flow of information through the systematic review



2.3.2 Included Studies Characteristics

Table 1 provides summary information about the 13 publications included in the final synthesis. All included publications described or documented the effect of PGHD and PROs on patient-clinician relationships. However, only one of the thirteen included articles had an explicit research objective to identify how PGHD and PROs have an effect on patient-clinician relationships [47]. One included publication concerned surgical patients [44] and the remaining twelve included publications pertained to medical conditions managed by primary care clinicians [45–56]. The publications were published between 2007 - 2017 and conducted in the United States (n=6), Denmark (n=1), South Korea (n=1), Finland (n=1), Slovakia (n=1), the United Kingdom (n=1), Italy (n=1), and Canada (n=1). The publications used qualitative (n=8) [44–46,49,51,53,54,56], quantitative (n=1) [50], or mixed methods (n=4) [47,48,52,55]. Sample sizes ranged from two to 800 patients and from one to 21 clinicians. All publications described or documented patients recording PGHD and/or PROs for personal use outside the clinic. Three publications [48,52,54] incorporated parents or caregivers tracking PGHD and PROs on behalf of a child. The PGHD collected by patients and/or caregivers can be found in table 1. The MMAT quality scores of the included publications ranged from one (one criteria component met) to four (all criteria met).

Table 1: Characteristics of included studies

Title	Authors	Quality	Year	Country	Patient Population	Provider Population	Caregiver Population	PGHD Types
Barriers and benefits to using mobile health technology after operation: A qualitative study	Abelson, Kaufman, Symer, Peters, Charlson, Yeo	4	2017	USA	800 phone survey respondents	N/A	N/A	N/A
Beyond Self-Monitoring: Understanding Non-functional Aspects of Home-based Healthcare Technology	Grönvall, Verdezoto	3	2013	Denmark	6 pregnant women with complications 7 Older adults with heart conditions 6 healthy self-monitoring older adults	1 midwife for pregnant patients Unknown number hospital nurses for patients with heart conditions	N/A	Pregnant women: Weight, blood pressure, pulse, CTG, urine protein levels, online questionnaire Heart condition patients: Weight, blood pressure, pulse, symptom survey, ECG data (subset of participants) Health older adults: Blood pressure
"My Doctor is Keeping an Eye on Me!": Exploring the Clinical Applicability of a Mobile Food Logger	Kim, Ji, Lee, Kim, Yoo, Lee	3	2016	South Korea	20 patients with lifestyle diseases (e.g. hypertension, diabetes, heart disease)	1 otorhinolaryngologist 2 family medicine physicians 1 OBGYN 1 rehabilitation physician 1 urologist	N/A	Food intake, perceptions of post-meal fullness, meal contexts, meal time, activity levels, and activity trackers

Title	Authors	Quality	Year	Country	Patient Population	Provider Population	Caregiver Population	PGHD Types
Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data	Chung, Dew, Cole, Zia, Fogarty, Kientz, Munson	3	2016	USA	211 surveyed patients who were overweight, obese, or diagnosed with IBS Interviews: 7 overweight/obese patients 2 patients diagnosed with IBS 9 overweight/obese patients diagnosed with IBS	6 family medicine physicians 5 gastroenterologists 7 dieticians 1 behavioral psychologist 1 APRN 1 health navigator	N/A	Food intake, calorie intake, physical activity levels, weight, heart rates, sleep quality, pain levels, medication use, bowel movement, stress, fatigue, nausea.
Evaluation of a web-based asthma self-management system: a randomised controlled pilot trial	Wiecha, Adams, Rybin, Rizzodepaoli, Keller, Clay, Jayanti	3	2015	USA	58 children ages 9-17 diagnosed with persistent asthma	Unknown number of primary care providers Unknown number of asthma nurses or asthma specialists	Parent or guardian of children participants	Peak flow readings, symptoms (e.g. cough, wheeze, shortness of breath), contextual data (e.g. activity limitations, missed school, ED visits), medication use
Information technology supporting diabetes self-care: a pilot study	Halkoaho, Kavilo, Pietilä	2	2007	Finland	3 type 1 diabetics 6 type 2 diabetics	3 nurses	N/A	Blood glucose levels and treatment goals

Title	Authors	Quality	Year	Country	Patient Population	Provider Population	Caregiver Population	PGHD Types
Yet Another Hypertension Telehealth Solution? The Rules Will Tell You	Lehocki, Kossaczky, Homola, Skalicky, Mydliar, Thurzo	2	2014	Slovakia	2 patients diagnosed with hypertension and unspecified co-morbidities	Unspecified providers	N/A	Blood pressure, pulse
Nurses' and community support workers' experience of telehealth: a longitudinal case study	Sharma, Clarke	4	2014	United Kingdom	Patients diagnosed with asthma, diabetes, COPD, or CHF (Not recruited for study participation)	Nurses treating patients with asthma, diabetes, COPD, or CHF Community support workers	N/A	Blood glucose level, weight, blood pressure, oxygen level and heart rate.
Using a Mobile App to Manage Type 1 Diabetes: The Case of TreC Diabetes	Miele, Eccher, Piras	2	2015	Italy	15 children ages 4-12 diagnosed with type 1 diabetes	Diabetes specialist	Parent or guardian of children participants	Blood glucose values, meal composition, carbohydrate content, and physical activity levels
Improving Diabetes Management With a Patient Portal: Qualitative Study of a Diabetes Self-Management Portal	Urowitz, Wiljer, Dupak, Kuehner, Leonard, Lovrics, Picton, Seto, Cafazzo	3	2012	Canada	1 patient diagnosed with type 1 diabetes 16 patients diagnosed with type 2 diabetes	Unspecified number of: General practitioners Dieticians APRNS Diabetes educators	N/A	All participants recorded blood glucose levels Additional data collected at provider discretion on a per patient basis (e.g. weight, blood pressure)
Integrating Patient-Generated Health Data Into Clinical Care Settings or Clinical Decision-Making: Lessons Learned From Project HealthDesign	Cohen, Keller, Hayes, Dorr, Ash, Sittig	3	2016	USA	Patients diagnosed with moderate to severe asthma Older adults at risk for cognitive decline Adolescent receiving behavioral health interventions Patient's diagnosed with Chron's Disease Premature infants with medical complications (Not recruited for study participation)	Primary care providers Nurses Gastroenterologists High-risk infant case managers	Parent or guardian of infant participants (Not recruited for study participation)	Asthma patients: medication use, peak flow measurements, environmental factors Older adults: task completion (data not shared with provider) Adolescents: Food intake, physical activity, mood Chron's disease patients: Weight, physical activity, mood, relevant symptoms Premature infants: infant weight, food consumption, elimination patterns

Title	Authors	Quality	Year	Country	Patient Population	Provider Population	Caregiver Population	PGHD Types
Using Patient-Generated Health Data From Mobile Technologies for Diabetes Self-Management Support: Provider Perspectives From an Academic Medical Center	Nundy, Lu, Hogan, Mishra, Peek	3	2014	USA	Unspecified number of type 1 or type 2 diabetic patients	Unspecified number of nurse-care managers 10 primary care providers 2 endocrinologists/diabetes specialists	N/A	Medication use, blood glucose levels, barriers to diabetes self-care
More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management	Chung, Cook, Bales, Zia, Munson	3	2015	USA	Patients who are overweight/obese and/or diagnosed with IBS	6 family medicine physicians 5 gastroenterologists 1 APRN 7 dieticians 1 behavioral psychologist 1 health navigator	N/A	Physical activity levels, food/diet data, stress logs, sleep logs, mood diaries

2.3.3 Synthesis of Findings

We identified three main themes and six sub-themes. The main themes were: 1) PGHD and PROs supported patient-clinician communication and health awareness, 2) patients desired for their clinicians to be involved with their PGHD, which clinicians had difficulty accommodating, and 3) PGHD platform features may support or hinder patient-clinician collaboration. Table 2 lists the analytical themes and subthemes.

Table 2: Major Analytical Themes and Sub-Themes

Major Analytical Theme	Sub-theme
PGHD supported patient-clinician communication and health awareness	PGHD fostered patient-clinician communication
	PGHD improved the clinicians understanding of their patients' health
Patients desired for their clinicians to be involved with their PGHD, which clinicians had difficulty accommodating	Patients desired clinician involvement with their PGHD
	Clinicians had varied interest, encountered barriers, and identified workarounds when integrating PGHD into clinical encounters
PGHD platform features may support or hinder patient-clinician collaboration	Trends, summary measures, and education supported PGHD clinical integration and use.
	Some PGHD platforms negatively impacted patient-clinician collaboration

2.3.4 Theme 1: PGHD Supported Patient-Clinician Communication and Health Awareness

PGHD Fostered Patient-Clinician Communication

Patients and clinicians in eight publications [44,46–49,51,54,56] viewed PGHD as a tool to enhance patient-clinician communication. Additionally in one publication, clinicians perceived PGHD as a tool to support clinician-clinician communication [53]. Five publications documented or described improved patient-clinician communication when collaboratively using PGHD as a discussion tool, such as identifying opportunities to improve patient health [46–48,55,56]. In two publications, clinicians explicitly informed patients when to expect communication from the healthcare team about their PGHD [53,54]. Specifically, the clinicians let them know they would not be contacted if their data appeared normal. In two publications, clinicians used PGHD to provide emotional support to patients [47,54], such as providing empathy regarding a patient's health experiences [47].

PGHD Improved the Clinicians Understanding of their Patients' Health

In seven publications, clinicians modified their patients' treatment plans after reviewing their PGHD [46,47,49,53–56]. In six publications, clinicians used PGHD to identify patient treatment or goal barriers [49,51,52,54–56]. PGHD was also utilized by clinicians in six publications to gain a greater understanding of a patient's health between clinic visits [45–48,54,56]. Patients in three publications had the option of recording additional PGHD to help illustrate the context of their health condition in relation to their daily lives for their clinicians [48,50,53]. Clinicians would use PGHD to set agendas during clinical encounters with patients in two publications [55,56]. Physicians in turn would use these agendas to focus clinical encounters on pertinent patient issues identified in PGHD such as poor blood sugar monitoring [55] or specific concerns patients have difficulty articulating [56]. Two publications reported clinicians used PGHD to identify if the patient's personally identified goals were being achieved [52,56].

2.3.5 Theme 2: Patients Desired for their Clinicians to be Involved with their PGHD, which Clinicians had Difficulty Accommodating

Patients Desired Clinician Involvement with their PGHD

In four publications patients wanted their clinicians to make PGHD review a central component of their clinic visits or expressed a desire for greater clinician involvement with their data [46,47,50,53]. In two publications patients wanted clinicians to provide empathy or emotional support after clinician PGHD review [46,47]. Additionally, in two publications, patients desired acknowledgement from clinicians for their efforts to record PGHD [46,50]. For example, some patients perceived clinician acknowledgement was the reward for collecting the data [46]. Clinician acknowledgement also had an impact on patient health management. In three publications patients had increased accountability and treatment adherence when clinicians asked about their tracking behaviors and emphasized the importance of tracking PGHD [46,47,56]. In three publications, some patients were unable to draw actionable insights from their PGHD because they were unable to make sense of their data. This prompted them to seek greater clinician involvement to aid in interpreting PGHD [46,47,53] and in one publication some patients wanted personalized treatment and action plans after clinician PGHD review [47]. While a large number of patients desired for their clinicians to be involved with their PGHD, in four publications some patients were able to make sense of their PGHD and generate actionable insights independent of clinician review [45,46,52,53].

Clinicians had Varied Interest, Encountered Barriers, and Identified Workarounds when Integrating PGHD into Clinical Encounters

However, across publications, clinicians had differing views about their roles regarding the collaborative PGHD use during clinical encounters with patients. In eleven publications, clinicians would review PGHD and discuss the data with their patients [45–51,53–56]. However, in eight publications, PGHD was identified by clinicians as an important educational tool to improve patient self-awareness about their health conditions, which in turn could promote patient self-care and support goal attainment, potentially without clinician involvement [45–47,49,50,53,54,56]. In three publications, physicians had varied levels of interest in using PGHD [46], with some delegating PGHD review to other clinicians [51,56] and others questioning whether additional health benefits would result from clinician review of PGHD [56].

Clinicians also encountered barriers when integrating PGHD into clinical encounters. In five publications, clinicians reported PGHD review barriers such as clinic appointment time constraints, a lack of formal workflow integration policies, information overload, or an absence of reimbursement incentives [46,48,53,54,56]. Some clinicians had varied ability in their confidence to interpret PGHD effectively during clinical encounters, which negatively affected the planning and suggestions they provided to their patients [46]. Some clinicians identified reviewing PGHD within an online portal reduced the amount of time they had to interact with patients in the clinic [53]. Clinicians reported the lack of reimbursement incentives and workflow issues has the potential to send mixed messages to patients. The patients are directed to collect PGHD because the data provides valuable insights. Yet the clinicians may not review the patients' PGHD because they lack sufficient time during clinical encounters and are not financially incentivized to make data review a greater priority, which could cause patients to question why they collected the data in the first place [56].

In an effort to Clinicians employed or identified different methods to overcome PGHD review barriers and facilitate the use of PGHD during clinical encounters. In two publications, clinicians would review brief summaries of PGHD prior to meeting with patients, which facilitated PGHD clinical integration [46,55]. Clinicians in three publications would ask patients to verbally summarize their PGHD to reduce the effort required to interpret PGHD during clinic appointments [47,54,56]. In six publications, clinicians identified certain types of patients who benefit from PGHD more than others, which would reduce the overall number of patients recording PGHD and potentially decrease the burden of assisting patients with data interpretation [45,48–50,53,55]. Examples of such patients are those who are starting new treatments, who travel frequently, or who have severe/chronic conditions.

2.3.6 Theme 3: PGHD Platform Design Choices Both Supported and Hindered Collaboration

Trends, Summary Measures, and Education Supported PGHD Clinical Integration and Use

In eight publications, clinicians and patients used trends and summary measures depicted in graphs or charts to help make sense of PGHD [46,48–50,52–55]. Two publications explicitly designed the trends and summary measures to be quickly interpreted by clinicians prior to seeing patients in the clinic [46,55]. Clinicians and patients in four publications received in-person training on how to use the PGHD platforms [45,48,49,53] and in one of those publications patients were financially incentivized to complete an online tutorial [48]. PGHD platforms in two publications incorporated patient focused educational materials to help educate patients about their conditions or data [48,53]. Patients and clinicians in four publications expressed a desire for automated PGHD analysis, the ability to interact with PGHD to highlight areas of interest, or additional data incorporated into trends and summary measure tools to facilitate data review [45–47,53].

Some PGHD Platforms Negatively Impacted Patient-Clinician Collaboration

Three publications identified specific PGHD tracking tools or preferred PGHD collection methods hindering patient-clinician collaboration [47,53,56]. For example, in one publication clinicians expressed concerns that using a diabetes PGHD platform could result in time-consuming, redundant work and decrease the time they had to spend with patients [53]. Another publication identified that PGHD platforms may lack enough flexibility to meet patient-clinician needs, standardized data presentation to make the data useful for patients and clinicians, and mechanisms for patients to easily share data with clinicians [56]. In four publications, some patients and clinicians perceived use of PGHD platforms would result in reduced face-to-face interaction and a negative impact on patient-clinician relationships [44,45,49,51]. In particular, some patients expressed a preference for in-person communication [44]. Additionally, clinicians stated not all patients are well suited for collecting PGHD [45], preferred interacting with patients not computers [45], identified they had to trust the authenticity of patient data [49], desired to set expectations for regular clinic visits in addition to remotely monitoring PGHD [49], feared social exclusion of patients [51], and perceived it would be challenging to accurately diagnose patient conditions without subjective information such as patient appearance [51].

2.4 Discussion

In this systematic review, we synthesized the existing literature to identify themes common across the published literature concerning the effect of PGHD and PROs on patient-clinician relationships within surgery and primary care and how patients engage in health information-seeking to understand their collected data. We identified that PGHD and PROs facilitated patient-clinician communication. These data provided additional context which improved the clinicians awareness of their patient's health states in between clinical encounters. Patients desired for their clinicians to be involved with their PGHD, which clinicians had difficulty accommodating. Finally, specific PGHD platform features either supported or hindered PGHD collaboration between patients and clinicians.

This research expands on previous work in a number of ways. First, the findings of this review demonstrate patients collecting PGHD desire for their clinicians to be involved with their PGHD, which clinicians have difficulty accommodating in addition to their existing work obligations and settings. This finding affirms the previous research we conducted, which identified the potential for this tension to exist was the motivation for this review and [12]. This finding also has implications for patient-clinician relationships. Unmet patient expectations during clinical encounters, such as information requests, can negatively affect patient satisfaction, treatment adherence, symptom improvement, and relationships with clinicians [59,60]. We build in this research by detailing how the unmet needs of patients collecting PGHD when collaborating with their physicians can contribute to relationship tensions for both parties. Second, this review details how PGHD contributes to enhanced communication and shared understanding between patients and clinicians. Patient-clinician communication has been associated with health outcomes [61–64]. Additionally, in general medical information is population focused, not on individual patients [65], and patients and clinicians often have different health and illness perspectives [66,67]. PGHD has the potential to bring shared understanding between patients and clinicians, which could increase satisfaction [68], promote patient participation during clinical encounters [69], patient trust of their clinicians [70], and treatment adherence [71]. Third, this review affirms previous research identifying workflow barriers exist when integrating PGHD into clinical encounters [5,12] and identifies specific opportunities to promote the use and integration of PGHD into clinical encounters.

This findings of this review establishes guidance for the integration of PGHD into clinical encounters and to mitigate the negative impacts on patient-clinician relationships.

2.4.1 Establishing Goals and Setting Expectations When Using PGHD

We recommend clinicians consider a baseline expectations for using PGHD. In our review, multiple publications suggested that explicit conversations concerning the goals of both parties for collecting and using PGHD may be needed before data collection is initiated. For example, two publications [53,54] developed a clinician communication algorithm to indicate when a clinician should reach out to patients. Patients were taught about this algorithm to set expectations concerning patient-clinician PGHD communication. Another publication [56] discovered that failing to set expectations may result in mixed messages from physicians to patients as to the purpose and role of PGHD in their relationships. Setting expectations can help both parties understand why and how PGHD will be used to address patient health concerns.

In addition to setting baseline expectations, we recommend patients and clinicians engage in a collaborative process incorporating input and agreement from both parties to achieve the full potential of PGHD. For example, Jahng and colleagues demonstrated that when patients and physicians have congruent beliefs about how involved a patient is in their health decision making, patients have better outcomes and higher levels of satisfaction [72]. Patient-physician disagreement about how involved patients are in their own care may result in lower rates of satisfaction [73]. Furthermore, research has demonstrated patient-clinician collaboration for expectations setting is desired by both groups and feasible [74,75]. Our review of the literature complements this research by demonstrating patients and clinicians may have differing viewpoints or preferences as to how PGHD could be collaboratively used to address health concerns. For example, one publication identified patient-clinician PGHD collaboration ceased due to a mismatch of each party being unable to agree which tool works best to meet their needs [47].

We recommend clinicians openly share their rationale for encouraging patients to record these data and the level of involvement patients can expect from their healthcare team. Multiple publications included in this review suggest patients and clinicians often do not explicitly discuss their roles and expectations when using PGHD to address a health concern. During these conversations, clinicians could ask questions to ascertain the patient's preferred role in healthcare according to the Match Model [76]. Patients are categorized as members of one of four categories depending on their health literacy and desired level of involvement in their health. Patient-clinician conversations may need to be revisited during subsequent clinical encounters and conversations because patients may shift to different Match Model quadrants over time [76] or develop higher levels of autonomy when using PGHD. Additionally, this process may result in contextualized and personalized patient plans of action, which can directly affect a patient's adherence [77–81]. This practice may help clinicians adapt their level of collaboration to be congruent with the patient or visa versa.

2.4.2 Integrating PGHD Into Clinical Encounters

Our review identified that patients who desire or need assistance with data review may encounter challenges or limitations when engaging with their healthcare team. For example, patients may desire greater involvement with their physicians to make sense of their data [46,47,50,53], which clinicians may have difficulties accommodating.

Publications in the review identified two strategies to improve patient-clinician PGHD collaboration. Two of the included publications [46,55] gave clinicians access to PGHD in the form of summarized reports that could be interpreted quickly immediately prior to patient encounters, which worked well for existing clinician workflows. In one of the publications [46], clinicians reported sufficient time for data review as a result of this process. Another strategy to improve PGHD collaboration, identified in three included publications [47,54,56], involved clinicians asking patients to verbally summarize their PGHD during clinic visits. We recommend PGHD platform designers considering developing trends and summary measures utilizing PGHD integration applications such as Apple Health Kit [82], Google Fit [83], or Samsung Health [84]. These applications have the potential to support electronic health record integration, which could facilitate clinician access to the data using their preferred platform. If clinicians or patients opt to use a PGHD platform that does not incorporate trends or summary measures, we recommend clinicians ask patients to verbally summarize their data during clinical encounters. Incorporating these three strategies into PGHD platforms may improve PGHD clinical integration and collaboration during patient clinic visits.

Augmenting the data review process with automated data analysis may reduce the burden on clinicians to perform data review tasks on behalf of patients. For example, exist.io is a web-based PGHD aggregation platform that supports the exploration of correlations between patient self-tracking attributes and behaviors [85]. Users are able to integrate multiple data sources, such as Apple Health Kit, activity trackers, email, calendar, social media, and weather. Exist.io is then able to identify how the data are interconnected and associated with health outcomes, such as weight gain. This in turn may reduce the need for clinicians to assist with patient PGHD review and could also be used to automatically identify patterns in PGHD to aid clinician interpretation.

2.4.3 Opportunities to Improve PGHD Data Collaboration

In this review we identified how PGHD platform components may support or hinder patient-clinician PGHD collaboration. One strategy to overcome the barriers of specific

PGHD platforms would be to create resources, such as physical spaces or websites, to help identify specific PGHD platform that meet the needs of both patients and clinicians. For example, the Ochsner Health System has created dedicated physical spaces in their hospitals where patients can learn more about the various clinician-preferred mHealth apps and devices to address their health concerns [86].

Another strategy to address PGHD platform barriers could be to conduct future PGHD platform design work using participatory design methods [87]. Participatory design incorporates all of the stakeholders (e.g. clinicians and patients) in the design process and has been previously used to design a clinician focused PGHD dashboard for use during clinic visits [88]. Involving patients and clinicians in the design process has the potential to create PGHD platforms that better meet the needs and preferences of both groups.

2.4.4 Recommendations for Future Research

While all of the publications included for final synthesis in this review described or documented the effect of PGHD and PROs on patient-clinician relationships, only one publication explicitly had an objective to study how these data affect patient-clinician relationships [47]. Additional research is needed to explicitly identify how these data and technological platforms can positively or negatively impact the relationships between patients and their healthcare team. For example, future research should consider how PGHD has an impact on patient-clinician communication, patient trust of clincians, satisfaction, and treatment adherence. Additionally, the majority of the included publications had a clinician perspective bias. For example, one included publication only recruited nurses and community support workers but reported changes to patient interactions when using a PGHD platform [51]. It is important for researchers to include patient perspectives when assessing these technologies to reduce the potential for informatics generated inequalities [89].

2.4.5 Limitations

Our review focused on primary care and surgery domains, which limits the generalizability of our findings. Despite this limitation, the patient and clinician participants included for final synthesis in this review represented a wide range of illnesses, diseases, and clinical roles. This provided a rich dataset for final synthesis.

Additionally, the use of a wide range of search terms in the research database queries reflects the lack of a unified language around PGHD and PROs within the biomedical literature. While the authors collaborated with a health sciences librarian to develop effective search strategies and queries; it is possible that articles fitting the inclusion criteria were not captured in our queries.

Finally, only one member of the research team assessed the included articles' quality. However, none of the included articles were excluded from the qualitative synthesis based on quality.

2.6 Conclusion

Using PGHD and PROs during clinical encounters may promote patients taking a more active role in their healthcare, improve patient-clinician communication, and support clinician work activities. However, patients and clinicians may disagree about how these data should be used to collaboratively address health concerns, which could be affected by how the PGHD platforms are designed. Future research needs to be conducted to explicitly improve the understanding of how PGHD and PROs affect patient-clinician relationships and identify opportunities to improve collaboration using these data.

2.7 Acknowledgements

We thank and acknowledge BHI Librarian Diana Louden for her assistance in developing the search strategies and queries used in this systematic review.

2.8 References

1 NW 1615 L. St, Washington S 800, Inquiries D 20036 U-419-4300 | M-419-4349 | F-419-4372 | M. Demographics of Mobile Device Ownership and Adoption in the United States. 2018.http://www.pewinternet.org/fact-sheet/mobile/ (accessed 20 Sep 2018).
2 Research2Guidance. mHealth App Economics 2017/2018 Current Status and Future Trends in Mobile Health. Berlin, Germany: 2017. https://research2guidance.com/wp-content/uploads/2017/11/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf
3 Rideout V, Fox S. Digital Health Practices, Social Media Use, and Mental Well-Being Among Teens and Young Adults in the U.S. *Hopelab and Well Being Trust* 2018.
4 Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR Mhealth Uhealth* 2015;**3**. doi:10.2196/mhealth.4924
5 Shapiro M, Johnston D, Wald J, *et al.* Patient-Generated Health Data. Research Triangle Park, NC: : RTI International 2012. https://www.rti.org/sites/default/files/resources/patientgeneratedhealthdata.pdf
6 Howie L, Hirsch B, Locklear T, *et al.* Assessing The Value Of Patient-Generated Data To Comparative Effectiveness Research. *Health Affairs* 2014;**33**:1220–8. doi:10.1377/hlthaff.2014.0225
7 Cochrane Patient Reported Outcomes Methods Group. https://methods.cochrane.org/pro/ (accessed 19 Aug 2019).

8 Cortez A, Hsii P, Mitchell E, *et al.* Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024. Accenture 2018. https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf
9 Deshpande PR, Rajan S, Sudeepthi BL, *et al.* Patient-reported outcomes: A new era in clinical research. *Perspectives in Clinical Research* Published Online First: 2011.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3227331/
10 Coons SJ, Eremenco S, Lundy JJ, *et al.* Capturing Patient-Reported Outcome (PRO) Data Electronically: The Past, Present, and Promise of ePRO Measurement in Clinical Trials | SpringerLink. *The Patient - Patient-Centered Outcomes Research* Published Online First: 10 October 2014.https://link.springer.com/article/10.1007/s40271-014-0090-z
11 Fox S, Duggan M. Tracking for Health | Pew Research Center. Pew Research Center 2013. http://www.pewinternet.org/2013/01/28/tracking-for-health/
12 Sanger P, Hartzler A, Armstrong C, *et al.* A patient-centered system in a provider-centered world: challenges of incorporating post-discharge wound data into practice. - PubMed - NCBI. *JAMIA* Published Online First: May 2016.https://www.ncbi.nlm.nih.gov/pubmed/26977103
13 Jones JB, Snyder CF, Wu AW. Issues in the design of Internet-based systems for collecting patient-reported outcomes. *Qual Life Res* 2007;**16**:1407–17. doi:10.1007/s11136-007-9235-z
14 West P, Kleek MV, Giordano R, *et al.* Information Quality Challenges of Patient-Generated Data in Clinical Practice. *Frontiers in Public Health* Published Online First: May 2017. doi:https://dx.doi.org/10.3389%2Ffpubh.2017.00284
15 Greenwod D, Gee PM, Fatkin K, *et al.* A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support - Deborah A. Greenwood, Perry M. Gee, Kathy J. Fatkin, Malinda Peeples, 2017. *Jounral of Diabetes Science and Technology* Published Online First: 31 May 2017. doi:https://doi.org/10.1177/1932296817713506
16 Basch E, Rogak L, Dueck A. Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. - PubMed - NCBI. *Clinical Therapeutics* Published Online First: April 2016. doi:10.1016/j.clinthera.2016.03.011
17 Mcnair A, Macefield R, Blencowe N, *et al.* “Trial Exegesis”: Methods for Synthesizing Clinical and Patient Reported Outcome (PRO) Data in Trials to Inform Clinical Practice. A Systematic Rev... - PubMed - NCBI. *PLoS One* Published Online First: 29 August 2016. doi:10.1371/journal.pone.0160998
18 Mercieca-Bebber R, Palmer M, Brundage M, *et al.* Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. - PubMed - NCBI. *BMJ Open* Published Online First: 15 June 2016. doi:10.1136/bmjopen-2015-010938
19 Antunes B, Harding R, Higginson IJ, *et al.* Implementing patient-reported outcome measures in palliative care clinical practice: A systematic review of facilitators and barriers. *Palliative Medicine* 2014;**28**:158–75. doi:10.1177/0269216313491619
20 Younossi Z, Henry L. Systematic review: patient-reported outcomes in chronic hepatitis C--the impact of liver disease and new treatment regimens. - PubMed - NCBI. *Aliment Pharmacol Therapuetics* Published Online First: 23 January 2015. doi:10.1111/apt.13090
21 Johnston D, Sung L, Stark D, *et al.* A systematic review of patient-reported outcome measures of neuropathy in children, adolescents and young adults. - PubMed - NCBI. *Support Care Cancer* Published Online First: 24 September 2016. doi:10.1007/s00520-016-3199-x
22 Mosa ASM, Yoo I, Sheets L. A Systematic Review of Healthcare Applications for

Smartphones. *BMC Medical Informatics and Decision Making* 2012;**12**:67. doi:10.1186/1472-6947-12-67
23 Arnhold M, Quade M, Kirch W. Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older. *Journal of Medical Internet Research* 2014;**16**:e104. doi:10.2196/jmir.2968
24 Mateo GF, Granado-Font E, Ferré-Grau C, *et al.* Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. *Journal of Medical Internet Research* 2015;**17**:e253. doi:10.2196/jmir.4836
25 McKay FH, Cheng C, Wright A, *et al.* Evaluating mobile phone applications for health behaviour change: A systematic review. *J Telemed Telecare* 2018;**24**:22–30. doi:10.1177/1357633X16673538
26 Matthews J, Win KT, Oinas-Kukkonen H, *et al.* Persuasive Technology in Mobile Applications Promoting Physical Activity: a Systematic Review. *J Med Syst* 2016;**40**:72. doi:10.1007/s10916-015-0425-x
27 Chiarini G, Ray P, Akter S, *et al.* mHealth Technologies for Chronic Diseases and Elders: A Systematic Review. *IEEE Journal on Selected Areas in Communications* 2013;**31**:6–18. doi:10.1109/JSAC.2013.SUP.0513001
28 Reynolds R, Dennis S, Hasan I, *et al.* A systematic review of chronic disease management interventions in primary care. *BMC Fam Pract* 2018;**19**:1–13. doi:10.1186/s12875-017-0692-3
29 Pham HH, O'Malley A, Bach PB, *et al.* Primary Care Physicians' Links to Other Physicians Through Medicare Patients: The Scope of Care Coordination. *Annals of Internal Medicine* 2009;**150**:236. doi:10.7326/0003-4819-150-4-200902170-00004
30 Lella A. U.S. Smartphone Penetration Surpassed 80 Percent in 2016. comScore, Inc. http://www.comscore.com/Insights/Blog/US-Smartphone-Penetration-Surpassed-80-Percent-in-2016
31 Obey DR. American Recovery and Reinvestment Act of 2009. 2009.https://www.congress.gov/bill/111th-congress/house-bill/1/text (accessed 19 Aug 2019).
32 111th Congress. Health Information Technology for Economic and Clinical Health (HITECH) Act - Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009 (ARRA). 2009.https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf
33 Medicare Promoting Interoperability Program Stage 3 Eligible Hospitals, Critical Access Hospitals, and Dual-Eligible Hospitals Attesting to CMS Objectives and Measures for 2018 Objective 4 of 6 Updated: July 2018. 2018;:11.
34 Covidence - Better systematic review management. https://www.covidence.org/home
35 Stata: Software for Statistics and Data Science. https://www.stata.com/
36 Dedoose. https://www.dedoose.com/
37 Tong A, Flemming K, McInnes E, *et al.* Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology* 2012;**12**:181. doi:10.1186/1471-2288-12-181
38 Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology* 2008;**8**:45. doi:10.1186/1471-2288-8-45
39 Ridd M, Shaw A, Lewis G, *et al.* The patient–doctor relationship: a synthesis of the qualitative literature on patients' perspectives. *Br J Gen Pract* 2009;**59**:e116–33. doi:10.3399/bjgp09X420248
40 Fereday J, Muir-Cochrane E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International*

Journal of Qualitative Methods 2006;**5**:80–92. doi:10.1177/160940690600500107
41 Holden RJ, Carayon P, Gurses AP, *et al.* SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics* 2013;**56**. doi:10.1080/00140139.2013.838643
42 Pluye P, Robert E, Cargo M, *et al.* Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. Published Online First: 2011.http://mixedmethodsappraisaltoolpublic.pbworks.com.
43 Pace R, Pluye P, Bartlett G, *et al.* Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies* 2012;**49**:47–53. doi:10.1016/j.ijnurstu.2011.07.002
44 Abelson JS, Kaufman E, Symer M, *et al.* Barriers and benefits to using mobile health technology after operation: A qualitative study. *Surgery* 2017;**162**:605–11. doi:10.1016/j.surg.2017.05.007
45 Grönvall E, Verdezoto N. Beyond self-monitoring: understanding non-functional aspects of home-based healthcare technology. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*. Zurich, Switzerland: : ACM Press 2013. 587. doi:10.1145/2493432.2493495
46 Kim Y, Ji S, Lee H, *et al.* "My Doctor is Keeping an Eye on Me!": Exploring the Clinical Applicability of a Mobile Food Logger. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2016. 5620–5631. doi:10.1145/2858036.2858145
47 Chung C-F, Dew K, Cole AM, *et al.* 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 - CSCW '16*. San Francisco, California, USA: : ACM Press 2016. 768–84. doi:10.1145/2818048.2819926
48 Wiecha JM, Adams WG, Rybin D, *et al.* Evaluation of a web-based asthma self-management system: a randomised controlled pilot trial. *BMC Pulmonary Medicine* 2015;**15**. doi:10.1186/s12890-015-0007-1
49 Halkoaho A, Kavilo M, Pietilä A-M. Information technology supporting diabetes self-care: a pilot study. *European Diabetes Nursing* 2007;**4**:14–7. doi:10.1002/edn.70
50 Lehocki F, Kossaczky I, Homola M, *et al.* Yet another hypertension telehealth solution? The rules will tell you. In: *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*. Kuala Lumpur, Malaysia: : IEEE 2014. 510–5. doi:10.1109/IECBES.2014.7047554
51 Sharma U, Clarke M. Nurses' and community support workers' experience of telehealth: a longitudinal case study. 2014;:7.
52 Miele F, Eccher C, Piras EM. Using a Mobile App to Manage Type 1 Diabetes: The Case of TreC Diabetes. In: *Proceedings of the 5th International Conference on Digital Health 2015 - DH '15*. Florence, Italy: : ACM Press 2015. 113–4. doi:10.1145/2750511.2750541
53 Urowitz S, Wiljer D, Dupak K, *et al.* Improving Diabetes Management With a Patient Portal: A Qualitative Study of Diabetes Self-Management Portal. *Journal of Medical Internet Research* 2012;**14**:e158. doi:10.2196/jmir.2265
54 Cohen DJ, Keller SR, Hayes GR, *et al.* Integrating Patient-Generated Health Data Into Clinical Care Settings or Clinical Decision-Making: Lessons Learned From Project HealthDesign. *JMIR Hum Factors* 2016;**3**. doi:10.2196/humanfactors.5919
55 Nundy S, Lu C-YE, Hogan P, *et al.* Using Patient-Generated Health Data From Mobile Technologies for Diabetes Self-Management Support. *J Diabetes Sci Technol* 2014;**8**:74–82. doi:10.1177/1932296813511727
56 Chung C-F, Cook J, Bales E, *et al.* More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management. *Journal of*

Medical Internet Research 2015;**17**:e203. doi:10.2196/jmir.4364
57 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009;**6**. doi:10.1371/journal.pmed.1000097
58 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;**33**:159–74. doi:10.2307/2529310
59 Bell RA, Kravitz RL, Thom D, *et al.* Unmet expectations for care and the patient-physician relationship. *J Gen Intern Med* 2002;**17**:817–24. doi:10.1046/j.1525-1497.2002.10319.x
60 Like R, Zyzanski SJ. Patient satisfaction with the clinical encounter: Social psychological determinants. *Social Science & Medicine* 1987;**24**:351–7. doi:10.1016/0277-9536(87)90153-5
61 Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995;**152**:1423–33.
62 Kaplan SH, Greenfield S, Ware JE. Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease. *Medical Care* 1989;**27**:S110–27.
63 Jackson JL. Communication about symptoms in primary care: impact on patient outcomes. *J Altern Complement Med* 2005;**11 Suppl 1**:S51-56. doi:10.1089/acm.2005.11.s-51
64 Stewart M, Brown JB, Donner A, *et al.* The impact of patient-centered care on outcomes. *J Fam Pract* 2000;**49**:796–804.
65 Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making* 2007;**27**:696–713. doi:10.1177/0272989X07307271
66 Klienman A. *Patients and Healers in the Context of Culture*. https://www.ucpress.edu/book/9780520045118/patients-and-healers-in-the-context-of-culture (accessed 23 Aug 2019).
67 Makoul G, Clayman ML, Lynch EB, *et al.* Four concepts of health in America: results of national surveys. *J Health Commun* 2009;**14**:3–14. doi:10.1080/10810730802592213
68 Davidson R, Mills ME. Cancer patients' satisfaction with communication, information and quality of care in a UK region. *Eur J Cancer Care (Engl)* 2005;**14**:83–90. doi:10.1111/j.1365-2354.2005.00530.x
69 Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. *Soc Sci Med* 1999;**49**:651–61. doi:10.1016/s0277-9536(99)00145-8
70 Gordon HS, Street RL, Sharf BF, *et al.* Racial differences in trust and lung cancer patients' perceptions of physician communication. *J Clin Oncol* 2006;**24**:904–9. doi:10.1200/JCO.2005.03.1955
71 DiMatteo MR, Sherbourne CD, Hays RD, *et al.* Physicians' characteristics influence patients' adherence to medical treatment: results from the Medical Outcomes Study. *Health Psychol* 1993;**12**:93–102.
72 Jahng KH, Martin LR, Golin CE, *et al.* Preferences for medical collaboration: patient–physician congruence and patient outcomes. *Patient Education and Counseling* 2005;**57**:308–14. doi:10.1016/j.pec.2004.08.006
73 Krupat E, Rosenkranz SL, Yeager CM, *et al.* The practice orientations of physicians and patients: the effect of doctor–patient congruence on satisfaction. *Patient Education and Counseling* 2000;**39**:49–59. doi:10.1016/S0738-3991(99)00090-7
74 American Medical Informatics Association. Redefining our Picture of Health: Towards a Person-Centered Integrated Care, Research, Wellness, and Community Ecosystem. https://www.amia.org/sites/default/files/API-2017-White-Paper-Redefining-our-Picture-of-Health.pdf

75 Cronin RM, Conway D, Condon D, *et al.* Patient and healthcare provider views on a patient-reported outcomes portal. *J Am Med Inform Assoc* 2018;**25**:1470–80. doi:10.1093/jamia/ocy111
76 Peters RM. Matching physician practice style to patient informational issues and decision-making preferences. An approach to patient autonomy and medical paternalism issues in clinical practice. *Archives of Family Medicine* 1994;**3**:760–3. doi:10.1001/archfami.3.9.760
77 Gómez G, Aillach E. Ways to improve the patient-physician relationship. *Curr Opin Psychiatry* 2013;**26**:453–7. doi:10.1097/YCO.0b013e328363be50
78 Tinsel I, Buchholz A, Vach W, *et al.* Implementation of shared decision making by physician training to optimise hypertension treatment. Study protocol of a cluster-RCT. *BMC Cardiovasc Disord* 2012;**12**:73. doi:10.1186/1471-2261-12-73
79 van Langenberg DR, Andrews JM. Satisfaction with patient-doctor relationships in inflammatory bowel diseases: examining patient-initiated change of specialist. *World J Gastroenterol* 2012;**18**:2212–8. doi:10.3748/wjg.v18.i18.2212
80 McCabe R, Bullenkamp J, Hansson L, *et al.* The Therapeutic Relationship and Adherence to Antipsychotic Medication in Schizophrenia. *PLOS ONE* 2012;**7**:e36080. doi:10.1371/journal.pone.0036080
81 Schoenthaler AM, Schwartz BS, Wood C, *et al.* Patient and physician factors associated with adherence to diabetes medications. *Diabetes Educ* 2012;**38**:397–408. doi:10.1177/0145721712440333
82 HealthKit - Apple Developer. https://developer.apple.com/healthkit/ (accessed 23 Aug 2019).
83 Google Fit. Google Fit. https://www.google.com/intl/en_us/fit/ (accessed 23 Aug 2019).
84 Samsung Health. Samsung Electronics America. https://www.samsung.com/us/support/owners/app/samsung-health/ (accessed 23 Aug 2019).
85 Exist. https://exist.io (accessed 23 Aug 2019).
86 Ochsner Health. The O Bar | Ochsner Health System. 2016.https://www.ochsner.org/io/the-o-bar
87 Schuler D, Namioka A. *Participatory Design: Principles and Practices*. CRC Press 1993.
88 Kim Y, Heo E, Lee H, *et al.* Prescribing 10,000 Steps Like Aspirin: Designing a Novel Interface for Data-Driven Medical Consultations. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. Denver, Colorado, USA: : ACM Press 2017. 5787–99. doi:10.1145/3025453.3025570
89 Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018;**25**:1080–8. doi:10.1093/jamia/ocy052

2.9 Appendix

Chapter 2 Supplemental table 1: Systematic review research database queries.

