# An Empirical Study of Older Adult's Voice Assistant Use for Health Information Seeking

ROBIN BREWER, University of Michigan, USA
CASEY PIERCE, University of Michigan, USA
POOJA UPADHYAY, TBD, USA
LEESEUL PARK, University of Michigan, USA

Although voice assistants are increasingly being adopted by older adults, we lack empirical research on how they interact with these devices for health information seeking. Also, prior work shows how voice assistant responses can provide misleading or inaccurate information and be harmful particularly in health contexts. Because of increased health needs while aging, this paper studies older adult's (ages 65+) health-related voice assistant interactions. Motivated by a lack of empirical evidence for how older adults approach information seeking with emerging technologies, we first conducted a survey of n = 201 older adults to understand how they engage voice assistants compared to a range of offline and digital sources for health information seeking. Findings show how voice assistants were used for confirmatory health queries, with users showing signs of distrust. As much prior work focuses on perceptions of voice assistant use, we conducted scenario-based interviews with n = 35 older adults to study health-related voice assistant behavior. In interviews, participants engaged with different health topics (flu, migraine, high blood pressure) and scenario types (symptom-driven, behavior-driven) using a voice assistant. Findings show how conversational and human-like expectations with voice assistants lead to information breakdowns between the older adult and voice assistant. This paper contributes a nuanced query-level analysis of older adults' voice-based health information seeking behaviors. Further, data provide evidence for how query reformulation happens with complex topics in voice-based information seeking. We use our findings to discuss how voice interfaces can better support older adults' health information seeking behaviors and expectations.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; **Human computer interaction (HCI)**.

Additional Key Words and Phrases: voice assistants, interactive systems, search, older adults, health

Authors' addresses: Robin Brewer, rnbrew@umich.edu, University of Michigan, 105 S. State Street, Ann Arbor, Michigan, USA, 48108; Casey Pierce, University of Michigan, 105 S. State Street, Ann Arbor, Michigan, USA, 48108; Pooja Upadhyay, TBD, Evanston, Illinois, USA, 60201; Leeseul Park, University of Michigan, 105 S. State Street, Ann Arbor, Michigan, USA, 48108.


2160-6455/2021/1-ART1 $15.00
https://doi.org/10.1145/3484507

**ACM Reference Format:**
Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2021. An Empirical Study of Older Adult's Voice Assistant Use for Health Information Seeking. *ACM Trans. Interact. Intell. Syst.* 1, 1, Article 1 (January 2021), 39 pages. https://doi.org/10.1145/3484507

## 1 INTRODUCTION

Recent reports estimate that nearly 53 million people in the United States own a voice assistant in the form of smart speakers, a growth of nearly 14 million owners since 2018 [43]. Most of the voice assistants [1] on the market today (e.g., Amazon Alexa, Google Home) were initially purposed for handling task-based queries, such as setting alarms, reporting weather forecasts, and playing music. As technology companies assess future uses of voice assistants, they have recently considered these devices' potential for handling more complex information seeking queries in the form of health information seeking. For example, Amazon announced that its Alexa device became HIPAA compliant in April 2019, which would allow healthcare companies to transmit private patient health data with its Echo voice assistants [16]. Moreover, health organizations are interested in leveraging these devices to mitigate physician shortages by allowing users – especially those who cannot use the internet to search by traditional means – to seek advice and answers to common health-related questions [2, 71]. Although millions of people are rapidly adopting voice assistants in their homes [44], the technology is in its infancy for handling how users interact with and understand health information.

However, existing research on health information seeking online has primarily examined screen-based modalities, such as using search engines, social media, or other medical websites to search for health information [30]. Using voice assistants for health information seeking raises two significant issues. First, it is important to understand how people make sense of and might act upon the health information they receive from voice assistants, as this reflects their mental models (expectations and beliefs) of voice-based health information seeking. Second, the quality of health information delivered via voice assistants can have potentially significant risks, especially of people misinterpreting information or acting upon inaccurate or incomplete advice [11]. Moreover, prior work has noted, understanding how people perceive health information is important for future decision-making (e.g. [19, 28, 29].

As we learn more about how people use voice assistants for health information seeking, it is important to consider how these devices can be useful for older adults, who account for 37% of voice assistant users [1] and also regularly engage in health information seeking online with computers and smartphones [10]. Recent studies have explored how voice assistants can be useful for older adults, particularly for those who face accessibility challenges in using screen-based devices, those with vision or motor impairments, or those with literacy challenges, as visual forms of search can be overwhelming [3, 47, 49, 69]. Digital forms of health information seeking such as voice assistants can also support disabled communities, such as neurodivergent people who face challenges communicating with medical professionals in-person, empowering them to be in control of their own health questions and needs [14]. As older adults

[1] In this paper, we focus on voice-only assistants or voice assistants, which we define as devices that allow input and output solely by voice without the use of a visual screen component (e.g. Amazon Echo Show).

age and may require more information to manage their health, voice assistants could thus empower and support them aging-in-place within their own homes, as being able to search for information independently can help people feel more in control of the information they receive.

Considering that voice assistants introduce a new way of interacting with technology, there is reason to suspect that older adults may employ different health information seeking strategies. While voice assistants can facilitate easier information access for older adult users, additional empirical research is needed to explore how older adults use voice assistants for health information seeking. To address these concerns, we investigate the following research questions:

- **RQ1:** How do older adults' health information seeking practices with voice assistants differ from other communication channels?
- **RQ2:** How do older adults engage in health information seeking with voice assistants?
- **RQ3:** What are older adults' expectations when using voice assistants for health information seeking?

To address these questions, we first surveyed older adults about their health information seeking practices using voice assistants and other information seeking sources. As the initial findings showed unique patterns of mistrust and skepticism with voice-based search, we continued investigating older adults' health information retrieval practices using voice assistants. To do so, we conducted scenario-based interviews with 35 older adults where we asked participants to perform queries in two health-related scenarios. We analyzed the query content and structure of the query-response exchanges between the participant and voice assistant. During the interviews, we also asked the participants to reflect on their experience using voice assistants for health information seeking.

Our findings contribute to existing computing research in several ways. First, we contribute an empirical study of how people use voice assistants for health information seeking. We focus on older adult's voice-based health search because of increased age-related health and accessibility needs for those with age-related disabilities. Our findings provide insights as to how voice assistants are and can be used with populations that face access challenges to medical professionals and visual communication channels. Second, we investigate information seeking to explore an emerging information context - voice assistants. In doing so, we build upon existing work that primarily considers health information seeking practices of younger adult voice assistant users [11] and extend recent work on older adult's perceptions of voice-based health information seeking [41]. Third, we investigate older adults' expectations *and* behaviors with an off-the-shelf voice assistant in a voice-based context to inform the design of better voice-based interactive systems for conversational health information seeking.

## 2 RELATED WORK

In this section, we begin with a general overview of how people use voice assistants for search activities and specifically how older adults use voice assistants. We then review previous literature on health information and the role of voice assistants for health information seeking.

## 2.1 Searching by Voice

Research points to the rising popularity of voice search, reducing barriers for people who face difficulties accessing graphical web search such as people with visual impairments, motor impairments, and digital skill [20, 49, 77]. However, projects like Google's Project Euphonia [2] highlight how searching by voice does not work well for everyone, such as people with atypical speech patterns. Large scale voice query studies [20, 56, 78] and lab studies [24] identify voice query characteristics that may lead to automatic speech recognition (ASR) errors. These studies also identify how people reformulate (or modify) their search when encountering an error. Beyond ASR errors, prior work suggests there is a need to better understand users' mental models of information search with voice systems to improve informational utility beyond factual search [27, 32, 77].