Database	Search Terms
PubMed	(("patient generated health data" OR "patient generated data" OR “person-generated health data” OR “person generated health data” OR ("patient generated" AND data) OR “ubiquitous health” OR “pervasive health” OR “quantified self” OR "patient reported outcome" OR "patient reported outcomes" OR “patient outcome assessment”[mesh] OR "Nurse-Patient Relations"[mesh] OR "Physician-Patient Relations"[mesh] OR "Professional-Patient Relations"[Mesh:NoExp] OR “patient provider” OR “doctor patient” OR “patient doctor” OR “patient physician” OR “physician patient” OR “nurse patient” OR “patient nurse” OR “patient surgeon” OR “surgeon patient”) AND ("communication"[mesh] OR communication OR communicating OR communicate*) AND (“Conflict (Psychology)”[mesh] or consensus[mesh] OR discord* OR concord* OR disaccord OR disagree* OR dissiden* OR dissonan* OR friction OR agree* OR consensus OR compromis*) AND ("clinical decision making"[mesh] OR "clinical decision making" OR "patient acceptance of health care"[mesh] OR "primary health care"[mesh] OR "patient care"[mesh] OR "perioperative care"[mesh] OR "primary care" OR "patient care" OR "perioperative care" OR "postacute care" OR "post-acute care" OR "Surgical Procedures, Operative"[Mesh] OR surgery OR surgical OR surgeries OR "specialties, surgical"[mesh])) NOT (“language”[mesh] OR “translations”[mesh] OR language*[ti] OR translator*[ti] OR interpreter*[ti] OR gender[ti] OR race[ti]) AND ("2006"[Date - Publication] : "2017"[Date - Publication])

Embase	(("patient generated health data" OR "patient generated data" OR “person-generated health data” OR “person generated health data” OR ("patient generated" AND data) OR “ubiquitous health” OR “pervasive health” OR “quantified self” OR "patient reported outcome" OR "patient reported outcomes" OR 'nurse patient relationship'/exp OR 'doctor patient relation'/exp OR 'human relation'/de OR “patient provider” OR “doctor patient” OR “patient doctor” OR “patient physician” OR “physician patient” OR “nurse patient” OR “patient nurse” OR “surgeon patient” OR “patient surgeon”) AND ('interpersonal communication'/exp OR communication OR communicating OR communicate*) AND ('conflict'/exp OR 'consensus'/exp OR discord* OR concord* OR disaccord OR disagree* OR dissiden* OR dissonan* OR friction OR agree* OR consensus OR compromis*) AND ('clinical decision making'/exp OR 'patient attitude'/exp OR ‘patient decision making’/exp OR 'primary health care'/exp OR 'patient care'/exp OR 'perioperative period'/exp OR 'postoperative period'/exp OR 'surgery'/exp OR ‘intensive care’/exp OR "clinical decision making" OR "primary care" OR "patient care" OR "perioperative care" OR "postacute care" OR "post-acute care")) NOT ('language'/exp OR language*:ti OR translator*:ti OR translation*:ti OR interpreter*:ti OR gender:ti OR race:ti) AND [2006-2017]/py

PsycINFO	("patient generated health data" OR "patient generated data" OR “person-generated health data” OR “person generated health data” OR ("patient generated" AND data) OR "patient reported outcome" OR "patient reported outcomes" OR “ubiquitous health” OR “quantified self” OR “pervasive health” OR “patient provider” OR “doctor patient” OR “patient doctor” OR “patient physician” OR “physician patient” OR “nurse patient” OR “patient nurse” OR “surgeon patient” OR “patient surgeon”) AND (DE "Communication" OR DE "Electronic Communication" OR DE "Interpersonal Communication" OR DE "Verbal Communication" OR communication OR communicating OR communicate*) AND (DE "Conflict" OR DE "Arguments" OR discord* OR concord* OR disaccord OR disagree* OR dissiden* OR dissonan* OR friction OR agree* OR consensus OR compromis*) AND (DE "Decision Making" OR DE "Choice Behavior" OR DE "Compliance" OR DE "Treatment Compliance" OR DE "Client Attitudes" OR DE "Client Satisfaction" OR DE “Client Participation” OR DE "Primary Health Care" OR DE "Treatment Planning" OR DE "Caring Behaviors" OR DE "Discharge Planning" OR DE "Postsurgical Complications" OR DE "Surgery" OR DE "Surgeons" OR DE "Surgical Patients" OR "clinical decision making" OR "primary care" OR "patient care" OR "perioperative care" OR "postacute care" OR "post acute care" OR surgery OR surgical OR surgeries) NOT (DE "Language" OR DE "Dialect" OR DE "Figurative Language" OR DE "Foreign Languages" OR DE "Form Classes (Language)" OR DE "Native Language" OR DE "Natural Language" OR DE "Phrases" OR DE "Profanity" OR DE "Rhetoric" OR DE "Sentences" OR DE "Sign Language" OR DE "Spelling" OR DE "Vocabulary" OR DE "Written Language" OR DE "Foreign Language Translation" OR DE "Interpreters" OR TI (language* OR translator* OR interpreter* OR gender OR race))
IEEE	((patient generated health data) OR (patient generated data) OR (person generated health data) OR (person-generated health data) OR (ubiquitous health) OR (pervasive health) OR (quantified self)) AND communication AND ((conflict OR consensus) AND (surgery OR (primary care)) OR (patient reported outcome) OR (patient reported outcomes))
ACM (via Google Scholar)	site:http://dl.acm.org (“patient generated health data” OR “patient generated data” OR “person generated health data” OR “person-generated health data”) OR (“ubiquitous health” OR “pervasive health” OR “quantified self”) OR (“patient reported outcome” OR “patient reported outcomes”) AND (communication) AND (conflict OR consensus) AND (surgery OR “primary care”)

Chapter 3. INVESTIGATING INGUINAL HERNIA SURGERY PATIENT INFORMATION NEEDS AND OPPORTUNITIES FOR HEALTH DIALOG SYSTEM SUPPORT: A PARTICIPATORY DESIGN STUDY

3.1 Background and Significance

Patients undergoing hernia repair routinely engage in health information-seeking, to obtain information from traditional media, social media, friends, family members, healthcare professionals, organizations, or websites [1]. This information-seeking is driven by specific post-operative information needs regarding wound monitoring, average pain levels, and expected time to return to regular life activities [2]. Researchers consider face-to-face communication between patients and clinicians as the ideal form of communication [1,3]. However, clinicians have limited availability to answer patient questions [4,5]. For example, patients encounter difficulties engaging with their care team when trying to make sense of health data they collect to address personal health concerns [6–8]. To mitigate these difficulties, patients can access online tools that provide up-to-date content customized to patient needs [1].

One online tool that has the potential to support inguinal hernia patients' information-seeking is a health dialog system (HDS). HDSs are "automated systems that can interview patients and consumers about their health and provide health education and behavior change interventions using natural language" [3]. Patient-centered HDSs have also been used to improve patient access to health information [9,10], promote behavior change [11–13], provide reminders [10], and assist with patient education [9]. The United Kingdom's National Health Service and Canada's Telus Health have incorporated an HDS into mobile health applications, providing patients with health information and triaging their concerns [14–17].

Given the recent increase in access to generalized dialog systems (DSs) such as the Google Assistant, Cortana, and Alexa, there is potential for patient-centered HDSs to be applied in various health contexts, including hernia surgery. DSs are "conversational agents that interact with users turn by turn using natural language" and are generally used in customer service, education, or leisure activities [18]. In 2018, 77% of U.S. adults owned a smartphone [19] and between 52% - 64% of smartphone owners reported using their phone's DS (e.g., Alexa, Cortana, Google Assistant, Siri) [20,21]. Additionally, almost 20% of U.S. adults have access to a DS using smart speaker technology (e.g., Google Home, Amazon Echo) [22]. A subset of these DS, such as the Google Assistant and Alexa, allow external innovators to extend these technologies through the creation of “Skills” or “Actions” to new domains such as supporting inguinal hernia patients’ information-seeking [23–25].

Given the wide reach of DSs that can be used to develop “Skills” or “Actions”, there is potential for to create patient-centered HDS “Skill” or “Action” to aid patients’ information-seeking activities. However, to realize this potential, care needs to be taken to avoid introducing unintended consequences with health information technology (HIT). Examples of HIT unintended consequences include information acquisition difficulties, decision making barriers, and delayed care delivery [26]. One way to avoid HIT unintended consequences is by involving users in HIT design. Participatory Design (PD) is a method incorporating stakeholders (e.g., users, those whose work will be affected by others’ use of a system, those responsible for deciding to adopt a system) as active members of the design process [27] and is a component of human-centered design (HCD) [28,29]. HCD is increasingly used in health informatics research [30–35] and PD has informed the design of systems such as prostate cancer dashboards [36], an eHealth solution for vulnerable women with perinatal depression [37], and infographics to support patient health literacy [38]. Our objective was to engage clinicians and inguinal hernia surgery patients, using PD, to identify when during a patient’s surgery journey a patient-centered HDS could support health information-seeking activities, and the desired topics and content for an inguinal hernia surgery focused HDS.

3.2 Methods

3.2.1 Sample, design, setting, and time period

We chose to recruit inguinal hernia surgery patients for a number of reasons. Inguinal hernias are a common health condition [39] with an estimated 500,000 - 600,000 people in the United States [40,41] and twenty million people worldwide [42] undergoing surgery to have their inguinal hernia repaired every year. Inguinal hernia surgery patients are typically male (96% of surgeries), middle-aged or older (mean 54 years old) [43–45], and undergo outpatient surgery to repair the hernia [46–48]. Outpatient surgery patients are generally discharged home from the hospital the same day as the surgery.

We determined we would be able to focus our efforts and achieve significant impact when creating Hernia Coach due to the high prevalence of inguinal hernias, and the patient population's outpatient hospital experiences, and homogenous demographics.

This descriptive PD study was completed June-October 2018 in the Northwest United States. To engage multiple stakeholders for an inguinal hernia HDS, we recruited patients who had previously undergone inguinal hernia surgery and clinicians who care for inguinal hernia surgery patients (i.e., users, people affected by HDS deployment, or those engaged in decision-making about a clinic's adoption of the HDS) . The University of Washington Human Subjects Division approved this study. All participants consented prior to starting the PD sessions.

3.2.2 Recruitment

Patients were eligible for the study if they were 18 years or older, were fluent in English, and had had hernia surgery at least two weeks prior to the PD session. Patients were recruited using online posts to public platforms (e.g., Craigslist, Reddit, and Facebook), and flyers posted in public places (e.g. libraries, community centers, coffee shops) [49]. We also directly approached patients in an academic medical center's hernia clinic during patient's surgery follow up appointments prior to being seen by their physician. Clinicians were eligible if they directly cared for inguinal hernia surgery patients as medical doctors (MDs) who specialized in surgery, registered nurses (RNs), or medical assistants (MAs). We mailed flyers to clinics that provide inguinal hernia surgery and identified additional clinicians via snowball sampling. Recruitment of patients and clinicians continued until thematic saturation was reached and no new concepts emerged in the PD session transcripts. Participants were not compensated for participating.

3.2.3 Design Session Procedures

RL led all PD sessions, which were held in person. Patient PD sessions were held in private meeting spaces on our university campus and public city libraries. Clinician PD sessions were conducted in private meeting spaces in a hospital and outpatient clinics. The sessions were conducted either one-on-one with a participant or as a group to accommodate participant schedules. PD session scripts were developed to accommodate individual or group sessions. During individual sessions, the lead researcher took notes; for group sessions, members of the research team took notes while the lead researcher conducted the session. All PD sessions were audio recorded and transcribed using a professional transcription service to generate text documents of the recordings for qualitative analysis. Artifacts generated by participants during the PD session were saved for analysis.

3.2.3.1 Patients

Surveys

Patient participants first completed a demographics survey assessing age, gender, post-surgery caregiver availability, education, home geographic region, hernia surgery details, healthcare industry employment, and DS use. Home geographic regions were classified according to U.S. census definitions, which are urbanized areas (population > 50,000 residents), urban clusters (population between 2,500 and 50,000 residents), or rural (population < 2,500) [50]. Participants had the option to selected multiple designated caregiver types and DSs. They then completed the 8-item eHealth Literacy Scale (eHEALS) “developed to measure consumers’ combined knowledge, comfort, and perceived skills at finding, evaluating, and applying electronic health information to health problems” [51]. eHEALS is scored on a five point Likert scale and each item ranges from 1-strongly disagree to 5-strongly agree, with a total score range of 8-40; higher eHEALS scores indicate greater eHealth literacy. Researchers have used the eHEALS in health informatics design research [52–54].

Warm-up Exercise

After completing the surveys, patient participants were given a brief description of PD and the goals of the design session. Participants then engaged in a five-minute warm-up exercise to help prepare them to engage in the HDS design activities with a creative mindset. For the exercise, participants designed the “pill bottle of the future” and described their design to the researcher.

Hernia Surgery Journey Mapping

After the warm-up exercise, patient participants completed a journey map [55], which is used to: depict a person’s experience of an event, act as a communication tool, and identify parts of the experience that can be improved [55–57]. The purpose of having participants generate the map at the early in the overall PD session was to assist them in recalling their surgery experience and for the map to serve as a reference throughout the rest of the design session. Our participants created a journey map illustrating major events from the first day of their hernia surgery journey -- seeking medical care that led to an inguinal hernia diagnosis -- through their recovery. After the participants completed their journey map, they verbally described the journey maps to the researcher(s).

Brainstorming Session

After completing their journey map, patient participants engaged in a brainstorming activity [58] to identify surgery related questions they had during their journey. Participants wrote down individual questions on individual sticky-notes and placed them on the table in front of them. Participants were encouraged to generate as many questions as possible.

When patient participants completed the brainstorming exercise, the researcher introduced the concept of DSs and described how these technologies can be used to create HDSs. The purpose of introducing DSs at this stage of the PD sessions, versus at the beginning, was to mitigate the potential of biasing what participants generated during the journey mapping exercise and brainstorming activities. Participants were provided with a handout depicting common smartphone and smart speaker DSs (e.g., Alexa, Siri). Then the participants were asked a series of probing questions to identify their use or non-use of DSs for everyday and health information-seeking activities. The researcher then asked participants to indicate which questions generated during the brainstorming exercise they felt were well suited to be answered by: (a) an HDS, (b) non-physician member of their healthcare team (e.g. registered nurse, medical assistant), or (c) physician. Participants were given three different colored dot stickers to indicate the three answer sources. The participants placed stickers on the sticky-notes (on which were the questions they generated) to indicate which source(s) they felt most comfortable for answering their questions; one question could have one or multiple answer sources indicated with the colored dots. After the participants completed this task, the researcher asked a series of probing questions to understand why some questions were better suited for HDSs or members of their healthcare team. Participants were also prompted to their share concerns or perceived benefits of using an HDS to help answer their surgical journey questions. Lastly, participants were asked to revisit their journey map and circle parts for which they felt an HDS could have supported them and verbally explain why.

3.2.3.2 Clinicians

At the beginning of the clinician PD sessions, clinician participants completed a brief survey collecting information about age, gender, employment geographic region, practice setting, time since training completion, time spent caring for inguinal hernia patients, typical number of inguinal hernia patients seen per week, and dialog system use. Afterwards, the clinicians engaged in a version of the patient PD session that we abbreviated to accommodate the clinicians' limited availability. The clinicians first completed the sticky-note brainstorming exercise, during which they identified common questions they received from patients. Then they indicated which questions were well suited for HDSs, non-physician healthcare team members, or physicians using the color coded dot stickers. Finally, they discussed perceived benefits or concerns they had regarding patients use of HDSs for surgery information-seeking.

3.2.4 Analysis

3.2.4.1 Comprehensive Model of Information-seeking

We identified a post-hoc theoretical model, the Comprehensive Model of Information-seeking (CMIS) (Figure 1) to inform our PD session transcript analysis [1]. This model was selected because it was developed to identify how factors influence patients perceptions and use of different information sources for health information-seeking activities. We include a description of the model and model component definitions in Table 1.

The CMIS by J.D. Johnson was developed to “explain the communication channel usage of information seekers” [1] (ref) and has been generally applied in oncology patient information seeking research [59–62]. The model has also been extended to other domains such as information scanning [63], prescription drug information seeking [60], and online health information seeking [64]. The CMIS consists of three different components (figure 1). The first component are the antecedents that influence how an individual searches for information from particular information carriers, which are the individual’s demographics, direct experience, salience, and beliefs. Demographics concerns aspects of the individual such as gender, age, race/ethnicity, education level, or socio-economic status. Direct experience is the degree of the individual’s experience with disease. Direct experience may be gained directly by the individual or through their friends/family experiences. Salience is the individual’s perceived applicability of information to a problem that an individual faces and the underlying motivating force to seek information. Beliefs are an individual’s belief in the efficacy of medical procedures. The antecedents activate individuals to seek information and determine the intensity of the search.

The second component of the CMIS are information carrier factors which focuses on the different mediums individuals may use to obtain information. The two carrier factors are characteristics and utilities. The characteristics are an information channel's attributes such as the editorial tone, accuracy, credibility, intentions, trustworthiness, competence, or the manner of information presentation. A channel’s utilities are a direct evaluation by an individual of a particular channel and relates the characteristics of a medium directly to the needs of an individual. Information carrier factors shape the intention of an individual's search and involves an individual's assessment of certain media mediums to meet their information needs.

The third and final component of the CMIS are the individual's information seeking actions. Information seeking actions concern the nature of the search itself and are determined by the first two components of the CMIS. Examples include seeking health information online, engaging with friends or family members, or asking healthcare professionals questions.

Figure 1: J.D. Johnson's Comprehensive Model of Information-seeking [1]. Reprinted with the permission of Hampton Press.

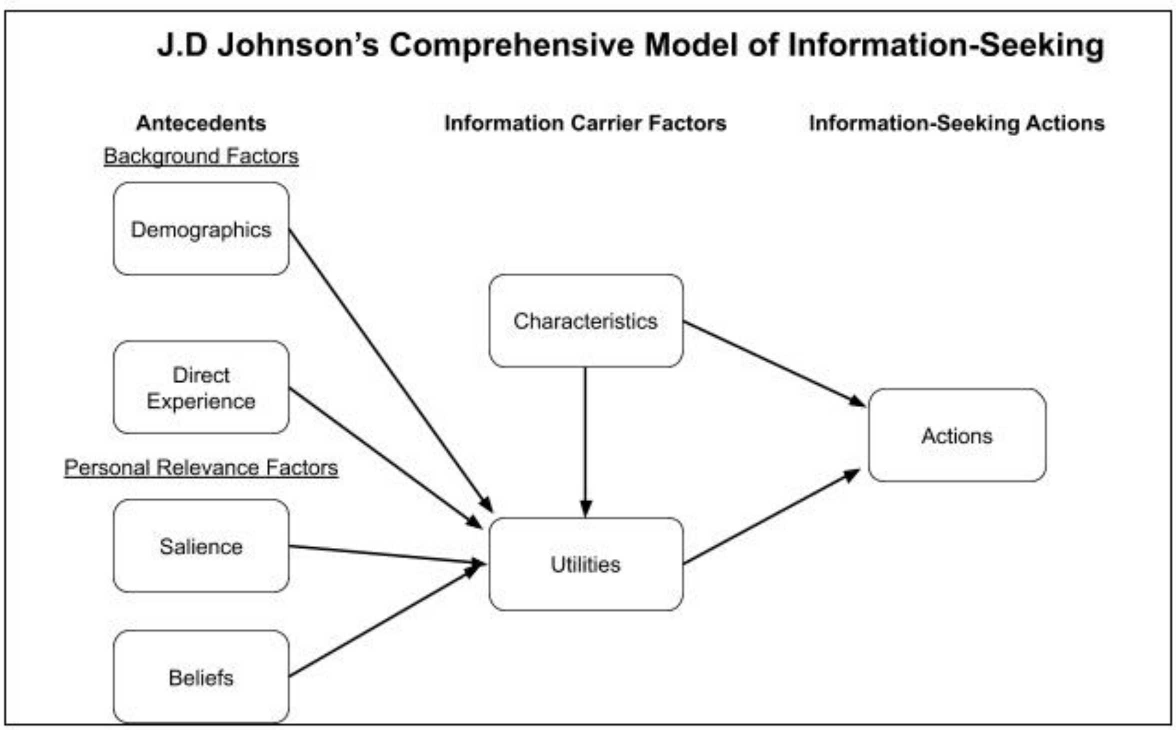


Table 1: J.D. Johnson's Comprehensive Model of Information-seeking component and subcomponent definitions [1].

Model Themes and Sub-Themes	Definition
Antecedents	Influence how an individual searches for information from particular information carriers.
Demographics	Aspects of the individual such as gender, age, race/ethnicity, education level, or socio-economic status.
Direct Experience	The degree of the individual's experience with disease, which may be gained directly by the individual or through their friends/family experiences.
Salience	The individual's perceived applicability of information to a problem that an individual faces and the underlying motivating force to seek information.
Beliefs	An individual's belief in the efficacy of medical procedures.
Information Carrier Factors	The different mediums individuals may use to obtain information. Information carrier factors shape the intention of an individual's search and involves an individual's assessment of certain media mediums to meet their information needs.
Characteristics	An information channel's attributes such as the editorial tone, accuracy, credibility, intentions, trustworthiness, competence, or the manner of information presentation.
Utilities	A direct evaluation by an individual of a particular channel and relates the characteristics of a medium directly to the needs of an individual.
Information-Seeking Actions	Concern the nature of the search itself and are determined by the first two components of the CMIS.

3.2.4.2 Qualitative Analysis

We performed directed content analysis of the design session transcripts [65]. We chose directed content analysis because it is a deductive qualitative analysis method used to validate or extend theoretical frameworks. The CMIS informed the initial development of the qualitative analysis codebook. The PD sessions and analysis occurred in parallel to allow for identification of thematic saturation. Initially, two researchers coded three design session transcript excerpts. After each of the three coding rounds, Ross Lordon (RL) and Uba Backonja (UB) discussed how the transcripts were coded. After the third round of coding, the two researchers agreed the code book adequately covered all of the themes. Topf [66] was used to calculate inter-rater reliability (IRR). Initial IRR was 76.5% indicating adequate agreement; the two researchers discussed coding discrepancies and came to consensus for a final IRR of 100%. The codebook was finalized and used by one of the two researchers to independently code remaining transcripts.

3.2.4.3 Survey Data and PD Artifacts

Survey data and participatory design artifacts were analyzed after thematic saturation of PD session transcripts was reached. Design session warm-up exercise artifacts were not analyzed because the data were not directly relevant for the purpose of this study. RL independently reviewed patient journey maps to identify patient participant motivations, actions, questions, barriers [67], and opportunities for HDS support during the hernia surgery journey. A modified affinity diagramming method was used to separately organize the patient and clinician sticky-note questions and generate themes [68]. Prior to conducting affinity diagramming, three distinct stages were identified from the patient journey maps: pre-surgery, day of surgery, and post-surgery. Questions were allocated to these stages and nested within high-level themes under each stage to align with the patients' surgery journey stages and identify when in the patient's surgery journey an HDS could support their information-seeking. The results of the affinity diagramming were transcribed into an Excel spreadsheet. We generated a table depicting the affinity diagramming themes and sticky-note colored dot frequencies to identify which types of questions are well suited for HDSs, non-physician clinicians, and physicians using Tableau.

3.3 Results

3.3.1 Participant Characteristics

Seven patients and eight clinicians participated in our study (n=15). One patient PD group session was conducted consisting of two participants, the rest were conducted individually. One patient participant subsequently joined the study and authorship team; they helped develop the codebook guided by the theory concepts but did not engage in the final analyses. The clinician sessions were conducted as two group sessions (n=5 and n=2) and a single one-on-one session. Clinicians and patients did not interact with each other during the study. Tables 2a and 2b provide demographics for patient and clinician participants, respectively.

During recruitment, due to a miscommunication with clinic staff and a misunderstanding by a patient, one incisional hernia [69] patient was also enrolled in the study. During the design session it became apparent this participant had surgery to address an incisional hernia, not inguinal hernia surgery. We had approval from our Human Subjects Division to recruit all types of hernia patients. However, due to our focus on inguinal hernia surgery patients, we excluded this participant from the final analysis for a final count of seven patient participants.

Table 2a: Patient participant characteristics (n=7).

Patient Characteristic	
Age in years, median (IQR)	61 (47.5-74.5)
Gender, *n*	
Male	6
Female	1
Designated post-surgery caregivers, *n*	
Spouse or significant other	5
Child/children	2
Parent(s)	1
Highest level of completed education, *n*	
Associate's degree	1
Bachelor's degree	2
Master's degree	2
Doctorate degree	2
Home geographic region, *n*	
Urban	5
Urban cluster	1
Rural	1
Number of hernia surgeries undergone, *n*	
One	6
Two	1
Inguinal hernia surgery was elective, *n*	6
Surgical approach, *n*	
Open	4
Laparoscopic	4
Healthcare industry experience, *n*	
Previously employed in healthcare	4
Currently employed in healthcare	1
No	2

Healthcare industry roles, *n*	
Nurse	3
Analyst	1
Physician	1
Dialog system use, *n*	
Alexa	3
Siri	3
Google assistant	1
None of the above	2
eHEALS score, median (IQR)	32 (32-39)

IQR=interquartile range; *n*=number

Table 2b: Clinician participant demographics (n=8).

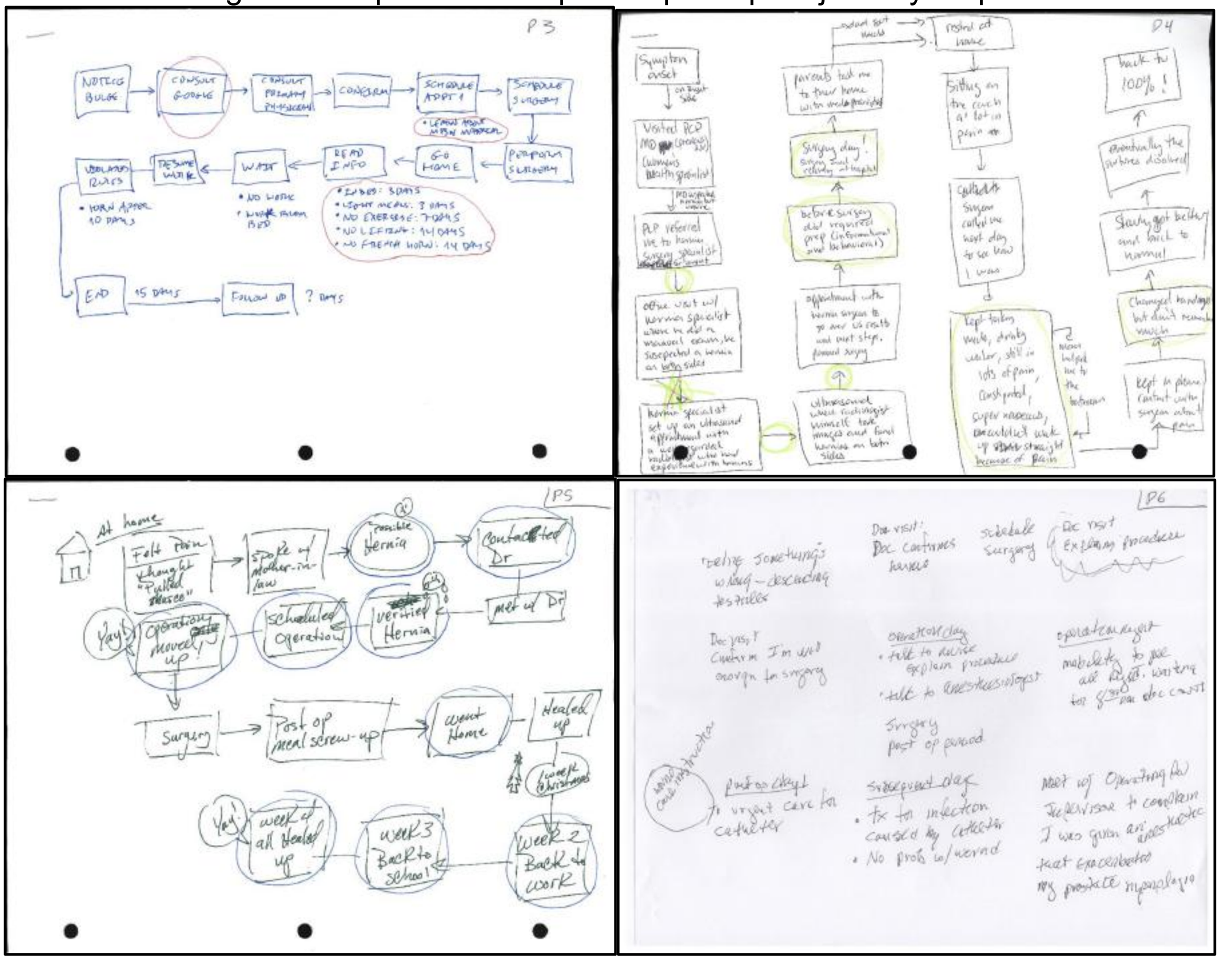
Clinician Characteristic	
Age in years, median (IQR)	39 (37.5-48.5)
Decline to respond, *n*	1
Gender, *n*	
Female	6
Male	2
Employment geographic region, *n*	
Urban	7
Urban cluster	1
Practice setting, *n*	
Non-profit community hospital	7
Independent surgical group	1
Clinician types, *n*	
Medical Doctor (MD) - Surgeon	6
Medical assistant (MA)	1
Registered nurse (RN)	1
Time since completing training in years, median (IQR)	10.5 (5.5-21.5)
Years professionally caring for inguinal hernia patients, median (IQR)	12 (2.75-16.0)
Typical number of inguinal hernia patients seen per week, median (IQR)	5.5 (4.25-15.0)
Dialog system use, *n*	
Alexa	3
Siri	2
Google assistant	1
None of the above	4

IQR=interquartile range; *n*=number

3.3.2 Patient Journey Maps

The patient journey maps captured the participants' surgery journey from their self-selected first day through recovery. All patient journey maps had three distinct stages: pre-surgery, day of surgery, and post-surgery. All seven patients noted motivation to initiate receiving medical care by noticing pain or a bulge in their groin. All patients reported seeking medical consultation pre-surgery, which resulted in scheduling hernia repair surgeries. Two participants noted on their journey maps that they sought additional information. One participant consulted a search engine; the other consulted a family member. All seven participants were discharged home the same day of surgery. Post-surgery, all participants reported returning to normal activities after a recovery period. One participant noted a barrier trying to contact his care team outside clinic hours about concerning symptoms, ultimately prompting him to seek and receive emergency care to address an urgent complication related to the surgery. No participants noted questions on their journey maps. Six participants did not note interaction with their care team outside of clinical appointments before, during, or after surgery. Two participants noted phone calls post-surgery with clinicians. Four participants identified opportunities for HDS information-seeking support pre-surgery, two participants marked day of surgery opportunities, and four participants indicated post-surgery opportunities during recovery. See Figure 2 for representative patient journey maps.

Figure 2: Representative patient participant journey maps.



3.3.3 Design Session Themes

Findings are organized below by themes based on CMIS concepts. Illustrative quotes for all themes are presented in Table 3.

3.3.3.1 Antecedents

Demographics

Two patient participants referenced their previous clinical training and that they were confident in their ability to locate and identify meaningful health information. One participant reported being the sole income earner in their family, prompting them to learn as much as possible pre-surgery to minimize time away from work during and post-surgery. The other patient participants did not state whether their households were single or dual income.

Direct experience

Four patient participants reported previous experience with hernias and/or hospital procedures in general. One patient participant stated their hernia surgery was their first hospital experience. Two patient participants noted they had family members with previous hernia knowledge or experience.

Beliefs

One patient participant noted their surgeon's explanation of using mesh beyond preventing hernia recurrence was to not see the patient again. Another patient participant spoke about the difficulty of identifying hernias using ultrasound.

Salience

One patient participant viewed information provided by information carriers as helpful if the source of the information was disclosed.

3.3.3.2 Information Carrier Factors

Characteristics

Four patient participants, three surgeons, and two non-physician clinicians expressed a desire for the HDS health content to be limited to topics concerning inguinal hernias and surgical repair in general and not about individual patient health states. These participants stated topics pertaining to patients on an individual basis should be answered by members of the healthcare team. In other words, the participants wanted the HDS to act an alternate standardized patient educational material such as a brochure or leaflet. Two patient participants and five surgeons were uncertain about HDSs ability to accurately understand questions asked by patients. Two patients, three surgeons, two non-physician clinicians were concerned about HDSs providing inaccurate information or failing to identify situations requiring emergency medical care. Four patient participants and three surgeons and voiced data privacy and trust concerns. Some of the most common HDS concerns among participants were the HDS's adherence to privacy laws (e.g., Health Insurance Portability and Accountability Act [HIPAA]), how external companies providing the HDS service would handle private information, and the source of the HDS's information.

Utilities

Five patient participants, three surgeons, and two non-physician clinicians stated an HDS would be useful for hernia surgery information-seeking either for themselves or in general. One patient participant and one non-physician clinician identified the HDS as an alternative for patients who do not want to interact with their healthcare team to address issues perceived as not important or urgent. One patient participant thought an HDS could help set agendas for clinic appointments: the HDS could record questions beyond its scope and send the questions to the patient's care team. Three patient participants expected that an HDS would not support their health information-seeking activities, and so they would prefer to not personally use an HDS. Two of these participants had received previous clinical training and perceived they could effectively seek out and identify relevant high quality health information without an HDS. The third participant was skeptical of the ability of an HDS to understand their questions accurately. However, two of the three skeptical participants saw potential value for an HDS to support patients other than themselves with health information-seeking activities. Two participants stated an HDS with only a voice interface may not meet their needs. They stated they either do not use voice-based device interaction methods often or identified there would be times when voice is not appropriate, such as in the middle of the night when others are sleeping.

3.3.3.3 Information-seeking Actions

Patient participants reported using online resources (e.g., Google, Mayo Clinic websites) (n=2), computers in general (n=1), the telephone (n=1), paper materials (e.g., discharge handouts) (n=2), their physician (n=1), and family members (n=2) during their surgery journeys. Two participants reported consulting multiple information carriers to triangulate information from multiple sources. One participant reported difficulties contacting their healthcare team outside normal clinic hours about an urgent concern. One non-physician clinician stated they used a semi-structured phone interview script to identify issues that could be affecting patients post-surgery.

Table 3: The Comprehensive Model of Information-seeking themes and illustrative participant quotes.

Theme	Sub-Theme	Participant Quote
Antecedents	Demographics	The big thing for me is we're a single income household. My wife is a stay at home mom. And there was no way she would be able to get any income coming in while I'm out. So it was important for us for me to take as little time off as possible. - P5
	Direct Experience	It was my first time ever to be in a hospital setting. So I guess I'm more interested in hearing more than anything else, what can I expect having never been a patient in a hospital before. - P3 I just remember mostly the post surgery part just being in a haze. - P4 But I wasn't able to urinate and all night long. I was in a really stressed condition not being able to urinate. When I called the number they gave me, somebody wouldn't be there until 8:30 in the morning which seemed interminable. Now I realize I should have called urgent care and just gone right in. - P6
	Salience	I still don't know why I had to stop eating at midnight as opposed to eight hours beforehand. - P5 I think it's just mostly for me it felt normal like I just want to make sure that it's looking how it should. I know to look for the obvious things and it's you know if there's any puffs or starts looking you know it gets irritated or I knew those things to look for. But you know never having seen that type of surgery I didn't know what to expect. - P5
Information Carrier Factors	Characteristics	Well it's a triage thing and it's a hierarchy and I think you get at a certain point you're not going to be comfortable with the answers from the chatbot [health dialog system]. So you need or are you not sure that the answer that you need to have reassurance. Maybe the healthcare team is the one that can provide that. Or maybe it's so specific or requires such expertise that only the surgeon can [answer]. - P2 So a good question for a chatbot is something that seems like a pretty straightforward fact. Or just some basic information. There doesn't need to be any sort of inference going on. So like what are the symptoms that I can expect post surgery that I could Google that myself … Things that are more just like fact list based but not necessarily where you have to make an inference or a decision. It's making the inference or the decision. - P4 If it was something more like diagnostic or beyond it’s [health dialog system] capabilities. It [health dialog system] could potentially send those questions to the to the healthcare provider in advance. Not like they'll read it. But at least it's in the record too. - P4

		Facts, you know people who've had this surgery in your age and gender group usually you fully heal by whatever I mean would be helpful. If you notice that this is happening with your wound call so and so right away. - P6 The wound is leaking what should I do? When is it OK to take the bandage off? ... When should I come in to have the stitches removed? These are fact-ey things. I can think of other things, logistical questions about my wife having to stay the whole time if she's taking me home. Where to park? How soon can I eat afterwards? I mean the device could help with that. - P6 I think we just need to limit it. We just can't have them say Alexa I have a fever after hernia surgery is this normal? Alexa has to say call your surgeon. - C2 I guess standardized questions. Like whether or not it's same day surgery. Yeah it's usually same day surgery. Yeah it's ones that can be standardized. - C8
	Utilities	The chatbot [health dialog system] could also be helpful during the post op period and the discharge period which can be short or rather long. Lots of questions come up that you may not want to bother the surgeon or health care team. - P2 I can see where it [health dialog system] will work. I mean we're not that far off from integrated computing ... I can see where that would be a very good benefit for patients that may not be as hospital familiar as I am. I had very little concern about this procedure. But I'm the outlier and I know I'm the outlier. Someone that has never even opened a biology book or taken a class in zoology has not got the faintest concept of what the inside of a human body looks like or how it works. So for the everyman I can see where this [health dialog system] would be very potentially a very good stress reducer. - P7 I think it's a cool idea. I like that. Because especially the real time nature of it for people who are up in the middle of the night. It might help offload providers that don't need to answer these kind of urgent questions let patients feel like they're getting any concern any time dispelled but offload us. - C4
Information-Seeking Actions		I noticed the bulge that is usually present with an inguinal hernia so I consulted Google. - P2 They [patients] Google a lot and I tell them make sure their sites are MD related and not like blogs. - C6

“P” indicates a patient quote and “C” indicates a clinician quote.

3.3.4 Affinity Diagramming

During the design sessions, 185 questions were generated, 113 by clinician participants and 72 by patient participants. We include the high level themes generated by the affinity diagramming process, the potential information carriers, and the sticky-note dot frequency counts in Table 4. The themes primarily concerned the pre-surgery and post-surgery time periods. The most preferred information resource were HDSs, non-physician clinicians were, and physicians third. This suggests both patients and clinicians are open to using HDSs to support patient health information-seeking.

Table 4: Affinity diagramming high-level themes, potential information carriers, and sticky-note dot frequency counts generated by patients (n=7) and clinicians (n=8).

	Potential Information Carrier			
Primary Theme	**Health Dialog System**	**Non-Physician Clinician**	**Physician**	**Total**
Hernia Surgery Details	43	23	51	**117**
Activity Restrictions	30	20	15	**65**
Logistics	22	24	9	**55**
Am I Normal?	12	14	6	**32**
Medications	5	5	4	**14**
Recovery	4	3	3	**10**
Clinician Experience	0	3	3	**6**
Total	**116**	**92**	**91**	**299**

3.5 Discussion

Patient-centered HDSs are an emerging medium to aid patients with health information-seeking activities. By using PD methods, we were able to engage direct stakeholders and inform the design of patient-centered HDS to support inguinal hernia patients. Our PD approach generated a rich dataset consisting of inguinal hernia surgery patient journey maps, the types of common questions inguinal hernia surgery patients have during their journey, and an understanding of how a patient-centered HDS could support future inguinal hernia surgery patients.

Our study expands on previous research to understand and address hernia patients' information-seeking in three ways. First, we describe patient information-seeking activities during all stages of surgery, whereas previous research focused mainly on post-surgery information-seeking. The "general day" surgery patient information needs assessment conducted by Bradshaw and colleagues [2] focused exclusively on the patients' post-operative information needs (e.g., restrictions, wound care). Our patient journey map data expands upon this work, indicating that patients may engage in health information-seeking activities before, during, and after surgery. Additionally, the patient journey maps show there are three clinical encounters where patients typically engage in direct information-seeking with clinicians: the pre-surgery consultation, the day of surgery, and the post-surgery follow up appointment. Therefore, if a patient has questions outside of those encounters, the patients are required to initiate information-seeking by contacting their care team, reaching out to friends or family members, or consulting other resources such as websites as described by the some of the patient participants in the PD sessions. Patient-centered HDSs could support hernia patients with their health information-seeking activities throughout the patients' journeys and between clinical encounters. For example, the HDS could act as an information resource between clinic visits with primary care physicians and surgical consults, in the lead up to and the day of surgery, and after surgery when the patient goes home the hospital.

Second, unlike previous research, we employed a variety of participatory methods and theory guided analyses to support participant idea generation and systematic understanding of those ideas. To help us understand the needs of inguinal hernia surgery patients we used the CMIS [1] to guide transcription analyses and grouped question topics using the established method of affinity diagramming. The results demonstrate the desired content of a patient-centered HDS as identified by clinicians and patients, which is an important step toward creating valuable patient education resources [70]. Generally, patients and clinicians identified topics about the health condition in general as the content best suited for an HDS. Patient participants frequently indicated that diagnostic questions or questions that are less fact-oriented, such as “is my wound infected”, were better suited for the healthcare team. The patient participants’ rationale for content focused on the health condition in general, as opposed to their personal health, is due to their trust of an HDS to accurately understand their questions and provide reliable information. In our study, patient participants identified the ways an HDS would be perceived as trustworthy. Examples included providing references for the source materials and links to multiple sources of information allowing for the patient to triangulate the most pertinent information. We recommend HDS developers and designers evaluate which HDS content is trustworthy to users because patient trust has been previously identified as an important component of HDSs [71], health technology acceptance [72–78] and technology in general [79–83].

Third, previous HDS research has generally focused on developing technologies without the use of existing DSs [3,84]. Our research focuses on extending existing DSs into the health domain. Johnson, who developed the CMIS model, described digital information carriers as having the potential for having greater impact than other information carrier channels [1]. However, Johnson also recognized that digital information carriers are subject to barriers such as access to technical equipment or technology incompatibility. Furthermore, intervention-generated inequalities are a recognized potential unintended consequence of health information technology [85]. We recommend designers consider using existing DSs (e.g. Alexa, Google Assistant) that have the greatest reach and least barriers of access for patients. These DSs have widespread adoption and use [19–22], indicating a robust infrastructure for innovators to leverage these technologies and potentially reach millions of patients using familiar interfaces now exists.

3.5.1 Limitations and Future Work

We recruited patients using methods that could have led to self-selection bias to those interested in HDSs and with healthcare employment. However, three out of seven patient participants stated they were not interested in using an HDS as a component of their healthcare, but some saw potential for an HDS to support others, suggesting that this bias was not pervasive. Given the geographic location and small sample, findings from this study may not be generalizable to other hernia patients. Specifically, no patients reported complications due to hernia recurrence and most did not report symptoms affecting their recovery, limiting the generalizability of our findings to patients with complications and recovery-impeding symptoms. Our sample size was small; however we were purposeful in our decision to stop recruitment after no new concepts emerged during concurrent analysis of the PD transcripts (i.e., saturation was achieved). Our sample size reached saturation consistent with previous research [86].

We recommend future work examine the information needs and opportunities for HDS support for: inguinal hernia surgery patients who have experienced complications or recurrence, other types of hernia surgery patients, and the caregivers supporting surgery patients. Researchers should also consider testing HDS prototypes for clinical implementation feasibility, adoption, and acceptance by clinicians and patients. Our research demonstrates the health content desired by patients and clinicians for an inguinal hernia-focused HDS. However, more research is needed to identify if patients would accept and use an HDS as a component of their surgery journey. In addition, HDS developers should use PD or HCD to engage potential end-users and stakeholders in the design process to create solutions that meets their needs. Using existing theories, frameworks, or models can help developers more efficiently and systematically understand, apply, and operationalize research findings to solution development, and evaluate the developed solutions that can be applied to new domains or patient populations [87].

3.6 Conclusion

Patients undergoing hernia surgery have information needs before, during, and after surgery and may have limited opportunities to obtain information from their clinicians. To address these needs and expand opportunities for information-seeking, we conducted a PD study and used an information-seeking theory to identify how HDSs could support patients throughout their surgery journey. Patients and clinicians may find utility in a patient-centered HDS to support inguinal hernia surgery patient information-seeking. Our findings demonstrate the potential of extending existing DSs into healthcare to create a patient-centered inguinal hernia HDS and expands upon prior research of information needs for inguinal hernia surgery patients.

3.8 Acknowledgements

We acknowledge the time and effort invested by the study participants, the clinic physicians and staff for recruitment assistance, and the contributions of Maher Khelifi and Shih-Yin Lin with this study’s design and data collection.

3.9 References

1 Johnson JD. *Cancer-related information seeking*. Cresskill, NJ: : Hampton Press 1997.
2 Bradshaw C, Pritchett C, Bryce C, *et al.* Information needs of general day surgery patients. *Ambul Surg* 1999;**7**:39–44. doi:10.1016/S0966-6532(98)00048-1
3 Bickmore T, Giorgino T. Health dialog systems for patients and consumers. *J Biomed Inform* 2006;**39**:556–71. doi:10.1016/j.jbi.2005.12.004
4 Tai-Seale M, McGuire TG, Zhang W. Time Allocation in Primary Care Office Visits. *Health Serv Res* 2007;**42**:1871–94. doi:10.1111/j.1475-6773.2006.00689.x
5 Weigl M, Müller A, Zupanc A, *et al.* Participant observation of time allocation, direct patient contact and simultaneous activities in hospital physicians. *BMC Health Serv Res* 2009;**9**:110. doi:10.1186/1472-6963-9-110
6 Urowitz S, Wiljer D, Dupak K, *et al.* Improving Diabetes Management With a Patient Portal: A Qualitative Study of Diabetes Self-Management Portal. *J Med Internet Res* 2012;**14**:e158. doi:10.2196/jmir.2265
7 Chung C-F, Dew K, Cole AM, *et al.* 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 - CSCW '16*. San Francisco, California, USA: : ACM Press 2016. 768–84. doi:10.1145/2818048.2819926
8 Kim Y, Ji S, Lee H, *et al.* "My Doctor is Keeping an Eye on Me!": Exploring the Clinical Applicability of a Mobile Food Logger. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2016. 5620–5631. doi:10.1145/2858036.2858145
9 Allen J, Ferguson G, Blaylock N, *et al.* Chester: Towards a personal medication advisor. *J Biomed Inform* 2006;**39**:500–13. doi:10.1016/j.jbi.2006.02.004
10 Piazza M, Giorgino T, Azzini I, *et al.* Cognitive Human Factors for Telemedicine Systems. ;:5.
11 Rosis F de, Novielli N, Carofiglio V, *et al.* User modeling and adaptation in health promotion dialogs with an animated character. *J Biomed Inform* 2006;**39**:514–31. doi:10.1016/j.jbi.2006.01.001
12 Bickmore TW. *Relational agents : effecting change through human-computer relationships*. 2003.http://dspace.mit.edu/handle/1721.1/36109 (accessed 19 Dec 2018).
13 Schulman D, Bickmore T, Sidner C. An Intelligent Conversational Agent for Promoting Long-Term Health Behavior Change using Motivational Interviewing. ;:4.
14 Barisevičius G, Coste M, Geleta D, *et al.* Supporting Digital Healthcare Services Using Semantic Web Technologies. In: Vrandečić D, Bontcheva K, Suárez-Figueroa MC, *et al.*, eds. *The Semantic Web – ISWC 2018*. Springer International Publishing 2018. 291–306.
15 Babylon Health. 2019.https://www.babylonhealth.com/product/ask-babylon (accessed 9 Jan 2019).
16 Babylon Health. 2019.https://www.babylonhealth.com/ai (accessed 9 Jan 2019).
17 Babylon by TELUS Health | TELUS. https://www.telus.com/en/health/ (accessed 6 Mar 2019).
18 Shawar BA, Atwell E. Different measurements metrics to evaluate a chatbot system. In: *Proceedings of the Workshop on Bridging the Gap Academic and Industrial Research in Dialog Technologies - NAACL-HLT '07*. Rochester, New York: : Association for Computational Linguistics 2007. 89–96. doi:10.3115/1556328.1556341
19 Demographics of Mobile Device Ownership and Adoption in the United States. Pew Research Center http://www.pewinternet.org/fact-sheet/mobile/ (accessed 19 Dec 2018).
20 Kinsella B. Over Half of Smartphone Owners Use Voice Assistants, Siri Leads the Pack. Voicebot. 2018.https://voicebot.ai/2018/04/03/over-half-of-smartphone-owners-use-voice-assistants-siri-leads-the-pack/ (accessed 19 Dec 2018).
21 Search Without Screens: Exploring the Rise of Amart Assistants & Voice Search. https://www.netelixir.com/wp-content/uploads/2018/04/search-without-screens-wp-2018.pdf (accessed 18 Jan 2019).
22 Smart Speaker Consumer Adoption Report. Voicebot.ai 2018.
23 Steps to Build a Custom Skill | Custom Skills. https://developer.amazon.com/docs/custom-skills/steps-to-build-a-custom-skill.html (accessed 9 Jan 2019).
24 Getting started with Cortana skills. https://docs.microsoft.com/en-us/cortana/skills/get-started (accessed 2 Feb 2019).
25 Actions on Google | Actions on Google. Google Dev. https://developers.google.com/actions/ (accessed 9 Jan 2019).
26 Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc* 2017;**24**:246–50. doi:10.1093/jamia/ocw154
27 Schuler D, Namioka A. *Participatory Design: Principles and Practices*. CRC Press 1993.

28 ISO 9241-210:2010 - Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. https://www.iso.org/standard/52075.html (accessed 21 Aug 2018).
29 User Centered System Design | New Perspectives on Human-computer Interaction. Taylor Francis. https://www.taylorfrancis.com/books/e/9781482229639 (accessed 2 Feb 2019).
30 Dopp AR, Parisi KE, Munson SA, *et al.* A glossary of user-centered design strategies for implementation experts. *Transl Behav Med* doi:10.1093/tbm/iby119
31 Sox CM, Gribbons WM, Loring BA, *et al.* Patient-centered design of an information management module for a personally controlled health record. *J Med Internet Res* 2010;**12**:e36. doi:10.2196/jmir.1269
32 Wolpin S, Stewart M. A deliberate and rigorous approach to development of patient-centered technologies. *Semin Oncol Nurs* 2011;**27**:183–91. doi:10.1016/j.soncn.2011.04.003
33 Brennan PF, Downs S, Casper G. Project HealthDesign: rethinking the power and potential of personal health records. *J Biomed Inform* 2010;**43**:S3-5. doi:10.1016/j.jbi.2010.09.001
34 Alexander G, Staggers N. A systematic review of the designs of clinical technology: findings and recommendations for future research. *ANS Adv Nurs Sci* 2009;**32**:252–79. doi:10.1097/ANS.0b013e3181b0d737
35 Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform* 2005;**38**:75–87. doi:10.1016/j.jbi.2004.11.005
36 Hartzler AL, Izard JP, Dalkin BL, *et al.* Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. *J Am Med Inform Assoc* 2016;**23**:38–47. doi:10.1093/jamia/ocv101
37 Gordon M, Henderson R, Holmes JH, *et al.* Participatory design of ehealth solutions for women from vulnerable populations with perinatal depression. *J Am Med Inform Assoc* 2016;**23**:105–9. doi:10.1093/jamia/ocv109
38 Arcia A, Suero-Tejeda N, Bales ME, *et al.* Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *J Am Med Inform Assoc* 2016;**23**:174–83. doi:10.1093/jamia/ocv079
39 Feliciano D, Hawn M, Strand N. Groin Hernia Repair Inguinal and Femoral. American College of Surgeons 2018. https://www.facs.org/~/media/files/education/patient%20ed/groin_hernia.ashx (accessed 31 Jan 2019).
40 Inguinal Hernia Repair Surgery Information from SAGES. SAGES. https://www.sages.org/publications/patient-information/patient-information-for-laparoscopic-inguinal-hernia-repair-from-sages/ (accessed 15 Mar 2018).
41 Cullen KA, Margaret H, Aleksander G. Ambulatory Surgery in the United States, 2006. *CDC - Natl Health Stat Rep* 2009;:28.
42 Bay-Nielsen M, Kehlet H, Strand L, *et al.* Quality assessment of 26,304 herniorrhaphies in Denmark: a prospective nationwide study. *Lancet Lond Engl* 2001;**358**:1124–8. doi:10.1016/S0140-6736(01)06251-1
43 Nordin P, Zetterström H, Gunnarsson U, *et al.* Local, regional, or general anaesthesia in groin hernia repair: multicentre randomised trial. *The Lancet* 2003;**362**:853–8. doi:10.1016/S0140-6736(03)14339-5
44 McCormack K, Scott NW, Go PM, *et al.* Laparoscopic techniques versus open techniques for inguinal hernia repair. *Cochrane Database Syst Rev* 2003;:CD001785. doi:10.1002/14651858.CD001785
45 Beadles CA, Meagher AD, Charles AG. Trends in emergent hernia repair in the United States. *JAMA Surg* 2015;**150**:194–200. doi:10.1001/jamasurg.2014.1242

46 Hewitt DB, Chojnacki K. Groin Hernia Repair by Open Surgery. *JAMA* 2017;**318**:764–764. doi:10.1001/jama.2017.9868
47 Minimally-Invasive Hernia Repair | Hernia Surgery Center | Dartmouth-Hitchcock. https://www.dartmouth-hitchcock.org/hernia/minimally_invasive_hernia_repair.html (accessed 1 Jul 2019).
48 Inguinal hernia repair - How it's performed. UK NHS. 2017.https://www.nhs.uk/conditions/inguinal-hernia-repair/what-happens/ (accessed 1 Jul 2019).
49 Goodman LA. Snowball Sampling. *Ann Math Stat* 1961;**32**:148–70. doi:10.1214/aoms/1177705148
50 US Census Bureau. 2010 Geographic Terms and Concepts - Urban and Rural. https://www.census.gov/geo/reference/gtc/gtc_urbanrural.html (accessed 4 Feb 2019).
51 Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006;**8**. doi:10.2196/jmir.8.4.e27
52 Stellefson M, Chaney B, Chaney D, *et al.* Engaging Community Stakeholders to Evaluate the Design, Usability, and Acceptability of a Chronic Obstructive Pulmonary Disease Social Media Resource Center. *JMIR Res Protoc* 2015;**4**. doi:10.2196/resprot.3959
53 Giunti G, Kool J, Rivera Romero O, *et al.* Exploring the Specific Needs of Persons with Multiple Sclerosis for mHealth Solutions for Physical Activity: Mixed-Methods Study. *JMIR MHealth UHealth* 2018;**6**. doi:10.2196/mhealth.8996
54 Brewer LC, Jenkins S, Lackore K, *et al.* mHealth Intervention Promoting Cardiovascular Health Among African-Americans: Recruitment and Baseline Characteristics of a Pilot Study. *JMIR Res Protoc* 2018;**7**:e31. doi:10.2196/resprot.8842
55 Howard T. Journey mapping: a brief overview. *Commun Des Q Rev* 2014;**2**:10–3. doi:10.1145/2644448.2644451
56 Temkin BD. Mapping The Customer Journey. Forrester 2010. https://www.forrester.com/report/Mapping+The+Customer+Journey/-/E-RES55987
57 Improving healthcare through the use of co-design - New Zealand Medical Journal. https://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2012/vol-125-no-1357/article-boyd (accessed 3 Dec 2018).
58 *The field guide to human-centered design: design kit*. 1st. ed. San Francisco, Calif: : IDEO 2015.
59 Van Stee SK, Yang Q. Online Cancer Information Seeking: Applying and Extending the Comprehensive Model of Information Seeking. *Health Commun* 2018;**33**:1583–92. doi:10.1080/10410236.2017.1384350
60 Hartoonian N, Ormseth SR, Hanson ER, *et al.* Information-Seeking in Cancer Survivors: Application of the Comprehensive Model of Information Seeking to HINTS 2007 Data. *J Health Commun* 2014;**19**:1308–25. doi:10.1080/10810730.2013.872730
61 Han JY, Kim J-H, Yoon HJ, *et al.* Social and Psychological Determinants of Levels of Engagement With an Online Breast Cancer Support Group: Posters, Lurkers, and Nonusers. *J Health Commun* 2012;**17**:356–71. doi:10.1080/10810730.2011.585696
62 Han JY, Wise M, Kim E, *et al.* Factors Associated with Use of Interactive Cancer Communication System: An Application of the Comprehensive Model of Information Seeking. *J Comput-Mediat Commun JCMC* 2010;**15**:367–88. doi:10.1111/j.1083-6101.2010.01508.x
63 Ruppel EK. Scanning Health Information Sources: Applying and Extending the Comprehensive Model of Information Seeking. *J Health Commun* 2016;**21**:208–16. doi:10.1080/10810730.2015.1058438
64 Basnyat I, Nekmat E, Jiang S, *et al.* Applying the Modified Comprehensive Model of Information Seeking to Online Health Information Seeking in the Context of India. *J Health Commun* 2018;**23**:563–72. doi:10.1080/10810730.2018.1493058

65 Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qual Health Res* 2005;**15**:1277–88. doi:10.1177/1049732305276687
66 Topf M. Three Estimates of Interrater Reliability for Nominal Data. *Nurs Res* July/August;**35**:253–5.
67 Richardson A. Using Customer Journey Maps to Improve Customer Experience. *Harv Bus Rev* 2010.
68 Holtzblatt K, Wendell J, Wood S. *Rapid Contextual Design - 1st Edition*. 1st ed. Elsevier https://www.elsevier.com/books/rapid-contextual-design/holtzblatt/978-0-12-354051-5 (accessed 2 Feb 2019).
69 Ventral Hernia Repair. https://www.facs.org/~/media/files/education/patient%20ed/ventral_hernia.ashx (accessed 4 Mar 2019).
70 Simply Put A Guide for Creating Easy-to-Understand Materials. Atlanta, GA: : Centers for Disease Control and Prevention 2009.
71 Wang W, Siau K. Trust in Health Chatbots. San Francisco, CA: 2018.
72 Wang W, Siau K. Trusting Artificial Intelligence in Healthcare. *AMCIS 2018 Proc* Published Online First: 16 August 2018.https://aisel.aisnet.org/amcis2018/TREOsPDS/Presentations/125
73 Holden RJ, Karsh B-T. The Technology Acceptance Model: Its past and its future in health care. *J Biomed Inform* 2010;**43**:159–72. doi:10.1016/j.jbi.2009.07.002
74 Jimison H, Gorman P, Woods S, *et al.* Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. *Evid ReportTechnology Assess* 2008;:1–1422.
75 Song J, Zahedi F “Mariam.” Trust in health infomediaries. *Decis Support Syst* 2007;**43**:390–407. doi:10.1016/j.dss.2006.11.011
76 Ash JS, Berg M, Coiera E. Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. *J Am Med Inform Assoc* 2004;**11**:104–12. doi:10.1197/jamia.M1471
77 Montague ENH, III WWW, Kleiner BM. Trust in medical technology by patients and healthcare providers in obstetric work systems. *Behav Inf Technol* 2010;**29**:541–54. doi:10.1080/01449291003752914
78 Heathfield H, Pitty D, Hanka R. Evaluating information technology in health care: barriers and challenges. *BMJ* 1998;**316**:1959. doi:10.1136/bmj.316.7149.1959
79 Lippert SK, Davis M. A conceptual model integrating trust into planned change activities to enhance technology adoption behavior. *J Inf Sci* 2006;**32**:434–48. doi:10.1177/0165551506066042
80 Merritt SM, Ilgen DR. Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Hum Factors* 2008;**50**:194–210. doi:10.1518/001872008X288574
81 Dzindolet MT, Peterson SA, Pomranky RA, *et al.* The role of trust in automation reliance. *Int J Hum-Comput Stud* 2003;**58**:697–718. doi:10.1016/S1071-5819(03)00038-7
82 Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Hum Factors* 2004;**46**:50–80. doi:10.1518/hfes.46.1.50_30392
83 Li X, Hess TJ, Valacich JS. Why do we trust new technology? A study of initial trust formation with organizational information systems. *J Strateg Inf Syst* 2008;**17**:39–71. doi:10.1016/j.jsis.2008.01.001
84 Laranjo L, Dunn AG, Tong HL, *et al.* Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018;**25**:1248–58. doi:10.1093/jamia/ocy072
85 Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc JAMIA* 2018;**25**:1080–8. doi:10.1093/jamia/ocy052

86 Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods* 2006;**18**:59–82. doi:10.1177/1525822X05279903
87 Lor M, Backonja U, Lauver DR. How Could Nurse Researchers Apply Theory to Generate Knowledge More Efficiently?: Applying Theory in Research. *J Nurs Scholarsh* 2017;**49**:580–9. doi:10.1111/jnu.12316

CHAPTER 4. HERNIA COACH - TOWARDS A PATIENT-CENTERED HEALTH DIALOG SYSTEM TO AID INGUINAL HERNIA SURGERY PATIENT HEALTH INFORMATION-SEEKING

4.1 Definitions

- ASQ: After Scenario Questionnaire, a three-item measure to assess participant satisfaction after completing a usability testing scenario [1].
- DS: Dialog system, a computer system designed to communicate with users in natural language. Common commercial dialog systems include Alexa, Cortana, the Google Assistant, and Siri [2].
- eHEALS: the eHealth Literacy Scale Survey, an eight-item survey developed to assess electronic health literacy [3].
- Hernia Coach: A patient-centered prototype Google Assistant “Action”, which was designed, developed, and evaluated in this research study. Hernia Coach serves as an HDS to aid inguinal hernia surgery patient health information-seeking.

- HDS: Health dialog system, an automated system that can interview patients and consumers about their health and provide health education and behavior change interventions using natural language [4].
- HIPAA: the Health Insurance Portability and Accountability Act, a set of Federal laws concerning patient privacy, data security, and sharing of protected health information [5,6].
- HIT: Health Information Technology, electronic systems healthcare professionals and patients use to store, share, and analyze health information [7].
- HITAM: Health information technology acceptance model, a theoretical model developed to extended Technology Acceptance Model [8] in health care by describing health consumers' behavioral intention of using health information technology [9].
- NLU: Natural language understanding, a form of natural language processing to identify the meaning of text or speech input into dialog systems by users [10].
- PSSUQ, Post-Study System Usability Questionnaire a 19 item survey to assess a user's satisfaction with a computer system [11].

4.2 Introduction

Dialog systems (DSs) (i.e., virtual assistants, conversational agents, chatbots) such as Alexa, Cortana, the Google Assistant, and Siri are computer systems designed to converse with users [2] and are being used by patients for health information-seeking [12]. However, these common DSs may provide poor or inaccurate health information which is a safety risk for patients [13]. Currently, it is possible to extend the functionality of some common DSs with “Skills” and “Actions” [14–16]. “Skills” and “Actions are created by persons who are external to the DS organization and are analogous to creating “apps” for smartphones. In 2018, there were 309 health and fitness Alexa “Skills” or Google Assistant “Actions” [12]. The ability to create DS health-focused “Skills” or “Actions” suggests it is possible to design a patient-centered resource to aid patient health information-seeking.

Patient-centered health dialog systems (HDSs) [4] are a type of DSs that aid patient information-seeking [17–19], promote behavior change [20–23], provide reminders [24], promote self-compassion [25], and triage patient concerns [26–29]. Although researchers have studied how HDSs can support patients, little to no research has been conducted to identify how surgery patient information-seeking may be facilitated with an HDS. In general, surgery patients have information needs throughout their surgery experiences, which include the times prior to surgery [30–37], during the day of surgery [38], and after undergoing surgery [39–43]. To meet patient information needs clinicians routinely provide health education to patients. However, research has shown patients have difficulty recalling the information provided to them by clinicians [44,45]. Clinicians will also typically provide patients with educational materials, such as leaflets and brochures, to aid patient recall and understanding [46]. While these education materials have been identified by patients as valuable sources of information [44], the materials have also been shown to have reading grade levels greater [45–49] than the average U.S. citizen [50] and Medicare beneficiaries [51]. The high reading grade levels negatively impact the ability of patient education materials to improve health understanding [45–49] and patient comprehension prior to surgery [52]. Multimedia, such as interactive computer programs and websites, have been shown to improve patient comprehension prior to surgery [52,53]. This research indicates surgery patients may benefit from additional information resources, such as an HDS, to aid their information-seeking.

To explore HDSs as an alternative information resource, we designed, developed, and evaluated a patient-centered prototype Google Assistant “Action”, called “Hernia Coach”, to serve as an HDS and aid inguinal hernia surgery patient information-seeking. Inguinal hernias are a common health condition [54] with an estimated 500,000 - 600,000 people in the United States undergoing surgery to have their inguinal hernia repaired every year [55,56]. In general, inguinal hernia patients are male (96% of surgeries), middle-aged or older (mean 54 years old) [57–59], and undergo outpatient surgery, where patients are discharged home the same day as the surgery [60–62]. This allowed us to focus our efforts while working towards achieving significant impact when creating Hernia Coach due to the high prevalence of inguinal hernias and the patient population’s generally homogeneous hospital experiences and demographics.

In this chapter, we present our study, which comprised of three stages. In the first stage, we evaluated common DSs to identify which one was best suited for developing Hernia Coach. In the second stage, we designed and developed Hernia Coach in four phases where we 1) generated the health content incorporated into Hernia Coach, 2) created personas to focus Hernia Coach development efforts on target users, 3) generated scenarios to envision typical Hernia Coach user activities, and 4) used these findings to develop the Hernia Coach system. In the third stage, we evaluated Hernia Coach in three phases where 1) design experts conducted heuristic evaluations, 2) previous inguinal hernia surgery patients engaged in usability testing sessions, and 3) we evaluated the system's ability to provide relevant information to participant queries in the first two evaluation phases. After our review of related work, we present the methods and results, grouped together for each task of the overall work. Then we end with a discussion Hernia Coach's ability to aid inguinal hernia surgery patient information-seeking and design recommendations for HDS designers and developers.

4.3 Related Work: Dialog Systems to Support Patient Health Information-Seeking

Health Dialog Systems

Within the last decade, HDSs have been developed to support patient-health information seeking. Goldenthal et al. (2019) created an HDS to support patient information-seeking after ureteroscopy. Ureteroscopy is a common outpatient procedure to treat kidney stones [17]. Vaira et al. (2018) created an HDS, called "MamaBot," to provide pregnant women, mothers, and families with young children information about hospitals, pharmacies, nutrition, emergency management, and pregnancy or child growth issues [18]. Crutzen et al. (2010) created an HDS, called "Bzz," to answer adolescents' questions about sex, drugs, and alcohol [19].

Descriptions of these HDS's development and evaluation efforts and ranged from limited to robust. The ureteroscopy HDS health content generation methods were described, but information about the underlying technology powering the HDS was very limited. The ureteroscopy HDS's evaluation efforts were also limited in scope. Seven of the twenty patients enrolled in the study used the HDS and participated in a semi-structured follow-up interview. Only three of the patients' perceived HDS benefits are described in the article. Two patients reported using the ureteroscopy HDS for information-seeking about concerning symptoms, which helped them understand their health conditions after surgery. The third patient identified the ureteroscopy HDS would be useful for efficient information-seeking for non-emergent questions. The other patient Mamabot's development efforts were described in detail. However, we could not identify literature evaluating or describing the use of Mamabot by patients. The researchers conducted a pilot implementation of Bzz with 929 participants. The participants rated Bzz with low ease of use, highly reliable information, and moderate information quality, which indicated Bzz's design could be improved but users found Bzz provided pertinent health information that had applicability to their questions. The participants also preferred Bzz more than phone hotlines and search engines for health information-seeking. This research demonstrates HDSs to support patient health information-seeking needs to be more rigorously reported, evaluated, and investigated further.

Commercial Dialog Systems

Researchers have also assessed commercial DSs themselves, not "Skills" or "Actions," to aid patient information-seeking. Miner et al. (2016) queried Siri, Google Now (Google Assistant precursor), Cortana, and S Voice (Samsung Bixby precursor) with statements and questions regarding mental health, interpersonal violence, and physical health to identify if the DSs were capable of identifying a crisis situation, responding with respectful language, and referred the user to medical services. The results were varied. Siri and Google Now were able to recognize suicide ideation and refer users to a suicide prevention hotline. Siri recognized statements concerning depression but did not refer users to medical resources. Cortana identified queries concerning rape and referred users to a sexual assault hotline. Siri identified physical health concerns and referred users to nearby hospitals [63].

Bickmore et al. (2018) recruited participants to query Alexa and the Google Assistant for medical, medication, and emergency situation information. The researchers identified 44 instances that potentially could have resulted in user harm. Thirty percent of these instances were attributed to incorrect information provided the DS, 46% to user error, and 25% consisting of both incorrect DS information and user error. In 21% of instances, the user-provided partial information in their query, which yielded an incorrect query response. In 16% of instances, the user-provided complete information in their query but the DS responded with a partially valid query response. Overall, Siri was identified to have caused the most potential harm and potential death, the Google Assistant was second, and Alexa was third [13].

Palancia et al. (2019) identified how effectively Alexa, the Google Assistant, and Siri identified the top 50 prescribed medication names. The Google Assistant had the highest accuracy rates, followed by Siri, and then Alexa. The researchers also found foreign accents affected Alexa's and Siri's medication name understanding [64].

Cho (2019) found when DS users engaged in non-sensitive health information-seeking using the Google Assistant, voice-based interaction modalities enhanced the social presence of the DS and users reported fewer privacy concerns. When users engaged in sensitive health-information seeking with high privacy concerns, the social presence of the Google Assistant did not differ based on voice-based or text-based interaction modalities. Cho also found the hardware used to interact with the DS (e.g., smart speaker, smartphone) did not affect user attitudes when engaging in health information-seeking [65].

This research indicates commercial DSs are poorly equipped to support patient-information seeking. These technologies have been developed to support accomplishing tasks, solving problems, or answering general questions. Supporting patient health information-seeking has not been a focus for the organizations developing these commercial DSs. The creation of health domain-specific DS "Skills" or "Actions" has the potential to mitigate the poor health information provided by DSs and better support patient health information-seeking.

Surgery Focused Dialog Systems "Skills" and "Actions"

We identified two surgery patient-centered HDS "Skills" or "Actions" available through DSs. The first HDS was "Hospital for Special Surgery," which was available as both an Alexa "Skill" and Google Assistant "Action." This HDS had limited functionality focused on connecting patients with resources at the New York-based hospital [23]. Additionally, the HDS only had one health information component about back injury pain, which then directed users to contact the hospital for an appointment. The second HDS we identified was “My Children’s Enhanced Recovery After Surgery” [66,67], which was an early Health Insurance Portability and Accountability Act (HIPAA) compliant Alexa “Skill.” This “Skill” was created by Boston’s Children Hospital to support parents and caregivers of children after cardiac surgery. Specifically, the “Skill” facilitated parents and caregivers providing updates to their child’s healthcare team concerning their child’s recovery progress after cardiac surgery. Parents and caregivers could also obtain information about post-surgery appointments using the “Skill”.

Of the two identified surgery “Skills,” one primarily supported connecting users with hospital resources and provided very limited health information. The other focused on collecting pediatric surgery recovery outcomes. This related work demonstrates the potential of DS to support health information-seeking and HDS built as “Skills” or “Actions.” However, this research leaves many questions unanswered. Little is known about which DS is best suited to build HDSs, the development process required to build HDSs as “Skills” or “Actions”, and if HDS “Skills” or “Actions” can effectively support patient health information-seeking.

4.4 Stage 1: Dialog System Evaluation

4.4.1 Stage 1: Dialog System Evaluation Methods

In the first stage of this study, we evaluated DSs to identify which was best-suited to design and develop Hernia Coach. We considered two potential health-focused software selection frameworks to facilitate the evaluation by reviewing peer-reviewed literature. The first is the Multiple Criteria Decision Analysis for Health Technology Assessment [68]. The second is the Framework for Selecting Digital Health Technology [69]. The Framework for Selecting Digital Health Technology was developed to guide the selection of digital health software and has been promoted as a method to evaluate emerging health information technology [70]. The Multiple Criteria Decision Analysis framework was adapted to guide organizational health technology assessments [68], which is defined by the World Health Organization as a systematic evaluation of health technology to inform policy decision making [71]. Given our need for a framework to identify software for building Hernia Coach, not for informing policy decision making, we selected the Framework for Selecting Digital Health Technology.

Guided by the Framework for Selecting Digital Health Technology, we performed the following three steps: 1) Select inclusion criteria to evaluate the technology options, 2) create a list of candidate technologies, 3) develop the evaluation process, evaluate, and rank the candidate technologies. We conducted the DS evaluation in January and February 2019.

Step 1) Select Inclusion Criteria to Evaluate Technology Options

We decided to focus on smartphone-oriented DSs as opposed to other DS interaction hardware (e.g., smart speakers) while selecting the DS inclusion criteria. This was motivated by a desire for Hernia Coach to be easily accessible for patients in a variety of settings such as the home, hospital, or other settings where they would need to access health information. U.S. adults keep their phones readily available throughout their whole day. In 2013, a survey of smartphone users aged 18-44 found 79% of the respondents have their smartphone on their person for all but two hours of their waking day and 25% could not recall the last time their smartphone was not in the same room as them [72]. Other forms of DS hardware (e.g., smart speakers) are less portable, making them less readily available to provide answers to questions. Additionally, smartphones have greater reach than smart speakers. In 2018, 77% of U.S. adults owned a smartphone [73] compared to 19% reporting interacting with a smart speaker [74].

In addition to developing Hernia Coach for smartphones; we opted to also focus on DSs that allow external innovators to extend the DSs to new domains through the creation of new “Skills” or “Actions”. “Skills” and “Actions” are additional functionality for DSs developed by persons who are external to the DS organization, which is analogous to creating “apps” for smartphones. We chose to build Hernia Coach with an existing DS because we wanted to create an information-seeking tool using technology that was already widely adopted for the greatest reach possible, leverage existing innovation in the DS space, and to build an information resource patients may already be familiar with using.

These considerations resulted in two inclusion criteria for the DS evaluation. First, the candidate DSs must be used by greater than 1% of U.S. adults who own a smartphone. Second, the DS must facilitate the creation of new functionality by external developers through “skills” or “actions.”

Step 2) Create a List of Candidate Technologies

We identified candidate DSs by reviewing DS adoption research and developer documentation, which contained the information necessary to determine if the candidate technologies met our inclusion criteria. First, we queried multiple peer-reviewed literature research databases including Google Scholar, IEEE, ACM Digital Library, PubMed and online search engines to identify candidate DSs. We recorded the identified research articles in a Zotero library [75] (i.e., citation manager). Second, we applied the inclusion criteria established in step one to all of the identified DSs, which resulted in a final list of candidate DSs for evaluation.

Step 3) Develop the Evaluation Process, Evaluate, and Rank the Candidate Technologies

We conducted a literature search using the research databases in step 1 to identify software selection criteria to evaluate the candidate DSs. During the literature search, we could not identify software selection criteria for patient-centered software, HDSs, or DSs. However, we identified a systematic review of software selection criteria for organizations that synthesized 52 software selection methodology studies and identified

67 individual software selection criteria comprising fifteen high-level criteria groups [76]. We identified the criteria in the systematic review as having the potential to facilitate a holistic DS evaluation.

We reviewed each individual criterion and identified 17 criteria with a focus on enterprise software implementation, which we determined not to be relevant for the DS evaluation and the 17 criteria were excluded. Additionally, the systematic review did not include health software selection methodology papers in the synthesis, which prompted us to adapt two of the criteria to better focus the criteria on healthcare applications. Specifically, we adapted the “security levels” criterion regarding the permitted use of protected health information as established by The U.S. Department of Health and Human Services [77] HIPAA regulations. HIPAA pertained to patient privacy, data security, and sharing of protected health information [5,6]. We adapted this criterion because four patient participants and three surgeons had voiced HDS data privacy and trust concerns in Aim 2. Second, we adapted the vendor’s “past business experience” to focus on “past healthcare experience.” In addition to the two criteria adapted for health, we also adapted the “interoperability” criterion to include a focus on integration with external natural language understanding engines (NLUs). NLUs are a form of natural language processing to identify the meaning of text or speech input by DS users [10]. We identified this was an important consideration prior to evaluating the DSs; some DSs could integrate only with a proprietary NLU specific to a DS and some permitted the integration with multiple third-party NLUs. Some third-party NLUs facilitated distributing a “skill” or “action” with multiple DSs simultaneously, which could extend the reach of

Hernia Coach during future development. These adaptations to the criteria resulted in a final set of 50 criteria comprising twelve high-level criteria groups. We include a complete table of the adapted DS evaluation criteria and results in Supplemental Table 1.

In addition to adapting the criteria, we categorized each of the 50 included criteria as low, medium, or high to indicate the criticality of each criterion and help evaluators distinguish between the software selection options more effectively [76]. We defined “high” as critical to DS development; “medium” as important, but not critical; and “low” as not important or critical, but relevant. We developed three criteria categories to minimize the ambiguity between which criteria are most important for HDS development. Each criterion was categorized based on the considerations used to generate the inclusion criteria in step 1 and as determined by the authors based on their experiences developing and using health information systems.

After we established and categorized the DS criteria, we then extracted information for each candidate technology for all criteria. The extraction process involved using the research databases in step 1 to identify documentation with the requisite information for each criterion and candidate DS. We archived all identified documentation archived in a Zotero library. We extracted pertinent information for each candidate DS and criterion from the documentation into a spreadsheet.

When data extraction for all of the candidate DSs and criteria was completed, we evaluated how well each of the DSs met each criterion categorized as high. Then we looked across all of the DSs, and the criteria categorized as high, to rank the candidate DSs. If a clear candidate was not identified from the criteria categorized as high (i.e., ranking tie), the process would continue using the medium and low criteria categories until a clear candidate DS was identified.

4.4.2 Stage 1: Dialog System Evaluation Results

Five candidate DSs were identified with adoption by greater than 1% of U.S. adults who own a smartphone [78]. Apple's Siri [79] had a 44% adoption rate, the Google Assistant [80] had a 30% adoption rate, Amazon's Alexa [81] had a 17% adoption rate, and Microsoft's Cortana [82] and Samsung's Bixby [83] each had a 4% adoption rate. Four of the five DSs allowed for outside developers to extend these technologies to new domains [14–16,84]. The exception is Apple's Siri, which only permits integration through separate iOS applications [85], which excluded it from the DS evaluation. We included Amazon's Alexa, Samsung's Bixby, Microsoft's Cortana, and the Google Assistant for the DS evaluation.

The Google Assistant was selected to implement Hernia Coach because it met the criteria categorized as high most effectively compared to the other DSs at the time of the evaluation. The Google Assistant: integrated with external NLUs; had the second most “skills” or “actions” developed; permitted some health-related data collection (excluding HIPAA protected data); was available on both Android and iOS smartphones;

had the largest adoption on smartphones by U.S. adults; supported voice, screen, and keyboard interaction modalities; permitted the incorporation of text, audio, video, and images for information presentation; and Google has previous experience in the healthcare industry. Additionally, Dialogflow was selected as Hernia Coach's NLU because Dialogflow was the officially supported NLU for Google Assistant Actions [86,87] and supported additional integration with Alexa and Cortana [88]. Dialogflow is a Google tool that analyzes text and audio inputs from DS users and facilitates the creation of conversational user interfaces for smartphones, smart speakers, or website [89]. We include a DS evaluation summary table of the criteria ranked as high in Table 1.

Table 1: Dialog System evaluation result summary table of the criteria categorized as high.

Criteria	Alexa	Bixby	Cortana	Google Assistant
NLU integration	Yes	No	Yes	Yes
Openness	50,000 “skills” developed	Unknown number of “capsules” developed	235 “skills” developed	2,400 “actions” developed
Health security and privacy policies (early 2019)	Yes, limited health functionality permitted, HIPAA protected data not permitted	No	Yes, limited health functionality permitted, HIPAA protected data not permitted	Yes, limited health functionality permitted, HIPAA protected data not permitted
Supported smartphones	Android, iOS	Samsung smartphones (limited)	Android, iOS	Android, iOS
U.S. adult smartphone adoption (late 2018)	17%	4%	4%	30%
User interaction modalities	Voice, screen (limited)	Voice, screen, keyboard	Voice, screen, keyboard	Voice, screen, keyboard
Data visualization	Text, audio, video, images	Text, audio	Text, audio, images	Text, audio, video, images
Previous vendor healthcare experience	Yes	Yes	Yes	Yes
Rank	2	4	3	1
Selected for study	**No**	**No**	**No**	**Yes**

4.5 Stage 2: Hernia Coach Design and Development

Stage two of this study involved the design and development of Hernia Coach, which comprised of four phases. First, we generated Hernia Coach's health content. Second, we created personas to guide Hernia Coach's development. Third, we generated scenarios to envision typical Hernia Coach user activities and guide Hernia Coach evaluation activities in stage 3. Fourth, we developed Hernia Coach itself.

4.5.1 Stage 2 - Phase 1: Hernia Coach Health Content Generation

4.5.1.1 Stage 2 - Phase 1: Hernia Coach Health Content Generation Methods

For the first development phase, we generated Hernia Coach's health content to provide key information for patients regarding inguinal hernia surgery. To generate Hernia Coach's content, we extracted health information from inguinal hernia surgery patient education materials published online by trusted institutions such as national healthcare organizations, medical centers, and hernia medical supply manufacturers. We used online search engines to identify the patient education materials. The patient education material identification and health information extraction occurred in February and March 2019. We archived all identified patient education materials in a Zotero library.

We generated and segmented the health information extracted from the identified patient education materials by surgery stage, topic, and sub-topic. We segmented the extracted health information for two reasons. First, we aimed to align the extracted health information with the three surgery stages (pre-, day of, and post-surgery) and topics identified during the participatory design sessions in Aim 2**.** Second, if a topic's character length was greater than 300 characters and we identified common concepts across the identified patient education materials within a topic, we then created subtopics for the concepts. We aimed for the length of each topic and subtopic to be 300 characters based on the Google Assistant design recommendations [90].

We also aimed for Hernia Coach's health content to support patient understanding by striving to write the health content at a 5th-grade reading level. We chose a 5th-grade reading level for several reasons. First, while the average U.S. resident reads at an 8th-grade reading level [50] the average Medicare beneficiary reads at a 5th-grade reading level [51]. Given that many inguinal hernia patients could be eligible for Medicare based on age [57–59], we aimed for text to align with their average reading level. Second, the Joint Commission states patient health information should be written at a 5th-grade reading level [91]. We evaluated the identified patient education materials and Hernia Coach's health content for readability using the "Flesch reading ease" and "Flesch-Kincaid reading grade level" tests [92] using Microsoft Word [93]. These two tests are widely used to evaluate the readability of health information [46,94–96]. Flesch reading ease is measured from 0-100. The higher the Flesch reading ease score, the easier the document is to read. The Flesch-Kincaid reading grade level ranges from 5th grade to college graduate (i.e., 13th grade or higher). The higher the Flesch-Kincaid reading grade level, the harder the document is to read. We generated readability test descriptive statistics using Tableau [97].

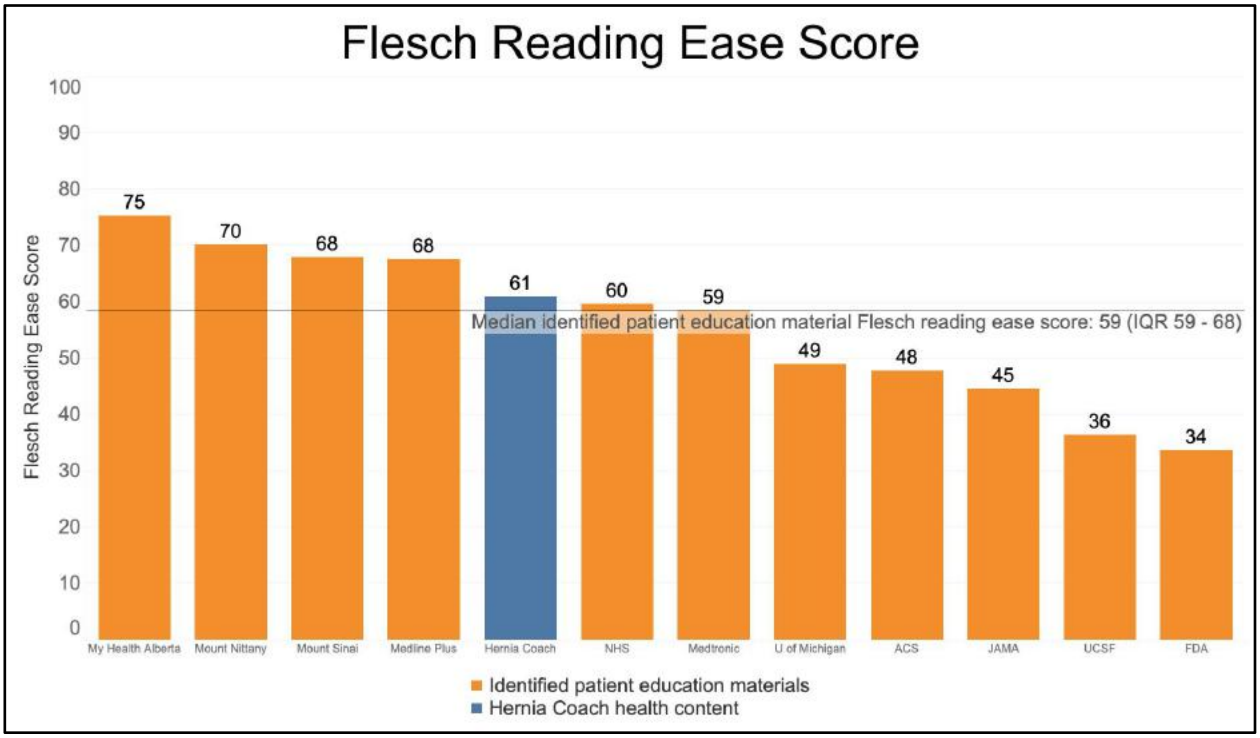
4.5.1.2 Stage 2 - Phase 1: Hernia Coach Health Content Generation Results

We primarily extracted Hernia Coach's health content from the American College of Surgeons inguinal hernia surgery patient information packet [54]. This particular patient education material was selected as the primary document for extraction because the organization is a highly regarded healthcare organization. Hernia Coach's health content was extracted from ten other patient educational materials, which were: the Federal Drug Administration [98], a U.S. Federal organization responsible for promoting health; the Journal for the American Medical Association [99], a peer-reviewed medical journal publishing original healthcare research; the MedlinePlus Medical Encyclopedia [100], a patient-oriented resource maintained by the U.S. National Institutes of Health; Medtronic [101,102], an inguinal hernia surgical supply vendor; the Mount Nittany Health System [103], a Pennsylvania based hospital group; the Mount Sinai Health System [104], a New York-based hospital group; My Health Alberta [105], a Canadian provincial health information resource the United Kingdom's National Health System [106]; the United Kingdom's single-payer healthcare organization, the University of Michigan Medical Center[107], an academic medical organization; and the University of California at San Francisco Medical Center [108], an academic medical organization.

We generated and segmented health content for pre-surgery, day of surgery, and post-surgery consisting of six high-level topics and 42 sub-topics. Hernia Coach's health content focused on inguinal hernia information without comorbities. The high-level topics concerned inguinal hernia surgery, inguinal hernias in general, activity restrictions after surgery, consuming food or beverages, pain, and tobacco use. The median topic and sub-topic lengths were 276 characters (IQR 225 - 326). We include Hernia Coach's topics, sub-topics, and health content in Supplemental Table 2.