Voice search is inherently different from visual search with keyboard input regarding length, keywords used, and reformulations. For instance, Guy (2016) found voice queries to be significantly longer than text queries. Voice search often uses terms common in spoken language, such as "what is", "in the", and "how is" [20, 56, 78]. Ryen et al (2015) show that these question-queries (e.g., Why is the sky blue?), when occurring in visual, text- based search, also tend to be 20-25% longer than keyword queries (e.g., blue sky reason), but that most search for information occurs with keywords [70]. Other researchers have studied voice query reformulations have found that repetition, refinement, hyperarticulation, simplification, and restarting as popular reformulation strategies [20, 40]. Jiang et al. (2013) studied reformulations specific to information search sessions, and reported a higher number of repetition, word substitution and reordering, with fewer instances of adding and removing words to improve relevance of results [24].

Yet, we lack research on how these reformulations change by context. There is a need to consider the efficacy of natural-language use across search task type [70]. Voice search handles fact-based or informational queries fairly well, including queries that can be answered by community question and answer websites [65], or by rich snippets on search result pages [20]. Research has identified an increasing prevalence of subjective queries that may not fit these criteria [64, 68]. In this paper, we extend research on voice search type by examining subjective voice search expectations and behaviors.

## 2.2 Older Adults and Voice Interfaces

Prior work on voice interfaces has investigated how older adults use voice-based online communities, Interactive Voice Response systems, and Spoken Dialogue Systems, showing their utility for an aging population [13, 73, 74]. For example, they can be an alternative to using a keyboard for information seeking for people with late-life vision loss [4] or for those with limited mobility [57]. Further, studies show that voice-based interfaces can be useful for any population for which access to a computer is a barrier [67].

With an increase in voice assistant ownership [43] and the popularity of using voice interfaces for information seeking [6], it is critical to understand how older adults are using them for their information needs. Trajkova and Martin-Hammond (2020) interviewed older adults who owned voice assistants, showing that many transitioned to

[2] https://sites.research.google/euphonia/about/

being non-users, but that they saw the potential for health-related uses [63]. Pradhan et al. (2019) study older adults' language expectations with voice assistants and show how they "personify and objectify" the devices [47] and that many use voice assistants for health-related information seeking [48].

The utility of voice assistants has also been studied for other marginalized populations. For example, Pradhan et al. (2018) analyzed product reviews of the Amazon Echo by people with disabilities and interviewed people with vision impairments using a voice assistant. Findings showed that use of voice assistants helped to increase independence and efficiency, but participants wanted richer voice-based interactions [49]. Other work has also studied how people with vision impairments use voice assistants [3, 69], showing people want to use them for complex tasks like sending email or longer information seeking queries. In this paper, we extend prior work on older adults and voice assistant use by studying their expectations and behaviors in a health information seeking context.

### 2.3 Voice Assistants and Health Information Seeking

Literature on health information seeking is vast, and primarily focuses on search behaviors on a computer when compared to in-person consultation with medical professionals. For example, Kanthawala et. al. (2016) found patients and caregivers search online for information about treatments, diet and exercise regimens, and follow up questions after meeting with medical professionals [26].

Voice assistants have not yet been studied in-depth in their ability to handle complex tasks. A recent analysis of log data using two voice assistants shows how they are predominantly used for task-based support (e.g. playing music, turning on lights) or general information seeking [6]. Yet, as the health industry searches for new and more convenient ways to engage people in sustainable and usable telehealth practices at home, researchers are looking to study their usage. Additionally, increased health demands (e.g. decreased mobility, decreased transportation access) limit the opportunity of querying medical professionals in-person, motivating people to engage in health information seeking online.

Communicating health information through voice interfaces presents a new set of communication challenges. For instance, prior work has studied speech recognition failures such as medical terms being mispronounced using voice assistants [45]. Misinterpreting health information from a voice assistant can have dire consequences. Bickmore et. al. (2018) conducted a study with adults (18+) on health information seeking with three common voice assistants. Participants were presented with several health scenarios, used an assistant to pose a question, and reflected on the information presented [11]. Medical professionals then evaluated the device output and participants' reflections, finding that many of the content interpretations could lead to potentially harmful or fatal outcomes if people perceived the information as credible enough to follow.

In this paper, we extend this initial work that examines how people engage in health information seeking with voice assistants by also focusing on how people interpret information they receive from the voice assistant. We are motivated by prior work to better understand the perceptions of one user group, older adults, as their health needs can be vast and access to medical professionals is often limited.

# 3 RQ1: HEALTH SEARCH BEHAVIORS

## 3.1 Methods

To better position how older adults use voice assistants amongst other online and offline sources for health information seeking, we used a third-party surveying company to recruit and survey older adults (ages 65+) in the United States on their search behaviors. All surveys were completed online. Extending existing research on health information seeking sources [35, 75], the survey defined and asked about the role of in-home and mobile digital assistants (e.g., Alexa, Siri, Cortana, Google Assistant).

The survey began by asking about search behaviors, specifically whether people had searched for health information online or offline within the last three months. If someone responded yes, they were asked to indicate what sources they used to do so (i.e., laptop/computer, cell phone, tablet, in-home voice assistant, mobile voice assistant, friend/family member, medical professional). We included friends, family members, and medical professionals as sources because prior work indicates how frequent older adults turn to these individuals for health advice and information seeking. For each source they used, respondents were asked to describe a specific example of something they searched for and whether the information was useful. Participants also ranked their preferred sources for finding health information.

Respondents who indicated using a computer, mobile device or in-home assistant to search for health information were asked to complete a modified version of the eHeals scale for digital health literacy [42]. The scale was modified by removing "on the Internet" and instead adding "when using a computer/laptop" and/or "when using a mobile or in-home digital assistant." Lastly, respondents answered a range of demographics questions (e.g., age, gender, income, ability, education level, race).

## 3.2 Participants

There were 201 respondents who completed the survey and passed attention checks in July of 2018. Respondents' ages ranged from 66 - 81 years old with an average age of 70 years old. The gender ratio reflects common patterns due to life expectancy and social roles in older adulthood with 144 women respondents and 57 men respondents. Racial diversity varied where 138 identified as Caucasian or white, 25 identified as Black or African American, 8 identified as American Indian or Alaska Native, 8 identified as Asian, 1 identified as Native Hawaiian or Pacific Islander, and 21 identified as mixed race or other. The annual household income of most participants was less than $60,000 (n = 129). Education status varied with nearly half receiving a college degree (n = 106) and a similar number of respondents with less than a high school diploma or some college education (n = 95).

## 3.3 Survey Analysis

We descriptively analyzed quantitative survey data. To analyze the open-ended responses in which participants described their recent searches/information seeking tasks for different sources, one member of the research team categorized each of the responses by source into different themes such as "treatment", "diagnosis", "advice". We primarily report on differences in themes by source, as this largely guides interviews on voice assistant use for health information seeking.

### 3.4 Findings: Health Info Seeking Across Sources

We provide a brief overview of older adults' qualitative survey responses to summarize their health search behaviors (RQ1). This survey allowed the research team to understand the ecosystem of resources people used for health information seeking and examples of how search patterns differed by source. From this survey, we observed how voice assistants were not used often and that people were often skeptical of the information provided from the devices.

From the survey data, we observe that older adults are using a variety of offline and digital sources for health information. More than half of participants (65.17%, n = 131) had searched or asked for health information within the last three months. The most common sources used (Figure 1) were laptop (n = 84), phone (n = 64), and medical professional (n = 51). These data differ from prior work [15] showing that older adults rely on doctors and other medical professionals for health information more than technology, although differences may exist because the survey was deployed online.

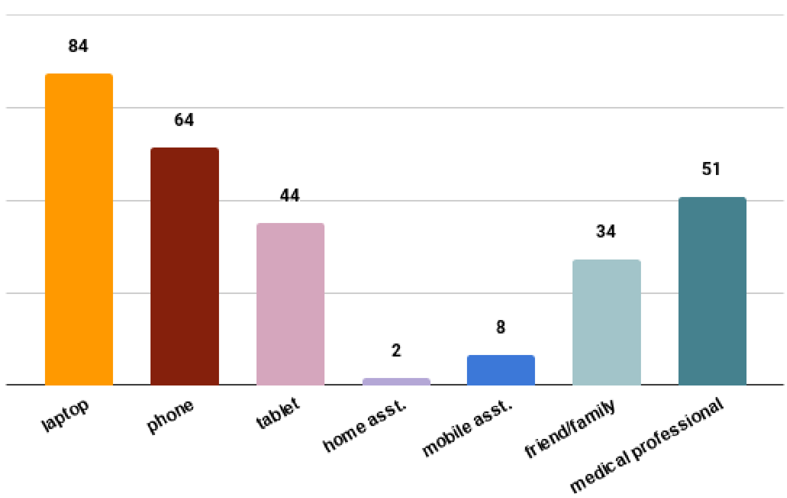


Fig. 1. Recent health search sources used

Older adult respondents were asked to rank their preferred sources for health information on a scale of 1-7 where 1 = most preferred and 7 = least preferred. Although phones and laptops were used most recently, we see that most older adult respondents prefer to engage in health information seeking with medical professionals (n = 95) and then laptops (n = 51) (Figure 2). This suggests access to medical professionals is a challenge that negatively impacts preferred methods of engaging with health information seeking, a challenge that is likely exacerbated by the COVID-19 pandemic.

The survey also asked respondents to describe the content of their most health search by source. Table 1 lists coded themes by source type and frequency of occurrence.

*3.4.1 Digital Tools for Health Information.* From a qualitative analysis of these themes (see Table 1), we find that older adults primarily use digital sources for objective health

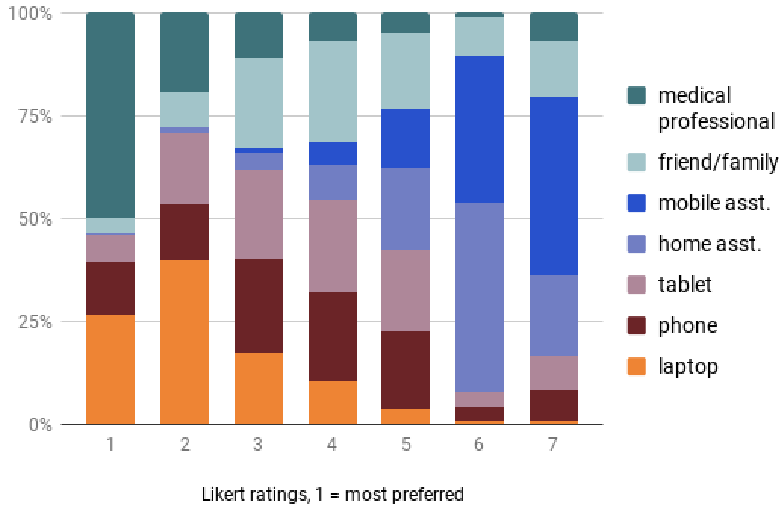


Fig. 2. Preferred health search sources

information and offline sources for subjective health information, aligning with prior work comparing behaviors of younger and older adults [37]. Participants frequently described how they used laptops, phones, tablets, and voice assistants to find information about conditions. For example, one person "*used my PC to research info on shoulder pain and how to heal shoulder injuries. I found some exercises and a heat treatment option. So far the exercises have been helpful*" and another "*was having shoulder pain from exercise and golf. I researched the possible cause of the pain by doing Google searches and researching YouTube.*"

Phones were used for similar condition research when not near the computer, in social settings, or needing to leverage smartphone features to improve their search. For example, one respondent "*used it when I was on the road and was having problems with my ankle sprain and felt I needed some more information*" and another used their phone in conversation to check "*symptoms of gout and bursitis for someone. We were all discussing the similarities and differences in order to do a 'home diagnosis' for persistent shoulder pain. The discussion also included basic arthritis as well as potential treatments.*" The smartphone camera was useful when needing to compare photos to images online. For example, one respondent "*used my smartphone to look for pictures of tick bites to compare the pictures to a picture my grandson sent me of a bite on his leg. With the information I gathered I was able to have some comfort that his bite was not a tick bite, and reassure my grandson.*"

Participants described using their tablet to supplement their search process and confirm responses from other sources. For example, one respondent "*I wanted information on recovery from a knee replacement and found duplicate information as I found on the computer*" and another "*googled dysautonomia, to find symptoms, as a doctor said*

**Table 1. Health search content themes by source. The frequency count is by survey response, not by participant.**

| Source | Search Content Themes |
|---|---|
| Laptop | Condition definition (30), treatment (13), nutrition/wellness (10), causes/symptoms (8), diagnosis (7), medication (7), doctor availability (7), insurance (3), advice for self (2) |
| Phone | Condition definition (23), medication (12), treatment (8), nutrition/wellness (7), causes/symptoms (5), doctor availability (5), advice for family (5), diagnosis (2) |
| Tablet | Condition definition (12), Treatment (10), causes/symptoms (6), nutrition/wellness (5), medication (4), doctor availability (3), advice for family (3) |
| Home Assistant | Condition definition (1), confirming information (1) |
| Mobile Assistant | Condition definition (3), confirming information (2), nutrition/wellness (1), medication (1) |
| Medical Professional | Medication (10), treatment (7), condition definition (7), causes/symptoms (6) diagnosis (3), nutrition/wellness (3), surgery details (3), advice for family (1), |
| Friend/Family Member | Advice for others (9), medication (5), advice for self (4), treatment (3), nutrition/wellness (2) condition definition (1), diagnosis (1), doctor availability (1) |

*many of my conditions are related to it. I was able to confirm that many things I was experiencing*".

Older adults primarily confirmed information using voice assistants such as when one person "*asked the same question [what results to expect from knee surgery]*" with a home assistant or "*was just looking to see if they would present any additional information than what I had received from the first searches I made*” with a mobile assistant.

*3.4.2 People for Health Information .* Non-digital sources such as medical professionals, family members, and friends not used for condition information as frequently as digital sources. Instead, older adults indicated leveraging other people for medication information, treatment, or advice for themselves or others. For example, one respondent described how they “*went to my cardiologist to further inquire about the side effects of blood thinners*” and another "*discussed Prevagen with a pharmacist today. She did not have much information. I will talk with a doctor tomorrow.*”

Friends and family were often sources of informal advice to supplement or replace, at times, advice from medical professionals. For example, one respondent "*talked with my friends about an itchy patch of skin. They thought it might be shingles. I went to the doctor and they were right.*” Another respondent discusses their "*clinical depression* *with my best friend. My antidepressants are not working very well and I spend days just laying in bed. My friend and I have discussed various ways for me to overcome this issue in my life and exercise seems to be a good answer*." Sometimes the information from non-medical professionals was unhelpful. For example, one respondent said, "*My daughter is pregnant and she had gone to the doctor and the heartbeat was 160 which is pretty fast and everyone was saying there's an old wives tale that a fast heartbeat means it's a girl but then a few days later she had a sonogram and it's a boy. So I guess the information from friends and family wasn't very useful*."