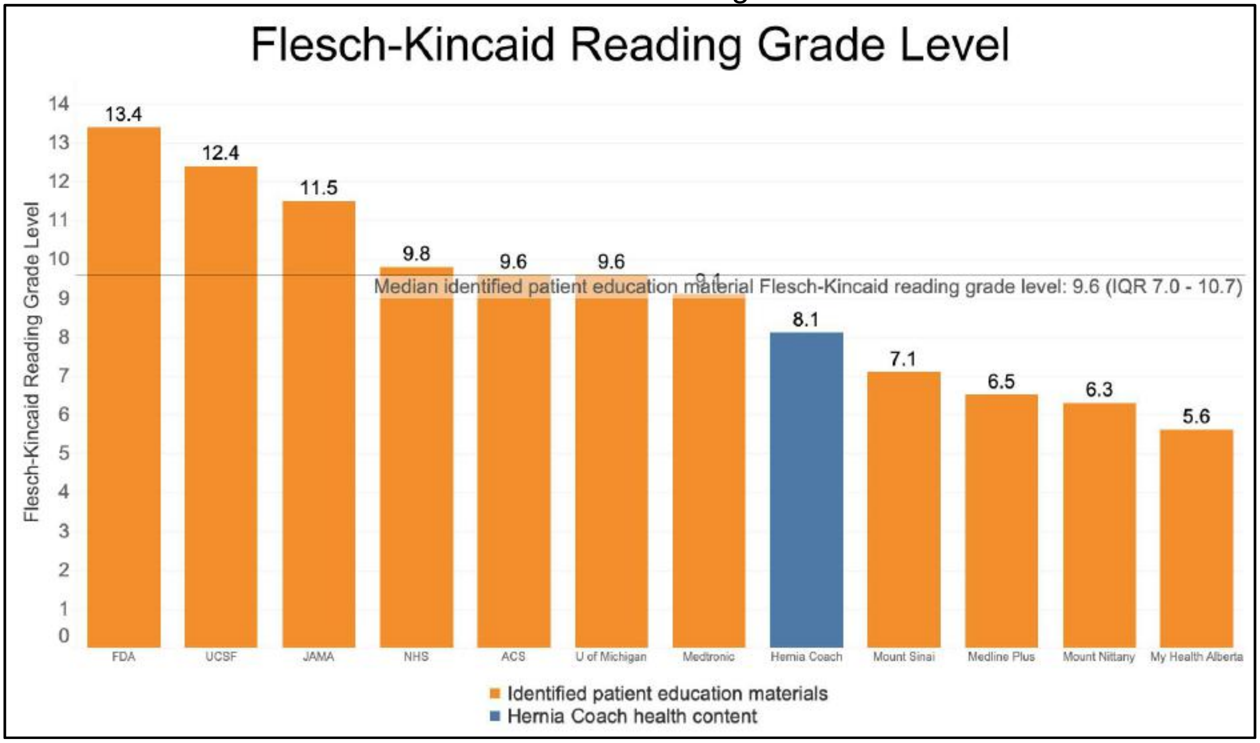
The identified patient education material median Flesch reading ease score was 59 out of 100 (IQR 47 - 68) and the median Flesch-Kincaid reading grade level was 9.4 (IQR 7.0 - 10.7). The content we created for Hernia Coach had a Flesch reading ease score of 61 out of 100 and Flesch-Kincaid reading grade level of 8.1. The largest contributing factor for Hernia Coach's reading level were due to long complex sentences distilling topics and sub-topics to approximately 300 characters. Breaking up the long sentences into multiple sentences would have decreased the reading grade level but simultaneously increased the topic and subtopic character counts. Comparisons between the identified patient education material and Hernia Coach's health content regarding Flesch reading ease scores are available in in Figure 1a and Flesch-Kincaid reading grade levels in Figure 1b.

Figure 1a: Identified patient education material Flesch reading ease scores, median identified patient education material reading ease score, and Hernia Coach health content Flesch reading ease score.



Flesch reading ease score range: 0-100. A higher score indicates easier readability. IQR = Interquartile range.

Figure 1b: Identified patient education materials, Flesch-Kincaid reading grade level, median identified patient education material reading grade level, and Hernia Coach health content Flesch reading ease score.



Flesch-Kincaid reading grade level range: 5th grade to college graduate (i.e., 13th grade or higher). A higher grade level indicates more difficult readability. IQR = Interquartile range.

4.5.2 Stage 2 - Phase 2: Personas to Guide Hernia Coach's Development and Evaluation

After we completed the DS evaluation and generated Hernia Coach's health content, we created personas, which are fictional people, to guide Hernia Coach's development. We chose to create personas because they are a tool to create a strong focus on target users during software design and development activities [109,110]. Human-computer interaction researchers have also argued personas integrated into scenarios make scenarios more compelling and effective [109]. Scenarios are described in greater detail in the following section. In prior health informatics research, personas have been used to inform the design of a mobile application to support patients with brain tumors [111], to evaluate personal health records for homebound older adults [112], and to develop a mobile health application to improve mental health [113]. Based on this research demonstrating the applicability of personas to aid software development, we created three personas that were integrated into subsequent parts of this research study: in stage 2 - phase 3 we used the personas to make the scenarios more effective, in stage 2 - phase 4 we used the personas in the generation of our scenarios described we used the personas to focus our Hernia Coach development efforts on target users, in stage 3, phases 1 and 2 we used during the evaluation of Hernia Coach.

We identified peer-reviewed literature describing characteristics of inguinal hernia surgery patients [57–59] to generate the three personas. Age and gender were the only persona-related characteristics we could identify in the peer-reviewed inguinal hernia literature. As a result, we created two male personas because inguinal hernias typically occur in males (96% of surgeries). We also created a female persona because women also undergo inguinal hernia surgery, but at a lower frequency than men. All three personas represent typical inguinal hernia surgery patients, a mean age of 54 years. However, personas consist of more characteristics than just age and gender, the two characteristics described in the literature. Personas also may include characteristics concerning technology use, education levels, or work activities that may impact their ability, desire, and access to technology like Hernia Coach [109]. We identified and incorporated additional characteristics into the personas we perceived as relevant to designing and developing Hernia Coach. These additional characteristics concerned education levels, occupation, home geographic region, hospital location geographic region, social environment, experience with technology, perspective of technology, DS use, and health information-seeking activities. We incorporated these characteristics for a number of reasons. Education and reading ability are key determinants of a person's health status [114] and lower levels of education and income negatively affect a person's use of information technology [115,116]. The geographic region of a patient and healthcare resources can also have an effect on health status. Patient proximity to healthcare resources has an impact on access to healthcare and health outcomes; for example patients who are further away from hospitals in rural areas are less likely to access hospital resources [117]. Additionally, rural residents seek health information at lower rates, are more likely to use traditional patient education materials such as brochures or online health websites [118,119], and are less likely to be confident when conversing with physicians about health concerns [120]. Social environments have an effect on a person's health status [114]. Finally, a U.S. citizen's digital literacy can range from accomplishing very simple tasks with one technological tool to solving complex problems with multiple technological tools [121]. We wrote content for each persona characteristic to represent the potential diversity of Hernia Coach users. We include the patient personas in 2.

Table 2: Patient persona characteristics generated to guide Hernia Coach development and usability testing scenarios

Persona Attribute	Persona 1	Persona 2	Persona 3
Photo			
Name	Linda Hernandez	John Brooks	Adam Campbell
Age in years	51	57	69
Gender	Female	Male	Male
Education level	Bachelor's degree	Master's degree	High School Graduate
Home geographic region	Urban cluster	Urban	Rural
Hospital geographic region	Urban	Urban	Urban cluster
Occupation	Sales manager	Teacher	Retired farmer
Family members	Linda is married with two kids, a son and daughter. Her son is a sophomore in college. Linda's daughter recently graduated with a bachelor's degree and is living abroad. Her husband is semi-retired from his job as a consultant and splits his time working and volunteering with children at the local aquarium.	John is married with one son. His kid is currently a senior in high school and preparing to start college in the fall in a neighboring state. His husband works full-time as a restaurant manager.	Adam is a widower and has three kids from his marriage, two daughters and one son. One son and one daughter currently live in other states but visit him often. His other daughter lives nearby in a different town.
Friends	Linda is a member of a local cookbook club and is an active leader within the group. Linda is also	John is a volunteer with a local non-profit. He assists with the fundraising and event organizing for the	Adam is an active member of his local 4-H program and assists with programs that

	still involved with her daughter's local girl scout troop even though her daughter left the organization a few years ago. Between these two groups Linda has been able to create lasting friendships.	group. John is also a regular attendee of community center events, where he gets to interact with his neighbors and other members of the community.	engage local kids with hands on agriculture activities. Adam also has fishing and hunting buddies, who regularly get together on the weekends.
Computer skills, knowledge, and abilities	Linda is considered a computer and smartphone power user at her job. She is an Excel and PowerPoint guru. Linda regularly assists her employees with troubleshooting and leads training sessions for new hires. Occasionally the IT department will ask her to try out new software or processes to provide feedback before making changes within her department.	John is a regular computer user for work and volunteering. He is not a power user nor does he struggle with using familiar technology. He uses Excel, PowerPoint, and Word on a regular basis. John owns a smartphone, which he uses primarily for calling/texting friends and family, email, social media, and games.	Adam has limited experience with computers and hasn't interacted with them on a regular basis. However, he did learn some basics from one of his younger farm hands as some of his business processes went digital over the past few years. He preferred to delegate those tasks to more technology familiar employees. Adam recently upgraded from a cell phone to a smartphone. He is still learning how to use all of the new features and primarily uses it to call, text, or email his kids and friends.
Perspective of technology	Linda views technology as either an enabler or a hindrance. Identifying how technology can improve efficiency and complement tasks is one of her favorite job responsibilities.	John views technology of a component of his daily life but does not make it a priority. He values both online and offline interactions.	Adam is a hesitant adopter of technology. He grew up during a time where computers were not commonplace and he was not exposed to them until he was middle aged. Even though Adam is skeptical and an infrequent user of technology, he still finds value and utility in using computers and smartphone for communication.
Dialog system use and experience	Linda is familiar with dialog systems and owns a smart speaker at	John has some experience with dialog systems. He has one built	Adam has access to a dialog system on his new smartphone but

	home. She also uses the dialog system included with her phone. She primarily interacts with dialog systems to control her smart home components, to seek out information, and transcribe emails/texts that would be cumbersome to type on her smartphone.	into his smartphone and has used it in the past, but not often. His teenage son uses the one on his smartphone regularly and would like to get a smartspeaker.	only used it a few times. He does like how the dialog system can aid him with performing tasks or locating information that would be tedious with normal smartphone interactions because he can speak to the device naturally.
Health information-seeking activities	Linda regularly searches for health information online and prefers to reference multiple sources. This allows Linda to triangulate and synthesize the information from the different sources. Linda is also a regular user of her hospital online patient portal. Within the portal she messages her care team, view lab test results, refill prescriptions, and schedules appointments. When needed or during emergencies she calls her care team over the phone.	John has consulted online sources in the past for health information. However, he was unsure if the content was accurate and had some difficulties making sense of the information. In light of this, he usually contacts his care team using the phone with questions. John also has some friends who work as clinicians and will occasionally engage them for advice.	Adam generally prefers to get his information from his doctor or care team. When he's underwent procedures in the past, he has referred to the paper handouts his doctor gave him. When he has questions, he will call his care team.

4.5.3 Stage 2 - Phase 3: Scenarios Generated for Hernia Coach's Evaluation

4.5.3.1 Stage 2 - Phase 3: Scenario Generation Methods

Scenarios are typically created to envision typical user activities during software development and are a common design tool used to evaluate software [122,123]. We generated scenarios to hypothesize how patients may use Hernia Coach. Then we used the scenarios to guide Hernia Coach's evaluation activities performed in stage 2 - phase 3. In general scenarios consist of four key components [122]. The first component is the scenario, which provides a setting for the user's activity. Second, the scenarios describe the people (i.e., personas, users) in the setting. The third component is the objectives of the people acting in the scenario. The fourth and final component describes the actions of the people in the scenario. With these four components, scenarios provide the framework for design-based human interaction research. Scenarios have been employed to inform consumer health informatics designs [124], to evaluate electronic health records [125,126], and to assess personal health records for older adults [112].

We created a persona-based inguinal hernia surgery patient journey map to depict common inguinal hernia surgery phases and patient health information seeking activities, which we include Figure 2. Journey maps depict a person's experience of an event [127–129]. We created the journey map to act as a foundation for the development of the scenarios. We generated the journey map by identifying common surgery phases and health information seeking activities documented by patients during the journey mapping exercise in Aim 2. We selected "John Brooks" to be the representative persona in the journey map because his age and gender represented a typical inguinal hernia patient as described in the literature [57–59], he had some familiarity with technology, but was not an expert user, and lives in an urban environment. Most U.S. adults have low to moderate levels of digital literacy [121] and the majority of U.S. citizens live in urban areas [130]. The other two personas represent people who either have great or very little familiarity with technology and live in suburban or rural areas. We wanted to use a persona in the scenarios which we perceived would be a typical user in regards to age, gender, digital literacy, and geographic region.

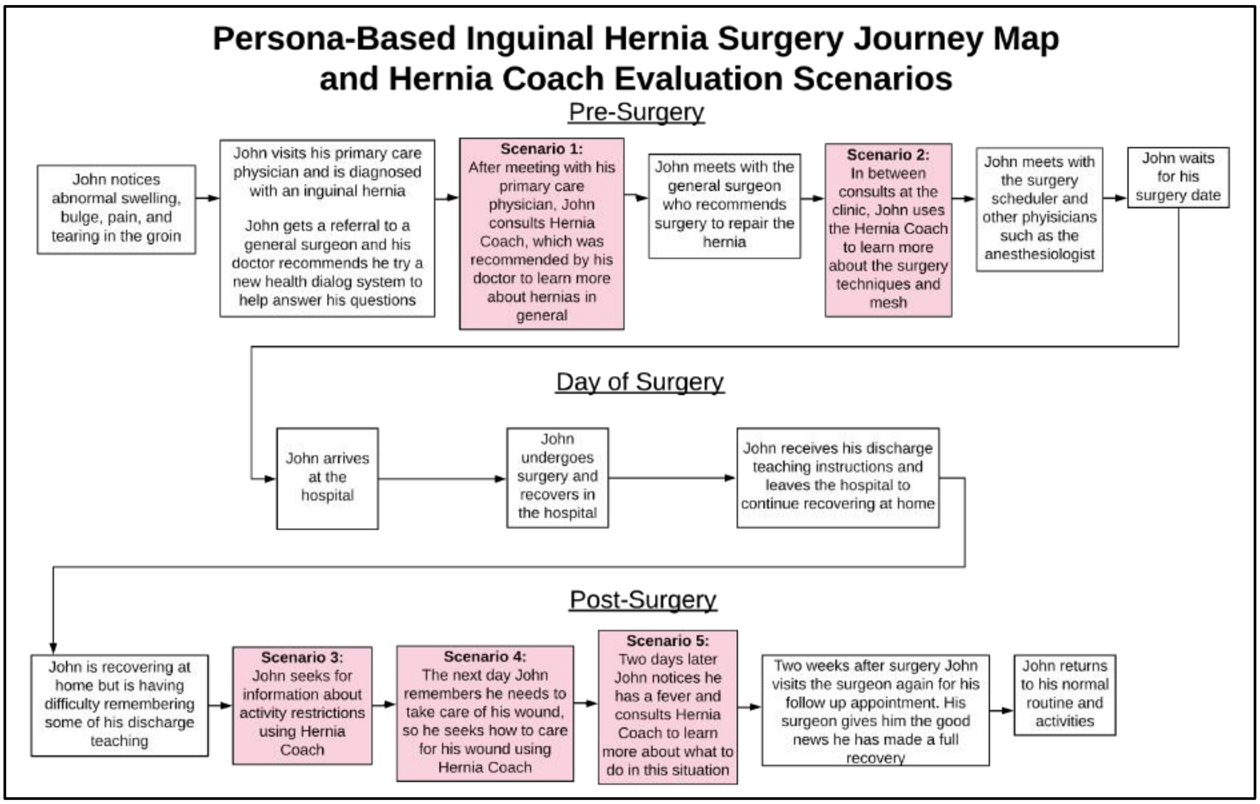
We selected two pre-surgery and three post-surgery fictional journey map components to act as the scenarios used in the phase 3 evaluations, which are highlighted in red in Figure 2. We selected these five scenarios and topics because the patient participants in Aim 2 generated questions during these two surgery phases the most. Additionally, the questions generated during the brainstorming sessions for all participants in Aim 2 generally concerned the topics incorporated into the highlighted scenarios.

4.5.3.2 Stage 2 - Phase 3: Scenario Generation Results

Details concerning the five highlighted scenarios, persona, settings, actions, goals, and example Hernia Coach queries for each of the five highlighted scenarios in Table 3. We also include the persona's anticipated Hernia Coach interaction modality (e.g., voice or keyboard) in the scenario actions. In two of the highlighted scenarios, the persona explicitly engages in information-seeking with Hernia Coach using voice and in two highlighted scenarios with a keyboard. The rationale for doing this was because we identified that not all situations are appropriate to engage in voice-based information-seeking using Hernia Coach, such as a clinic waiting room. In the final scenario, we gave the persona the option to select an interaction modality based on his personal preference. We opted to design the scenario interaction methods this way because the patients who participated in the usability testing sessions (stage 3 - phase 2) were instructed to query Hernia Coach based on the modality incorporated in the scenarios.

We wanted the participants to interact with Hernia Coach using both voice and keyboard twice. Then we wanted to give the participants the option to engage in information-seeking with their personally preferred modality in scenario five to later identify why they prefer a voice or a keyboard when interacting with Hernia Coach.

Figure 2: Persona-based inguinal hernia surgery journey map with the five Hernia Coach phase 3 evaluation scenarios highlighted in red.



Journey map components highlighted in red indicate scenarios used to guide Hernia Coach phase 3 evaluation activities.

Table 3: The five highlighted journey map scenario details; scenario people and settings; setting details and objective; actions; and example Hernia Coach queries.

Scenario	People and Setting	Setting Details and Objective	Actions	Example Hernia Coach Queries
Scenario 1	John at home	John is at home and in between clinic appointments on different days with his primary care physician and surgeon and wants to learn more about hernias.	John learns more about hernias in general using voice-based interaction methods with Hernia Coach.	What is a hernia? Why do hernias happen? How did I get a hernia? Do I need surgery?
Scenario 2	John in the hernia clinic waiting room	John is in between consults at the hospital with his surgeon and the anesthesiologist. While sitting in the waiting room, John decides he wants to learn more about surgery details while waiting for his next appointment.	John learns more about inguinal hernia surgery details using text-based interaction methods with Hernia Coach.	What types of anesthesia are used during surgery? What are the differences in surgical approaches? What is the mesh used to repair the hernia?
Scenario 3	John at home	John has recently returned home from the hospital after surgery, but he is having difficulties remembering the activity restrictions he is supposed to follow during his recovery. This prompts John to identify what he should avoid doing after surgery.	John learns more about post-surgery activity restrictions using voice-based interaction methods with Hernia Coach.	When can I resume my athletic hobbies (e.g. golf, going to the gym)? How soon can I have intimate relations with my significant other?
Scenario 4	John at home	That day after coming home from the hospital, John remembers he needs to take care of his wound and wants to learn more about the specifics of wound care.	John learns more about wound care using text based interaction methods with Hernia Coach.	How do I take care of my wound? How do I prevent a scar?
Scenario 5	John at home	John wakes up in the morning a two later and thinks he might have a fever. He seeks out information about what to do in this situation.	John identifies what to do with a fever after having surgery using his personally preferred interaction method with Hernia Coach.	Is a fever after surgery normal? Should I go to the hospital if I have a fever?

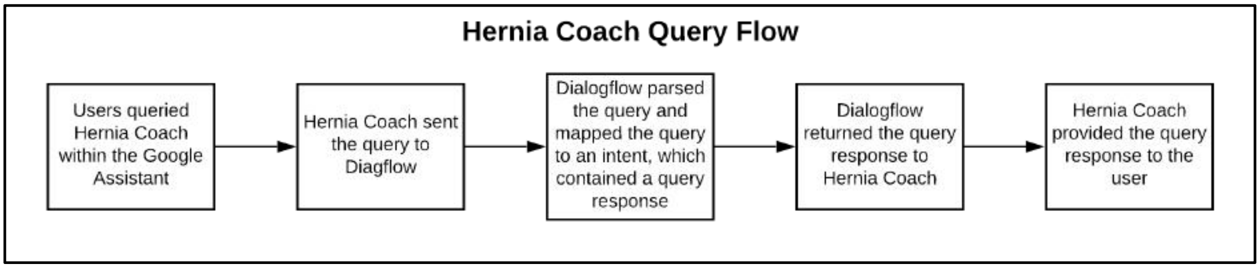
4.5.4 Stage 2 - Phase 4: Hernia Coach Development

After we evaluated the DSs, generated Hernia Coach's health content, and created personas and scenarios, we developed Hernia Coach itself. We present Hernia Coach's development in three parts: first a technical overview of Dialogflow and the Google Assistant, second Hernia Coach's conversation design and flow, and third descriptions of Hernia Coach's visual components.We referenced the personas previously described in stage 2 - phase 2; the Google Assistant Actions developer documentation [131]; the Google Assistant Action design documentation [132]; and the Dialogflow developer documentation [133] throughout Hernia Coach's development process.

4.5.4.1 Stage 2 - Phase 4: Hernia Coach Technical Overview

At the technical level, there were five components to Hernia Coach conversation process [134]. First, the user-submitted queries to Hernia Coach within the Google Assistant using voice or a keyboard. Second, the Hernia Coach sent the query to Dialogflow. Third, Dialogflow (i.e., the NLU) then parsed the query and mapped the query to an intent [134]. Intents were the distinct Hernia Coach health content topics and sub-topics described in stage 2 - phase 1 and included in Supplemental Table 2 (e.g., inguinal hernia description, the types of inguinal hernia surgery). Each intent had example training phrases of what a user might say or ask relating to the intent (e.g., "what is an inguinal hernia?"). These training phrases were used to help guide Dialogflow as to which intent matched a user's query. We incorporated the questions participants generated during the brainstorming activities in Aim 2 as training phrases where appropriate**.** Additionally, each intent contained a query response consisting of text and visual conversation components. Fourth, Dialogflow then returned the query response from the matched intent to Hernia Coach. Fifth, Hernia Coach provided the query response to the user. We include a figure depicting this process in Figure 3.

Figure 3: The flow of a user's query between all of the technical systems when using Hernia Coach.



We incorporated two different types of intents into Hernia Coach. The first were patient education intents, which are the health education information returned to a user's health related query (e.g., "an inguinal hernia is a weakness or tear of your abdomen inside your groin" which is returned with the query "what is a hernia?"). The second type of intents were system intents. Examples of a system intents are when the user gives input that system does not understand or if the user asks a question that is beyond the scope of Hernia Coach (e.g., "I missed what you said, could you please repeat the question?", "I do not know the answer to that question, ask your doctor"). For each intent query response, the ideal character length was 300 characters, with a maximum of 640 characters [90].

In total, we created intents for pre-surgery, day of surgery, and post-surgery consisting of six high level topics and 42 intents overall. The high level topics pertained to surgery information, hernia information, activity restrictions, eating or drinking, pain, and tobacco use. The median patient education query response was 276 characters (IQR 225 - 326).

Nine system intents were included in Hernia Coach, with a median character length of 71 (IQR 54 - 158). We include Hernia Coach's topics, sub-topics, and query responses in Supplemental Table 2.

We reviewed user queries on the Dialogflow (i.e., the NLU) website to identify if the correct responses were returned to the users and to improve Hernia Coach's query results. During this review process, we annotated correct responses as correct and incorrect responses with the correct response. This annotation process was then used to update intent training phrases used by Hernia Coach and improved Hernia Coach's query results during later use.

We conducted pilot testing with four friends and family members to improve Hernia Coach's query results before the evaluation activities conducted in phase 3. The friends and family members were prompted to ask questions about the topics in Hernia Coach. We recruited the friends and family members to identify semantic and syntactic understanding gaps in Dialogflow (i.e., the NLU) and to improve Hernia Coach before usability testing. Semantics relate to the meaning of speech and syntactics concern the structure of speech. Two of these friends and family members had learned English as a second language, which proved to be effective for improving system understanding. They asked questions that were semantically valid but syntactically confusing to Dialogflow, which we then annotated to improve Hernia Coach's query responses.

4.5.4.2 Stage 2 - Phase 4: Hernia Coach Conversation Design and Flow

To load Hernia Coach in the Google Assistant, users would say a phrase that would launch Hernia Coach, such as “Hey Google, I want to talk to Hernia Coach.” Then the Google Assistant would load Hernia Coach and present users with a disclaimer stating Hernia Coach is not a replacement for direct medical care and to call 911 in the event of an emergency. Some DSs required a disclaimer like this for health-related “Actions” or “Skills” [135–137]. After the disclaimer, a follow-up prompt would be presented to the user saying “to get started, you can ask me questions like ‘what is a hernia?’” We included this prompt to indicate to the user that they can begin interacting with Hernia Coach and asking questions.

At this point, the user interaction with Hernia Coach was intentionally open-ended to encourage users asking questions they wanted to be answered and facilitate the exploration of desired inguinal hernia topics. When users asked questions or submitted input to Hernia Coach, Hernia Coach responded using text on the screen and in some circumstances also voice, which was determined by modality the user employed to when interacting with Hernia Coach. If a user interacted with Hernia Coach using the keyboard, Hernia Coach would only display information on the screen with no audio. If a user interacted with Hernia Coach using voice, Hernia Coach responded with information on the screen and audibly read the text aloud to the user. The Google Assistant Action development team implemented these different types of interaction-based responses and could not be modified.

When users asked health-related questions, Hernia Coach would provide health content to help answer their questions. Some responses had associated details not initially presented to the user to keep the responses succinct, which was encouraged in the Google Assistant design documentation for all Google Assistant Actions [90]. For these types of responses, follow up prompts would be conveyed to the user that indicated there is more to learn about this topic. For example, if a patient asked "how is a hernia fixed?" Hernia Coach responded, "there are two different types of surgery to fix hernias, open surgery and laparoscopic surgery. Your surgeon will select the one that is best to fix your hernia" Then there would be a brief pause and a follow up prompt saying "if you want to learn more, say 'open surgery,' 'laparoscopic surgery,' or 'differences between surgery types.'"

In addition to health content responses, we also included system responses. Examples of system responses include error messages or responses to users asking for help. When a user was ready to exit Hernia Coach, they said phrases like "goodbye" or "stop," which quit Hernia Coach.

4.5.4.3 Stage 2 - Phase 4: Hernia Coach Visual Components

Visual components were incorporated into Hernia Coach to help guide the user's conversation with Hernia Coach, link to the external information sources, and aid user understanding of health information presented by Hernia Coach. The major motivation for including these visual components into Hernia Coach were the "John Brooks" and "Adam Campbell" personas described in phase 2, stage 2 and included in Supplemental

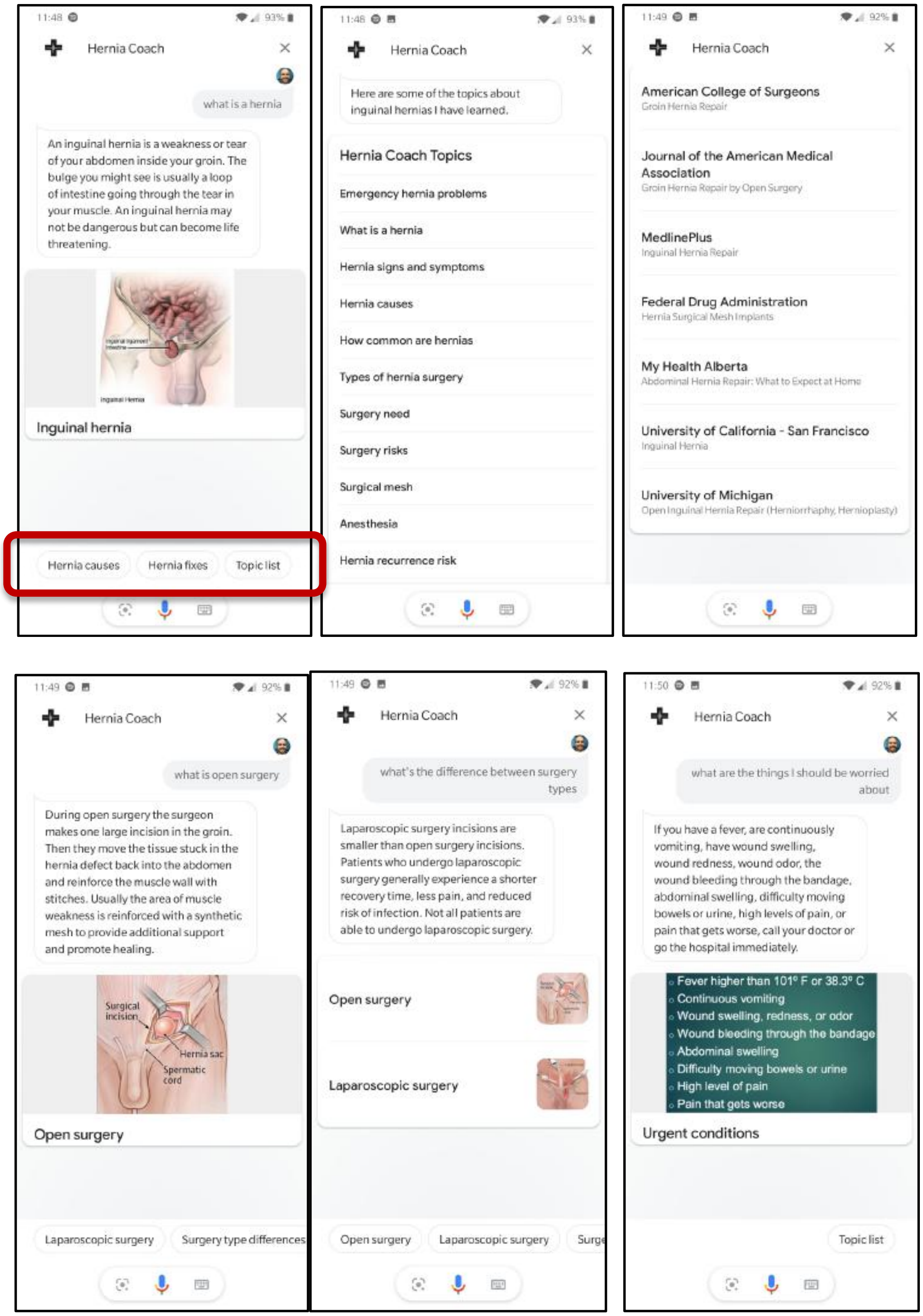
Table 3. Both of these personas represented users who have experienced difficulty understanding health information and using simple visuals can support communicating health information and patient comprehension [138–140]. Also, of the health information sources we reviewed for this project, many included visuals to support navigation through and understanding of health information. We incorporated the visual components to facilitate their health content understanding.

Two visual components of Google Assistant Actions were utilized to help guide the user's conversation. The first component to guide the conversation were instances of the Google Assistant "suggestion chip" [141] object. Suggestion chips are intended to help users discover related topics or pivot the conversation to new topics. The suggestion chips act as tappable queries that appeared at the bottom of each response in the conversation on the smartphone's screen. We incorporated suggestion chips into every intent, which are depicted in Figure 4a. The second component to guide the conversation was a "topic list," which was an instance of a "list" Google Assistant Action object [142]. We intended the topic list to act as a tool to convey Hernia Coach's scope and "knowledge" to the user. We incorporated the suggestion to load the topic list in error messages and as the final suggestion chip in each intent to help improve the discoverability of the topic list. We include a screenshot of the topic list in Figure 4b. At the time of development, the Google Assistant Action design guide and developer documentation provided conflicting information on how to utilize list objects. The design guide stated the string value for each list item would act as a tappable query; whereas the developer documentation stated list objects needed to be handled by a standalone intent, but did not provide further guidance. Google Assistant and Dialogflow help desk requests did receive replies and extensive online research to resolve this issue was unsuccessful. This lack of support and documentation resulted in a situation where the topic list would successfully load, but tapping on an option would result in an error message stating during the expert heuristic evaluation described in stage 3 - phase 1 "I did not understand what you said, please try saying your question again in a different way." Instructions were included in separate documentation informing the design experts this is a known issue and to say or type their selected option as a workaround. A minor guidance improvement was discovered in the Google Assistant developer documentation while iterating on the design between the heuristic evaluation and patient usability testing in stage 3 - phase 2. The error message was changed, stating "the topic list is being fixed, please say or type your selected option."

Links to the websites and materials used to generate Hernia Coach's content were provided with an instance of the Google Assistant Action "browse carousel card" object [143] to help build user Hernia Coach trust. This object is how users can select and interact with external content within the Google Assistant. The reason for including this feature was because in Aim 2 the participants voiced trust concerns and desired to know the sources of the information incorporated into the HDS. Users accessed these links by saying phrases like "how did you learn what you know?" or "what sources do you reference?" We include a screenshot of Hernia Coach's information source object in Figure 4c.

We incorporated visuals into 16 of the 42 (38%) of the patient education intents to aid user understanding of inguinal hernia concepts. Ten intents incorporated visuals related to the content of the intent using instances of Google Assistant Action “basic card” object [144]. For example, we incorporated a medical diagram of open surgery in the open surgery intent, which is depicted in Figure 4d. Two intents utilized the browse carousel card object previously described, which allowed for the presentation of laparoscopic surgery and open surgery medical drawings side by side to depict the differences in surgical techniques along with a text-based description, which is depicted in Figure 4e. Finally, four intents contained text information comprised of long compound sentences, which presented a long block of text on the screen and negatively affected readability. We created slides with summarized bullet points of the text to break down the concepts into smaller, more succinct sentences to aid user health content understanding. The slides were presented using the basic card object previously described. Initially, we attempted to design the intents to read the compound sentence aloud but only present the bullets on the slide without the corresponding text. However, this was not possible and was a limitation of Dialogflow and the Google Assistant. Given this limitation, we included both the text and slides into these four intents. We include an example of these four intents in Figure 4f.

Figures 4a-f (right to left): Screenshots of Hernia Coach's suggestion chips (highlighted in red), topic list, information sources, medical visuals, comparison medical visuals, and slide visuals.



4.6 Stage 3: Hernia Coach Evaluation

Hernia Coach was evaluated in three phases to identify and address Hernia Coach's usability problems using a holistic information technology evaluation approach. First, design experts engaged in a heuristic evaluation of the system. Hernia Coach's design was then improved based on the results of the heuristic evaluation. Second, patients who had previously undergone inguinal hernia surgery engaged in usability testing sessions to evaluate Hernia Coach. We evaluated Hernia Coach using both heuristic evaluations and user testing because employing both methods to evaluate the same software has been shown to identify the majority of usability problems [145]. Additionally, heuristic evaluations and usability testing identify different types of usability problems, which generate greater insight as to how software can be improved [145]. Third, we evaluated Hernia Coach's query responses from the heuristic evaluation and patient usability testing sessions for accuracy using a 2x2 confusion matrix. A confusion matrix is a tool that facilitates analyzing the ability of an information retrieval system to provide relevant information to a query [146]. The University of Washington Human Subject's Division approved all evaluation procedures involving human participants.

4.6.1 Stage 3 - Phase 1: Design Expert Heuristic Evaluation

For the initial round of usability testing, design experts conducted a heuristic evaluation of Hernia Coach. A heuristic evaluation is an expert based evaluation of user interfaces to identify usability problems [147]. We chose to perform a heuristic evaluation with the design experts because it is the most widely used expert-based evaluation method [148,149], efficiently evaluates the usability of a system [145,150], and generates severity ratings of usability issues to prioritize changes during design iterations [151]. We chose to have the design experts apply Nielsen's Heuristics [148] during Hernia Coach's evaluation because these heuristics are commonly applied in expert software evaluation and have been used to evaluate health information technology [152–157].

Design experts were eligible to participate in our evaluation if they were 18 years or older, fluent in English, and had previous experience in design (e.g., a design degree or professional experience as a designer). The design experts were recruited by directly emailing individuals within our professional network and snowball sampling [158]. We recruited design experts until three to five enrolled in the study and completed their evaluation, which is considered sufficient for heuristic evaluations [147,151,159,160].

The design expert evaluations occurred in the Paficic Northwest region of the United States in March and April 2019. Expert evaluation participants completed evaluations by themselves at a location of their own choosing. We provided consent forms to design expert participants to review before beginning the study. We did not document design expert consent because we conducted the heuristic evaluations remotely. The University of Washington Human Subjects Division granted a waiver of consent documentation for remote participants. We compensated the design experts with a $20 gift card.

4.6.1.1 Stage 3 - Phase 1: Nielsen's Heuristics Adapted for the Heuristic Evaluation

Despite the wide adoption of Nielsen’s heuristics, these heuristics may not be generalizable to new computing interfaces such as DSs. Nielsen’s heuristics were initially published in 1990 before the advent of DSs and were generated from the evaluation of simple display terminals depicting information without graphics or color [161]. Nielsen also stated the heuristics are not specific usability guidelines but “broad rules of thumb” [162], which has prompted researchers to adapt the heuristics to more specifically evaluate smartphone software [163–165] or consider different user groups such as older adults [163,166]. Human-centered design professionals have also identified Nielsen’s heuristics may need to be adapted to more effectively evaluate DSs [167,168]. Additionally, the Nielsen Norman Group has published considerations when assessing DS user experiences [169]. However, the Nielsen Norman Group has not published guidance on how to adapt Nielsen’s original heuristics to DSs. Furthermore, we could not identify peer-reviewed research where Nielsen’s heuristics were adapted to evaluate HDSs or DSs to guide our heuristic evaluation.

Given the heuristics are not specific usability guidelines and were developed for other types of computing interfaces, we adapted Nielsen’s heuristic definitions [149] for this study. Specifically, we adapted Nielsen’s heuristic definitions to recognize that users interact with Hernia Coach using voice and keyboard, to evaluate the health content in response to user queries, and to assess the conversation guidance components incorporated into Hernia Coach. We did not adapt the high level heuristics for this study. We include Nielsen’s original heuristics and our adaptations in Table 4.

Table 4: Nielsen's heuristics and definitions [149], Nielsen's Heuristic definitions with adaptations for Hernia Coach's evaluation in bold, and the rationale for adapting the Heuristic definitions.

Nielsen's Heuristic	Nielsen's Heuristic Definition	Heuristics Definition Adaptation	Rationale for Adaptation
Visibility of system status	The system should allow the user to request information about what is going on, through appropriate feedback within reasonable time	The system should allow the user to **visually or audibly** request information or identify what is going on, through appropriate feedback within reasonable time	Hernia Coach is designed for use on smartphones, which have the potential for voice or graphical user interface interaction methods.
Match between system and the real world	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The system should **understand and** speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms **or with confusing medical jargon**. **Use common terminology making information natural, logical, accessible.**	Hernia Coach needs to understand the users questions, which could be asked at a different levels of health literacy depending on the user. Hernia Coach should be able to understand and convey information for varying levels of user health literacy.
User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.	Users often choose system functions by mistake and will need **an option to effortlessly leave the unwanted state without having to go through an extended dialogue.**	Hernia Coach's purpose does not involve users to perform actions but rather ask questions. Users should be able to leave an unwanted state with little effort.
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	Users should not have to wonder whether different words, situations, or actions mean the same thing. **The dialog system should stick to a single style of consistent language**	The main purpose of Hernia Coach is to answer user questions about inguinal hernias in a conversation. The language should be consistent throughout the conversation.
Error prevention	Reduce the likelihood of error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Reduce the likelihood of error-prone conditions with **conversation design and GUI elements**	The main purpose of Hernia Coach is to answer user questions about inguinal hernias. The conversation and interaction design incorporated into Hernia Coach should provide

			markers to guide the conversation and convey what Hernia Coach knows about inguinal hernias. Hernia Coach's scope does not require users to "commit actions."
Recognition rather than recall	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate. Avoid overwhelming the user with a wall of text	Minimize the user's memory load by making objects, actions, and options visible **and/or audible**. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate. Avoid overwhelming the user with a wall of text.	Users will have the option to interact with Hernia Coach using voice or keyboard interaction methods.
Flexibility and efficiency of use	Accelerators-unseen by the novice user-may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	**Allow users the ability to interact with the system using their preferred modality for each type of hardware such as voice, text, or graphical elements**	Hernia Coach's scope does not involve users performing frequent actions, which negates the need for this heuristic definition. This definition was changed to asses user to interacting with Hernia Coach using their preferred interaction method, voice or keyboard, to identify if there are limitations or usability issues for either option.
Aesthetic and minimalist design	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	No adaptation	This heuristic may be generalizable to HDSs.
Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	No adaptation	This heuristic may be generalizable to HDSs.

Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.	No adaptation	This heuristic may be generalizable to HDSs.

4.6.1.2 Stage 3 - Phase 1: Design Expert Heuristic Evaluation Methods & Procedures

After we recruited the design experts, we emailed them a copy of the heuristic evaluation consent form. The design expert participants were instructed to ask questions or for clarification before proceeding with the heuristic evaluation. In addition to the consent form, the design expert participants were sent instructions on how to access Hernia Coach on their personal smartphone, a spreadsheet with the heuristics and adapted definitions for the evaluation, and copies of the personas and scenarios described previously. The purpose of providing the personas and scenarios was to give the design experts context about the background and scope of Hernia Coach, but not to require it to be used during the heuristic evaluation. Sharing personas and scenarios is generally not done for heuristic evaluations, but is the norm for cognitive walkthroughs. Cognitive walkthroughs are a different expert-based usability inspection method where experts evaluate an interface by analyzing the cognitive processes required for accomplishing tasks [151] and are similar to usability testing methods with representative end users. However, we opted to have the design experts perform a heuristic evaluation because of the benefits previously described. Additionally, the patient participants would be engaging with Hernia Coach using methods similar to a cognitive walkthrough in the subsequent evaluation phase.

We conducted the heuristic evaluations remotely and unmoderated to accommodate their limited availability for in-person evaluations, which is a recognized evaluation method [170]. The design expert participants were not asked to video or audio record their sessions because their interactions with Hernia Coach were automatically logged in Dialogflow (i.e., NLU). While completing the evaluation, participants were prompted to interact with Hernia Coach using their voice making using the think-aloud method impractical. Think aloud is a common usability evaluation method where participants verbalize their thoughts as they move through the user interface [163]; we identified it would not be possible for a participant to simultaneously voice their thoughts aloud while submitting queries to Hernia Coach using their voice because Hernia Coach would interpret their thoughts as a component of the query. Therefore instead of video or audio recording, the design experts manually recorded their usability notes in the heuristic spreadsheet we provided. The spreadsheet included Nieslen's heuristics and the heuristic definitions adapted for this study. The design experts recorded notes, either positive or negative, in the spreadsheet for each heuristic.