These examples show how older adults engaged in health information seeking online and offline. Although descriptive analysis suggests that voice assistants are not a common technology older adults use for health information seeking, qualitative data show interesting patterns with home and mobile voice assistant use. Participants seemed to primarily use voice assistants to confirm information they heard from other sources and/or after already searching using other sources. In contrast to recent work which suggests some 'digitally inactive' older adults trust voice assistants for information more than asking other people [48], our data signals a lack of trust with voice assistants for health information seeking. Further, voice assistants afford conversational, human-like interactions through voice input, yet are internet-enabled devices, which means mental models for health information seeking may indeed be a combination of human and digital source search patterns [47]. As such, we continue our research by further exploring older adults' expectations of voice assistants for different health information needs.

## 4 RQ2 AND RQ3: VOICE ASSISTANT BEHAVIORS AND EXPECTATIONS

### 4.1 Methods

We used a semi-structured elicitation interview approach [39, 72] to elicit participants' expectations of interacting with a voice assistant when searching for health information. To do this, we presented three information seeking scenarios to participants and asked how they would interact with the voice assistant to get information and address each prompt - one open-ended and two health scenarios. Participants then engaged in the search process with a Google Home Mini for each of the three scenarios. The Google Home Mini was chosen because of the popularity of the Google search engine and due to its performance at handling health-related queries compared to other voice assistants on the market (e.g., Amazon Alexa, Siri) [45]. Also, prior work comparing Amazon's Alexa, Apple's Siri, and the Google Assistant consistently ranks the Google Home device as a more neutral and favorable voice assistant for health information seeking [11]. No voice assistant applications (i.e. Google Actions [3]) were enabled during the searches. Additionally, we asked interview questions about participants' general technology use and any prior experiences with a voice assistant.

First, the researcher prompted participants with an open-ended information seeking scenario in which they could query about any topic to mimic a natural and unscripted interaction and to ensure that every participant would know how to interact with the Google Home Mini prior to engaging with the health scenarios. During this initial open-ended information seeking scenario, participants could ask clarification questions on using the device. Next, the participants were randomly assigned to one of three health

[3] https://developers.google.com/actions/

information seeking topics (flu, migraine, high blood pressure). We intentionally chose topics common to the general public [4] for the health information seeking scenarios and randomized topics to mimic the range of health information seeking needs. In many cases, our participants described having searched for or taken an action recently related to the topic presented in their interview, which validated their relevance. For the remainder of the interview participants received two scenarios to engage with the voice assistant: symptom-driven and behavior-driven (see Table 2). The first "symptom-driven" scenario had participants seek information to better understand the symptoms of their assigned health topic. The second "behavior-driven" scenario had participants seek information about what actions they should take concerning their assigned health topic. Participants informed the researcher when they were done with each search task to indicate when to proceed to the next task. Our findings herein only include analyses of the health scenarios (not the open-ended scenario). We used a scenario-based approach to avoid asking participants about health information they were uncomfortable disclosing to the researcher, although we recognized that the topic relevance could impact their query process. For instance, a participant with a recent personal experience with high blood pressure was randomly assigned to the high blood pressure topic and requested to switch topics for this reason.

After engaging in the three scenarios (open-ended, symptom-driven, behavior-driven), the interviewer replayed participants' interactions with the voice assistant with a voice recorder and asked reflective questions including: 1) how well did this response match their expectations, 2) whether they needed any additional information and where they would go for such information, 3) how well they thought the assistant understood them and why, and 4) how the evaluated the quality of the response they received.

**Table 2. Health Information Seeking Prompts**

| | **Prompt 1: Symptom-Driven** |
|---|---|
| **Flu** | You're not sure if you have the flu. You want to learn more about the symptoms. |
| **Migraine** | You have a migraine and want to know the key causes of migraines. |
| **HBP** | You're not sure if you are experiencing high blood pressure. You want to know more about the symptoms/causes. |
| | **Prompt 2: Behavior-Driven** |
| **Flu** | You now think you have the flu and want to know if you should get a flu shot. |
| **Migraine** | You have a migraine and want to know the best way to decrease the pain. |
| **HBP** | You want to know more about healthy diets to help you manage high blood pressure. |

*Note.* HBP - high blood pressure

[4] https://www.cnn.com/2018/12/21/health/health-questions-2018-google-explainer/index.html

Interviews lasted between 35-60 minutes. All interviews were conducted in person either on our university's campus or a local public library, where written consent was obtained before the start of the interviews. During the debrief after completing the interview, participants were reminded that the information given by the voice assistant was not from a medical professional. They were advised to speak to a medical professional before taking any medical advice from the search results. Interviews were audio recorded and each participant was compensated $20 in cash.

After the interview, participants were asked to complete a short demographics questionnaire after being interviewed. They were offered the choice to take the questionnaire on paper or online. The purpose of the questionnaire was to elicit demographic information about participants, and not to support causal claims. Using established scale items, the questionnaire also asked about computer self-efficacy [7], health self-efficacy [8], and online health information seeking efficacy [42].

### 4.2 Participants

We recruited participants using two research pools from local organizations with older adult members (e.g., senior centers). The primary recruitment was done through a university participant recruitment pool, which resulted in a sample with higher education and income than would be expected in the broader older adult population. To try to balance this, we used a research pool from a center for minority adults in a large city in the Midwestern United States. Eligible participants were at least 60 years old and able to travel to the interview site. In total, we recruited 35 people (female = 22, male = 12; ages 60 - 90; average age = 70.3 years old). Although all participants reported having used the internet for more than 10 years, a wide variety of voice assistant experience and internet use was reported in the interviews. Eight participants had used a mobile voice assistant, 12 used an in-home voice assistant or smart speaker, 5 had been exposed to voice assistants through advertising or commercials, 7 had observed others using a voice assistant, and 3 had no voice assistant experience or prior exposure. While prior experience could have affected how participants interacted with the voice assistant during the interviews, interviewing older adults with a range of voice assistant experience reflects the diversity of older adult's tech use. Participants also reported relatively high computer (M=3.73, SD=0.72), eHEALs (M=4.13, SD=0.83), and health (M=4.20, SD=.57) self-efficacy. One interview was excluded from this study due to the voice assistant not working.

### 4.3 Query Analysis

We coded participants' queries made in the interviews in three ways. First, we analyzed initial queries and reformulations as changes between two queries within a participant's search session using categories defined in existing literature (e.g., [25, 40, 53]) (Table 3: 3A). Second, we analyzed query characteristics such as natural-language related attributes, and multi-part queries [20] (Table 3: 3B). Third, we labelled voice assistant responses based on their relevance to the query [31, 33, 36, 66] (Table 3: 3C). To simplify analysis, we assigned binary classifications of voice assistant responses as being favorable (answer on topic) or unfavorable (all others). Note, this was a researcher-designated classification and not a rating by participants.

To categorize participants' queries and the voice assistants responses, we developed a codebook of query and response identifiers (Table 3). The codebook was iteratively revised based on group discussions and alignment with prior literature on information retrieval and voice interfaces. Two members of the research team coded the queries in each search session. Because the goals of this paper are to understand health information seeking practices, we focus on the queries and reformulations in the two health-related scenarios, analyzing 194 queries and 127 reformulations across 66 health search sessions. Any discrepancies in the codes were resolved by two other team members using the codebook and through discussion until agreement was reached.

We split queries and reformulations by voice assistant response types (Table 3:3C), similar to [21], but identify user strategies in response to system success and error. We analyzed trends in symptom-driven, behavior-driven search scenarios across topics (flu, migraine, high-blood pressure). See Table 4 for a summary.

We analyzed our data by interpreting it as two parts: 1) analysis of query reformulation strategies and outcomes 2) analysis of voice assistant responses. To analyze reformulation strategies and outcomes, we calculated the total number of reformulation types (n=127) that led to different voice assistant responses for each health topic (flu, high blood pressure, migraine) across 66 search sessions. We calculated the likelihood of using different reformulation strategies and their outcomes. To analyze effects of health tasks on reformulation behavior, we calculated the percentage of total and average of reformulations across health tasks (three health topics in symptom-driven and behavior-driven scenarios) and reformulation types (switch, specify, repeat, generalize, elaborate) used in the tasks.

To analyze voice assistant responses, we labelled the voice assistant's response type.To interpret how prior favorable and unfavorable voice assistant responses affect search behavior of older adult participants, we calculated the ratio of voice assistant responses that were followed by a certain reformulation type in a session. This showed how the five reformulation types (switch, specify, repeat, generalize, elaborate) were used after different voice assistant response types (on topic, wrong topic, can't, generic).

### 4.4 Interview Analysis

All interviews were transcribed by a professional transcription service and then validated by a member of the research team. Following Tracy [62], we used an iterative coding approach based to analyze the transcripts. In the initial round of coding, two researchers developed descriptive codes and identified common themes across the transcripts. Then, a final list of codes was developed based on the common themes from the primary cycle of coding and included codes about older adults' expectations when interacting with voice assistants for health information seeking (e.g., vetting device, lists, difficulty (re)formulating) and their mental models of health information seeking and voice-based information seeking (e.g., human expectations, personalization). Two researchers used these codes to do the second cycle of coding. Any discrepancies in coding were resolved by the two members of the research team who did not code the transcripts.

**Table 3. Classifications of queries and voice assistant responses**

| 3A - Reformulations | 3B - Query Details | 3C - Responses |
|---|---|---|
| **Initial**: User's first query | **Query length**: Average num. characters for voice queries 23.9 [20] | **Answer on topic** (favorable): Response includes information directly related to topical intent of immediate query |
| **Switch topic**: when users initiated a query with a topic different than the preceding query [identified as follow up questions, clarifying questions, further details in [50]] | **Multi-part**: Instances where users expressed their information need asking more than one question in a single query | **Answer wrong topic**: (unfavorable.): Response is related to an unrelated topic. |
| **Repeat**: when users iterate on the immediate prior query, reformulating with identical terms , similar terms, or substituting terms keeping intent same [24] | | **Generic** (unfavorable): Response is related to topic, but does not fulfil topical intent of query. |
| **Generalize**: when users broadened the scope of the information need of preceding query [53] | | **Can't** (unfavorable): Voice assistant does not provide a response, or suggests that it does not know or can't answer. |
| **Specify**: when users narrowed the scope of the information need of preceding query [53] | | |
| **Elaboration**: when users initiated a query with a topic that was similar to their previous query, but included additional details to clarify what has already been said. | | |

### 4.5 Findings: Voice Assistant Search Behaviors

In this section, we analyze and report findings from older adult's symptom- and behavior-driven health queries made to the voice assistant during the interview (RQ2). We review their query structure, search and query reformulation strategies, and differences by health topic and scenario type. Based on interview data, we find that participants

**Table 4. Summary of data across each task**

| Scenarios | Topics | # Sessions | Total Reformulations |
|---|---|---|---|
| Open ended (S1) | Varied | 33 | 87 |
| Health - symptom driven (S2) | Flu | 12 | 19 |
| | Migraine | 12 | 15 |
| | HBP | 9 | 25 |
| Health - behavior driven (S3) | Flu | 12 | 36 |
| | Migraine | 12 | 20 |
| | HBP | 9 | 12 |
| Totals across health scenarios | | 66 | 127 |

expected the voice assistant to respond to their requests as if they were engaging in conversation with another human.

*4.5.1 Conversational Query Structure.* From this data, we observe how participants engaged in requests that included objective and subjective health-related queries and queries that had multi-part questions. These types of conversational turns, which are typical in natural language, presented challenges for the voice assistant to respond to in ways that matched users expectations for a human-like conversational experience.

**Subjective Health Queries Difficult to Parse:** During the two health-related scenarios, participants made queries to the voice assistant that were objective and subjective health requests. For objective health queries (fact-based), participants asked for health information such as “What are some flu symptoms? (P35)” and “What medications would decrease the pain levels of a migraine headache? (P21)” The voice assistant responded to such objective questions stating facts referenced from reputable online sources (e.g., Mayo Clinic, WebMD). However, for subjective health queries (subject to interpretation, opinion-based), users asked for more personalized health information and often for behavioral advice on what they should do about a particular concern, rather than facts defining what the health concern is. For such subjective health queries, the voice assistant was unable to provide personalized feedback or recommendations for actions. For example, the voice assistant was unable to provide a response to subjective health information queries such as “Is it good to get a flu shot if you think you already have the flu? (P13)” or “What should I do if I suspect I have high blood pressure? (P11)” or “I’m experiencing intense anger because I’m in such pain. What can I do?(P9)”. In one scenario about whether to have the flu shot, the voice assistant could not answer 83% of participants’ subjective, advice-seeking queries. As illustrated in the Table 5, P35 makes a query for personal advice concerning an immunization. However, the voice assistant is unable to answer this type of subjective health query. As the exchange below details, the participant tries to ask the same query multiple times, but does not receive the desired response.

From the interview data, we observed how older adults wanted the voice assistant to be able to understand natural language and subjective requests. However, from analyzing query data, we find that need is not being met, in part because of the subjective nature of queries. In other instances, we observed how the voice assistant

**Table 5. Example of subjective health query**

| **P35 Query** | **Voice Assistant Response** |
|---|---|
| Hey, Google. I think I have the flu. Should I get a flu shot now? *(Initial)* | `I don't understand.` *(Can't answer)* |
| Hey, Google. Should I get a flu shot if I currently have the flu? *(Repeat)* | `On the website CDC.gov, they say, "CDC recommends a yearly flu vaccine for just about everyone six months and older, even when the viruses the vaccine protects against have not changed from the previous season."` *(Generic answer)* |

was not able to answer longer health queries. Next, we discuss one component of queries - multi-part queries (by clause complexity).

**Multi-Part Queries Unanswered:** Next, our data show how some participants asked multi-part queries, that is, queries that had more than one clause or question, similar to conversations one might have with another human. Although this conversational tendency was not frequent (n = 8), we observed some trends in the voice assistant's responses to multi-part queries. When parts of participants' queries could be separated into distinct independent clauses without losing context, the voice agent responded to both parts of the query (Table 6).

**Table 6. Multi-part query with independent clauses**

| **P5 Query** | **Voice Assistant Response** |
|---|---|
| Hey Google, I have the flu. What remedies to cure the flu, and should I take a flu shot? *(Initial)* | `On the website Universityhealthnews.com, they say, "Use lots of honey, lemon, and tea as home remedies for flu. While you're hydrating with water, keep a nice, hot cup of herbal tea at hand, and add organic honey and lemon, not only for taste and vitamin boost, but to soothe the sore throat."`<br><br>`Here's a summary from the website Mayoclinic.org, "Flu shot: Your best for avoiding influenza. Getting a flu shot often protects you from coming down with the flu. The Centers for Disease Control and Prevention recommends that everyone six months of age or older be vaccinated annually against influenza."` *(Answer on topic)* |

However, when components of the multi-part query could not be separated into distinct independent clauses, the voice assistant was not able to answer each clause in the question. Table 7 shows how the voice assistant only answered the first clause and Table 8 shows how the assistant answered the second clause. In the first example below, P14 is providing personal health information that is crucial to the device recommending appropriate medication. In the following example, P24 provides contextual information in the first query that included personal health information concerning having a kidney disease. However, the voice assistant ignored the first part of the query containing this key contextual information and answered the second part of the query that asked for migraine medication suggestions. Yet, the assistant suggested migraine medications that are harmful for those with kidney disease. We found that most multi-part queries (n= 5) were answered unfavorably by the voice assistant, which is a reason for concern since misinformation can be harmful or fatal [11]. This suggests that more research on parsing complex, multi-part queries is needed. We return to this argument in the discussion section.