After the design experts completed their heuristic evaluations and emailed us their completed spreadsheet, we aggregated their findings by identifying commonalities across each of the heuristics and counting the number of design experts who identified each commonality. We shared the aggregated findings back to the design experts in an online follow-up survey using Microsoft Forms [171]. In the survey, the design experts were asked to rate the severity of the aggregated findings using Nielsen's heuristic severity ratings [172]. The ratings consist of a five-point scale from zero to four. The definitions of the numerical severity ratings are: 0) I don't agree that this is a usability problem at all, 1) Cosmetic problem only: need not be fixed unless extra time is available on project, 2) Minor usability problem: fixing this should be given low priority, 3) Major usability problem: important to fix, so should be given high priority, 4) Usability catastrophe: imperative to fix this before product can be released. Summary severity rating scores for each heuristic were calculated using a mean and standard deviation. We defined a heuristic violation as an aggregated finding with a mean severity score of 1.0 or greater and a Hernia Coach strength as an aggregated finding with a mean severity score of less than 1.0.

In addition to conducting the heuristic evaluation, the design experts also completed an online exit survey, which consisted of three parts. The first part is the Post-Study System Usability Questionnaire (PSSUQ) [11]. We chose to include this survey for the heuristic evaluation because we also include the survey during the patient testing sessions in Stage 3- Phase 2. We wanted to identify if patient testing participant satisfaction increased or decreased compared to the design experts. The PSSUQ is a 19-item instrument to assess a user's satisfaction with a computer system. Each item is rated by the participants on a seven-point scale, from 1 (strongly agree) to 7 (strongly disagree). Scores were summed across items to generate sub-scores and an overall score. The lower the score, the better the user's satisfaction with Hernia Coach. There are three PSSUQ sub-scores, which are system usability (eight items; score range 8-56), information quality (seven items; score range 7-49), and interface quality (four items; score range 4-28). The sum of the three sub-scores that reflect all 19 items results in an overall score (range 19-133). The PSSUQ has been used to assess satisfaction regarding wearable sensors for Parkinson's disease [173], a physical activity mobile health application [174], and the influence of design aesthetics in usability testing [175].

The second component of the design expert exit survey was the eHealth Literacy Scale (eHEALS) survey. eHEALS is an eight-item survey "developed to measure consumers' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying electronic health information to health problems" [3]. eHEALS is scored on a five-point scale and each item ranges from 1 (strongly disagree) to 5 (strongly agree), with a total score range of 8-40; higher eHEALS scores indicate greater eHealth literacy. Researchers have previously used the eHEALS in health informatics design research [176–178].

Finally, the design experts completed a demographics survey to assess age, gender, education, home geographic region, healthcare design experience, previous surgery experience, the type of smartphone used to conduct the heuristic evaluation, and personal DS use. Home geographic regions were classified according to U.S. census definitions, which are urbanized areas (population > 50,000 residents), urban clusters (population between 2,500 and 50,000 residents), or rural (population < 2,500) [179].

Participants were able to indicate using more than one personal DS and more than one DS interaction hardware type.

Descriptive statistics for the heuristic evaluation severity ratings, PSSUQ, eHEALS, and demographics surveys were generated using Tableau.

4.6.1.3 Stage 3 - Phase 1: Design Expert Heuristic Evaluation Results

Five design experts completed a heuristic evaluations of Hernia Coach and four of the five completed the severity rating follow up survey. One design expert declined to complete the follow-up survey. We include the design expert participant characteristics in Table 5.

Table 5: Design expert characteristics.

Design Expert Characteristics	
Age in years, median (IQR)	32 (30-37)
Declined to respond, *n*	1
Gender, *n*	
Female	3
Male	2
Highest level of completed education, *n*	
Bachelor's degree	1
Master's degree	3
Doctorate degree	1
Home geographic region, *n*	
Urban	4
Declined to respond	1

Healthcare design experience, *n*	
Yes	4
Declined to respond	1
Previous surgery experience as a patient, *n*	
Yes	4
No	1
Personal dialog system use, *n*	
Alexa	2
Google assistant	2
Siri	2
Cortana	1
Personal dialog system interaction hardware, *n*	
Smartphone	4
Smart speaker	4
Most preferred dialog system interaction modality, *n*	
Voice	3
Keyboard	1
Declined to respond	1
Smartphone operating system used for evaluation, *n*	
iOS	3
Android	2
eHEALS score, median (IQR)	37 (34-40)

Design expert participants could indicate using more than one dialog system and dialog system hardware type. The eHEALS score ranges from 8-40; a higher score indicates greater health literacy. IQR = interquartile range. *n* = number.

The design experts noted a number of Hernia Coach strengths: five said Hernia Coach had clear and consistent content, five identified Hernia Coach was successful at supporting both voice and text inputs, four stated Hernia Coach had good natural language understanding, three said it was easy to recover from mistakes, two said it was easy to ask questions again using alternate wording, two said Hernia Coach's opening disclaimer and conversation guidance were helpful, two stated the suggestion chips were useful, two said the topic list was beneficial, and two noted good guidance within Hernia Coach. These strengths were preserved for the patient user testing sessions while addressing the usability issues identified by the design experts .

Design experts identified nine heuristic violations across eight heuristics. The design experts did not note violations for "consistency and standards" and "flexibility and efficiency of use." We include the aggregated heuristic evaluation findings, frequency counts, mean severity ratings, Hernia Coach design changes resulting from the heuristic evaluation findings, limitations, and notes in Table 6.

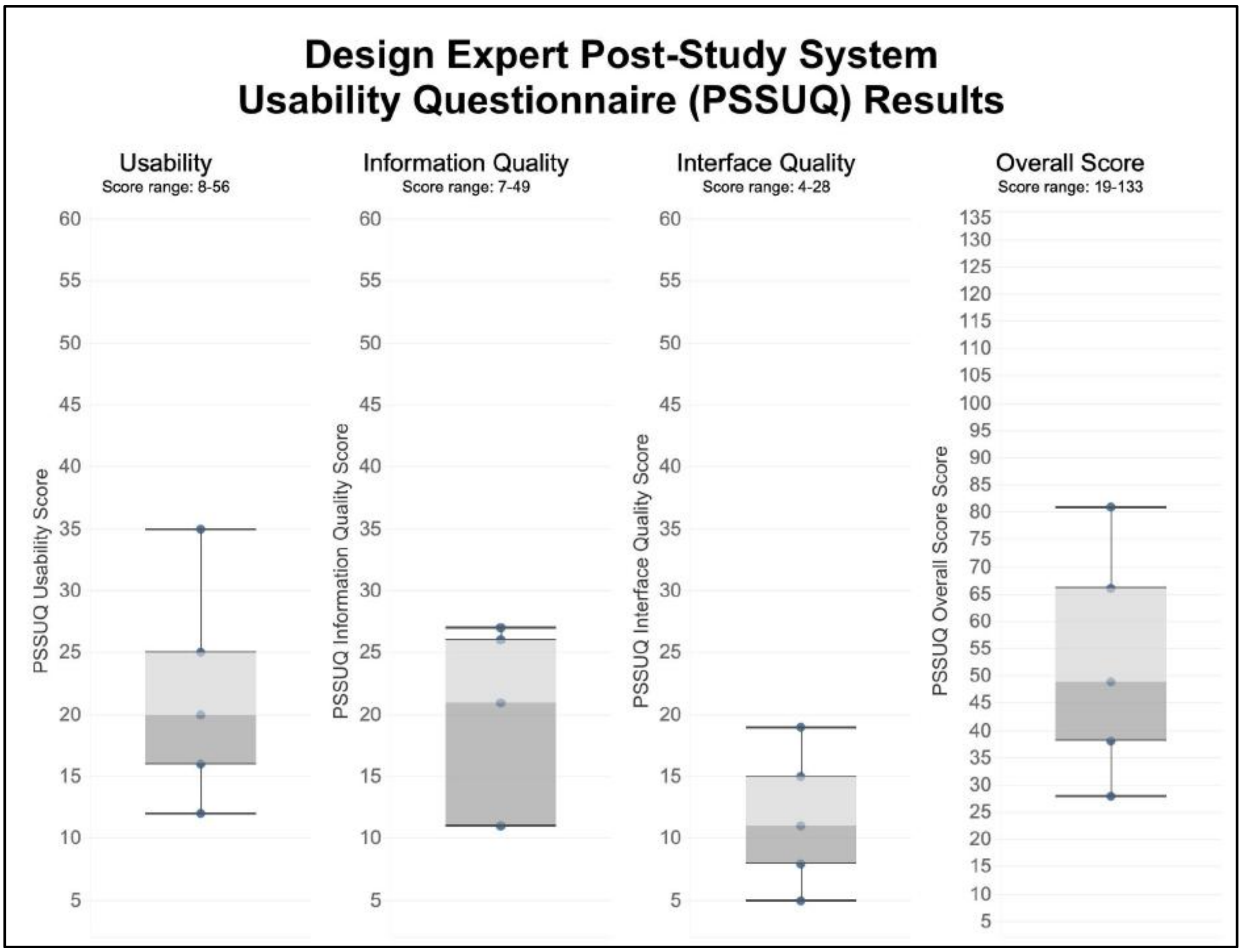
Table 6: Design expert heuristic evaluation violations, frequency counts, mean severity ratings, iterative design changes, limitations, and notes.

Heuristic	Heuristic Violation	Frequency Count	Severity Rating, mean (SD)	Design Changes, Limitations, and Notes
Visibility of system status	Issues opening Hernia Coach or knowing if the Hernia Coach is open	4	2.8 (1.3)	Changed Hernia Coach's invocation from "Inguinal hernia info bot" to "Hernia Coach" to be easier to say and remember. It is not possible to change visibility within the action. The only indication the action is open is the DS logo, which changes from the Google Assistant logo to the Hernia Coach icon.
Match between system and the real world	Incorrect system response to query	3	2.0 (1.4)	We annotated the training data to improve system responses.
	Provides broad content or suggests poor solutions due to lack of context	3	1.5 (0.6)	This is a limitation of the Google Assistant and not possible to change.
User control and freedom	Issues leaving Hernia Coach when undesired or unexpectedly	3	1.5 (1.3)	This is not possible to change and a limitation of the Google Assistant. Saying "stop" results in exiting Hernia Coach. We attempted to fix this issue by creating a new intent incorporating the training phrase "stop" but were unsuccessful.
Error prevention	Topic list issues	4	1.8 (1.3)	The Dialogflow/Google Assistant developer documentation and Google Assistant design documentation provided conflicting information about list functionality. Help desk requests for assistance did not receive replies.
Recognition rather than recall	Previous content disappears when asking a new question	2	1.0 (0.8)	This is a limitation of the Google Assistant and not possible to change.
Aesthetic and minimalist design	Unhelpful supplementary visuals	3	1.0 (0.8)	Three design experts did not find the supplementary visuals helpful, but two did find the supplementary visuals helpful. The supplementary visuals were left unchanged to explore patient perceptions and attitudes towards the supplementary visuals during the stage 3 - phase 2 user testing.
Help users recognize, diagnose, and recover from errors	Unhelpful error messages	2	2.3 (1.0)	Error messages were developed according to the design documentation guidelines. This was left unchanged for patient testing to explore further during patient user testing.
Help and documentation	Suggestion chips not relevant to question or topic	2	1.0 (0.8)	Suggestion chips were reviewed and some were updated with more relevant options.

The heuristic violation severity score ranges from 0-4; a higher score indicates a more severe usability issue. Four of the five recruited design experts completed the follow up survey to assign heuristic evaluation severity ratings.

We assessed the design experts satisfaction using Hernia Coach with the PSSUQ. A lower score indicates greater satisfaction. The overall median satisfaction score was 49 out of 133 (IQR 38-66), the median usability score was 20 out of 56 (IQR 16-25), the median information quality score was 26 out of 49 (IQR 11-26), and the median interface quality score was 11 out of 28 (IQR 8-15). We include box and whisker plots depicting the design expert PSSUQ scores in Figure 5.

Figure 5: Design expert box and whisker plot Post-Study System Usability Questionnaire (PSSUQ) scores for information quality, interface quality, and overall satisfaction results.



A lower score indicates greater satisfaction.

4.6.2 Stage 3 - Phase 2: Patient Usability Testing

After the design expert completed the heuristic evaluations and Hernia Coach's design was improved, we recruited patients who had previously undergone inguinal hernia surgery to evaluate Hernia Coach during usability testing sessions. We recruited this type of patient because they already had the experience of inguinal hernia surgery and were able to assess Hernia Coach's potential to aid information-seeking. We also chose to conduct user testing because this method has been shown to identify different types of usability problems not identified during heuristic evaluations [145], which gave us additional findings to improve Hernia Coach.

Patient participants were eligible to participate in our evaluation if they were: 18 years or older, fluent in English because Hernia Coach was written in the English language and we skeptical of the Google Assistant's ability to automatically translate Hernia Coach into other languages, did not have a physical impairment preventing them from interacting with Hernia Coach, and had been discharged from the hospital for at least two weeks after having surgery to allow them sufficient time to recover from recent surgery. Patient participants were re-recruited from Aim 2**;** new participants were recruited through online posts to public platforms (e.g., Craigslist, Reddit, and Facebook), posting flyers in public places (e.g., libraries, community centers, YMCAs, hospital waiting rooms), and snowball sampling. We recruited patient participants until five to eight enrolled in the study and no new concepts were added to the qualitative analysis (i.e., we reached saturation); reaching these criteria are considered sufficient for user testing [180–182]. We analyzed the first four patient usability testing transcripts which recruitment was ongoing, then the subsequent sessions were analyzed to allow for the identification of thematic saturation.

The patient usability testing sessions occurred in the Paficic Northwest region of the United States in May and June 2019. Patient participants provided documented consent before conducting the usability testing sessions and were compensated with a $20 gift card.

4.6.2.1 Stage 3 - Phase 2: Patient Usability Testing Methods & Procedures

Patient participant usability testing sessions were in-person, moderated by a researcher, and conducted in reserved private rooms at the University of Washington or in city libraries. We provided participants with a device Hernia Coach installed to conduct the usability testing sessions. We provided the device because giving participants access to Hernia Coach on their personal device would have required substantial effort and detracted from the usability testing sessions. The specific device provided was a OnePlus 6t smartphone using the Android operating system version 9 (i.e., Pie) featuring a 6.4-inch display [183]. We used the Google Assistant version 0.1.187945513 for Android during the patient usability testing sessions. All sessions were audio recorded. Transcriptions were generated from the audio recording using a professional transcription service.

At the beginning of the session, we gauged participant familiarity with DSs to identify how much experience they have had with DSs before beginning the evaluation. Then the participants were given a brief overview of Hernia Coach to give context for Hernia Coach's purpose. The patient participants were then asked to query Hernia Coach based on the context and information presented for each of the five scenarios previously described in stage 2 - phase 3. The patient participants were encouraged to ask as many questions as possible related to each scenario until they felt they had asked all the questions they could think of, or until seven minutes had elapsed during the scenario, whichever came first. We used these two parameters to allow the participants an opportunity to ask as many questions relating to the scenarios, but also provide sufficient time to complete each scenario. We opted to structure the testing sessions this way to test the breadth of Hernia Coach and its ability to answer the common questions and health information-seeking contexts identified in Aim 2. At the end of each scenario, participants were asked to complete the After Scenario Questionnaire (ASQ) [1] in a paper packet, which is a three-item measure to assess participant satisfaction for each scenario. The three items assess participant satisfaction with the ease of task completion, satisfaction with the time to complete a task, and the satisfaction of support information. Each item is scored on a seven-point scale from 1-strongly agree to 7-strongly disagree and summed together for each task. The ASQ has been promoted in health informatics research for patient-centered design and interactive health technologies [184,185], applied in the assessment of home healthcare devices [186], and used in the evaluation of mobile health technology for lung transplantations [187].

When the participants had completed the five scenarios, they were asked a series of questions in a semi-structured interview to understand their perceptions of Hernia Coach. We completed a semi-structured interview because participants were not prompted to think aloud while engaging with Hernia Coach due to reasons previously described in section 4.5.1.2. The researcher conducting the evaluation sessions took notes to follow up with questions during the interview to mitigate this methodological limitation. Additionally, participant queries were automatically logged in Dialogflow (i.e., NLU) for review after the sessions. In addition to the follow up questions, patient participants were asked what they liked about Hernia Coach, what they disliked, Hernia Coach’s quality of information, satisfaction using Hernia Coach, suggestions for improving Hernia Coach, if they would recommend Hernia Coach to others, and if they thought Hernia Coach would replace communication with the medical care team. Finally, the patient participants completed PSSUQ survey to assess satisfaction using Hernia Coach and the eHEALS survey to assess health literacy. We described these surveys in greater detail in the design expert heuristic evaluation procedures section of this chapter (section 4.5.1.2). In addition to the PSSUQ and eHEALS surveys, the patient completed a demographics survey to assess their age, gender, education, home geographic region, hospital geographic region, hernia surgery details, current or previous healthcare industry employment, and DS use. Participants were able to indicate using more than one personal DS and more than one DS interaction hardware type. All surveys were completed by patients using paper, which were transcribed digitally using Microsoft Forms.

Descriptive statistics of the ASQ, PSSUQ, eHEALS, and demographics responses were generated using Tableau.

4.6.2.2 Stage 3 - Phase 2: Patient Usability Testing Qualitative Analysis Theoretical Model and Procedures

We identified a post-hoc theoretical model, the Health Information Technology Acceptance Model (HITAM) [9], to inform usability testing qualitative analysis. This model was selected because it was developed to extended Technology Acceptance Model [8] in health care by describing health consumers' behavioral intention of using health information technology (HIT). The Technology Acceptance Model was developed to predict individual adoption and use of new technology [8]. The HITAM has been used in research to assess video games promoting health activities with seniors [188], health consumer behavior using self-tracking [189], mobile health apps to facilitate self-care [190], understand fitness tracker use [191], and electronic health record patient portal adoption [192]. A depiction of the HITAM and the model definitions are included in Figure 6.

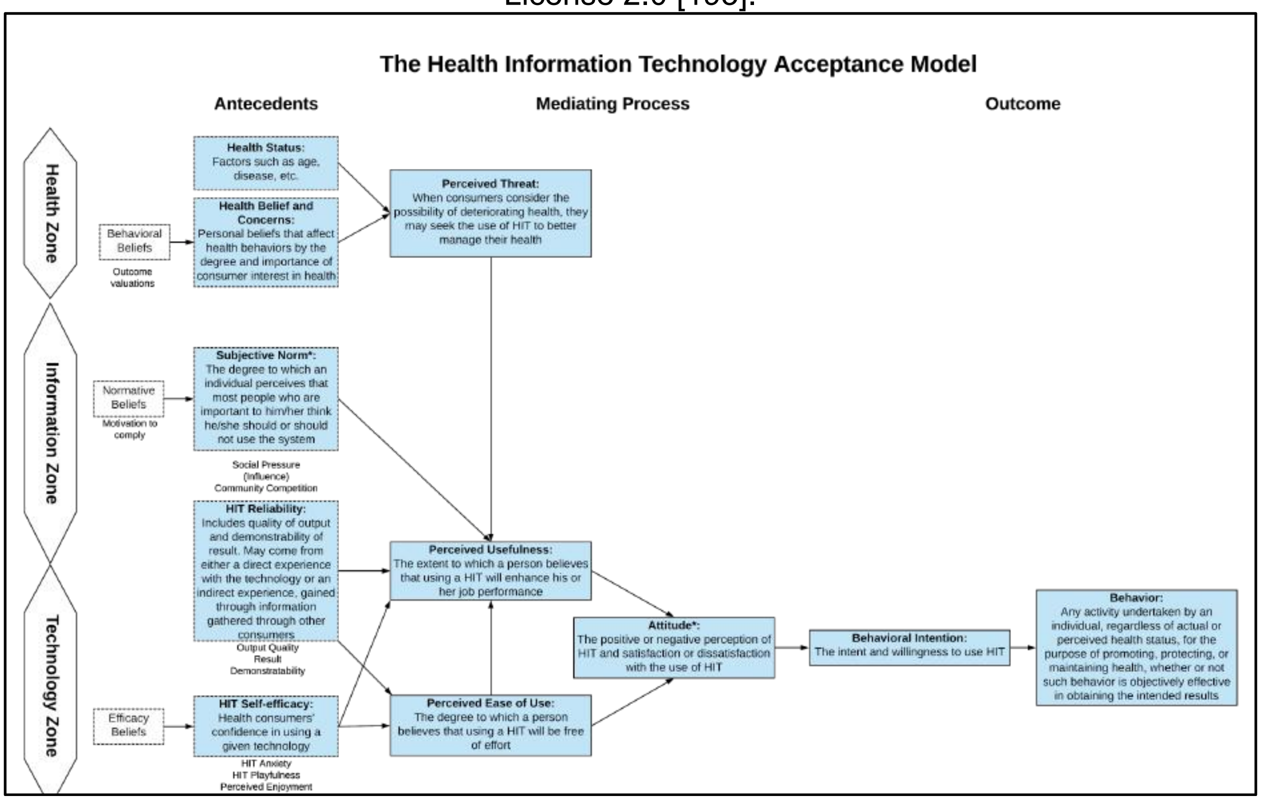
The HITAM extends the Technology Acceptance Model into three domains. The first domain is the health zone, which is composed of two antecedents and one mediating process. The first antecedent is health status, which refers to the demographics and the health condition of the patient. The second antecedent is the patient's health beliefs and concerns, which are the personal beliefs of the patient that affect health behaviors by the degree and importance of their interest in health. The mediating process of perceived threat is affected by the two antecedents, which is when patients have the possibility of deteriorating health, they may use HIT to improve their health management.

The second domain is the information zone, which consists of one antecedent and one mediating process. The antecedent of subjective norm concerns how others important to the patient perceive how the patient should not use Hernia Coach. This antecedent affects the mediating process of perceived usefulness, which is the patient's perception of using HIT will aid with their health information-seeking activities.

The third domain is the technology zone, which concerns HIT itself. This zone is comprised of two antecedents, three mediating processes, and two outcomes. The first antecedent is HIT reliability, which is the quality of the technology's output and the result of using the technology. This is ascertained from direct experience with the technology or indirect experience through the experience of others using the technology. HIT reliability overlaps into the information zone. The second antecedent is HIT self-efficacy, which is the patient's confidence in using technology. Both of these antecedents affect the three mediating processes in the technology zone. The first is perceived usefulness, which we previously described in the information zone. The second is the perceived ease of use, which is the patient's perception that using HIT will be free of effort and affects perceived usefulness. The third mediating process is attitude, which is the positive or negative perception of HIT and the satisfaction or dissatisfaction of associated with using HIT. Attitude is affected by perceived usefulness and perceived ease of use. The two outcomes within the technology zone are behavioral intention and behavior, which are affected by attitude. Behavioral intention is the intent and willingness of the patient to use HIT. Behavior is any activity undertaken by the patient for the purpose of promoting, protecting, or maintaining their health.

We adapted two definitions of the HITAM for this study. First, the definition of "subjective norm" was adapted to identify how the perceptions of people important to the patient should or should not use Hernia Coach. Second, the definition of "attitude" was adapted to include the assessment of negative perceptions of HIT and dissatisfaction of HIT in addition to positive perceptions and satisfaction of HIT.

Figure 6: The Health Information Technology Acceptance Model [9]. This figure was initially published and may be reproduced under the terms of the Creative Commons License 2.0 [193].



An asterisk (*) indicates the model component was adapted for the qualitative transcript analysis.

We conducted directed content analysis of the patient usability testing session transcripts [194]. We chose directed content analysis because it is a deductive qualitative analysis method used to validate or extend theoretical frameworks. The HITAM informed the qualitative analysis codebook; a code was created for each HITAM concept. We also included a “design suggestion” code to capture insights provided by participants regarding ways to improve Hernia Coach.

The first four patient usability testing sessions were analyzed while recruitment was ongoing, then the subsequent sessions were analyzed to allow for the identification of thematic saturation indicating completion of recruitment. Initially, two researchers coded two usability testing session transcript excerpts. After each of the two coding rounds, the researchers discussed how the transcripts were coded including instances of concordant and discordant coding of the same text and new codes. After the two rounds of coding, the two researchers agreed the codebook adequately covered all of the themes. Topf [179] was used to calculate inter-rater reliability (IRR). Initial IRR was 87.5% indicating adequate agreement; the two researchers discussed coding discrepancies and came to a consensus for a final IRR of 100%. The codebook was finalized and used by one researcher to independently code the remaining transcripts.

4.6.2.3 Stage 3 - Phase 2: Patient Usability Testing Quantitative Results

Six patients participated in the usability testing sessions. Two of the participants were re-recruited from Aim 2, three were recruited using online posts, one with a recruitment flyer, and one using snowball sampling. We include the patient participant characteristics in Table 7.

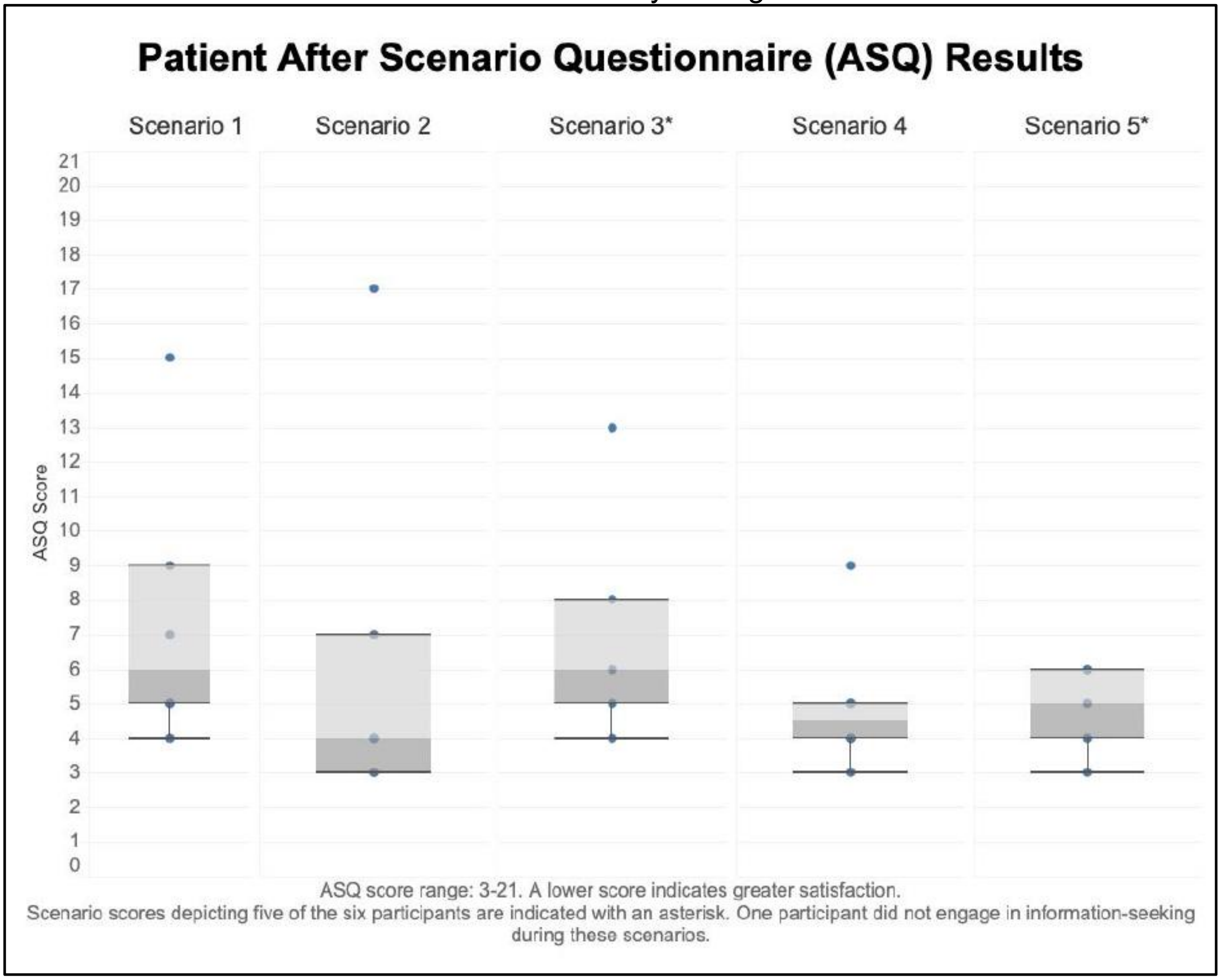
Table 7: Patient participant scenario-based usability testing characteristics

Patient Characteristics	
Age in years, median (IQR)	60 (54-76)
Gender, *n*	
Male	6
Highest level of completed education, *n*	
Graduated high school	1
Associate's degree	1
Bachelor's degree	1
Master's degree	3
Home geographic region, *n*	
Urban	5
Urban cluster	1
Designated post-surgery caregiver	
Spouse or significant other	5
Other	1
Number of inguinal hernia surgeries undergone, *n*	
One	6
Healthcare industry experience, *n*	
No	4
Previously employed in healthcare	2
Healthcare industry roles, *n*	
Social worker	1
Computer software development	1
Personal dialog system use, *n*	
Google Assistant	4
Alexa	2
Siri	2
Personal dialog system interaction hardware, *n*	
Smartphone	4
Smart speaker	2
Computer	2
Most preferred dialog system interaction modality, *n*	
Voice	5
No strong preference	1
eHEALS score, median (IQR)	32 (30-33)

Patient participants could indicate using more than one dialog system and dialog system hardware type. The eHEALS score ranges from 8-40, a higher score indicates greater health literacy. IQR = interquartile range. *n* = number.

We measured patient participant satisfaction after every usability testing scenario with the ASQ. A lower score indicates greater satisfaction. Five of the six participants completed all of the ASQs. One participant could not think of any questions during scenarios 3 and 5 and chose not to engage information-seeking. The median score for scenario 1 was 6 out of 21 (IQR 5-9), the median score for scenario 2 was 4 out of 21 (IQR 3-7), the median score for scenario 3 was 6 out of 21 (IQR 5-8), the median score for scenario 4 was 4.5 out of 21 (IQR 4-5), and the median score for scenario 5 was 5 out of 21 (IQR 4-6). We include box and whisker plots of the patient usability ASQ scores in Figure 7.

Figure 7: Patient usability testing After Scenario Questionnaire (ASQ) satisfaction results for the five usability testing scenarios.

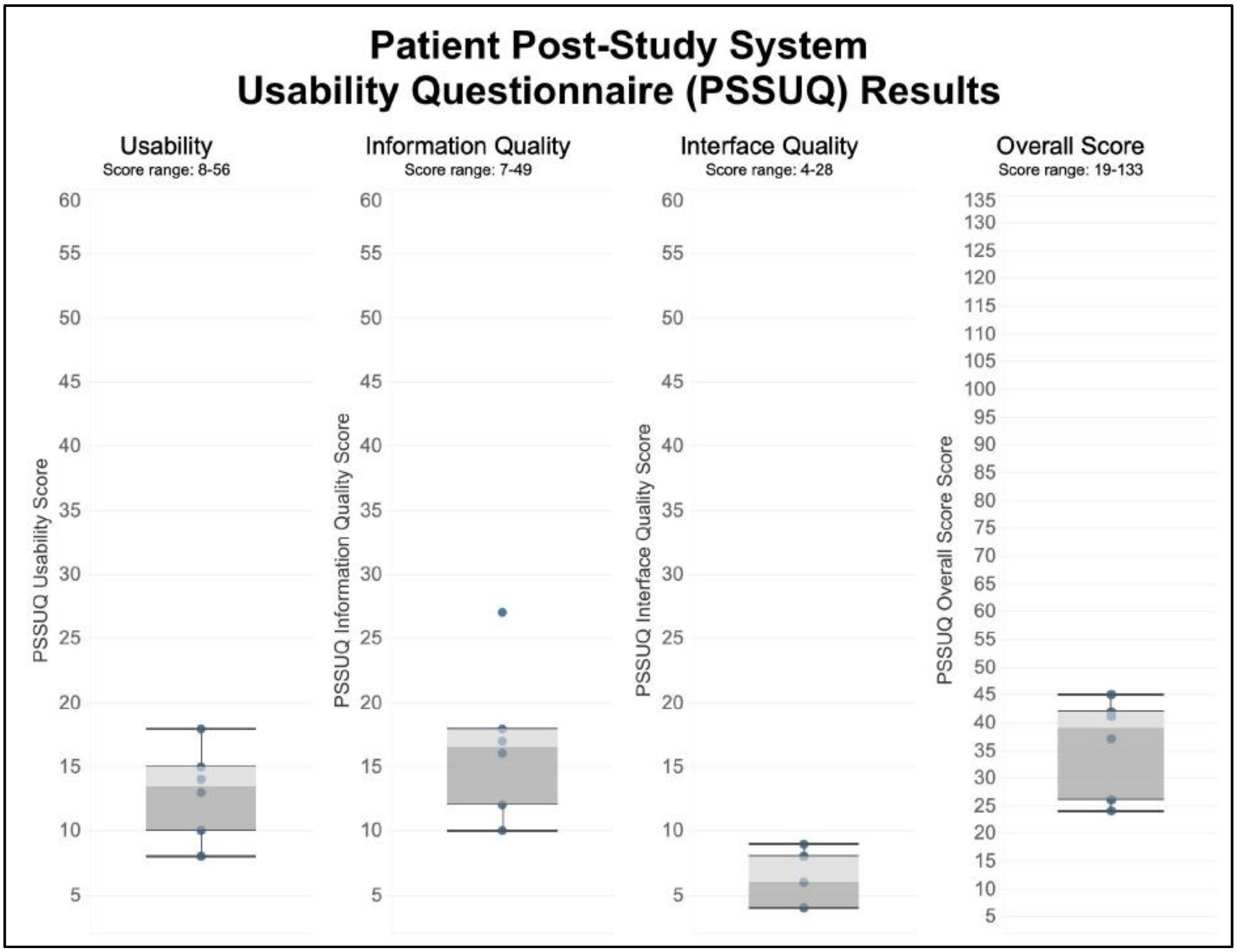


ASQ score range: 3-21. A lower score indicates greater satisfaction. ASQ scores depicting five of the six patient participants are indicated with an asterisk. One of the five patient participants did not engage in health information-seeking during these scenarios.

We assessed patient participant satisfaction using Hernia Coach with the PSSUQ. A lower score indicates greater satisfaction. The overall median satisfaction was 39 out of 133 (IQR 26-42), the median usability score was 13.5 out of 56 (IQR 10-15) , the median information quality score was 16.5 out of 49 (IQR 12-18), and the median interface quality score was 6 out of 28 (IQR 4-8). All of the patient PSSUQ scores improved compared to the design expert PSSUQ scores, indicating the design iterations between the expert heuristic evaluation and the patient usability testing improved Hernia

Coach. We include box and whisker plots depicting the patient usability testing PSSUQ scores in Figure 8. and the median overall score was.

Figure 8: Patient box and whisker plot Post-Study System Usability Questionnaire (PSSUQ) scores for information quality, interface quality, and overall satisfaction scores.



A lower score indicates greater satisfaction.

4.6.2.4 Stage 3 - Phase 2: Patient Usability Testing Qualitative Results

We report results where we applied HITAM codes during the patient usability testing qualitative analysis. We did not apply codes for health status, health belief and concerns, perceived threat, subjective norm, and behavior during the analysis. We include representative participant quotes in Table 8.

Antecedents: HIT Reliability

When querying Hernia Coach, five participants identified interaction issues and took action to workaround these issues. Four of these participants identified that at times they did not get answers to the questions they asked. Two stated they would shorten their responses to be more precise or focus on topic keywords to increase the chances of successful responses. One initially asked about “what activities I can do after surgery?”, then followed up with “what can’t I do after surgery?”, which prompted the same response from Hernia Coach. Another participant was unsure if they were asking the right questions to get the response desired from Hernia Coach. One of the participants likened interacting with Hernia Coach to be similar to speaking with someone who had recently learned English as a non-native language.

Four participants explicitly noted Hernia Coach had helpful education content. Three of the participants stated the graphics used in conjunction with the text and verbal responses aided understanding of the health information. Two participants reported getting all of the information they wanted from Hernia Coach. One said Hernia Coach was generally informative at answering their questions and improved their general knowledge about the surgery. Another stated the education content was straightforward and perceived the level of detail as appropriate for most patients. Two participants desired Hernia Coach to have even more content to answer more of their questions. In these instances the participants had new questions based on Hernia Coach’s query response, which Hernia Coach was unable to provide. These participants noted they did not ask follow up questions because they were aware they had exhausted Hernia

Coach's knowledge about a topic and the participants identified Hernia Coach would not be able to provide additional information if pressed for more. One participant reported the content about general recovery activities did not match their situation recovering at home after surgery. Specifically, he noted he was not being able to walk for a few days due to high levels of pain, which contradicted the suggestion of trying to walk for a few minutes every hour if pain levels permit.

Two participants specifically mentioned Hernia Coach's error messages. One found the error messages helpful, while the other voiced the prompts could use more context or guidance.

Two participants also identified Hernia Coach is unable to process nonverbal communication such as body language, which they felt is an important component of health conversations.

Antecedents: HIT Self-Efficacy

Three participants voiced confidence concerns with Hernia Coach. One participant would ask similar questions about a topic and would receive the same response from Hernia Coach, which affected their confidence of Hernia Coach to answer their questions effectively. Another participant who presented themselves as not a frequent user of technology noted it was somewhat challenging to become acquainted with Hernia Coach. Finally, one participant wondered if they were asking the right questions of Hernia Coach to get their desired query response but noted they had confidence Hernia Coach would provide the majority of information he wanted. This participant was asking very specific detailed questions and receiving responses from Hernia Coach describing all the information about that particular topic or subtopic. They were expecting a focused answer to their specific question.

Mediating Process: Perceived Usefulness

All six of the participants perceived Hernia Coach could serve as a potentially useful component of a patient's hernia surgery journey to aid health information-seeking activities. However, all of the participants stated they did not foresee Hernia Coach replacing communication with their healthcare team entirely. All six participants identified there would still be instances where Hernia Coach would not be able to handle all of their questions or concerns and it would still be necessary to call their care team. Five of the participants identified Hernia Coach as being able to field most of their questions, which would help them get answers quickly. They also identified Hernia Coach would reduce the burden of clinicians needing to answer common questions and allow clinicians to focus on other clinical duties. Three of the participants stated Hernia Coach has the potential to replace the paper education materials they received during their surgery journey. One of these three said he would not even use the paper packets and only use Hernia Coach as their preferred information resource at home.

Mediating Process: Perceived Ease of Use

Three participants stated Hernia Coach was easy to use. One said Hernia Coach was able to anticipate questions they thought of beforehand. Another noted that as long as anyone has a smartphone, they have access to using Hernia Coach. One stated they were already familiar with the Google Assistant, which made it easy for them to interact with Hernia Coach.

Three participants reported input modality issues affecting their use of Hernia Coach. Two had difficulties using the keyboard and said they preferred using voice-based interaction methods with Hernia Coach. Another said the keyboard was their preferred interaction method but found the unfamiliar keyboard on the testing device slowed them down while typing.

Mediating Process: Attitude

All six participants voiced positive satisfaction using Hernia Coach. One participant said it was cool, neat, and he would want to see it do more. Another said Hernia Coach performed better than he anticipated. An additional participant said it would have been great to have access to Hernia Coach to make access to information easier during their hernia surgery. One participant reiterated their preference for using this tool instead of reading through the paper education materials provided at discharge. This participant felt the paper education materials were verbose, which made it difficult to identify specific information to answer a question. They enjoyed being able to ask their question to Hernia Coach and getting a response that did not require sifting through pages of documents.

Two of the six participants also voiced some dissatisfaction with Hernia Coach. One had some negative perceptions of Hernia Coach because of incorrect responses to questions. Another said some of the responses seemed canned and noted Hernia Coach would direct users to contact a medical professional regular basis.

Outcomes: Behavioral Intention

All six participants indicated they would use Hernia Coach if they were having surgery or recommend it to others having inguinal hernia surgery. One participant would use it as a component of an overall information package. Another stated technology is interwoven with their life and would use it if they were to have surgery again. One participant said the information was quite thorough and he would ask the same kinds of questions of his clinicians.

Hernia Coach Design Suggestions

All six participants provided suggestions to improve Hernia Coach content. Two participants did not locate the information sources intent wanted links to external information resources for each response to learn more about each topic. One participant wanted outcome milestones after having surgery. Another desired more specific examples of allowed and not allowed activities after surgery. One wanted the caveats for certain restrictions (i.e., stop walking if you are experiencing pain) and additional multimedia content. A different participant wanted better guidance or context for error messages. One participant voiced that Hernia Coach should contain all of the information contained in the paper information packets.

Three participants had recommendations to improve the visuals provided with query responses. Two participants desired a bigger screen to view the visuals. The other participant suggested incorporating gender specific visuals matched to the gender of the user.

Two participants expressed a desire for additional functionality or interaction with Hernia Coach. One participant wanted Hernia Coach to pass notes or questions automatically to their care team. Another stated they would like to have access to a physical keyboard when typing queries.

Table 8: Representative qualitative analysis patient participant quotes.

HITAM Themes and Sub-Themes		Representative Patient Participant Quotes
Antecedent	HIT Reliability	“Getting access to the basic information seemed quite straightforward. And I felt like that the level of detail was probably appropriate for most patients.” - P3 “I think that you could go through the list when you got home and in probably like I say 9 times out of 10 is going to give you the answer you want but for that 10 percent you then you might have to call back in.” - P6 “I had to dumb down my questions. I had to parse my question so that would be understandable … I had to be more precise, as if I were talking to somebody for whom English isn't their first language. And of course, working with machine you lose the ability to communicate in ways that are verbal ways.” - P1
	HIT Self-Efficacy	“Maybe to some point I realized I was just getting back paragraphs that were set paragraphs. And so when I was sensing loops I sort of lost confidence.” - P1 “So I have confidence that so far what I saw that it would take care of most anything I thought during the day. But there are those little niche areas that well it didn't help me there and I'm going to have to call a doctor, which might save everybody a lot of time by the way.” - P6 “I guess it leads to how you got maybe how you asked the questions I don't know how the software is all set up but maybe it's. I'm not asking the right questions to get to the area the exact pinpoint target that I want.” - P6
Mediating Process	Perceived Usefulness	Yeah. It would've been cool if I had something like that to get information quickly so I didn't have to google it or call my doctor. Yeah definitely would have helped me when I was having my surgery … It's all kind of collated right there for you just all in one place.” - P4 “I didn't actually call and I just used the papers. So for that it would replace [the papers]. And sometimes people get frustrated because they can't find what they want in their papers. Whereas here generally you can find it quickly because you can tap on the topic or ask it. I think it could replace a lot of interaction with the doctor's when at home.” - P4

		"So I don't think this would replace a person 100%. But it could certainly answer some questions. So it would filter. The purposes would reduce the time a professional would spend. And this could do that to get some stuff out of the way. It should be [used] in conjunction with a person I think." - P1 "Well any time someone goes to surgery you know there's always fear. There's always the idea of like oh well what can happen. So a system like this because doctors and nurses are busy, they got stuff going on in their own lives and all that stuff. So if you were able to sit this down in front of the patient and he's got you know a thousand questions he or she at least go through most of the questions. This is decent information, the same information that doctors can give you." - P2
	Perceived Ease of Use	"It was easy and anticipated questions I could come up with cleary." - P1 "It's pretty easy to use as long as you got someone who's got a smartphone they can use it." - P2
	Attitude	"I think it's cool. I think it's neat. I'd like to see it do more." - P1 "I think that I would go to that before I would even look at paper. You know like they give you the flyers stapled packet. I would go to that before I went to the stapled packet. I think after that I would go right to the doctor or the nurse. I wouldn't even go back to the packet. I think everything that a packet can supply that should have." - P6 "Yeah it's better than what I was anticipating coming in here." - P3
Outcome	Behavioral Intention	"I would use it as a step. And I would use it in conjunction with being able to make inquiries of trained persons. I don't want [to use Hernia Coach exclusively] otherwise I would feel shunted off ... So as part of an information package that would include live people, I think has a place." - P1 "I would [use Hernia Coach]. You know me, personally I'm used to technology. I've kind of grown up with it." - P2 "If this is a mechanism that can be structured so that the surgeon is able to see more people without me getting the impression that he's trying to just push me off to something, then yes it frees the surgeon up … So I think where you're going here there is some there's some utility and some value here. And yes I'd recommend it." - P3

4.6.3 Stage 3 - Phase 3: Hernia Coach Query Response Evaluation

While annotating the design expert heuristic evaluation query logs and conducting the patient usability testing sessions, we identified there were instances when Hernia Coach query responses provided incorrect information to a question. The ability of Hernia Coach to understand patient queries and then provide relevant health information was also a concern raised by participants in Aim 2. To quantify the correct and incorrect query responses, we used a 2x2 confusion matrix to evaluate the accuracy of Hernia Coach's query responses during the design expert heuristic evaluation and patient usability testing sessions. Additionally, we wanted to identify if there was an improvement in query accuracy between the heuristic evaluation and patient usability testing sessions.