**Table 7. Multi-part query with only first clause answered**

| **P14 Query** | **Voice Assistant Response** |
|---|---|
| Hey google. What foods are good for high blood pressure and what we should avoid? *(Initial)* | `I found 11 on the website Clevelandclinic.org. Here are the first seven Skimmed or 1 % milk, yogurt, Greek yogurt, lean meat, skinless turkey and chicken, low salt ready to eat cereals, cooked hot cereal, low fat and low salt cheeses and fruits. By the way if you want to see the whole list, you can check out the link I just sent to your phone.` |

**Table 8. Multi-part query with only second question answered**

| **P24 Query** | **Voice Assistant Response** |
|---|---|
| Hey, Google, what if I have kidney disease? What medications are available to me for treating a migraine headache? *(3rd query)* | `I found six on the website, Mayoclinic.org. Pain relievers, aspirin or ibuprofen, may help relieve mild migraines. Triptans, these medications are often used in treating migraines. Urgets, anti-nausea medication, opioid medications, and finally, glucocorticoids.` |

**Table 9. Reformulation after the voice assistant answered on topic**

| **P31 Query** | **Voice Assistant Response** |
|---|---|
| Hey Google, how do I treat a migraine? *(Initial)* | `According to Mayo Clinic, preventive and pain relieving medications can help manage migraine headaches.` *(Answer on topic)* |
| Hey Google, what type of medication for the migraine? *(Switch topic)* | `On the website rxlist.com they say, antidepressants, anti-convulsants, antihistamines, and narcotic pain relievers are all used in migraine therapy. Do you want a little more context?` *(Answer on topic)* |
| Yes. | `Over the counter migraine medications used to address pain include analgesics, NSAIDS, and caffeine including Aspirin, Naproxen, Ibuprofen.` |
| Hey Google, what are side effects of those drugs? *(Switch topic)* | `I found nine on the website medicalnewstoday.com. Here are the first five. Constipation, skin rash or dermatitis, diarrhea, dizziness, and drowsiness. Want me to tell you more?` *(Answer on topic)* |
| Yes. | `Okay, here's three more.Dry mouth, headache, and insomnia.` *(Answer on topic)* |

*4.5.2 Search Strategies.* In this section, we report findings on older adults' voice-based health search strategies. Although prior work often associates query reformulation with unsuccessful searches and speech recognition errors [21, 24, 58, 76], our data show older adult participants used query reformulations to extend and improve their search. Older adults reformulated their search queries when the voice assistant responded both favorably (on topic, 53%) and unfavorably (wrong topic, generic, can't answer, 47%). Even though older adults' initial queries were answered on topic in 76% of search sessions, most (75%) reformulated their queries to extend the search session. A similar number of reformulations were made after favorable (53%) and unfavorable voice assistant responses (47%). We observed how one participant (P31) continued to reformulate their search, even after the voice assistant answered on topic (see table 9).

Across a total of 127 reformulations, elaboration topics was the most popular strategy (61%), followed by repeating (18%), specifying (13%), generalizing (6%), and topic switching (2%) (Fig. 3). Participants elaborated when they received favorable responses (70% of elaborations) and unfavorable responses (30% of elaborations). Repeat was mostly used when response to prior query was unfavorable (>80%, Fig. 4).

Next, we present details of reformulation-level analysis showing how older adults engaged in topic switching and repeating queries for health information exploration.

| Reformulations in health topics, n=127 | | | VA response to reformulation | |
|---|---|---|---|---|
| | | | % favorable | % unfavorable |
| Elaboration | 61% | 77 | 53% | 47% |
| Repeat | 18% | 23 | 35% | 65% |
| Specify | 13% | 16 | 56% | 44% |
| Generalize | 6% | 8 | 38% | 62% |
| Topic Switch | 2% | 3 | - | - |

Fig. 3. n=127 Reformulations and voice assistant (VA) response outcomes

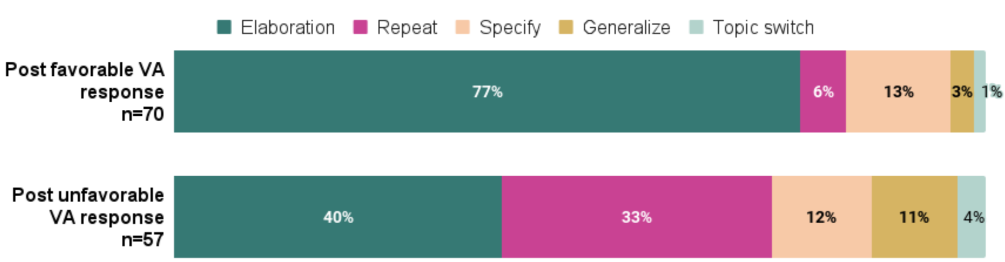


Fig. 4. Older adults' reformulation behavior based on voice assistant(VA) response to prior query

We do not focus on generalization, specification, or topic switching as they did not occur frequently. Figure 5 shows how reformulation strategies differed across task controls. We identified that apart from one outlier scenario (behavior-driven flu), the average reformulations and reformulation types were similar across the other health tasks (Fig. 5), as well as open ended task (Fig. 6).

**Elaboration for exploration and search improvement:** Participants elaborated their queries in 61% of all reformulations, using it extensively in the symptom and behavior-driven scenarios. Elaborations were equally likely to receive favorable (53%) or unfavorable responses (47%) from the voice assistant(Fig. 3). When a prior query was answered favorably, participants elaborated topics frequently, nearly 75% of the following reformulations (Fig. 4).

Participants elaborated upon their previous queries to incrementally learn about additional details of health conditions, such as with P8:

| Health tasks | | #sessions | % queries with favorable VA response | Avg reformulations per session | Total reformulations | Proportion of reformulation types used in task: Elaboration | Repeat | Specify | Generalize | Topic switch |
|---|---|---|---|---|---|---|---|---|---|---|
| Symptom driven | Flu | 12 | 68% | 1.58 | 19 | 79% | 11% | 5% | 5% | 0% |
| | Migraine | 12 | 89% | 1.25 | 15 | 80% | 7% | 7% | 7% | 0% |
| | HBP | 9 | **74%** | **2.78** | 25 | 64% | 20% | 12% | 0% | 4% |
| Behavior driven | Flu | 12 | **17%** | **3.00** | 36 | 44% | 28% | 14% | 14% | 0% |
| | Migraine | 12 | 63% | 1.67 | 20 | 45% | 20% | 25% | 0% | 10% |
| | HBP | 9 | 71% | 1.33 | 12 | 75% | 8% | 8% | 8% | |
| | | | | Total | 127 | | | | | |

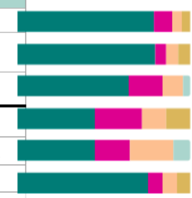

Fig. 5. Reformulation behavior across health tasks (red, bolded text corresponds to notable high or low figures)