Fawcett (2006) describes the use of a 2x2 confusion matrix to depict classification outcomes, which are the basis for measures such as true positive rate or positive predictive value [146]. These measures are created by classifying the results of a clinical test or information retrieval result (i.e., instances) with a known true outcome (i.e., classifier). The four categories constituting a 2x2 confusion matrix are: a true positive where the instance and classifier are both positive, a false negative where an instance is negative and classifier is positive, a true negative where an instance and classifier are both negative, and false positive where an instance is positive but a classifier is negative. Confusion matrices are a common method to assess the accuracy of diagnostic clinical tests [195]. Confusion matrices are also commonly applied in computer science information retrieval research to assess the recall (i.e., true positive rate or sensitivity) and precision (i.e., positive predictive value) [196] of search engines or machine learning classification programs. Additionally, Walker and colleagues demonstrated confusion matrices can be used to analyze DS system query result accuracy [197]. Given we had participant queries submitted when interacting with Hernia Coach, Hernia Coach's query responses, and the ability to annotate the query responses with the correct outcome; we used Fawcett's confusion matrix classification method to assess the accuracy of Hernia Coach query results from the heuristic evaluation and patient usability testing sessions.

4.6.3.1 Stage 3 - Phase 3: Query Response Evaluation Methods

We analyzed Hernia Coach's query logs generated by design expert and patient. The query logs contained participant questions and Hernia Coach's query responses. RL reviewed the query logs and first annotated user queries using the classifier "within scope" (i.e., Hernia Coach is capable of answering the question) or "out of scope" (i.e., Hernia Coach is not capable of answering the question). Second, query responses (i.e., instances) were annotated as "within scope or "out of scope" (i.e., Hernia Coach generated an incorrect response). This process resulted in the classification outcomes for each query, which could be a true positive, false positive, true negative, false negative. In the subsequent two paragraphs, we describe the classification definitions used in the annotation process in detail and include a 2x2 confusion matrix depicting the definitions in Table 9.

For ***within scope user queries***, there were two different possible outcomes from the chatbot: within scope and out of scope responses. The first outcome were true positives (similar to the concepts of sensitivity or recall), which were defined as when a participant asked a within scope question and received the correct health content from Hernia Coach (i.e., within scope response). For example, a participant asked "what is a hernia?" and Hernia Coach correctly explains what is a hernia. The second outcome for within scope user queries were false negatives (similar to the concept of Type II error or β/beta). There were two different types of false negatives used in this analysis. False negative type A's were defined as Hernia Coach incorrectly returning a fallback intent to a within scope query (i.e., out of scope response). A fallback intent is a type of system response effectively saying "I don't know the answer to the question" or "ask your doctor." For example, a participant asked "what are the different types of surgery to fix a hernia?" and Hernia Coach says "I don't know what you said, could you please phrase the question in a different way." False negative type B's we defined as when a participant asked a within scope query but got a response consisting of irrelevant health content the question. For example, a participant asked "what are the risks of surgery?" (a known Hernia Coach topic) and a description of the mesh that could be implanted in the body during surgery is returned by Hernia Coach. A false negative type B could easily be misconstrued as a false positive. However, using the logic described in the preceding paragraph describing the annotation process, in this scenario the participant first asked a within scope question and second the participant received irrelevant health content from Hernia Coach for that particular question. Therefore this is a false negative because the participant asked a question Hernia Coach be capable of answering, but Hernia Coach provided irrelevant information to answer the question.

The reason why Hernia Coach provided incorrect responses to within scope questions is due to semantics and syntactics. The participants were asking semantically correct questions. In other words, the meaning of their questions were within scope. But the syntactics of their questions were confusing to Hernia Coach. In other words, the participants' questions were structured in a way that Hernia Coach did not understand.

For ***out of scope user queries***, there were two possible outcomes from the chatbot: within scope and out of scope responses. First, true negatives (similar to the concept of specificity) were defined as when the participant asked an out of scope query and Hernia Coach returned a fallback intent (i.e., within scope response). For example, a participant asked, “where is the nearest hospital?” (a topic not currently capable of being answered by Hernia Coach) and Hernia Coach says “I don’t know the answer to that question.” Second, false positives (similar to the concept of Type I error or ɑ/alpha) we defined as when participants asked an out of scope question, then Hernia Coach provided health content when in actuality Hernia Coach should have returned a fallback (i.e., out of scope response). For example, the participant might ask “where is the hospital?” and Hernia Coach states “on the day of surgery before going to the hospital wear loose clothing and don’t bring any valuables with you.” In this instance, Hernia Coach should have returned a response saying, “I don’t know the answer to that question.”

Table 9: Hernia Coach query annotation 2x2 confusion matrix definitions.

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Participant Query</th></tr>
<tr><th>Within Scope</th><th>Out of Scope</th></tr>
<tr><td rowspan="2">Hernia Coach Query Response</td><td>Within Scope</td><td>True Positive:
When a participant asked a within scope query and Hernia Coach returned the correct response.

Related terms: sensitivity or recall</td><td>False Positive:
When the participant asked an out of scope query, then Hernia Coach incorrectly returns health content and did not correctly return a fallback intent.

Related terms: Type I error or α/alpha</td></tr>
<tr><td>Out of Scope</td><td>False Negative:
Type A: When a participant asked a within scope query and Hernia Coach incorrectly returns a fallback intent.

Type B: When a participant asked a within scope query and Hernia Coach incorrectly returned irrelevant health content.

Related term: Type II error or β/beta</td><td>True Negative:
When the participant asked an out of scope query and Hernia Coach correctly returned a fallback intent.

Related term: specificity</td></tr>
</table>

Using the definitions in Table 9, we generated frequency counts of the classification outcomes and recorded the counts into a spreadsheet. Then we calculated true positive rates, true negative rates, false positive rates, and false negative rates, positive predictive values (i.e., precision), and negative predictive values using Google Sheets [198]. To calculate these measures, we first summed each type of classification outcome (e.g., true positives, false positives) and then applied the mathematical equations using the requisite classification sums. We used the sum of false negative types A and B in these calculations. True positive rate is the proportion of query responses correctly identified as within scope by Hernia Coach to all within scope

Hernia Coach queries. False negative rate is the proportion of query responses incorrectly identified as out of scope by Hernia Coach to all within scope Hernia Coach queries. False positive rate is the proportion of query responses incorrectly identified by Hernia Coach as within scope to all out of scope Hernia Coach queries. True negative rate is the proportion of query responses correctly identified as out of scope by Hernia Coach to all out of scope Hernia Coach queries. Positive predictive value is the proportion of queries correctly identified as within scope by Hernia Coach to all queries Hernia Coach predicted to be within scope. Negative predictive value is the proportion of queries correctly identified as out of scope to all queries Hernia Coach predicted to be out of scope.

We also performed a sub-analysis of patient false negative type B query responses (i.e., within scope query, irrelevant health content response) to identify if patients recognized they had received incorrect query responses. We performed this sub-analysis by reviewing the subsequent queries after the first instance of a false negative type B. If patients asked a subsequent query that rephrased their initial query, we used this as an indicator the participant potentially recognized an incorrect system response. If the participant asked a subsequent query that did not rephrase the initial query, we used this an indicator they may not have identified the incorrect system response.

4.6.3.2 Stage 3 - Phase 3: Query Response Evaluation Results

In total, design experts submitted 119 queries to Hernia Coach during the heuristic evaluation, and patients submitted 190 queries to Hernia Coach during the usability testing sessions, for a total of 309 queries. All of the query result measures improved between the heuristic evaluation and usability testing except for positive predictive value. The reason for this decrease is because design experts received less false positives from Hernia Coach (n=9) to true positives (n=86) compared to the patients who received more false positives from Hernia Coach (n=35) to true positives (n=125). We include Hernia Coach query result measures in Table 10.

Table 10: Hernia Coach query result confusion matrix measures for the design expert heuristic evaluation and patient usability testing sessions.

Measure	Design Experts	Patients
True Positive Rate	79%	93%
True Negative Rate	10%	26%
False Positive Rate	90%	74%
False Negative Rate	21%	7%
Positive Predictive Value	91%	78%
Negative Predictive Value	4%	57%

The sub-analysis results indicated in instances during which two patients may have recognized that they did not receive the correct response from the chatbot after they made a within scope query; this instance was marked by the participants re-asking their question, likely in an attempt to get the correct information. Four patient participants may not have recognized they did not receive a correct query response because their subsequent questions did not attempt to try and elicit a correct system response (e.g., by re-asking the previous question). Additionally, the first two patients who had re-asked questions to get a correct response also did not re-ask questions in other instances of incorrect system responses. This finding indicates all six patient participants, who have previously experienced inguinal hernia surgery, may not have realized they had gotten incorrect information to some of their queries. This suggests a potential exists that naive inguinal hernia surgery patients may not readily identify incorrect query responses.

4.7 Discussion

Patient-centered HDSs have the potential to provide an innovative platform to facilitate surgery patient information-seeking and reduce the burden of clinicians answering common patient questions. Guided by the principles of human centered design, we engaged in a rigorous multi-step HDS development and evaluation process. The development process resulted in a patient-centered Google Assistant “Action” prototype, called Hernia Coach, to serve as an HDS and aid inguinal hernia surgery patient information-seeking.

This research expands on previous work in a number of ways. First, we were able to demonstrate HDSs have the potential to support surgery patient information-seeking. Previous research has focused on developing HDSs to support patient information-seeking for patients undergoing ureteroscopy [17], women who are pregnant or parents of young children [18], and adolescents wanting to learn more about sex, drugs, and alcohol [19]. We build upon this existing research by detailing our methodologies for selecting a DS to build Hernia Coach, describing Hernia Coach’s design and development process, and engaging in a multi-step evaluation of Hernia Coach. The three HDSs supporting health information-seeking we identified provided varied descriptions or results regarding design, development, and evaluation activities. We described our process and results in great detail so other researchers may use our findings and methods to inform their own work. In addition to describing our methodologies in great detail, we used the Aim 2 needs assessment results, which were generated by patients and clinicians, to inform our design and development work. Two of the three identified HDSs also involved stakeholders in their design and development process [17,19]. This contrasts with the majority of mobile health applications, which are typically developed without the input of stakeholders such as patients and clinicians [199–202] or do not adhere to evidence based guidelines [203]. We also expanded on previous work by demonstrating it is possible to create an HDS built as a DS “Skill” or “Action.” Previous research has assessed the ability of DSs themselves to support patient information-seeking, not “Skills” or “Actions” [13,63,64]. While the specific form of technology assessed (i.e., DS) is slightly dissimilar to Hernia Coach, we reaffirmed the previous research findings by demonstrating there is a potential to provide patients with irrelevant or inaccurate health information with these technologies [13,63,64]. Finally, the patient participants in this study stated Hernia Coach could provide accurate and readily available health information compared to search engines or patient education materials, which aligns with the perceptions of other HDS users [19].

We also gained a number of insights during Hernia Coach’s design, development, and evaluation. First, the questions generated by participants in Aim 2 and recruiting friends and family members created a solid foundation for Hernia Coach’s NLU model in Dialogflow. However, during Hernia Coach’s evaluation activities in stage 3 - phases 1 and 2, we identified incorrect query responses indicating more work was needed to improve the NLU model in Dialogflow. Second, we identified Nielsen's Heuristics were effective for evaluating the Google Assistant "Skill" user interface. However, we desired more explicit results regarding the quality and usability of the conversation between Hernia Coach and the design experts. Third, it was challenging to distill patient education materials into conversation components to fit within the design constraints of a DS "Skill" or "Action." Guided by the topics participants desired in Aim 2 and the identified patient education materials to create Hernia Coach's health content, we were able to create topics and subtopics to fit the content within the constraints. Additionally, while creating Hernia Coach's health content, we identified consistent and inconsistent information across the identified patient education materials. Additionally, one of the nurses in Aim 2 stated she would vary her answers to patient questions depending on the surgeon who operated on the patient. This demonstrated a potential for practice variation to be misaligned with Hernia Coach's current health content. Fourth, the conflicting design and development documents published by Dialogflow and the Google Assistant at times provided conflicting guidance. Dialogflow was initially an independent company but was acquired by Google to facilitate the creation of Google Assistant "Actions." As we were designing and developing Hernia Coach, we encountered instances where Dialogflow's documentation guided us to perform a certain set of steps but the Google Assistant documentation had differing steps and was incomplete. The documentation was also regularly updated with new information or capabilities, indicating these technologies are constantly improving which will require HDS designers and developers to regularly reference the documents. Finally fifth, we encountered a number of Google Assistant "Skill" limitations such as differentiating between the

Google Assistant and Hernia Coach, displaying visuals, and discoverability of Hernia Coach features.

The evaluation of Hernia Coach generated design insights concerning HDS scope and capabilities, evaluation, and functionality. These contributions establish guidance for the design and development of patient-centered HDSs to facilitate health information-seeking which are described in detail below and summarized in Table 11.

Table 11: Design recommendations for developing patient-centered health dialog systems to facilitate health information-seeking

Health Dialog System Design Aspect	Recommendation
Scope and capabilities	Explicitly establish the health dialog system scope and capabilities.
Evaluation heuristics	Adapt or generate health dialog system heuristics considering both user interaction and conversation.
Google Assistant “Actions”	Consider Google Assistant Action limitations for health dialog systems.

4.7.1 Health Dialog System Design Recommendations

Our findings resulted three design recommendations. First, we recommend HDS designers and developers explicitly establish the HDS scope and capabilities. Second, we recommend adapting or generating heuristics considering both the HDS user interaction and conversation with the user. Third, we recommend considering Google Assistant “Action” limitations for HDSs and opportunities to improve Google Assistant “Actions.”

4.7.1.1 Design Recommendation 1: Explicitly Establish Health Dialog System Scope and Capabilities

Our evaluation of Hernia Coach with design experts and patients revealed Hernia Coach is generally capable of providing patient-centered health content to aid patient information-seeking activities. However, there is a risk for HDSs to answer patient questions incorrectly, which could lead to patient misunderstanding at best, patient harm at worst [13,63,64]. Establishing trust with users has been found to be an important component of HDSs [204], health technology acceptance [205–211] and technology in general [212–216]. If HDSs are going to serve as a trustworthy health information resource, it is paramount to ensure HDSs provide accurate health information within their scope and capabilities.

During the design and development process of Hernia Coach, we focused the majority of our efforts on establishing the information-seeking Hernia Coach should be capable of supporting. While we did some work to establish health content that should be beyond the scope of Hernia Coach; these efforts should have made a greater priority and was a significant blind spot during the design and development of Hernia Coach. Currently, little to no research has been published regarding HDSs conversation design recommendations or guidelines. Furthermore, the DSs evaluated in stage 1 published documentation exclusively focusing on developing what a system should know or be capable of accomplishing, and do not explicitly include considerations for establishing what a “Skill” or “Action” should not know or cannot accomplish [217–222]. This blind spot contributed to us focusing our efforts on developing system capabilities for what

Hernia Coach should know or accomplish and overlooking making Hernia Coach's knowledge limits more explicit to the user.

We recommend researchers creating HDSs for patient information-seeking establish the scope and capabilities of their system during the design and development process. Then the scope and capabilities should be explicitly conveyed to the HDS users. One method of establishing the scope and capabilities would be to train the HDS to know what questions are beyond its capabilities. The technology used to create Hernia Coach, Dialogflow and the Google Assistant, was designed to always provide a response to a user's query. If we created intents with training phrases consisting out of scope questions, Hernia Coach would have been able to better identify and inform the user it is not capable of providing the desired information. By programming an HDS to know what types of topics and questions it is unable to answer, this will reduce the chance patients are presented with incorrect information. Other Hernia Coach components that proved useful for conveying the Hernia Coach's scope were the topic list, error messages suggesting potential alternate queries, and suggestions chips for related topics.

After establishing an HDSs scope and capabilities, we suggest researchers recruit a diverse group of participants to ask questions of their HDS to provide rich training data for the system's natural language understanding engine prior to use by patients. A major factor that contributed to Hernia Coach failing to provide correct answers to within scope participant queries was due the participants asking questions that were

semantically correct but **syntactically** confusing to Hernia Coach. In order words, the **meaning** of the participants queries were capable of being answered by Hernia Coach, but the **structure** of the query was incorrectly recognized by Hernia Coach. This misunderstanding resulted in an incorrect system response. For example, if a user asks “what are the risks of surgery?” and Hernia Coach responds with a description of people who generally develop an inguinal hernia. By using these design and development recommendations, HDS designers and developers will be able to establish and convey the knowledge and capability limits of an HDS to the user.

In addition to establishing and testing an HDSs scope and capabilities, we recommend HDS designers and developers select and build technology that supports users who have less socioeconomic advantages than others. When creating health information technology, it is possible to inadvertently create informatics generated inequalities [223]. This occurs when an intervention is more accessible, highly adopted, and effective for individuals who have greater socioeconomic advantages [224,225]. To mitigate the potential of creating greater health informatics inequalities, we considered how to make Hernia Coach accessible in a way that would promote adoption and use by individuals with less socioeconomic advantages. To accomplish this goal, we built Hernia Coach using widely-adopted technology. In 2019 eighty-one percent of U.S. adults owned a smartphone [226], in 2018 sixty-six percent reported using a smartphone based DS, and twenty-five percent reported using the DS daily [78]. We also selected the Google Assistant because it offered the ability for users to interact with Hernia Coach using voice or a screen, supported presenting visual information, incorporated buttons, and presented information both visually and audibly, and supported nineteen different languages. We perceived these characteristics as having the ability to support users who may have physical impairments that would prevent them from using other types of DSs. Additionally, we aimed to write Hernia Coach's health content to be accessible to individuals with lower rates of health literacy and education. We encourage other researchers to consider these factors when designing and developing HDSs.

4.7.1.2 Design Recommendation 2: Adapt or Generate Health Dialog System Heuristics Considering both User Interaction and Conversation.

While preparing to conduct Hernia Coach's evaluation activities, we identified research studies and online articles by human-centered design professionals describing how Nielsen's Heuristics may need to be adapted to better evaluate new computing hardware such as smartphones [163–165] or DSs [167–169]. In light of these considerations, we adapted Nielsen's Heuristics in an effort to improve the evaluation of Hernia Coach. We found the adaptations to be an acceptable first attempt, but more research is needed to generate generalizable HDS heuristics. We recommend HDS designers use our heuristics and other recommendations from the literature to inform their adaptations for evaluating HDSs. The heuristics we created may not generalizable to other HDSs, such as HDSs focused on accomplishing tasks (e.g., requesting medication refills, modifying clinic appointment times). Additionally, we identified the heuristics were not as effective for evaluating the conversations between users as we had hoped, which we described in greater detail in the rest of this section.

Our adapted heuristics were most effective at identifying the Google Assistant's user interface. However, the heuristics were not as effective for evaluating Hernia Coach's conversation with users. We included adaptations to evaluate the Hernia Coach's conversation, but these adaptations could have been more explicit and rigorous. For example, we asked the design experts to evaluate the consistency of the language within the conversation and identify if the medical terminology is accessible. However, we did not ask the design experts to consider aspects such as the quantity, quality, manner, and relevance of the conversation components. This most notably was a limitation when the design experts reported the error messages were acceptable but in some circumstances could be improved due to lack of context. We had written the error messages according to the Google Assistant "Action" design recommendations [227] and we considered Nielsen's error heuristic, but we were unable to ascertain how the error messages could be improved solely based on the design expert's feedback. We were later able to identify the error message's relevance could be improved depending on the types of questions users were asking during the patient user testing sessions.

Given these findings, we recommend HDS designers adapt or develop evaluation heuristics considering both user interface and conversation as distinct units of analysis in an overall set heuristics. This type of approach was presented as an effective evaluation method for Google Assistant "Actions" at Google I/O 2019 [228]. In particular, we recommend considering H. Paul Grice's (1975) conversation "Cooperative Principle" [229] to inform conversation heuristics. Grice's cooperative principle is an encouraged consideration when building Google Assistant "Actions" [217], but we could not identify literature incorporating the principle into a heuristic evaluation. Grice's Cooperative Principle consists of four maxims, which are quantity, quality, manner, and relevance. Quantity concerns the amount of information in a conversation, which is to say as much as the conversation requires but no more. Quality refers to the information accuracy or truthfulness in a conversation, which is to not make false statements or make a claim when lacking evidence. Manner pertains to clear communication, which is to be brief and orderly in a conversation. Avoid being obscure or ambiguous when conversing. Finally relevance concerns the context of the conversation, which is to provide applicable information during discourse. For example, "aesthetic and minimalist design" could be adapted to say "only provide interactional elements that are necessary to engage the user and fit within the goal of the system. voice interfaces should support short interactions and expand on the conversation if the user chooses," which aligns with the maxim of quality. Or a new heuristic could be developed called "veracity," which states the HDS should "be honest with the user by providing accurate information within the dialogue." While these recommendations are a good initial set of considerations for evaluating HDS user interfaces and conversation, more work is needed to develop generalizable HDS evaluation heuristics.

4.7.1.3 Design Recommendation 3: Consider Google Assistant "Action" Limitations for Health Dialog Systems

Hernia Coach's evaluation findings generated five specific Google Assistant "Action" design recommendations. We include descriptions of the limitations and design recommendations for Google Assistant Actions in Table 12. First, despite our best efforts to generate Hernia Coach health content at the recommended fifth grade reading level [91], we were ultimately unsuccessful. The Google Assistant “Action” 300 character length design guideline for conversation responses [90] was a limiting factor. This guideline made it challenging for us to create succinct content that also was easy to read. Writing some sub-topic concepts around 300 - 400 characters in length produced long compound sentences, which increased reading difficulty. We attempted to mitigate the increased reading difficulty of these sentences by distilling the concepts into smaller chunks as bullet points on a slide image. The non-compound sentences on the slides we created had easier readability than the compound sentences. Our goal was for Hernia Coach to read the compound sentence aloud but only present the slide image. In the end these efforts broke Dialogflow and in turn Hernia Coach. We recommend HDS designers and developers consider this Google Assistant “Action” limitation when generating HDS health content. We also recommend the Dialogflow and Google Assistant development team allow “Action” creators to facilitate new methods of displaying content that is not dependent on including the query response text. We include a description of Google Assistant “Action” limitations and design recommendations in Table 12.

Second, all of the design expert participants reported difficulty opening Hernia Coach or identifying if Hernia Coach was active. Initially Hernia Coach’s invocation name (i.e., how the “Action” is loaded within the Google Assistant) during the heuristic evaluation was “Inguinal Hernia Info Bot.” This name was reported by the design experts as difficult to remember and say. As a result, we changed the invocation name before patient testing to “Hernia Coach.” We recommend choosing an easy to remember invocation name to aid HDS discoverability. Additionally, the design experts noted it was difficult to identify when they were interacting with the Google Assistant itself or Hernia Coach. The only differences between the Google Assistant and Hernia Coach was a change from the Google Assistant logo and voice to Hernia Coach’s logo and voice. We recommend HDS designers and developers create a distinctive logo to help aid a user’s ability to identify if the HDS is active to mitigate this limitation. Additionally, we recommend the Google Assistant “Action” development group improve system visibility to better indicate if the HDS is active. Some examples include visually altering the user interface by augmenting colors and making the logos larger.

Third, the evaluation participants also noted at times it was difficult to view the visuals incorporated in Hernia Coach. We recommend HDS designers consider alternatives for visuals, such as the use of basic cards instead of browsing carousels to increase image size on smartphone screens. We also recommend the Google Assistant “Action” development group give HDS designers and developers greater flexibility to modify image sizes within query responses.

Fourth, some evaluation participants were unable to discover the information sources list or desired greater information about a topic than what was provided by Hernia Coach. We recommend HDS designers and developers consider incorporating links to external information sources for each query response, prompting users with follow up intents to learn more, and creating an information source suggestion chip for each intent.

Finally, we used a confusion matrix to evaluate Hernia Coach's query responses to identify if Hernia Coach was providing relevant information to user questions. This method of analysis has been previously used to evaluate task-oriented DSs, but not question and answering focused DSs [230]. Dialogflow incorporated the ability to annotate query logs and indicate correct or incorrect system responses. We recommend HDS designers and evaluators use this method to identify the effectiveness of their system to provide relevant health information to user queries. We also recommended the Dialogflow development group extend this functionality to include the automatic generation of the metrics used to analyze Hernia Coach's query responses, such as true positive rate and positive predictive value. It is also important to note query review presents a data privacy, security, and scalability limitation. This practice assumes someone with specific health condition domain expertise is available to regularly review user queries. Additionally, query review is a regular component of DS quality improvement processes, but has been subject to a large amount of negative news articles. The articles focus on how technology organization employees are reviewing and listening to potentially sensitive DS user interactions and the users may not be aware of the practice [231–233].

Table 12: Google Assistant “Action” Limitations and Design Recommendations

Limitation	Google Assistant Design Recommendation
Query response character limit	Facilitate alternate methods of displaying “Action” content that is not dependent on including query response text
Visibility of Google Assistant “Actions”	Make “Actions” more distinct visibly and audibly from the Google Assistant itself
Visual content sizing	Provide “Action” designers and developers more flexibility when incorporating visual content into query responses
Query response accuracy	Provide “Action” designers and developers query response accuracy metrics

4.7.2 Limitations and Future Work

Our work is subject to a number of limitations. First, the patient participant recruitment methods could have resulted in self-selection bias. Given the sample and geographic location of this study, our findings may not be generalizable to other hernia patients. Second, the patient-education materials used to generate Hernia Coach’s health content was not adapted for use by a specific surgeon or clinic. The purpose of this study was not to create clinic ready health information for Hernia Coach. Rather, the objective was to identify if a prototype Google Assistant “Action” has the potential to aid patient health information-seeking and opportunities to improve the prototype. Third, the patient participants may have been affected by acquiescence bias (i.e., friendliness bias) because the researcher who developed Hernia Coach also lead the evaluation activities in phase 3 - stages 1 and 2, not an independent party. However, the design expert heuristic evaluations were conducted remotely, which mitigated the potential of this bias in their feedback. Fourth, our design, development, and evaluation efforts was limited to the English language. Given this focus, our findings may not be generalizable to other languages.

We recommend future clinic oriented work incorporate additional focus group or brainstorming activities with surgeons to better identify questions and topics that should be answered by surgeons or nurses. Additionally, we suggest conducting a pilot implementation with a copy of Hernia Coach with partnering physicians and clinics. During the pilot implementation we recommend adapting Hernia Coach's health content to align with the physicians and clinics practices

We also recommend the Google Assistant "Action" development team improve how HDS designers and developers can create Google Assistant "Actions." Specifically, more flexibility for information presentation, improved differentiation of the Google Assistant itself versus "Actions," and providing metrics for query response accuracy. In addition to these improvements, we recommend the development team include considerations for HDS designers to create intents pertaining to questions beyond the scope of their system within the design and development documentation. Conveying to the user what the HDS does not know is equally or more important than conveying what it can accomplish. In addition to revising the "Action" design and development documentation, we recommend the "Action" design and development team explore methods for conveying HDS scope and capabilities more effectively to users. Finally, we recommend the Google Assistant "Action" team explore potential methods for creating copies of HDSs that allow clinicians or health educators to adapt the HDS content to account for practice variation. For example, future functionality could support the HDS team creating an HDS copy and providing a link for different stakeholders to adjust the content with a user friendly website.

4.8 Conclusion

Inguinal hernia surgery patients have information needs before, during, and after surgery. However, patient education materials and clinician interactions may not sufficiently address these information needs. We approached this problem by designing, developing, and evaluating a patient-centered prototype Google Assistant "Action," called Hernia Coach, to serve as an HDS and facilitate patient health information-seeking. Our user-centered Hernia Coach provides a promising alternate information resource for patients to consult during their hernia surgery journey from diagnosis through recovery. Our findings generated design recommendations to guide the development of future patient-centered HDSs.

4.9 Acknowledgements

We acknowledge the time and effort invested by the study participants; Will Kearns and Shefali Haldar for their feedback and insights with this research; and Claire Lordon, who created the HDS Prototype logo.

4.10 Appendix

Chapter 4 Supplemental Table 1: The dialog system evaluation criteria, definitions, ranks, and results for each dialog system.

Criteria	Definition	Rank	Google Assistant	Alexa	Cortana	Bixby
Included functionality	Areas or functions of the research study that the software has to serve.	High	Provide inguinal hernia surgery patients with relevant and easily accessible health information before surgery, during the day of surgery, and after surgery.	Provide inguinal hernia surgery patients with relevant and easily accessible health information before surgery, during the day of surgery, and after surgery.	Provide inguinal hernia surgery patients with relevant and easily accessible health information before surgery, during the day of surgery, and after surgery.	Provide inguinal hernia surgery patients with relevant and easily accessible health information before surgery, during the day of surgery, and after surgery.

Openness	Level of openness to additional development (internal and external) and to other existing applications (i.e. build content or "skill" that runs as additional functionality in the distribution platform)	High	Google supports the creation of "actions" by third parties. In January 2018 there were 2400 google assistant actions. "Actions on Google is the platform that allows developers to build Actions for the Google Assistant on Google Home and other voice-activated speakers, eligible Android phones, iPhones, Android TVs, headphones and soon everywhere else the Assistant is available."	Alexa supports skill creation and distribution. Additionally Alexa is supports integration with other devices created by external organizations and enterprise integration. In September 2018 Amazon announced 50,000 skills had been created for Alexa.	Cortana supports the creation and distribution of skills. However the skills must be reviewed and approved by Microsoft.	Bixby supports the creation of additional content by external parties with the use of "capsules" (i.e. skills) but capsules are limited to Bixby.

Main target	Functional area(s) for which software is specially oriented or strong	Medium	"The Google Assistant allows you to have a natural conversation between you and Google. It's one Assistant that's ready to help you throughout your day. It is already accessible on more than 100M devices across Android phones, the iPhone, Google Home and other voice-activated speakers, Android TVs, wearables and our messaging app Allo but it is soon coming to additional devices and contexts." The google assistant can be used for information-seeking, accomplishing tasks, and smart home integration. The Google Assistant supports the following languages: English, Danish, Dutch, French, German, Hindi, Indonesian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, Thai, and Turkish. Dialogflow supports	One of Alexa's major differentiators is it's widespread adoption compared to Cortana and Bixby. Amazon states "tens of thousands of developers have built skills using the Alexa Skills Kit." Compared to other smart assistants, Alexa has a much stronger commerce orientation. Another focal point of Alexa is smart home integration Alexa supports the following languages: English (AU), English (CA), English (IN), English (UK), English (US), French (CA), French (FR), German (DE), Italian (IT), Japanese (JP), Spanish (ES), Spanish (MX)	Cortana provides answers to questions, schedule reminders, directions to locations, or accomplish tasks (e.g. playing music). Recently Microsoft announced Cortana will no longer be pursued as a standalone smart assistant. Rather, Cortana is going to be incorporated into other smart assistants or accessed through existing smart assistant infrastructure. Current Cortana Skills Kit is in public preview and available in the United States only. The following languages are available in these countries: Australia: English, Brazil: Portuguese,	Assists with completing tasks, learning routines to provide appropriate and timely content (e.g news, messages, app data), and providing reminders. Can identify objects (e.g. landmarks, QR codes) with a device's camera , stored images, or using a web browser. Assists users with smartphone navigation and interactions. Bixby only recognizes English (UK), English (US), French (France), German (Germany), Italian (Italy), Korean (South Korea), Mandarin Chinese (China), and Spanish (Spain). Other languages will be supported in the future.

			action development for multiple languages.		Canada: English/French, China: Chinese (Simplified), France: French, Germany: German, India: English, Italy: Italian, Japan: Japanese, Mexico: Spanish, Spain: Spanish, United Kingdom: English, United States: English	

Adaptability	Possible level of customization in general and for this research study	Medium	Developers have the ability to create custom voice and visual content. Google offers templates to aid novice designers.	Amazon offers an extensive level of pre-built "off the shelf" Alexa features to assist developers with the creation of Alexa skills. Amazon also offers the ability for developers to create custom skills as well. There is an "Alexa Cookbook" repository on Github.	Cortana is designed to function on devices with or without a screen. However designers are encouraged to consider design that is not screen dependent and use visuals as an enhancement. Designers have the option to integrate "Cortana Cards" to customize components of the visual interface. Designers are also able to customize the content and purpose of a Cortana Skill.	Bixby supports the creation of additional content by external parties with the use of "capsules" (i.e. skills) using their Bixby Developer's studio.

Interoperability	Capability to integrate with other tools and applications, including NLUs	High	Dialogflow supports integration with the Google Assistant. Converstation.One supports skill creation.	Compatible with Amazon Lex Framework. Dialog supports exporting "agent" to Alexa and importing of Alexa skill to Dialogflow. Converstation.One supports skill creation.	Cortana is compatible with the Microsoft Bot Framework. Cortana is available as an Amazon Alexa skill. Dialogflow supports exporting "agent" to Cortana, but Cortana skill cannot be imported into Dialogflow.	Bixby currently does not support integration with other tools, applications, or external NLUs.
Security levels	Breadth of security policies, including HIPAA, supported by the software package (user identification, auditing, data encryption)	High	Google Assistant Actions must be approved prior to publication. Google has extensive privacy and security policies published for to guide their developers. Google Assistant Actions that transmit data covered by HIPAA are not allowed. Actions providing health information must include a disclaimer when the Action is loaded and in the description of the Action in the Action Directory. However,	Alexa skills must be approved prior to publication. To be approved skills must adhere to policy guidelines and pass security requirements. To collect personal information, developers are required to have a published privacy, obtain user consent to collect the data, and adhere to privacy policies	Cortana skills must be approved before publication. Health related data are not permitted to be collected using Cortana, except for activity and fitness related data. The Cortana skills kit explicitly states it is not HIPAA compliant. Developers must provide users with a data privacy policy.	Bixby capsules must be approved for publication. Bixby does not have explicit documentation regarding the permitted or unpermitted use of health related capsules. The Samsung privacy policy states it may retain data and share it with third parties.

			Actions concerning fitness and activity monitoring data such as weight and calories burned are permitted.	and laws. Alexa health related skills are not permitted to collect health related data, claim to provide lifesaving assistance, contain false or misleading health claims in the content of the skill. provide information about black market prescription drugs. Health related skills are required to have a disclaimer published in the description and may be required to have a disclaimer when the skill is loaded. Alexa skills may not connect users to 911 or other emergency response services.		

Number of simultaneous users	Number of simultaneous users that can be linked and served by the system	Medium	Not explicitly listed in the developer documentation or on the web.	Not explicitly listed in the developer documentation or on the web.	Not explicitly stated in the developers documentation	Not explicitly stated in the developers documentation.
Customizability	Ability to personalize the layout and other components of software interface	Medium	Developers are able to create custom visual content for screen enabled Google Assistant devices. Developers can incorporate pictures, video, audio, and text content into screen based content.	Developers are able to create custom visual content for screen enabled Alexa devices. Developers can incorporate pictures, video, audio, and text content into screen based content.	Designers have the option to integrate "Cortana Cards" to customize components of the visual interface and the content within a Cortana skill.	Bixby developers have the ability to choose from different interface designs for capturing user information, presenting information, and user interactions.
Interface type(s)	Interface type(s) of the software	Medium	Users may interact with the Google Assistant using voice and text based methods	Alexa users may only interact with Alexa using voice, not text. This is a differentiating feature of Alexa from other smart assistants.	Cortana supports speech interface and text based interfaces on devices that support either or both of those inputs. For example it is possible to text or talk with Cortana on Windows 10 devices or with the mobile app.	Supports both speech and text based interaction on mobile devices. Future support for a proprietary smart speaker is noted.

Programming languages	Ability to personalize modules by programming languages	Medium	The Google Assistant supports the speech synthesis markup language. Google also has an Assistant SDK and API that supports Node.js, Python, Go, C++, and Java.	The Alexa Skills kit supports Node.js, Java, and Python. Visual content can be created using the Alexa Presentation Language. Alexa skills can access cloud services with JSON requests/respon ses. Alexa's speech synthesis markup language is used to adjust how Alexa pronounces terms and sentences.	Cortana skill developers may create customized skills using C# or Node.js	Bixby supports personalization within capsules with their proprietary software developers kit.
DBMS standards and integration	Breadth of database management systems that can be accessed by software package (SQL server, Oracle, DB2, Sybase, Informics)	Medium	The Assistant SDK and API supports Node.js, Python, Go, C++, and Java.	Alexa skills kit has a built in web or DB search function. Alexa also supports other DBs such as dynamo db and lambda functions to query relational databases.	Not explicitly stated in the developers documentation	The developer documentation does not specifically list supported DBMS integration. However, it does have documentation regarding Javascript API integration.

Communication standards	Inter-organizational data exchange standards that are supported by software package (EDI, XML)	Medium	Not listed in the developer documentation	Alexa supports JSON.	Cortana supports XML	Javascript API
Platform variety	Capability of the software package to run on a wide variety of computer platforms	Medium	Alexa is supported on Android devices, iOS devices, and Google Assistant enabled hardware.	Alexa is supported by an extremely large number of devices and is available as an Android or iOS download.	Cortana is supported on a variety of devices, including Windows 10 PCs, Android and iOS as mobile app downloads, and a smart speaker. Cortana is also capable of running as a skill within Amazon's Alexa.	Bixby is currently capable of being implemented on a small subset of handheld devices. Currently supported: Galaxy S9 / S9+, Galaxy S8 / S8+, Galaxy Note8, Galaxy Note9, Samsung Smart Fridge, Samsung Smart TVs Future device: Galaxy home (Samsung smart speaker), Samsung washing machines
Scalability	Ability of the software package to handle increasing number of users and higher load of transaction	Medium	Not listed in the developer documentation	Amazon has published online documentation as to handle provisioning for increased server loads but does not list an upper limit	Not explicitly stated in the developers documentation	Not explicitly stated in the developers documentation.

User interface	Ease with which users can use interface of the software package	High	Users may interact with the Google Assistant using voice and text based methods	Users can easily interact with Alexa through a variety of devices or smartphone applications. However, Alexa is limited to voice interactions and cannot used text based methods,	Users can easily access Cortana by using the wake word "Hey Cortana" or by launching Cortana within the mobile app or Windows 10 devices. Users may interface with Cortana using text or voice methods. Cortana is designed to provide a cohesive experience across devices logged into the same account.	Generally not explicitly stated in reviews of Bixby. Reviewers generally dislike the dedicated Bixby button and mentioned it is easily disabled. Bixby is easily accessible using a dedicated button on smartphones or by using the wakeword "Hi Bixby". Users may interface with Cortana using text or voice methods.

User types	Ability of the software package to support beginners, intermediate, and advanced users or a combination of user types	Medium	At this time, explicit information regarding support a diverse set of user types is not included in the developers documentation. An ILL scan request has been placed for a formal usability assessment of Alexa with other smart assistants. This will be updated with new applicable information once it becomes available.	At this time, explicit information regarding support a diverse set of user types is not included in the developers documentation. An ILL scan request has been placed for a formal usability assessment of Alexa with other smart assistants. This will be updated with new applicable information once it becomes available.	At this time, explicit information regarding support a diverse set of user types is not included in the developers documentation. An ILL scan request has been placed for a formal usability assessment of Cortana with other smart assistants. This will be updated with new applicable information once it becomes available.	Bixby incorporates gamification to incentivize users to use Bixby, but it was not well received by one reviewer. The gamification component encourages users to interact with Bixby using all of the different features available.
Data visualization	Capability of the software package to present data effectively	High	The Google Assistant has the ability to present data using text, audio, and screens	Alexa has the ability to convey information to users using voice and screen methods. However, the limitation of only using voice is a concern for use in public areas such as clinic waiting rooms.	Cortana skill developers may incorporate visual information for screen based devices that may contain a combination of text, speech, images, buttons, and input fields. Developers may also customize how Cortana presents information	Not explicitly stated in reviews of Bixby. Developers are able to customize data presentation visually and verbally in Bixby.

					verbally to the user.	
Error reporting	Error reporting and messaging ability of the software package	Medium	The Google Assistant supports error logging	Alexa has an error handling interface built into the skills kit. It includes messaging and a data dictionary for error values,	Cortana supports error logging and debugging during skill testing prior to deployment	Bixby supports error logging and reporting

Domain variety	Capability of the software package to be used in different industries to solve different kinds of business problems	Low	The Google Assistant supports enterprise applications	Alexa has the capabilities to be incorporated in enterprise environments. There are two main use cases. First concerns using Alexa during meetings to coordinate logistics and to provide audio conference capabilities. Second is for employee productivity.	Cortana currently has limited ability to support enterprise applications to assist with information retrieval tasks, complete tasks or workflows with voice, or assist with user task focus	Bixby is currently limited to a subset of capabilities concerning smartphones and other Bixby supported devices. The ability of Bixby to be used in other domains has not been shown at this time.