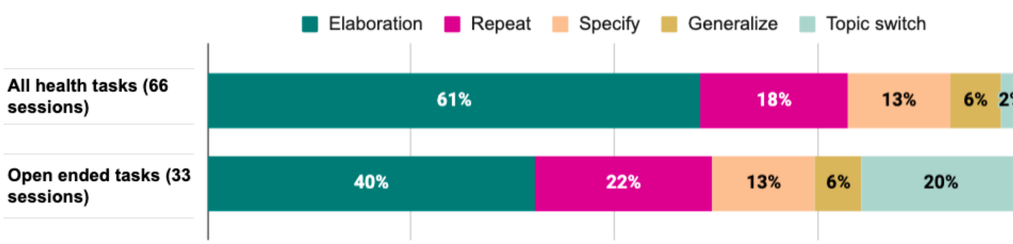


Fig. 6. Comparing reformulation behavior in health tasks to open ended task

**Table 10. Elaboration to learn about additional details of health conditions**

| P8 Query | Voice Assistant Response |
|---|---|
| Hey Google, what are the symptoms of flu? *(Initial)* | `According to Mayo Clinic symptoms include fever, chills, muscle aches, cough, congestion, runny nose, headaches, and fatigue.` *(Answer on topic)* |
| Hey Google, what do you recommend for treatment for the flu? *(Switch topic)* | `On the website mayoclinic.org they say, but in some cases your doctor may prescribe an antiviral medication such as Oseltamivir, Zanamivir.` *(Answer on topic)* |
| Hey Google, does taking Tylenol help with the flu? *(Switch topic)* | `Here's a summary from the website webmd.com. Pain relief. When it comes to pain relief, you have a couple of options and nonsteroidal anti-inflammatory drugs such as Ibuprofen or Naproxen sodium. Another option is Acetaminophen.` *(Answer on topic)* |

| | |
|---|---|
| Hey Google, how long does the flu usually last? *(Switch topic)* | `On the website everydayhealth.com they say, on average the flu typically lasts one to two weeks with severe symptoms subsiding in two to three days. Do you want a little more context?` *(Answer on topic)* |
| Yes, I do. Hey Google, yes. It asked a question. *(N/A)* | `On average, the flu typically lasts one to two weeks with severe symptoms subsiding in two to three days. However, weakness, fatigue, dry cough, and a reduced ability to exercise can linger for three to seven days.` *(Answer on topic)* |
| Hey Google, should I get more rest when I have the flu? *(Switch topic)* | `Sorry, I don't know how to help with that yet.` *(Can't answer)* |

As the exchange between P8 and the voice assistant illustrates (Table 10), query elaboration was used to slowly request components of information; in this case the participant asks about recommendations, medications, timeline, and behavioral advice.

Additionally, we observed how participants used words in the voice assistants' responses to iteratively create new queries [46]. For example, P25 elaborated to continue seeking additional information about flu symptoms in the voice assistant's response:

**Table 11. Elaboration to create new queries and seek additional information**

| **P25 Query** | **Voice Assistant Response** |
|---|---|
| Hey, Google. What are symptoms of the flu? *(Initial)* | `According to Mayo Clinic, symptoms include fever, chills, muscle aches, cough, congestion, running nose, headaches, and fatigue.` *(Answer on topic)* |
| Oh. Hey, Google. If I don't have a fever, can I still have the flu? *(Switch topic)* | `Here's a summary from the website helpline.com,"The flu and fever. A fever is a common symptom of the flu virus, but not everyone who gets the flu will have one. Treat a case of the flu seriously, even if you don't have a fever."` *(Answer on topic)* |

When a prior query was answered unfavorably (answer off topic, generic response, can't answer), query elaborations comprised more than one-third of following reformulations (40%, fig. 4). This suggests that elaboration was, at times, perceived as a tactic to overcome obstacles.

**Table 12. Repeat queries by keeping search terms identical**

| P24 Query | Voice Assistant Response |
|---|---|
| Hey, Google, what are the adverse effects of a . . . *(Switch topic)* | `According to Wikipedia, in medicine, an adverse effect is undesired harmful effect resulting from a medication or other interventions such as surgery.` *(Answer wrong topic)* |
| What are the mean ... I'm sorry. Hey, Google, what are the adverse effects of an enlarged mean corpuscular . . . *(Repeat)* | `According to Wikipedia, in medicine, an adverse effect is undesired harmful effect resulting from a medication or other interventions such as surgery.` *(Answer wrong topic)* |
| Okay. Hey, Google, what are the adverse effects of mean corpuscular volume being large in red blood cells? *(Repeat)* | `Sorry. I don't understand.` *(Can't answer)* |

*Note. Example of repeat queries when a prior query was answered unfavorably, and it led to unfavorable outcomes*

**Repeating to improve responses:** When a prior query was answered unfavorably, participants repeated their query often (avg. 33%, fig. 4). This was evident in the behavior-driven flu scenario (28% repeat reformulations), where most users' queries were answered unfavorably (fig. 5). Prior literature defines query repeats as queries with identical terms repeated in different manners, such as emphasizing certain terms [24]. Similarly, we observed that older adults repeated their queries by 1) keeping search terms identical or 2) substituting or reordering some terms with identical intent. It may be possible that repeating with identical terms led to a higher proportion of unfavorable outcomes. Most (65%) repeat reformulations led to unfavorable outcomes.

We observed that participants tended to repeat identical queries only if the voice assistant could not respond favorably (see table 12). In the following example, P13 asked the same question changing query terms in five unfavorable attempts. Across each of the participants' queries, the terms "have the flu" and "flu shot" were similar, and other ways of question-asking were explored. For example:

**Table 13. Repeat queries by substituting or reordering terms with identical intent**

| **P13 Query** | **Voice Assistant Response** |
|---|---|
| Hey Google, I think I have the flu, can I get a flu shot? *(Initial)* | `It's important to do everything you can to stay healthy.` *(Generic answer)* |
| Well that was non-committal. Hey Google, should I get a flu shot when I think I already have the flu? *(Repeat)* | `It's important to do everything you can to stay healthy.` *(Generic answer)* |
| Hmm. Hey Google, is it good to get a flu shot if you think you already have the flu? *(Repeat)* | `Sorry, I don't understand. Another option is Acetaminophen.` *(Can't answer)* |
| Hey Google, if you already have the flu should you get a flu shot? *(Repeat)* | `My apologies, I don't understand.` *(Can't answer)* |
| Hey Google, I might already have the flu, should I go get a flu shot? *(Repeat)* | `It's important to do everything you can to stay healthy.` *(Generic answer)* |

*Note. Example of repeat queries leading to a favorable outcome even when a prior query was answered unfavorably*

Not all repeated queries led to unfavorable outcomes. In the following example, P4 successfully changed their query terms to receive an on-topic response from the voice assistant. This suggests that voice assistants need better ways to model similar query structures, particularly in health contexts when a non-response or unfavorable response can be detrimental to one's health behaviors.

**Table 14. Repeat queries by substituting or reordering terms with identical intent**

| P4 Query | Voice Assistant Response |
|---|---|
| Hey Google, what causes a migraine? *(Initial)* | `On the website mayoclinic.org, they say, "It's often accompanied by nausea,vomiting and extreme sensitivity to light and sound. Migraine attacks can cause significant pain for hours to days and can be so severe that the pain is disabling."` *(Answer wrong topic)* |
| Hey Google, why do people get migraines? *(Repeat)* | `On the website medicalnewstoday.com, they say, "Hormonal changes. Women may experience migraine symptoms during menstruation due to changing hormone levels. Emotional triggers, stress, depression, anxiety, excitement and shock can trigger a migraine."` *(Answer on topic)* |

*Note. Example of repeat queries when a prior query was answered unfavorably, but led to favorable outcome*

These data suggest a need to study how and why people elaborate or repeat queries as these strategies seem to be more effective than when generalizing and specifying topics for additional information seeking. Rather than analyzing queries with different topics as the start of a new session, our data demonstrates how participants may switch topics to build on their search activity, particularly for subjective health queries.