Ease of use	Ease with which users can learn and operate the software package	Medium	The Nielsen Norman Group found the Google Assistant, Alexa, and Siri provided very poor user experiences. "Our ideology has always been that computers should adapt to humans, not the other way around. The promise of AI is exactly one of high adaptability, but we didn't see that when observing actual use. In contrast, observing users struggle with the AI interfaces felt like a return to the dark ages of the 1970s: the need to memorize cryptic commands, oppressive modes, confusing content, inflexible interactions — basically an unpleasant user experience." The group found smart assistants have poor usability in regards to voice input, natural language, voice output, intelligent interpretation, agency (little to no use of external resources), and integration with other technologies.		Cortana usability testing research could not be located using scholarly and general web searching.	Gamification promotes users learning Bixby's capabilities.
Robustness	Capability of the software package to run consistently without crashing	Medium	There are forum reports of the Google Assistant crashing	Not explicitly listed in the developers documentation or on the web	Anecdotally Cortana reliability issues seem to pertain to Windows PCs. Information regarding Cortana reliability is not included in the developers documentation	Not explicitly listed in the developer documentation and on the web.
Backup and recovery	Capability of the software package to support backup and recovery feature	Medium	Not explicitly listed in the developer documentation or on the web.	The skills kit offers the capability to clone skills to create backups	Not explicitly listed in the developer documentation or on the web	Not explicitly listed in the developer documentation and on the web.

Time behavior	Ability of the software package to produce results in a reasonable amount of time relative to data size	Medium	Not explicitly listed in the developer documentation or on the web.	Not explicitly listed in the developer documentation or on the web	Not explicitly listed in the developer documentation or on the web	Not explicitly listed in the developer documentation and on the web.
User manual	Availability of user manual with indexes, with important information and the main commands	Medium	Google has published extensive developer documentation online	Extensive developer documentation is available on the web	Cortana Skill Kit developer documentation is available online and as a PDF download. Important developer information, including commands, are included in the documentation.	The Bixby developers guide is available online and in the development software. Pertinent information and command documentation are available in both locations.
Tutorial	Availability of tutorial to learn how to use the software package	Medium	Google has published tutorials online for beginners	Amazon offers online tutorials for beginners and advanced skill builders. There are also many free tutorial videos on the web.	Text and video tutorials for novice Cortana Skill developers are available in the developer documentation and online.	Gamification promotes users learning Bixby's capabilities.
Troubleshooting guide	Availability of troubleshooting guide	Medium	An action troubleshooting guide is included in the developers documentation	A custom skill test and debug chapter is included in the developers documentation	A troubleshooting guide is included in the developers documentation.	Troubleshooting guide for natural language training is incorporated into the developers documentation.

Training	Availability of training courses to learn the package	Medium	Google has published tutorials online for beginners	Amazon offers online tutorials for beginners and advanced skill builders. There are also many free tutorial videos on the web. Amazon also has extensive voice ui design documentation and guides.	A Cortana Skills virtual class is available online.	Text based training materials and some online videos are available for developers to learn Bixby. Samsung hosted a Bixby specific developers conference in 2018.
Maintenance and upgrading	Vendor support for upgrading and maintenance of the software	Medium	Google has support communities, a known issues KR, and direct contact information published on their support pages	Alexa hardware automatically updates and offers updates for mobile applications.	Not explicitly stated in the developers documentation, but press releases and news articles point to Microsoft regularly updating and maintaining the software on the developers behalf.	Not explicitly listed in the developer documentation and on the web.
Consultancy	Availability of technical support and consultancy by the vendor	Medium	Google has support communities, a known issues KR, and direct contact information published on their support pages	Amazon hosts developer forums for crowdsourced technical support	The Cortana support portal in the developers documentation website links to a blog, a FAQ page, and stack overflow sites pertaining to the Microsoft Bot Framework and Cortana.	Samsung offers a helpdesk system, feature requests/feedback channels, and developer forums on their website.

Communication	Communication between user and vendor	Medium	Google has support communities, a known issues KR, and direct contact information published on their support pages	Amazon hosts developer forums for crowdsourced technical support	Microsoft affiliated support engineers engage on the stack overflow sites. A direct contact option is not included in the Cortana Skills developers portal.	Samsung offers a helpdesk system, feature requests/feedback channels, and developer forums on their website.
Demo	Availability of on-site demo and free-trial version	Low	Users have the ability to try out the google assistant's mobile app as the lowest cost option	Developers have the ability to try the smartphone app.	Developers have the ability to demo Cortana by using a device capable of running Cortana.	Developers have the ability to demo Bixby by using a supported Samsung device.
Number of installations	Number of installations of the software package	High	Google claims the Assistant works with 10,000 devices and is installed on more than 400 million devices in 2017	Amazon claims 100 million Alexa enabled devices have been sold	Microsoft state in June 2018 there were 150 million Cortana users worldwide	Over 100 million devices have been sold with Bixby - Android Central
Response time	Level of service rendered by the vendor	Medium	Not explicitly listed in the developer documentation or noted on the web.	Not explicitly listed in the developer documentation or noted on the web.	Not explicitly listed in the developer documentation or noted on the web.	Not explicitly listed in the developer documentation or noted on the web.

Length of experience	Experience of vendor about development of the software product	Medium	The Google Assistant was launched in 2017. Google is heavily investing in the Google Assistant and currently is used by 30% of US adult smartphone users	In November 2014 Alexa was launched at the same time as the Echo smart speaker device. In 2017 Amazon stated 5000 employees were working on Alexa. Amazon regularly conducts contests to promote the development of the Alexa platform.	Cortana was launched in April 2014 and has been regularly updated since its release. Cortana has also been extended to numerous devices and software platforms since its inception.	Bixby was launched on March 29, 2017 and is generally considered less refined than other smart assistant distribution platforms.
Product history	Popularity of vendor product in the market	High	The Google Assistant currently is used by 30% of US adult smartphone users and on 400 million devices	Amazon claims 100 million Alexa enabled devices have been sold and is used by 17% of smartphone owners	Cortana represented 4% of smart assistants on smartphones in November 2018.	Bixby represented 4% of smart assistants on smartphones in November 2018

Vendor popularity	Popularity of vendor in the market	Low	In 2018 Google Assistant's parent company, Alphabet, was the 3rd most valuable company that year. In 2018 Android comprised 85% of all smartphones worldwide.	On January 8th 2019 Amazon was the most valuable company on the planet. Amazon does not manufacture smartphones.	The final update for Windows Phone OS was in June 2015, indicating Microsoft's exit from smartphone software development. Cortana is used by 4% of US adult smartphone owners. Microsoft has reported 150 million active Cortana and has windows software installed on 83% of desktops worldwide.	Samsung ended 2017 with 21.9% of the smartphone market
Technical and business skills	Technical and business skills of the vendor	Medium	Google's core products are search, translation, web browsing, maps, online video, and productivity software for personal use, enterprise, and developers.	Amazon is a leader in ecommerce and cloud computing	Microsoft is a personal computer software and hardware manufacturer and its products are used by people all over the globe.	Samsung is a consumer electronics and durable goods manufacturer focusing on cell phones, smartphones, tablets, wearable technology, durable home goods such as televisions and washing machines, and computers

Past healthcare experience	Past healthcare experience of the vendor, if any	High	Google currently states it is involved with healthcare and biosciences using AI. Currently has tools to diagnose retinal diabetes, detecting cancer, and genomics. Their website lists partnerships with UCSF, University of Chicago Medicine, and Stanford Medicine. Google at one point offered a PHR called Google Health. Google also recently hired the former CEO of Geisinger Health, Dr. David Feinberg to run it's health team. Google also has new health APIs to support interoperability and health focused cloud services.	Amazon has initiated a partnership with Berkshire Hathaway and JP Morgan for an employer health initiative. It has acquired Pill Pack and explored at home medical diagnostics. A patent has been filed for Alexa to identify a cold or cough. Amazon also started an internal health clinic and mines patient records.	Microsoft has been involved with healthcare in the past. Specifically Microsoft has developed an activity tracker (Microsoft Band) and actively engages in health related research to help achieve the quadruple aim. Microsoft is focusing on security/complian ce, personalized care using clinical analytics/genomic , care coordination, operational analytics, and patient engagement.	Bixby integrates with Samsung Health application in it's interface. The computer vision feature will assist with nutrition label interpretation and calorie counting. Samsung also has healthcare industry oriented products including ultrasound, radiography, in-vitro diagnostics, and portable ct scanners. Bixby has not been previously used in health.

License cost	License cost of the product in terms of number of users	Medium	Dialogflow offers a free tier of support as long as the chatbots usage is less than a certain limit, which is up to 180 requests per minute. Higher usage chatbots can purchase additional computing time for a nominal fee.	The Free AWS subscription offers one million lambda requests and 750 of elastic compute times per month to run Alexa skills. Once those thresholds are reached, it's possible to purchase additional computing resources as needed	Not explicitly listed in the developer documentation and on the web. A windows 10 home license is $140 and the mobile app is available for free for iOS and Android.	Not explicitly listed in the developer documentation and on the web. The cheapest Bixby supported smartphone (i.e. Galaxy S8) is currently $450 on Amazon.
Training cost	Cost of training to the users of the system	Medium	Developers documentation, access to the SDK, and testing environments are all free	Not explicitly listed in the developer documentation and on the web	Not explicitly listed in the developer documentation and on the web	Not explicitly listed in the developer documentation and on the web.

Installation and implementation cost	Cost of installation and implementation of the product	Medium	Dialogflow offers a free tier of support as long as the chatbots usage is less than a certain limit, which is up to 180 requests per minute. Higher usage chatbots can purchase additional computing time for a nominal fee.	The Free AWS subscription offers one million lambda requests and 750 of elastic compute times per month to run Alexa skills. Once those thresholds are reached, it's possible to purchase additional computing resources as needed	Not explicitly listed in the developer documentation and on the web	Not explicitly listed in the developer documentation and on the web.
Maintenance cost	Maintenance cost of the product	Medium	Dialogflow offers a free tier of support as long as the chatbots usage is less than a certain limit, which is up to 180 requests per minute. Higher usage chatbots can purchase additional computing time for a nominal fee.	The Free AWS subscription offers one million lambda requests and 750 of elastic compute times per month to run Alexa skills. Once those thresholds are reached, it's possible to purchase additional computing resources as needed	Not explicitly listed in the developer documentation and on the web	Not explicitly listed in the developer documentation and on the web.

Upgrading cost	Cost of upgrading of the product when new version will be launched	Medium	Google regularly updates the Google Assistant	Alexa hardware automatically updates and offers updates for mobile applications.	Not explicitly listed in the developer documentation and on the web	It appears new versions of Bixby are announced in tandem with new smartphone announcements. New versions of Bixby are backwards compatible with a small subset of older devices.
Cost of hardware	Cost of machinery used to support the system, including processor, memory and terminals	Medium	Dialogflow offers a free tier of support as long as the chatbots usage is less than a certain limit, which is up to 180 requests per minute. Higher usage chatbots can purchase additional computing time for a nominal fee.	The Free AWS subscription offers one million lambda requests and 750 of elastic compute times per month to run Alexa skills. Once those thresholds are reached, it's possible to purchase additional computing resources as needed	Not explicitly listed in the developer documentation and on the web	Not explicitly listed in the developers documentation and on the web

Compatibility	Compatibility with the existing software and hardware	High	The Google Assistant is available on Android, iOS devices, and via Google Assistant supported devices	Alexa is compatible with numerous devices in addition to iOS and Android devices	Cortana is supported on a variety of devices, including Windows 10 PCs, Android and iOS as mobile app downloads, and a smart speaker. Cortana is also capable of running as a skill within Amazon's Alexa.	Bixby is a proprietary component of Samsung devices
Hardware platform	Hardware platform required to run the software	High	The Google Assistant is available on Android, iOS devices, and via Google Assistant supported devices	The list of hardware running Alexa goes from smart speakers, to microwaves, to the car, to toilets, and mountain bikes.	Cortana is supported on a variety of devices, including Windows 10 PCs, Android and iOS as mobile app downloads, and a smart speaker. Cortana is also capable of running as a skill within Amazon's Alexa.	Currently supported: Galaxy S9 / S9+ Galaxy S8 / S8+ Galaxy Note8 Galaxy Note9 Samsung Smart Fridge Samsung Smart TVs Future device: Galaxy home (Samsung smart speaker) Washing machines
Communication protocols	Communication protocols supported by the package	Medium	The Assistant SDK and API supports Node.js, Python, Go, C++, and Java.	JSON and lambda requests	XML and APIs	Javascript API

Network configuration	Network technology needed to run the software package e.g. LAN, WAN	Medium	A Google Assistant supported device with an internet connection	An Alexa supported device with an internet connection	A mobile or desktop device with an internet connection	A Samsung account and data network connection (e.g. Wi-Fi, cellular data network) are needed to use Bixby

Opinions-technical sources	Opinions about the software package from – Potential vendors/sales representatives – In-house experts – External consultants – Computer/IS trade magazines, software product leaflets	Medium	"Between Assistant and Siri, I always feel like Google's software pays better attention to what I'm saying because Assistant's slightly better at answering my follow-up questions. But Siri is sassier and has more personality, especially when we switch to speaking in Cantonese ... I'm convinced that Assistant is just as capable as Siri, and I would be happy should I decide to leave Apple and fully commit to Google." -Engadget "Its voice recognition is significantly more accurate, and it supports threaded conversations, so you can actually ask follow-up questions after making a query." - Engadget Bixby article	"I've come to regard Alexa as a sort of cloud-powered child. One moment she's hyper-capable and in the next, she can't grasp what I'm saying ... I have three Philips Hue bulbs in lamps around my home, and Alexa can always understand when I ask her to turn them on. Anything more complicated than that is a crapshoot: Sometimes she'll bring the brightness down to 50 percent, and sometimes she has no idea what I'm asking her. Same goes for leaving myself reminders. If I give Alexa a simple command like "remind me to buy eggs," she's fine. Almost	"Cortana was just as handy at transcribing my spoken notes and dictating driving directions to me as Google's own Assistant. I mean, if you take the time to slow down, enunciate and really spell out your request for Cortana, sure, the program is quite competent. ... Cortana was just as handy at transcribing my spoken notes and dictating driving directions to me as Google's own Assistant. I mean, if you take the time to slow down, enunciate and really spell out your request for Cortana, sure, the program is quite competent ... Cortana had tried hard to fulfill its role as my new digital assistant, but it's hit-or-miss reaction times, scattershot search results and	"In reality there's very little that Bixby Home does that's unique or useful" - Pocket Lint "Nobody wants an assistant they can't trust. And no matter how hard I tried, I couldn't put my faith in Samsung's Bixby ... After living with Bixby on the Galaxy S8 for a week, I couldn't wait to kick it to the curb ... When all of Bixby's features click -- when it listens and interprets what I'm saying properly and then translates that into a specific function -- it can feel like magic. But that's easily outweighed by the frustration I had getting it to work properly ... The bigger problem with Bixby: It just feels unnecessary because the Google Assistant is also installed on the Galaxy S8 and S8+" - Engadget

				anything more complex than that is hit-or-miss to the point where her mangled interpretations have become a running joke ... I don't mean to make Alexa sound bad; most of the time she handles my commands just fine, so long as they're reasonably simple. Every once in awhile, she's even able to remember the context around what I ask so that I didn't need to ask a slew of repetitive questions." - Engadget	shoehorned integration with the rest of my apps left me searching for more competent help. At this point, I can only hope that Cortana quickly outgrows its current skin and sheds it like a tarantula's carapace in search of a more spacious host on another device" - Engadget General queries submitted to Cortana were answered correctly 64% of the time, but improved by 7% compared to a research firms previous tests.	"Bixby isn't perfect, but it's slowly improving ... Bixby's ability to understand voice commands has improved exponentially over the last year." - Digital Trends "After finally spending some dedicated time with Bixby, I want to say I am sorry for being so harsh about it in past writing. Bixby is often compared to Apple's Siri, Google Assistant, Microsoft Cortana, and Amazon Alexa, but it is really not focused on being a voice-controlled search engine. Bixby will not always provide you with the next date your favorite team plays, the current score of a ballgame, or answers to trivia questions. It is an assistant that is designed to optimize the use of

						your hardware while also providing some of the other search capabilities of other assistants." - ZDnet

Completeness	It is defined as the degree to which software satisfies functional requirements	High	The Google Assistant has the second largest adoption on smartphones by U.S. adults second to Siri, integrates with external NLUs, permits some health related functionality, supports both voice and text based interaction methods, is available on a variety of different non-smartphone hardware platforms, and Google is actively engaged in healthcare related activities.	Alexa has the third largest adoption on smartphones by U.S. adults, integrates with external NLUs, does not permit health related functionality, only supports voice based interaction methods, is available on a variety of different non-smartphone hardware platforms, and Amazon is actively engaging in health related activities.	Cortana is tied for the smallest adoption on smartphones by U.S. adults, integrates with external NLUs, permits some health related functionality, supports both voice and text based interaction methods, is available on a limited number of non-smartphone hardware platforms, and Microsoft is actively engaged in health related activities.	Bixby is tied for the smallest adoption on smartphones by U.S. adults, does not integrate with external NLUs, does not have explicit information regarding health related functionality, Bixby supports voice based interaction methods but it is not known if Bixby supports text based methods at this time, is available on a limited number of non-smartphone hardware platforms, and Samsung is actively engaged in some health related activities.

Chapter 4 Supplemental Table 2: Hernia Coach health content surgery stages, topics, sub-topics, example training phrases, and query responses.

Surgery Stage	Topic	Sub-Topic	Example training phrases (i.e., what the user asks)	Response (i.e., the prototype query response to a question)
All surgery stages	Pain	Pain	What drugs am I on? Can I take more pain pills? I want a prescription refill	Everyone reacts to pain in a different way. Follow your doctor's directions and take your medications as instructed. If you have questions, are experiencing pain that will not go away, or have pain that gets worse, call your doctor or go to the hospital right away.
All surgery stages	Tobacco	Tobacco	Can I smoke?	If you smoke, be sure to tell your surgery team. You should quit smoking before surgery and use this as an opportunity to start living a smoke free life. Smoking increases the risk of complications before and after the operation.
All surgery stages	Surgery information	Urgent Concerns	Phrases related to fever, nausea, chills, strangulation, incarceration etc.	This could be a sign of a very serious problem. Please call your doctor or go to the hospital right away.
Pre-surgery	Surgery information	Anesthesia	Tell me more about the anesthesia What kind of anesthesia is used?	There are two types of anesthesia used during inguinal hernia surgery. One type is general anesthesia, where the patient sleeps during surgery. The other is local anesthesia, which numbs the area of your body near the hernia. Your doctor will choose the one that's best for you.
Pre-surgery	Surgery information	Anesthesia risks	Are there any risks to the anesthesia?	Anesthesia risks are influenced by your health. Your doctor will help answer questions about the risks before your surgery.
Pre-surgery	Surgery information	Difference between surgery types	What are the differences between open and laparoscopic surgery?	Laparoscopic surgery incisions are smaller than open surgery incisions. Patients who undergo laparoscopic surgery generally experience a shorter recovery time, less pain, and reduced risk of infection. Not all patients are able to

				undergo laparoscopic surgery.
Pre-surgery	Eating or drinking	Eating before surgery	When do I have to quit drinking before surgery? Is it ok to eat before surgery?	You can eat a diet that is normal for you before surgery. However, for safety you need to stop eating or drinking before surgery. You may be told to stop eating and drinking at midnight the night before surgery. Or for at least six hours before surgery. You can take your medications with a small sip of water.
Pre-surgery	Activity restrictions	Exercising before surgery	Can I exercise before surgery?	Avoid any activity that causes pain. Exercise is okay as long as it does not put undo force on the hernia causing it to pop out uncomfortably. Exercises such as heavy lifting or abdominal crunches may cause the hernia to pop out or cause pain.
Pre-surgery	Hernia information	Hernia causes	How did I get a hernia? Why did I get a hernia? What causes a hernia?	Inguinal hernias may be caused by many factors. Some examples include older age, obesity, sudden twists or pulls or strains, chronic straining when passing bowels or urine, family history of hernias, connective tissue disease, or pregnancy.
Pre-surgery	Hernia information	Hernia Incidence	How common are hernias? Who gets hernias?	Hernias are a very common health condition. About 20 million people worldwide undergo inguinal hernia surgery every year. An inguinal hernia may occur at any time from infancy to adulthood. However, it is far more common in males. A person’s lifetime risk is 27% for men and 3% for women.
Pre-surgery	Hernia information	Hernia signs and symptoms	Is it a hernia? How do I know if I have a hernia?	You might have an inguinal hernia if you have a bulge in your groin that is painful, burns, aches, or causes discomfort. These feelings might become worse when bending over, coughing, or lifting. You should consult your doctor to know for sure if you have an inguinal hernia.

Pre-surgery	Hernia information	Incarcerated/ Strangulated Hernia	What is an incarcerated hernia? What happens when my guts get into the hernia?	An incarcerated hernia happens when intestines become trapped inside the hernia and can not be pushed back. If blood flow is cut off to the intestines, this is called a strangulated hernia. This can be life-threatening. If this happens, you will need emergency surgery. If you think your hernia may be incarcerated or strangulated, call your surgeon or go to the hospital right away
Pre-surgery	Surgery information	Incision location	Where will I be cut open?	It depends on the type of surgery you're having. For open surgeries the surgeon will make an incision in the groin near the hernia. For laparoscopic surgery the surgeon will make small incisions in the abdomen and groin.
Pre-surgery	Surgery information	Laparoscopic surgery info	What is laparoscopic surgery? Tell me more about laparoscopic surgery	During laparoscopic surgery, the surgeon makes several small cuts in the lower abdomen. Then the surgeon inserts a laparoscope, which is a thin tube with a tiny video camera. The camera helps the surgeon repair the hernia. The surgeon may repair the hernia with mesh or sutures.
Pre-surgery	Surgery information	Length of stay	Is it same day surgery? How long will I be in the hospital?	Typically inguinal hernia surgery is a same day surgery. Most patients go home after having the surgery. Patients having a complicated surgery, or who experience concerning side effects, may need to spend a night or more in the hospital.
Pre-surgery	Surgery information	Mesh info	What is mesh? Can you tell me more about mesh?	The mesh used during surgery is made of a flexible material that is left between the layers of tissue in the abdominal wall. Mesh is generally used to reinforce the weakness of the abdominal wall and to prevent the hernia from coming back. The risk of a hernia coming back occurs half as often for patients who have their hernia repaired with mesh. However, each hernia is different and your surgeon may or may not choose to use mesh.

Pre-surgery	Surgery information	Mesh risks	Are there any risks with using mesh to repair a hernia?	In the past there have been some complications related to the mesh used to repair hernias. However, many complications related to surgical mesh that have been reported to the FDA have been associated with recalled mesh products that are no longer used. Talk to your surgeon to learn more.
Pre-surgery	Surgery information	Open surgery info	What is open surgery? Tell me more about open surgery	During open surgery the surgeon makes one large incision in the groin. Then they move the tissue stuck in the hernia defect back into the abdomen and reinforce the muscle wall with stitches. Usually the area of muscle weakness is reinforced with a synthetic mesh to provide additional support and promote healing.
Pre-surgery	Surgery information	Operation day	What happens on operation day?	Once you are in the operating room you will be given anesthesia, which will put you to sleep or make the hernia area numb. Then the surgeon will repair your hernia. After the surgery you will be taken to the recovery room where you will be monitored for a little while. If everything goes well, you will receive some discharge teaching and sent home. If you had some complications from the surgery, you may need to spend some more time in the hospital before going home.
Pre-surgery	Surgery information	Recurrence risk	What is the potential for hernia recurrence? How often do hernias come back?	About 1% to 17% of all hernia patients have their hernia return. The risk of a hernia coming back occurs half as often for patients who have their hernia repaired with mesh. Prior hernia repairs, smoking, obesity, or chronic cough can increase the risk of a hernia coming back.

Pre-surgery	Surgery information	Surgery need	Why do I need surgery? Does a hernia go away? Do I need to have surgery?	In some situations a surgeon may say you do not need surgery. However, a hernia will not go away by itself. The only way to fix a hernia is to have surgery. The weakness in your groin will increase with time. This can cause higher amounts of pain in other parts of your body. Sometimes intestines become trapped inside the hernia, can not be pushed back, and blood flow is cut off. This is called a strangulated hernia and can be life-threatening. If this happens, you will need emergency surgery.
Pre-surgery	Surgery information	Surgery Prep	How do I prepare for surgery?	To prepare for your surgery bring a list of your medications, insurance card, government identification, comfortable clothes, slip on shoes, and a list of questions. Leave your jewelry and valuables at home.
Pre-surgery	Surgery information	Surgery Risks	Is there a danger to have surgery? Are there any risks?	Inguinal hernia surgery does have some risks. The biggest risk are anesthesia complications. Other risks include infection, bleeding at the surgery site, damage to the tissues and organs near the surgery site, long term pain, and the hernia returning.
Pre-surgery	Hernia information	Truss/athletic support	Can I wear a jock strap? Do I need to wear a truss? Can I prevent surgery with a jock strap? If I wear a truss, do I need surgery?	Your surgeon may recommend that you wear a truss or athletic support undergarment. These types of undergarments apply gentle pressure on the hernia to help keep it in place. This may ease discomfort or pain. Only use these garments if directed by your surgeon because they can make your hernia worse.
Pre-surgery	Surgery information	Types of surgery overview	What are the different types of surgery? Tell me more about surgery options	Your surgeon will choose between two different types of surgery. One type is laparoscopic surgery. The other is open surgery. The surgeon usually selects a method depending on the size of the hernia and your health history.

Pre-surgery	Hernia information	What is a hernia	Help me learn more about hernias What is a hernia?	An inguinal hernia is a weakness or tear of your abdomen inside your groin. The bulge you might see is usually a loop of intestine going through the tear in your muscle. An inguinal hernia may not be dangerous but can become life threatening.
Post-surgery	Eating or drinking	Alcohol	Can I drink after surgery?	Do not drink alcohol for at least two days after surgery. Talk to your doctor before drinking alcohol again.
Post-surgery	Surgery information	Breathing exercises	What are the breathing exercises?	While recovering from surgery, take 5-10 deep breaths and hold for 3-5 seconds every hour. This helps reduce the chances of getting a blood clot, fluid in your lungs, and pneumonia.
Post-surgery	Eating or drinking	Eating after surgery	What can I eat after surgery? Are there foods I should eat after surgery? When can I eat after surgery?	Your doctor will tell you when you can eat again. A diet high in fiber, fresh fruits, vegetables, and lots of fluids will help reduce constipation. Constipation can be caused by pain medication or resting. This can put extra strain on your intestines. Ask your doctor if you might benefit from using a stool softener or laxative. If you are having a hard time passing gas or having a bowel movement, call your doctor or go to the hospital right away.
Post-surgery	Activity restrictions	General activity restrictions	How long should I rest after having surgery? When can I do chores?	Give yourself time to heal and rest after surgery. You can ask family members or friends to help with chores and errands. Some patients recover within a week after surgery, but most are able to return to normal activities about 2-4 weeks. Talk with your doctor before resuming activities or exercise requiring a lot of effort.
Post-surgery	Eating or drinking	High fiber foods	What are some recommend high fiber foods?	Foods high in fiber include beans, bran cereals, whole-grains, peas, fruit, sweet corn, broccoli, baked potatoes with skin, greens, and nuts.

Post-surgery	Activity restrictions	Lifting restrictions	Can I pick up my kid? When can I start carrying heavy things?	Avoid lifting anything for at least a few days after having surgery. Then start lifting only very light objects that are not heavier than ten pounds for the first 4-6 weeks after your surgery. For example, a gallon of milk weighs about ten pounds. If you start to feel pain while lifting, stop immediately. Ask your doctor when it is okay to start lifting heavier items.
Post-surgery	Activity restrictions	Return to bathing	When can I shower? Can I soak in the bathtub?	Depending on the type of surgery you had, you may be able to take short showers about two days after surgery. Do not take a bath until your stitches, staples, or steri-strips are removed. Ask your doctor how long you need to wait before bathing.
Post-surgery	Activity restrictions	Return to driving	Can I drive? When can I ride my motorcycle?	You should arrange for someone else to drive you to and from the hospital the day of surgery. After surgery, the effects of anesthesia can affect you for up to one week. This can impair your ability to drive safely. You should never drive while taking prescription pain medication. Depending on your type of surgery, driving may strain the wound where you had surgery. Make sure you talk to your doctor about when it's safe for you to drive.
Post-surgery	Activity restrictions	Return to intimacy	When is it safe to have sex again? When can I have sex with my partner?	Most sexually active patients before surgery are able to return to normal activity after about 14 days following surgery. Your physical comfort will be a good guide. Talk to your doctor before resuming sexual activity.
Post-surgery	Activity restrictions	Return to sports	Can I play [insert popular sport here]	After surgery you will not be able to play sports or do intense exercise for a few weeks. The type of surgery you had and the type of sports you play will affect how long you need to wait before starting again. You should ask your doctor when you can start exercising or playing sports.

Post-surgery	Activity restrictions	Return to walking	Is it ok for me to walk after surgery? When can I start walking?	Walking can help increase blood flow and help you heal after surgery. If possible, try to walk about once an hour. Make sure walking does not cause you pain or discomfort. Stop if you do not feel well.
Post-surgery	Activity restrictions	Return to work	Do I have to take time off of work? How long will I be off of work?	After surgery some people return to work within a week. Others may need 2-4 weeks before going back to their job. It depends on the type of surgery you had and the type of work you do. It’s important to remember that you’ll become tired more easily after having surgery. If your job involves a lot of physical activity, you may not be ready to go back for a few weeks. Talk with your doctor when it is ok to start working again and if you need forms filled out for your company.
Post-surgery	Surgery information	Scarring	Will I have a scar?	Your scar should heal in about 4 to 6 weeks and continue to fade. Protect your skin from the sun and rough clothing. Also follow your wound care instructions.
Post-surgery	Surgery information	Urgent concerns list	How do I know if I need to go to the hospital?	If you have a fever, are continuously vomiting, have wound swelling, wound redness, wound odor, the wound bleeding through the bandage, abdominal swelling, difficulty moving bowels or urine, high levels of pain, or pain that gets worse, call your doctor or go the hospital immediately.
Post-surgery	Surgery information	Wound care	How often do I need to change my bandages? How do I take care of my cut? How do I prevent an infection?	Follow your surgeon’s instructions on when to change the bandages. Always wash your hands before and after touching near your wound. Avoid wearing tight or rough clothes, which can rub your cut and make it harder to heal. A little bit of liquid draining from the wound is normal. If the bandage is soaked with blood, smells bad, or has a large amount of redness or swelling, call your doctor or go to the hospital right away.

Post-surgery	Surgery information	Wound swelling	My wound is swollen My wound is bumpy	Swelling over the incision is common after hernia surgery. It doesn't mean that the surgery was unsuccessful. To reduce swelling and pain, put ice or a cold pack on the area for 10-20 minutes at a time. Do this every 1-2 hours. Put a thin cloth between the ice and your skin. Call your doctor or go to the hospital if your wound is bleeding through bandages, has a large amount of swelling, feels firm, has an odor, or feels warm.

4.11 References

1 Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Int J Hum-Comput Interact* 1995;**7**:57–78. doi:10.1080/10447319509526110
2 Jurfasky D, Martin JH. *Speech and Language Processing*. Stanford: 2018. https://web.stanford.edu/~jurafsky/slp3/24.pdf
3 Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006;**8**. doi:10.2196/jmir.8.4.e27
4 Bickmore T, Giorgino T. Health dialog systems for patients and consumers. *J Biomed Inform* 2006;**39**:556–71. doi:10.1016/j.jbi.2005.12.004
5 45 CFR Parts 160 and 164 Standards for Privacy of Individually Identifiable Health Information; Final Rule. *Fed Regist* 2001;**65**.https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/privacyrule/prdecember2000all8parts.pdf (accessed 21 Jun 2019).
6 45 CFR Parts 160, 162, and 164 Health Insurance Reform: Security Standards; Final Rule. *Fed Regist* 2003;**68**:49.
7 The Office of the National Coordinator for Health Information Technology. Health IT: Advancing America’s Health Care. https://www.healthit.gov/sites/default/files/pdf/health-information-technology-fact-sheet.pdf (accessed 1 Jul 2019).
8 Technology Acceptance Model 3 and a Research Agenda on Interventions - Venkatesh - 2008 - Decision Sciences - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-5915.2008.00192.x (accessed 29 Apr 2019).
9 Kim J, Park H-A. Development of a Health Information Technology Acceptance Model Using Consumers’ Health Behavior Intention. *J Med Internet Res* 2012;**14**:e133. doi:10.2196/jmir.2143
10 What Is Natural Language Understanding? https://developer.amazon.com/alexa-skills-kit/nlu (accessed 18 May 2019).
11 Lewis JR. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. *Proc Hum Factors Ergon Soc Annu Meet* 1992;**36**:1259–60. doi:10.1177/154193129203601617
12 Chung AE, Griffin AC, Selezneva D, *et al.* Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis. *JMIR MHealth UHealth* 2018;**6**:e174. doi:10.2196/mhealth.9705
13 Bickmore TW, Trinh H, Olafsson S, *et al.* Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *J Med Internet Res* 2018;**20**:e11510. doi:10.2196/11510
14 Steps to Build a Custom Skill | Custom Skills. https://developer.amazon.com/docs/custom-skills/steps-to-build-a-custom-skill.html (accessed 9 Jan 2019).
15 Getting started with Cortana skills. https://docs.microsoft.com/en-us/cortana/skills/get-started (accessed 2 Feb 2019).
16 Actions on Google | Actions on Google. Google Dev. https://developers.google.com/actions/ (accessed 9 Jan 2019).
17 Goldenthal S, Portney D, Steppe E, *et al.* Assessing the feasibility of a chatbot after ureteroscopy. *mHealth* 2019;**5**.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6458345/ (accessed 30 Apr 2019).
18 Vaira L, Bochicchio MA, Conte M, *et al.* MamaBot: A System Based on ML and NLP for Supporting Women and Families During Pregnancy. In: *Proceedings of the 22Nd*

International Database Engineering & Applications Symposium. New York, NY, USA: : ACM 2018. 273–277. doi:10.1145/3216122.3216173
19 Crutzen R, Peters G-JY, Portugal SD, *et al.* An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. *J Adolesc Health* 2011;**48**:514–9. doi:10.1016/j.jadohealth.2010.09.002
20 Rosis F de, Novielli N, Carofiglio V, *et al.* User modeling and adaptation in health promotion dialogs with an animated character. *J Biomed Inform* 2006;**39**:514–31. doi:10.1016/j.jbi.2006.01.001
21 Bickmore TW. *Relational agents : effecting change through human-computer relationships*. 2003.http://dspace.mit.edu/handle/1721.1/36109 (accessed 19 Dec 2018).
22 Schulman D, Bickmore T, Sidner C. An Intelligent Conversational Agent for Promoting Long-Term Health Behavior Change using Motivational Interviewing. ;:4.
23 Fadhil A, Wang Y, Reiterer H. Assistive Conversational Agent for Health Coaching: A Validation Study. *Methods Inf Med* Published Online First: 22 May 2019. doi:10.1055/s-0039-1688757
24 Piazza M, Giorgino T, Azzini I, *et al.* Cognitive Human Factors for Telemedicine Systems. ;:5.
25 Lee M, Ackermans S, van As N, *et al.* Caring for Vincent: A Chatbot for Self-Compassion. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2019. 702:1–702:13. doi:10.1145/3290605.3300932
26 Barisevičius G, Coste M, Geleta D, *et al.* Supporting Digital Healthcare Services Using Semantic Web Technologies. In: Vrandečić D, Bontcheva K, Suárez-Figueroa MC, *et al.*, eds. *The Semantic Web – ISWC 2018*. Springer International Publishing 2018. 291–306.
27 Babylon Health. 2019.https://www.babylonhealth.com/product/ask-babylon (accessed 9 Jan 2019).
28 Babylon Health. 2019.https://www.babylonhealth.com/ai (accessed 9 Jan 2019).
29 Babylon by TELUS Health | TELUS. https://www.telus.com/en/health/ (accessed 6 Mar 2019).
30 Hörmann S, Doering S, Richter R, *et al.* Patients' need for information before surgery. *Psychother Psychosom Med Psychol* 2001;**51**:56–61.
31 Newell R, Ziegler LE, Stafford N, *et al.* The information needs of head and neck cancer patients prior to surgery. *Ann R Coll Surg Engl* 2004;**86**:407–10.
32 McGaughey J, Harrisson S. Understanding the pre-operative information needs of patients and their relatives in intensive care units. *Intensive Crit Care Nurs* 1994;**10**:186–94. doi:10.1016/0964-3397(94)90019-1
33 Sjöling M, Norbergh K-G, Malker H, *et al.* What information do patients waiting for and undergoing arthroplastic surgery want? Their side of the story. *J Orthop Nurs* 2006;**10**:5–14. doi:10.1016/j.joon.2005.09.014
34 De GJ, Abbotts S, Collins B, *et al.* Same day surgery: results of a patient satisfaction survey. *J Qual Clin Pract* 1996;**16**:165–73.
35 Ivarsson B, Larsson S, Lührs C, *et al.* Extended written pre-operative information about possible complications at cardiac surgery—do the patients want to know? *Eur J Cardiothorac Surg* 2005;**28**:407–14. doi:10.1016/j.ejcts.2005.05.006
36 Aasa A, Hovbäck M, Berterö CM. The importance of preoperative information for patient participation in colorectal surgery care. *J Clin Nurs* 2013;**22**:1604–12. doi:10.1111/jocn.12110
37 Pager CK. Randomised controlled trial of preoperative information to improve satisfaction with cataract surgery. *Br J Ophthalmol* 2005;**89**:10–3. doi:10.1136/bjo.2004.048637

38 Suhonen R, Leino-Kilpi H. Adult surgical patients and the information provided to them by nurses: A literature review. *Patient Educ Couns* 2006;**61**:5–15. doi:10.1016/j.pec.2005.02.012
39 Pieper BA, Sieggreen MY, Freeland BS, *et al.* Discharge information needs of patients after surgery. *J Wound Ostomy Cont Nurs Off Publ Wound Ostomy Cont Nurses Soc* 2006;**33**:281–9; quiz 290–1.
40 Bradshaw C, Pritchett C, Bryce C, *et al.* Information needs of general day surgery patients. *Ambul Surg* 1999;**7**:39–44. doi:10.1016/S0966-6532(98)00048-1
41 Yiu HYM, Chien W-T, Lui MH-L, *et al.* Information needs of Chinese surgical patients on discharge:a comparison of patients' and nurses' perceptions. *J Adv Nurs* 2011;**67**:1041–52. doi:10.1111/j.1365-2648.2010.05528.x
42 Galloway S, Graydon J. Uncertainty, symptom distress, and information needs after surgery for cancer of the colon. https://ovidsp.tx.ovid.com/sp-3.33.0b/ovidweb.cgi?QS2=434f4e1a73d37e8c163b3d0d52d889b103368af7d714733736581ed5dcdf04c2b1ee3a51911cf2b8271926549a89548cd66fc62a00111c8e98fa78fbc75c179ac98286330e78d93642e54a05be9b20792c53e86fe807fe342b7ae67c4705f330400637dd5869e0c43c24b4035638ddfadfac258e845bbd162bb5ccbcdc89f340c993bae02e59cdedf1005ba47857bc4421d073d362ab029345a3891a9c304bbc153f511d009633d65cc0f5875b9951dbe54569726a8baa52833305af25f321614252ba1fbdd82c63454b28e0adc0ef340f877469c24b649315452d27bbc3a6a0dd4eb02a905dfe503a5a689c50488f5da6eeb560b6bc7cb26d32808f7ba82625a89652af7c679153 (accessed 6 Jun 2019).
43 Henderson A, Chien WT. Information needs of Hong Kong Chinese patients undergoing surgery. *J Clin Nurs* 2004;**13**:960–6.
44 Moerenhout T, Borgermans L, Schol S, *et al.* Patient health information materials in waiting rooms of family physicians: do patients care? *Patient Prefer Adherence* 2013;**7**:489–97. doi:10.2147/PPA.S45777
45 Agarwal N, Hansberry DR, Sabourin V, *et al.* A Comparative Analysis of the Quality of Patient Education Materials From Medical Specialties. *JAMA Intern Med* 2013;**173**:1257–9. doi:10.1001/jamainternmed.2013.6060
46 Walsh TM, Volsko TA. Readability Assessment of Internet-Based Consumer Health Information. *Respir Care* 2008;**53**:1310–5.
47 Johansson K, Nuutila L, Virtanen H, *et al.* Preoperative education for orthopaedic patients: systematic review. *J Adv Nurs* 2005;**50**:212–23. doi:10.1111/j.1365-2648.2005.03381.x
48 Hansberry DR, Agarwal N, Shah R, *et al.* Analysis of the readability of patient education materials from surgical subspecialties. *The Laryngoscope* 2014;**124**:405–12. doi:10.1002/lary.24261
49 Protheroe J, Estacio EV, Saidy-Khan S. Patient information materials in general practices and promotion of health literacy: an observational study of their effectiveness. *Br J Gen Pract* 2015;**65**:e192–7. doi:10.3399/bjgp15X684013
50 Davis TC, Wolf MS. Health literacy: implications for family medicine. *Fam Med* 2004;**36**:595–8.
51 United States Government Accountability Office Report to Congressional Requesters. Medicare: Communications to Beneficiaries on the Prescription Drug Benefit Could Be Improved. http://www.gao.gov/new.items/d06654.pdf. (accessed 18 Jun 2019).
52 Mulsow JJW, Feeley TM, Tierney S. Beyond consent—improving understanding in surgical patients. *Am J Surg* 2012;**203**:112–20. doi:10.1016/j.amjsurg.2010.12.010
53 Özkan S, Mellema JJ, Nazzal A, *et al.* Online Health Information Seeking in Hand and Upper Extremity Surgery. *J Hand Surg* 2016;**41**:e469–75. doi:10.1016/j.jhsa.2016.09.006
54 Feliciano D, Hawn M, Strand N. Groin Hernia Repair Inguinal and Femoral.

American College of Surgeons 2018.
https://www.facs.org/~/media/files/education/patient%20ed/groin_hernia.ashx (accessed 31 Jan 2019).
55 Inguinal Hernia Repair Surgery Information from SAGES. SAGES. https://www.sages.org/publications/patient-information/patient-information-for-laparoscopic-inguinal-hernia-repair-from-sages/ (accessed 15 Mar 2018).
56 Cullen KA, Margaret H, Aleksander G. Ambulatory Surgery in the United States, 2006. *CDC - Natl Health Stat Rep* 2009;:28.
57 Nordin P, Zetterström H, Gunnarsson U, *et al.* Local, regional, or general anaesthesia in groin hernia repair: multicentre randomised trial. *The Lancet* 2003;**362**:853–8. doi:10.1016/S0140-6736(03)14339-5
58 McCormack K, Scott NW, Go PM, *et al.* Laparoscopic techniques versus open techniques for inguinal hernia repair. *Cochrane Database Syst Rev* 2003;:CD001785. doi:10.1002/14651858.CD001785
59 Beadles CA, Meagher AD, Charles AG. Trends in emergent hernia repair in the United States. *JAMA Surg* 2015;**150**:194–200. doi:10.1001/jamasurg.2014.1242
60 Hewitt DB, Chojnacki K. Groin Hernia Repair by Open Surgery. *JAMA* 2017;**318**:764–764. doi:10.1001/jama.2017.9868
61 Minimally-Invasive Hernia Repair | Hernia Surgery Center | Dartmouth-Hitchcock. https://www.dartmouth-hitchcock.org/hernia/minimally_invasive_hernia_repair.html (accessed 1 Jul 2019).
62 Inguinal hernia repair - How it's performed. UK NHS. 2017.https://www.nhs.uk/conditions/inguinal-hernia-repair/what-happens/ (accessed 1 Jul 2019).
63 Miner AS, Milstein A, Schueller S, *et al.* Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Intern Med* 2016;**176**:619–25. doi:10.1001/jamainternmed.2016.0400
64 Palanica A, Thommandram A, Lee A, *et al.* Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. *Npj Digit Med* 2019;**2**:55. doi:10.1038/s41746-019-0133-x
65 Cho E. Hey Google, Can I Ask You Something in Private? In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2019. 258:1–258:9. doi:10.1145/3290605.3300488
66 Amazon.com: MyChildren's Enhanced Recovery After Surgery: Alexa Skills. https://www.amazon.com/MyChildrens-Enhanced-Recovery-After-Surgery/dp/B07QB7PQYW (accessed 3 Jul 2019).
67 My Children's Enhanced Recovery After Surgery (ERAS) Alexa Skill. Voice.Health. https://www.voice.health/news/my-childrens-enhanced-recovery-after-surgery-eras-alexa-skill (accessed 3 Jul 2019).
68 Thokala P, Duenas A. Multiple Criteria Decision Analysis for Health Technology Assessment. *Value Health* 2012;**15**:1172–81. doi:10.1016/j.jval.2012.06.015
69 Ostrovsky A, Deen N, Simon A, *et al.* A Framework for Selecting Digital Health Technology. Cambridge, MA: : Institute for Healthcare Improvement 2014. www.ihi.org (accessed 7 Jan 2019).
70 Soriano R, Weberg D. Incorporating New Evidence from Big Data, Emerging Technoloyg, and Disruptive Practies into your Innovation Ecosystem. In: *Leadership for Evidence-Based Innovation in Nursing and Health Professions*. Burlington, MA: : Jones & Bartlett Publishers 2017. https://books.google.com/books?id=26DOCwAAQBAJ&lr=&source=gbs_navlinks_s (accessed 26 Jun 2019).

71 WHO | Health technology assessment. WHO. https://www.who.int/medical_devices/assessment/en/ (accessed 26 Jun 2019).
72 IDC. Always Connected - How Smartphones And Social Keep Us Engaged. IDC 2013. https://www.nu.nl/files/IDC-Facebook%20Always%20Connected%20(1).pdf (accessed 17 Jun 2019).
73 Demographics of Mobile Device Ownership and Adoption in the United States. Pew Research Center http://www.pewinternet.org/fact-sheet/mobile/ (accessed 19 Dec 2018).
74 Voicebot.ai. Smart Speaker Consumer Adoption Report - March 2018. Voicebot.ai 2018. https://voicebot.ai/wp-content/uploads/2018/10/voicebot-smart-speaker-consumer-adoption-report.pdf (accessed 17 Jan 2019).
75 Zotero | Your personal research assistant. https://www.zotero.org/ (accessed 13 May 2019).
76 Jadhav AS, Sonar RM. Evaluating and selecting software packages: A review. *Inf Softw Technol* 2009;**51**:555–63. doi:10.1016/j.infsof.2008.09.003
77 Health Information Privacy. HHS.gov. 2015.https://www.hhs.gov/hipaa/index.html (accessed 21 Jun 2019).
78 Voicebot.ai. Voice Assistant Consumer Adoption Report 2018. Voicebot.ai 2018.
79 Siri. Apple. https://www.apple.com/siri/ (accessed 19 May 2019).
80 Google Assistant, your own personal Google. https://assistant.google.com/ (accessed 19 May 2019).
81 Echo & Alexa - Amazon Devices - Amazon Official Site. https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011 (accessed 19 May 2019).
82 Personal Digital Assistant - Cortana Home Assistant - Microsoft. Microsoft Cortana Your Intell. Assist. 2019.https://www.microsoft.com/en-us/cortana (accessed 16 Jan 2019).
83 Samsung Bixby: Your Personal Voice Assistant | Samsung US. Samsung Electron. Am. /us/explore/bixby/ (accessed 19 May 2019).
84 Bixby Developers. https://bixbydevelopers.com (accessed 9 Jan 2019).
85 SiriKit - Apple Developer. 2019.https://developer.apple.com/sirikit/ (accessed 8 Jan 2019).
86 Google acquires API.AI, a company helping developers build bots that aren't awful to talk to | TechCrunch. https://techcrunch.com/2016/09/19/google-acquires-api-ai-a-company-helping-developers-build-bots-that-arent-awful-to-talk-to/ (accessed 15 Feb 2019).
87 Dialogflow. Dialogflow. https://dialogflow.com/ (accessed 18 May 2019).
88 Integrations | Dialogflow. https://dialogflow.com/docs/integrations (accessed 16 Jan 2019).
89 Dialogflow documentation | Dialogflow. Google Cloud. https://cloud.google.com/dialogflow/docs/ (accessed 23 Jul 2019).
90 Overview - Conversational components - Conversation design. https://designguidelines.withgoogle.com/conversation/conversational-components/overview.html (accessed 23 Feb 2019).
91 The Joint Commission: Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care: A Roadmap for Hospitals. Oakbrook Terrace, IL: : The Joint Commission 2010.
92 Flesch R. *How to write plain English : a book for lawyers and consumers*. 1st ed. New York: : New York : Harper & Row 1979.
93 Test your document's readability. https://support.office.com/en-us/article/test-your-document-s-readability-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2 (accessed 13 May

2019).
94 Ley P, Florio T. The use of readability formulas in health care. *Psychol Health Med* 1996;**1**:7–28. doi:10.1080/13548509608400003
95 Williamson JML, Martin AG. Analysis of patient information leaflets provided by a district general hospital by the Flesch and Flesch–Kincaid method. *Int J Clin Pract* 2010;**64**:1824–31. doi:10.1111/j.1742-1241.2010.02408.x
96 Paasche-Orlow MK, Taylor HA, Brancati FL. Readability Standards for Informed-Consent Forms as Compared with Actual Readability. *N Engl J Med* 2003;**348**:721–6. doi:10.1056/NEJMsa021212
97 Tableau - Business Intelligence and Analytics Software. https://www.tableau.com/ (accessed 2 Feb 2019).
98 Health C for D and R. Hernia Surgical Mesh Implants. https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ImplantsandProsthetics/HerniaSurgicalMesh/default.htm (accessed 15 Mar 2018).
99 Hewitt DB. Groin Hernia. *JAMA* 2017;**317**:2560–2560. doi:10.1001/jama.2017.1556
100 Inguinal hernia repair: MedlinePlus Medical Encyclopedia. https://medlineplus.gov/ency/article/007406.htm (accessed 15 Mar 2018).
101 Medtronic. Inguinal Hernia Repair Surgery Recovery. https://www.medtronic.com/us-en/patients/treatments-therapies/hernia-surgery/about-recovery.html 5/5 (accessed 5 Feb 2019).
102 Medtronic. Inguinal Hernia Surgery Procedure Repair Options. https://www.medtronic.com/us-en/patients/treatments-therapies/hernia-surgery.html (accessed 5 Feb 2019).
103 Discharge Instructions for Open Hernia Repair | Articles | Mount Nittany Health System. https://www.mountnittany.org/articles/healthsheets/2283 (accessed 24 Jan 2019).
104 Inguinal hernia - discharge Information | Mount Sinai - New York. Mt. Sinai Health Syst. https://www.mountsinai.org/health-library/discharge-instructions/inguinal-hernia-discharge (accessed 24 Jan 2019).
105 Abdominal Hernia Repair: What to Expect at Home. https://myhealth.alberta.ca:443/Health/aftercareinformation/pages/conditions.aspx?hwid=bo1696 (accessed 12 Feb 2019).
106 Can I eat or drink before an operation? nhs.uk. 2018.https://www.nhs.uk/common-health-questions/operations-tests-and-procedures/can-i-eat-or-drink-before-an-operation/ (accessed 14 Feb 2019).
107 Open Inguinal Hernia Repair (Herniorrhaphy, Hernioplasty) | Michigan Medicine. https://www.uofmhealth.org/health-library/hw170381 (accessed 24 Jan 2019).
108 General Surgery - Inguinal Hernia. https://general.surgery.ucsf.edu/conditions--procedures/inguinal-hernia.aspx (accessed 31 Jan 2019).
109 Grudin J, Pruitt J. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. ;:8.
110 Pruitt J, Grudin J. Personas: practice and theory. In: *Proceedings of the 2003 conference on Designing for user experiences - DUX '03*. San Francisco, California: : ACM Press 2003. 1. doi:10.1145/997078.997089
111 Hazen R. *Designing and Evaluating a Patient-Driven Application for Patients with Primary Brain Tumors*. 2016.
112 Kneale L. *Using Personal Health Records to Promote Patient Activation in the Homebound Older Adult Population*. 2017.
113 Sakata J, Zhang M, Pu S, *et al.* Beam: A Mobile Application to Improve Happiness and Mental Health. In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: : ACM 2014. 221–226.

doi:10.1145/2559206.2580931
114 Kickbusch IS. Health literacy: addressing the health and education divide. *Health Promot Int* 2001;**16**:289–97. doi:10.1093/heapro/16.3.289
115 Porter CE, Donthu N. Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. *J Bus Res* 2006;**59**:999–1007. doi:10.1016/j.jbusres.2006.06.003
116 Selwyn N, Gorard S, Williams S. Digital Divide or Digital Opportunity? The Role of Technology in Overcoming Social Exclusion in U.S. Education. *Educ Policy* 2001;**15**:258–77. doi:10.1177/0895904801015002002
117 Graves BA. Integrative Literature Review: A Review of Literature Related to Geographical Information Systems, Healthcare Access, and Health Outcomes. *Perspect Health Inf Manag AHIMA Am Health Inf Manag Assoc* 2008;**5**.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2500173/ (accessed 23 Jul 2019).
118 Powe BD. Health Information Seeking Among Rural African Americans, Caucasians, and Hispanics: It Is Built, Did They Come? *Nurs Clin North Am* 2015;**50**:531–43. doi:10.1016/j.cnur.2015.05.007
119 JoAnn SO, Underwood SM, editors. *Rural and Other Medically Underserved Populations, An Issue of Nursing Clinics of North America 50-3, E-Book*. Philadelphia, PA: : Elsevier 2015.
120 Dean CA, Geneus CJ, Rice S, *et al.* Assessing the significance of health information seeking in chronic condition management. *Patient Educ Couns* 2017;**100**:1519–26. doi:10.1016/j.pec.2017.03.012
121 OECD. *Skills Matter: Further Results from the Survey of Adult Skills*. OECD 2016. doi:10.1787/9789264258051-en
122 Carroll JM. Five reasons for scenario-based design. *Interact Comput* 2000;**13**:43–60. doi:10.1016/S0953-5438(00)00023-0
123 Campbell RL. WILL THE REAL SCENARIO PLEASE STAND UP? *ACM SIGCHI Bull* 1992;**24**:6–8. doi:10.1145/142386.1054872
124 Zayas-Cabán T, Marquard JL, Radhakrishnan K, *et al.* Scenario-based User Testing to Guide Consumer Health Informatics Design. *AMIA Annu Symp Proc* 2009;**2009**:719–23.
125 McDaniel JG. *Advances in Information Technology and Communication in Health*. IOS Press 2009.
126 Belden JL, Grayson R, Barnes J. Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating. Healthcare Information and Management Systems Society (HIMSS) 2009. https://mospace.umsystem.edu/xmlui/handle/10355/3719 (accessed 22 Jun 2019).
127 Howard T. Journey mapping: a brief overview. *Commun Des Q Rev* 2014;**2**:10–3. doi:10.1145/2644448.2644451
128 Temkin BD. Mapping The Customer Journey. Forrester 2010. https://www.forrester.com/report/Mapping+The+Customer+Journey/-/E-RES55987
129 Improving healthcare through the use of co-design - New Zealand Medical Journal. https://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2012/vol-125-no-1357/article-boyd (accessed 3 Dec 2018).
130 Bureau UC. U.S. Cities Home to 62.7% of Population but Comprise 3.5% of Land Area. U. S. Census Bur. https://www.census.gov/newsroom/press-releases/2015/cb15-33.html (accessed 23 Jul 2019).
131 Overview | Actions on Google. Google Dev. 2019.https://developers.google.com/actions/dialogflow/ (accessed 18 Jan 2019).
132 Conversation design | Actions on Google | Google Developers. 2019.https://developers.google.com/actions/design/ (accessed 18 Jan 2019).

133 Dialogflow. https://dialogflow.com/docs (accessed 18 May 2019).
134 Intents overview | Dialogflow. https://dialogflow.com/docs/intents (accessed 18 May 2019).
135 Policy Testing | Custom Skills. 2019.https://developer.amazon.com/docs/custom-skills/policy-testing-for-an-alexa-skill.html (accessed 18 Jan 2019).
136 JohnBucyMSFT. Cortana Skills Publishing Review Guidelines - Cortana. 2019.https://docs.microsoft.com/en-us/cortana/skills/skill-review-guidelines (accessed 16 Jan 2019).
137 Policies for Actions on Google | Actions on Google. Google Dev. 2019.https://developers.google.com/actions/policies/general-policies (accessed 18 Jan 2019).
138 Houts PS, Doak CC, Doak LG, *et al.* The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006;**61**:173–90. doi:10.1016/j.pec.2005.05.004
139 United States, United States, editors. *Research-based web design & usability guidelines*. Version 2. Washington, D.C: : U.S. Dept. of Health and Human Services : U.S. General Services Administration 2006.
140 Katz MG, Kripalani S, Weiss BD. Use of pictorial aids in medication instructions: a review of the literature. *Am J Health-Syst Pharm AJHP Off J Am Soc Health-Syst Pharm* 2006;**63**:2391–7. doi:10.2146/ajhp060162
141 Chips - Conversational components - Conversation design. https://designguidelines.withgoogle.com/conversation/conversational-components/chips.html#chips-requirements (accessed 23 Feb 2019).
142 List - Visual components - Conversation design. https://designguidelines.withgoogle.com/conversation/visual-components/list.html#list-guidance (accessed 23 Feb 2019).
143 Browsing carousel - Visual components - Conversation design. https://designguidelines.withgoogle.com/conversation/visual-components/browsing-carousel.html (accessed 23 Feb 2019).
144 Basic card - Visual components - Conversation design. https://designguidelines.withgoogle.com/conversation/visual-components/basic-card.html#basic-card-requirements (accessed 23 Feb 2019).
145 Jeffries R, Desurvire H. Usability Testing vs. Heuristic Evaluation: Was There a Contest? *SIGCHI Bull* 1992;**24**:39–41. doi:10.1145/142167.142179
146 Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;**27**:861–74. doi:10.1016/j.patrec.2005.10.010
147 Nielsen J. Enhancing the Explanatory Power of Usability Heuristics. Human Factors in Computing Systems 1994. http://delivery.acm.org/10.1145/200000/191729/p152-nielsen.pdf?ip=205.175.118.188&id=191729&acc=ACTIVE%20SERVICE&key=B63ACEF81C6334F5%2EF43F328D6C8418D0%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1557776379_0db379376c87fa64a91a92c602d7ec03 (accessed 13 May 2019).
148 Nielsen J. *Usability Engineering.* Cambridge, MA: : AP Professional 1993.
149 Nielsen J. Usability Inspection Methods. In: Nielsen J, Mack RL, eds. . New York, NY, USA: : John Wiley & Sons, Inc. 1994. 25–62.http://dl.acm.org/citation.cfm?id=189200.189209 (accessed 14 Nov 2018).
150 Guerrilla HCI: Article by Jakob Nielsen. Nielsen Norman Group. https://www.nngroup.com/articles/guerrilla-hci/ (accessed 14 Nov 2018).
151 Jaspers MWM. A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *Int J Med Inf* 2009;**78**:340–

53. doi:10.1016/j.ijmedinf.2008.10.002
152 Agnisarman S, Narasimha S, Madathil KC, *et al.* Toward a More Usable Home-Based Video Telemedicine System: A Heuristic Evaluation of the Clinician User Interfaces of Home-Based Video Telemedicine Systems. *JMIR Hum Factors* 2017;**4**:e11. doi:10.2196/humanfactors.7293
153 Georgsson M, Staggers N. Heuristic Evaluation of a mHealth Diabetes Self-Management System Using Disease. ;:6.
154 Atashi A, Khajouei R, Azizi A, *et al.* User Interface Problems of a Nationwide Inpatient Information System: A Heuristic Evaluation. *Appl Clin Inform* 2016;**7**:89–100. doi:10.4338/ACI-2015-07-RA-0086
155 Walsh L, Hemsley B, Allan M, *et al.* The E-health Literacy Demands of Australia's My Health Record: A Heuristic Evaluation of Usability. *Perspect Health Inf Manag* 2017;**14**.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5653954/ (accessed 14 May 2019).
156 Savoy A, Patel H, Flanagan ME, *et al.* Systematic Heuristic Evaluation of Computerized Consultation Order Templates: Clinicians' and Human Factors Engineers' Perspectives. *J Med Syst* 2017;**41**:129. doi:10.1007/s10916-017-0775-7
157 Hartzler AL, Izard JP, Dalkin BL, *et al.* Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. *J Am Med Inform Assoc* 2016;**23**:38–47. doi:10.1093/jamia/ocv101
158 Goodman LA. Snowball Sampling. *Ann Math Stat* 1961;**32**:148–70. doi:10.1214/aoms/1177705148
159 Nielsen J. Finding Usability Problems Through Heuristic Evaluation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: : ACM 1992. 373–380. doi:10.1145/142750.142834
160 Ebbe N, Hertzum JM, John BE. *The Evaluator Effect in Usability Tests*.
161 Molich R, Nielsen J. Improving a human-computer dialogue. *Commun ACM* 1990;**33**:338–48. doi:10.1145/77481.77486
162 10 Heuristics for User Interface Design: Article by Jakob Nielsen. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/ (accessed 28 Jun 2019).
163 Calak P. Smartphone Evaluation Heuristics for Older Adults. ;:138.
164 Marcus A, editor. *Design, User Experience, and Usability: Users and Interactions: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II*. Cham: : Springer International Publishing 2015. doi:10.1007/978-3-319-20898-5
165 Reynaga G, Chiasson S, van Oorschot PC. Heuristics for the evaluation of captchas on smartphones. In: *Proceedings of the 2015 British HCI Conference on - British HCI '15*. Lincoln, Lincolnshire, United Kingdom: : ACM Press 2015. 126–35. doi:10.1145/2783446.2783583
166 Schmorrow DD, Fidopiastis CM, editors. *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems: 8th International Conference, AC 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*. Cham: : Springer International Publishing 2014. doi:10.1007/978-3-319-07527-3
167 Marler S. Do you understand me? Code Girl. 2017.https://code.likeagirl.io/do-you-understand-me-39268b210bd8 (accessed 7 Feb 2019).
168 Scott K. Usability Heuristics For Bots. https://thekevinscott.com/usability-heuristics-for-bots/ (accessed 7 Feb 2019).
169 The User Experience of Chatbots. Nielsen Norman Group. https://www.nngroup.com/articles/chatbots/ (accessed 7 Feb 2019).

170 Remote Usability Tests: Moderated and Unmoderated. Nielsen Norman Group. https://www.nngroup.com/articles/remote-usability-tests/ (accessed 15 Mar 2018).
171 Microsoft Forms - Easily create surveys, quizzes, and polls. https://forms.office.com/ (accessed 23 Jul 2019).
172 Severity Ratings for Usability Problems: Article by Jakob Nielsen. Nielsen Norman Group. https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ (accessed 14 May 2019).
173 Ferreira JJ, Godinho C, Santos AT, *et al.* Quantitative home-based assessment of Parkinson's symptoms: The SENSE-PARK feasibility and usability study. *BMC Neurol* 2015;**15**:89. doi:10.1186/s12883-015-0343-z
174 Ayubi SUA, Parmanto B, Branch R, *et al.* A Persuasive and Social mHealth Application for Physical Activity: A Usability and Feasibility Study. *JMIR MHealth UHealth* 2014;**2**:e25. doi:10.2196/mhealth.2902
175 Sonderegger A, Sauer J. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Appl Ergon* 2010;**41**:403–10. doi:10.1016/j.apergo.2009.09.002
176 Stellefson M, Chaney B, Chaney D, *et al.* Engaging Community Stakeholders to Evaluate the Design, Usability, and Acceptability of a Chronic Obstructive Pulmonary Disease Social Media Resource Center. *JMIR Res Protoc* 2015;**4**. doi:10.2196/resprot.3959
177 Giunti G, Kool J, Rivera Romero O, *et al.* Exploring the Specific Needs of Persons with Multiple Sclerosis for mHealth Solutions for Physical Activity: Mixed-Methods Study. *JMIR MHealth UHealth* 2018;**6**. doi:10.2196/mhealth.8996
178 Brewer LC, Jenkins S, Lackore K, *et al.* mHealth Intervention Promoting Cardiovascular Health Among African-Americans: Recruitment and Baseline Characteristics of a Pilot Study. *JMIR Res Protoc* 2018;**7**:e31. doi:10.2196/resprot.8842
179 US Census Bureau. 2010 Geographic Terms and Concepts - Urban and Rural. https://www.census.gov/geo/reference/gtc/gtc_urbanrural.html (accessed 4 Feb 2019).
180 Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*. Amsterdam, The Netherlands: : ACM Press 1993. 206–13. doi:10.1145/169059.169166
181 Virzi RA. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Hum Factors J Hum Factors Ergon Soc* 1992;**34**:457–68. doi:10.1177/001872089203400407
182 Monk A, Monk A. *Improving your human-computer interface : a practical technique*. Hemel Hempstead, Herts.: : Hemel Hempstead, Herts. : Prentice Hall International 1993.
183 OnePlus 6T Tech Specs. https://www.oneplus.com/6t/specs (accessed 19 May 2019).
184 De Vito Dabbs A, Myers BA, Mc Curry KR, *et al.* User-Centered Design and Interactive Health Technologies for Patients. *Comput Inform Nurs CIN* 2009;**27**:175. doi:10.1097/NCN.0b013e31819f7c7c
185 Harte R, Glynn L, Rodríguez-Molinero A, *et al.* A Human-Centered Design Methodology to Enhance the Usability, Human Factors, and User Experience of Connected Health Systems: A Three-Phase Methodology. *JMIR Hum Factors* 2017;**4**:e8. doi:10.2196/humanfactors.5443
186 Gao M, Kortum P. The Relationship between Subjective and Objective Usability Metrics for Home Healthcare Devices. *Proc Hum Factors Ergon Soc Annu Meet* 2015;**59**:1001–5. doi:10.1177/1541931215591286
187 Jiang Y, Sereika SM, DeVito Dabbs A, *et al.* Using mobile health technology to

deliver decision support for self-monitoring after lung transplantation. *Int J Med Inf* 2016;**94**:164–71. doi:10.1016/j.ijmedinf.2016.07.012
188 Lee H-Y, Kim J, Kim KS. The Effects of Nursing Interventions Utilizing Serious Games That Promote Health Activities on the Health Behaviors of Seniors. *Games Health J* 2015;**4**:175. doi:10.1089/g4h.2014.0124
189 Kim J. Analysis of Health Consumers' Behavior Using Self-Tracker for Activity, Sleep, and Diet. *Telemed E-Health* 2014;**20**:552–558.
190 Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. *PLOS ONE* 2016;**11**:e0156164. doi:10.1371/journal.pone.0156164
191 Becker M, Kolbeck A, Matt C, *et al.* Understanding the Continuous Use of Fitness Trackers: A Thematic Analysis. 2017;:13.
192 Tavares J, Oliveira T. Electronic Health Record Patient Portal Adoption by Health Care Consumers: An Acceptance Model and Survey. *J Med Internet Res* 2016;**18**:e49. doi:10.2196/jmir.5069
193 Creative Commons — Attribution 2.0 Generic — CC BY 2.0. https://creativecommons.org/licenses/by/2.0/ (accessed 28 Jun 2019).
194 Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qual Health Res* 2005;**15**:1277–88. doi:10.1177/1049732305276687
195 Parikh R, Mathai A, Parikh S, *et al.* Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008;**56**:45.
196 Brodersen KH, Ong CS, Stephan KE, *et al.* The Binormal Assumption on Precision-Recall Curves. In: *2010 20th International Conference on Pattern Recognition*. 2010. 4263–6. doi:10.1109/ICPR.2010.1036
197 Walker MA, Litman DJ, Kamm CA, *et al.* PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*. Madrid, Spain: : Association for Computational Linguistics 1997. 271–280. doi:10.3115/976909.979652
198 Google Sheets: Free Online Spreadsheets for Personal Use. https://www.google.com/sheets/about/ (accessed 16 May 2019).
199 Cook VE, Ellis AK, Hildebrand KJ. Mobile health applications in clinical practice: pearls, pitfalls, and key considerations. *Ann Allergy Asthma Immunol* 2016;**117**:143–9. doi:10.1016/j.anai.2016.01.012
200 Wallace LS, Dhingra LK. A systematic review of smartphone applications for chronic pain available for download in the United States. *J Opioid Manag* 2014;**10**:63–8. doi:10.5055/jom.2014.0193
201 Rivera J, McPherson A, Hamilton J, *et al.* Mobile Apps for Weight Management: A Scoping Review. *JMIR MHealth UHealth* 2016;**4**. doi:10.2196/mhealth.5115
202 Lalloo C, Jibb LA, Rivera J, *et al.* "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. *Clin J Pain* 2015;**31**:557–63. doi:10.1097/AJP.0000000000000171
203 Subhi Y, Bube SH, Rolskov Bojsen S, *et al.* Expert Involvement and Adherence to Medical Evidence in Medical Mobile Phone Apps: A Systematic Review. *JMIR MHealth UHealth* 2015;**3**:e79. doi:10.2196/mhealth.4169
204 Wang W, Siau K. Trust in Health Chatbots. San Francisco, CA: 2018.
205 Wang W, Siau K. Trusting Artificial Intelligence in Healthcare. *AMCIS 2018 Proc* Published Online First: 16 August 2018.https://aisel.aisnet.org/amcis2018/TREOsPDS/Presentations/125
206 Holden RJ, Karsh B-T. The Technology Acceptance Model: Its past and its future in health care. *J Biomed Inform* 2010;**43**:159–72. doi:10.1016/j.jbi.2009.07.002
207 Jimison H, Gorman P, Woods S, *et al.* Barriers and drivers of health information

technology use for the elderly, chronically ill, and underserved. *Evid ReportTechnology Assess* 2008;:1–1422.
208 Song J, Zahedi F “Mariam.” Trust in health infomediaries. *Decis Support Syst* 2007;**43**:390–407. doi:10.1016/j.dss.2006.11.011
209 Ash JS, Berg M, Coiera E. Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. *J Am Med Inform Assoc* 2004;**11**:104–12. doi:10.1197/jamia.M1471
210 Montague ENH, III WWW, Kleiner BM. Trust in medical technology by patients and healthcare providers in obstetric work systems. *Behav Inf Technol* 2010;**29**:541–54. doi:10.1080/01449291003752914
211 Heathfield H, Pitty D, Hanka R. Evaluating information technology in health care: barriers and challenges. *BMJ* 1998;**316**:1959. doi:10.1136/bmj.316.7149.1959
212 Lippert SK, Davis M. A conceptual model integrating trust into planned change activities to enhance technology adoption behavior. *J Inf Sci* 2006;**32**:434–48. doi:10.1177/0165551506066042
213 Merritt SM, Ilgen DR. Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Hum Factors* 2008;**50**:194–210. doi:10.1518/001872008X288574
214 Dzindolet MT, Peterson SA, Pomranky RA, *et al.* The role of trust in automation reliance. *Int J Hum-Comput Stud* 2003;**58**:697–718. doi:10.1016/S1071-5819(03)00038-7
215 Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Hum Factors* 2004;**46**:50–80. doi:10.1518/hfes.46.1.50_30392
216 Li X, Hess TJ, Valacich JS. Why do we trust new technology? A study of initial trust formation with organizational information systems. *J Strateg Inf Syst* 2008;**17**:39–71. doi:10.1016/j.jsis.2008.01.001
217 Learn about conversation - Conversation design. https://designguidelines.withgoogle.com/conversation/conversation-design/learn-about-conversation.html (accessed 13 Feb 2019).
218 Write sample dialogs - Conversation design process - Conversation design. https://designguidelines.withgoogle.com/conversation/conversation-design-process/write-sample-dialogs.html# (accessed 11 Jun 2019).
219 Alexa Design Guide | Alexa Design Guide. 2019.https://developer.amazon.com/docs/alexa-design/intro.html (accessed 18 Jan 2019).
220 Write Out a Script with Conversational Turns | Alexa Skills Kit. https://developer.amazon.com/docs/alexa-design/script.html (accessed 11 Jun 2019).
221 Bucy J. Principles of Cortana skill design - Cortana. https://docs.microsoft.com/en-us/cortana/skills/design-principles (accessed 11 Jun 2019).
222 Writing Dialog - Bixby. 2019.https://bixbydevelopers.com/dev/docs/dev-guide/design-guides/writing (accessed 9 Jan 2019).
223 Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc JAMIA* 2018;**25**:1080–8. doi:10.1093/jamia/ocy052
224 White M, Adams J, Heywood P. How and why do interventions that increase health overall widen inequalities within populations? *Health (N Y)* 2009;:12.
225 Lorenc T, Oliver K. Adverse effects of public health interventions: a conceptual framework. *J Epidemiol Community Health* 2014;**68**:288–90. doi:10.1136/jech-2013-203118
226 Demographics of Mobile Device Ownership and Adoption in the United States | Pew Research Center. Pew Research Center 2019. https://www.pewinternet.org/fact-sheet/mobile/ (accessed 11 Jul 2019).
227 Errors - Conversational components - Conversation design.

https://designguidelines.withgoogle.com/conversation/conversational-components/errors.html#errors-no-match (accessed 23 Feb 2019).
228 Google Design. Prototyping Voice Experiences: Design Sprints for the Google Assistant (Google I/O'19). https://www.youtube.com/watch?v=FrGaV4wzgeo&list=PLOU2XLYxmsILVTiOlMJdo7RQS55jYhsMi&index=50&t=0s (accessed 12 Jun 2019).
229 Grice HP. Logic and conversation. In: Cole P, Morgan JL, eds. *Syntax and Semantics*. Academic Press 1975. 41–58.
230 Deriu J, Rodrigo A, Otegi A, *et al.* Survey on Evaluation Methods for Dialogue Systems. *ArXiv190504071 Cs* Published Online First: 10 May 2019.http://arxiv.org/abs/1905.04071 (accessed 20 May 2019).
231 Vincent J. Yep, human workers are listening to recordings from Google Assistant, too. The Verge. 2019.https://www.theverge.com/2019/7/11/20690020/google-assistant-home-human-contractors-listening-recordings-vrt-nws (accessed 11 Jul 2019).
232 Washenko A. Siri records fights, doctor's appointments, and sex (and contractors hear it). Ars Tech. 2019.https://arstechnica.com/gadgets/2019/07/siri-records-fights-doctors-appointments-and-sex-and-contractors-hear-it/ (accessed 30 Jul 2019).
233 McKay T. Amazon's Human Helpers Are Quietly Listening in on Some Alexa Recordings. Gizmodo. https://gizmodo.com/amazons-human-helpers-are-quietly-listening-in-on-some-1833960052 (accessed 10 Apr 2019).

CHAPTER 5. DISSERTATION CONCLUSION

5.1 Aim 1. To identify the effect of patients collecting health data outside of the hospital on patient-clinician relationships and how patients engage in health information-seeking to understand their collected data.

In Aim 1, I systematically reviewed the literature to better understand the effect of primary care and surgery patients collecting patient-generated health data (PGHD) and patient-reported outcomes (PROs) on patient-clinician relationships. I also identified how these patients engage in health information-seeking with their clinicians to make sense of their collected data.

This work identified three key themes regarding PGHD and PROs. First, PGHD and PROs supported health awareness and communication for both patients and clinicians. These data increased patient self-awareness and motivation. Additionally, patients and clinicians had improved understanding of the patients' everyday lives. Second, patients and clinicians often differed in their perspective regarding how to integrate these data into clinical care. Patient understanding of their collected PGHD was varied, where some were able to interpret the PGHD and act upon it, and others sought interpretation assistance from their clinicians. While some clinicians assisted with data review, others were less inclined to collaborate with patient data. Third, the technology used to collect

PGHD promoted or hindered patient-clinician data collaboration. PGHD technology supporting collaboration incorporated trends and summary features and to aid patient and clinician understanding of the data. PGHD technology hindering collaboration duplicated physician work, lacked the flexibility to collect data meaningful to patients, and had the potential to reduce the frequency of patient-clinician face-to-face interactions.

Overall, this work generated an improved understanding of patient and clinician experiences using PGHD and PROs to better patient health conditions. Additionally, this research identified that some patients encounter barriers trying to make sense of their recorded data, which prompted them to engage in health information-seeking to make sense of their collected data. This finding was the major motivation for my subsequent work, which was to identify, design, develop, and evaluate a tool to aid independent surgery patient health information-seeking.

5.2 Aim 2. To identify how health dialog systems can support inguinal hernia surgery patient health information-seeking.

In Aim 2, I engaged both patients and clinicians in participatory design sessions to identify inguinal hernia surgery patient information needs and how health dialog systems (HDSs) can aid these patients' independent health information-seeking.

Inguinal hernia patients have information needs throughout their surgery journeys, spanning the time before, during, and after surgery. During their surgery journeys, patients consulted search engines, health organization websites, patient education materials, physicians, and family members for inguinal hernia health information. The majority of patients and clinicians identified that an HDS would have utility to aid patient information-seeking for non-urgent questions and/or for patients who do not desire to interact with their clinicians. Additionally, the participants expressed their desire that an inguinal hernia-focused HDS should be able to provide answers to questions about inguinal hernias and surgical repair in general, but not about their care as individual patients. In other words, the participants wanted the HDS to act as an alternate standardized patient educational material such as a brochure or leaflet. While the participants identified potential benefits for an inguinal hernia HDS, they also expressed concerns about using the HDS. Specifically, the participants were concerned about the HDS misunderstanding questions, providing inaccurate information, failing to identify situations requiring emergency medical care, and data privacy.

Overall, this work provided an improved understanding of inguinal hernia surgery patient information needs, how an HDS could support these patients' information-seeking, and the type of health content should be incorporated into an HDS. These findings were used to inform the design, development, and evaluation of a prototype HDS in Aim 3.

5.3 Aim 3. To design, develop, and evaluate a health dialog system prototype to aid inguinal hernia surgery health patient-information seeking.

In Aim 3, I designed, developed, and evaluated a patient-centered HDS prototype, called “Hernia Coach,” to aid inguinal hernia surgery patient information-seeking in three different stages. In the first stage, evaluated the top five most widely adopted smartphone-based DSs to identify the best-suited DS for building Hernia Coach as a “Skill” or “Action.” In the second stage, I generated Hernia Coach’s health content, generated personas to focus development efforts on Hernia Coach target users (i.e., inguinal hernia surgery patients), created scenarios to guide Hernia Coach’s evaluation activities, and developed Hernia Coach itself. In the third phase, I evaluated Hernia Coach by recruiting design experts to perform heuristic evaluations, patients to engage in usability testing sessions, and Hernia Coach’s ability to provide relevant information to participant queries.

I selected to build Hernia Coach as a Google Assistant “Action” because it had high levels of adoption amongst smartphone owners, supported multiple user interaction modalities (e.g., voice, screen, keyboard), and permitted a variety of health content visualization options. Hernia Coach’s health content had marginally better readability than the median identified patient education materials due to Google Assistant limitations regarding content length and health content visualization. Design experts applied an adapted version of Nielsen’s widely accepted heuristics to evaluate Hernia

Coach. These heuristics were effective for evaluating Hernia Coach's user interface but only somewhat effective for evaluating Hernia Coach's conversations with the design experts. All of the patient participants perceived Hernia Coach could serve as a potentially useful component of a patient's hernia surgery journey to aid health information-seeking activities. However, all of the participants also stated they did not foresee Hernia Coach replacing communication with their healthcare team entirely. Hernia Coach's query responses generally provided relevant information to the participant's queries. However, we identified in some instances Hernia Coach provided irrelevant information to queries Hernia Coach should have correctly answered or failed to inform the participants some questions were beyond its capabilities or knowledge. Additionally, we identified situations when the patient participants may not have recognized the irrelevant or incorrect information provided by Hernia Coach.

Overall, this work demonstrates HDSs have the potential to provide an innovative platform to facilitate surgery patient information-seeking and reduce the burden of clinicians answering common patient questions. The evaluation of Hernia Coach generated design recommendations concerning HDS scope and capabilities, evaluation, and functionality. Specifically, HDS designers and developers need to consider explicitly identifying their intended HDS scope and capabilities while creating the system. Then we recommend the designers and developers convey their HDS scope and capabilities to users, which will provide the users a better understanding of what they can accomplish when using the HDS. During heuristic evaluations, evaluators should consider adapting or creating heuristics accounting for both user interactions and conversations with the HDS. Finally, while Google Assistant “Actions” are a promising method for supporting surgery patient health information-seeking, there are limits to the Google Assistant’s capabilities, which designers and developers should consider when using this technology to create HDSs.

5.4 Limitations and Future Work

Limitations

The work in this dissertation had a number of limitations. I recruited most of the patient participants using flyers and posts to online forums such as Craigslist and Reddit. A minority of patients were recruited by directly approaching them in a hernia surgery clinic. While a substantial effort was made to bolster recruitment efforts in the hernia clinic, I encountered barriers preventing me from accomplishing this goal. Overall the patient recruitment methods could have resulted in self-selection bias, confirmation bias, and recall bias. The patients could have had a pre-disposition for interest in HDSs, had already experienced inguinal hernia surgery, and mistakenly recalled their perceptions of their pre-surgery experiences. Additionally, a minority of patients reported complications related to their surgery experiences in Aim 2, which resulted in the design, development, and evaluation activities in Aim 3 focusing less on patient complications. The clinician population I recruited in Aim 2 was also small, primarily consisting of physicians from the same clinic. However, I was purposeful in my decision to stop recruitment after no new concepts emerged during the qualitative analysis. In sum, given the sample and geographic location of this dissertation, the findings may not be generalizable to other inguinal hernia patients.

Future Work

In future work, I recommend researchers identify the information needs and opportunities for HDS support for inguinal hernia surgery patients who have experienced complications or recurrence, other types of hernia surgery patients, and the caregivers supporting surgery patients. The participatory design methods I employed in Aim 2 could inform the future needs assessment activities involving these additional stakeholders. Additionally, I suggest future research identify which questions patients and clinicians perceive are not well suited for Hernia Coach. Furthermore, I recommend investigating how the Enhanced Recovery After Surgery framework could be incorporated into Hernia Coach to improve the effectiveness of Hernia Coach. The findings from the additional needs assessment research could then be incorporated into Hernia Coach and evaluated guided by the methods I selected in Aim 3. While conducting the evaluation activities, I recommend revisiting the adapted version of Nielsen's Heuristics I generated for the design experts to apply in Aim 3. Specifically, I recommend incorporating H. Paul Grice's "Cooperative Principle" as an addition to my adapted version of Nielsen's heuristics to identify if the four maxims constituting the Cooperative Principle are effective for evaluating human-DS conversations. Second, I also recommend researchers consider how to improve the focus of the adapted heuristics on healthcare to improve the evaluation of HDSs. Third, I recommend research identify alternate methods for creating health content. For example, patient-clinician clinical encounters, electronic messaging, or phone calls could be analyzed to identify patient health information-seeking before, during, and after surgery. Fourth, I recommend researchers identify if clinicians desire HDSs content be adapted to account for practice variation or if they prefer to incorporate generic health content. Fifth, I recommend researchers identify how patients and clinicians perceive the utility of generic health context versus practice based or individual patient focused content. Sixth, I recommend research identify the effectiveness of HDSs to support patient health information-seeking compared to traditional patient education materials. Finally seventh, I recommend researchers conduct a pilot implementation with a copy of Hernia Coach with partnering physicians or clinics after the additional needs assessment work has been completed. During the pilot implementation, I recommend adapting the health content generated for Hernia Coach in Aim 3 to align the practice norms of the partnering physicians or clinics. Additionally, I suggest researchers assess Hernia Coach's query responses using the methods I outlined in Aim 3 to continue quantifying Hernia Coach's ability to provide relevant health information to surgery patients. The findings from the pilot implementation could be used to further improve Hernia Coach's design.

5.5 Contributions

In this dissertation, I first in Aim 1 contributed an improved understanding of how patients collecting data outside of clinical settings impacts patient-clinician relationships and how patients engage in health information-seeking to make sense of their collected data. I identified the data collected by patients supported patient-clinician communication. However, the specific technologies used by patients had varying abilities to support patient-clinician collaboration. Additionally, I recognized some patients could encounter difficulties making sense of their collected data, which prompted them to seek interpretation assistance from their clinicians, which their clinicians may not be able to provide.

Second, in Aim 2 I also contributed to the body of research assessing surgery patient information needs. An extremely limited amount of research has been conducted to identify the information needs of inguinal hernia surgery patients. Of the literature I identified, it exclusively focused on the post-operative period. My research demonstrates these patients have information needs beyond after surgery, which includes before surgery and the day of surgery. I also contributed by identifying patients and clinicians perceive health dialog systems have the potential to support health information-seeking throughout a patient's surgery experiences. Additionally, I generated a novel methodology for engaging clinicians and patients in participatory design sessions to inform the creation of patient-centered information technology to support patient information-seeking. Researchers have identified most mobile health apps and the health content incorporated into the apps were not developed with the input of stakeholders or medical experts. My methodology directly involves stakeholders such as patients and medical experts as co-creators in the design process. These methods could be used by other researchers to inform their patient-centered design activities.

Finally, in Aim 3 I demonstrated the potential for designing, developing, and evaluating HDSs, built as DS "Skills" or "Actions," to aid patient-information seeking. In particular, the methods I used to select a DS for developing Hernia Coach could be used by other researchers to select commercial software for patient-centered applications. For example, researchers developing an HDS called “Cocobot” to support caregivers of children with asthma used my research and methods to inform their DS selection for developing Cocobot. While reviewing related HDSs, I identified a very small number supporting surgery patient information-seeking and a slightly larger number to support other types of patient information-seeking. The existing literature demonstrated HDSs need to be more rigorously evaluated and the design and development efforts detailed more thoroughly. My research contributes to a greater understanding of how HDSs can be built using DSs by detailing the design and development process so other researchers can use my efforts to inform future work. Furthermore, I contributed by showing that Nielsen’s Heuristics can be used to evaluate DS user interfaces and that there are opportunities to improve the evaluation of DS conversations. Finally, I also contributed by extending an existing method to quantify information retrieval accuracy to assess how effective HDSs are at providing relevant health information.

5.6 Closing Remarks

The increasing use of DSs is fundamentally augmenting the way people interact with technology and search for information. Healthcare is no exception to this paradigm shift. As this technology gains greater utility and matures, patients are going to increasingly expect their hospital and clinicians to keep up with these trends. While we cannot predict what the future holds, this technology has the potential to significantly impact healthcare and human-computer interaction. I am excited to see how DSs and HDSs technologies grow, develop, and are adopted in the coming years.