*4.5.3 Differences by Search Topic and Goal.* Our analysis suggests there are differences in how people search for health information depending on their information goal and health topic. Below, we describe differences between symptom- and behavior-driven scenarios and unique patterns when older adults searched for information about the flu.

**Behavior-driven goals different from symptom-driven goals:** Although query elaboration was frequent across all scenarios, our data show how it was more common in the symptom-driven health scenarios (73%) than behavior-driven health scenarios (50%). In behavior-driven scenarios, participants used other reformulation strategies such as repeating, specifying and generalizing more often (Fig. 7). We also observed differences based on the health topic(Fig.8). Older adults who searched for behavior-driven information about the flu had distinct search patterns in comparison to high-blood pressure and migraine behavior searches; flu searchers repeated their queries more often than in high blood pressure and migraine searches.

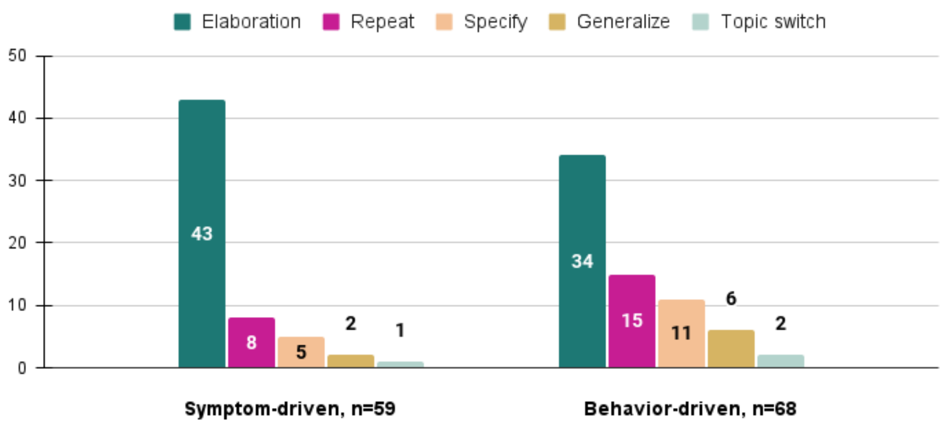


**Fig. 7. Reformulation behavior across symptom and behavior driven health scenarios**

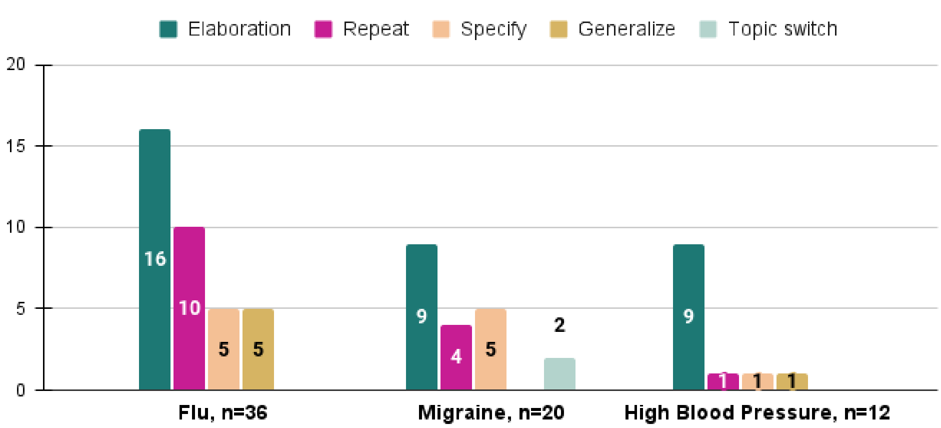


**Fig. 8. Reformulation strategies across health topics in behavior-driven scenarios.**

Lastly, we observed how participants' reformulation strategies were more successful in symptom-driven scenarios than in behavior-driven scenarios (Fig.9, Fig.10), suggesting that voice assistants may currently be better suited for objective health topics rather than subjective information. In most scenarios, elaborating upon queries led to favorable responses (Fig.10).

| Elaboration reformulations that were favorable | | |
|---|---|---|
| **Symptom-driven** | Flu | 53% |
| | Migraine | 100% |
| | HBP | 63% |
| **Behavior-driven** | Flu | 13% |
| | Migraine | 44% |
| | HBP | 56% |

**Fig. 9. Proportion of reformulations that led to favorable VA response (red text corresponds to notable data points)**

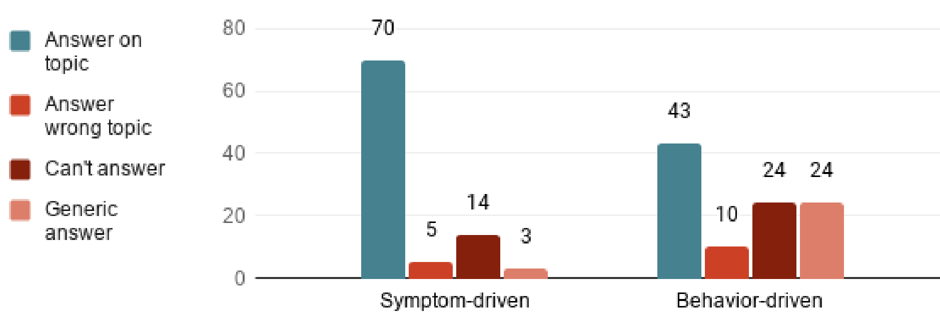


**Fig. 10. VA responses across symptom and behavior driven scenarios**

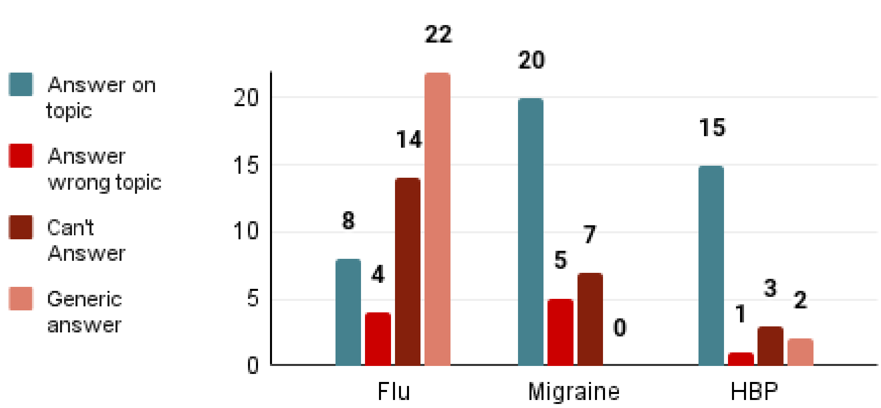


**Fig. 11. VA responses for Behavior-driven health tasks**

**Flu as an Outlier:** In this section, we describe how the voice assistant responded differently to older adults' queries regarding the behavior-driven flu task. The voice assistant was more likely to provide favorable answers for the migraine and high blood pressure topics, but not for the flu (Fig.9, Fig. 11). 83% of the voice assistants' responses were unfavorable when older adults asked about the flu, compared to 28% for high blood pressure queries and 37% for migraine queries. Table 13 shows an exchange between P13 and the voice assistant, demonstrating how the voice assistant often provided unfavorable responses even after older adults reformulated their queries in the behavior-driven flu task. Figure 11 shows how unfavorable "can't answer" and "generic responses" were much higher for the flu topic.

In the symptom- and behavior-driven tasks, the voice assistant generally responded favorably when participants elaborated upon their previous query. However, few (13%) of elaborations were answered on topic in the behavior-driven flu scenario. This may have been due to the level of subjectivity inherent in seeking advice for whether to get a flu shot and the controversy surrounding the effectiveness of vaccinations.

## 4.6 Findings: Voice Assistant Search Expectations

Our third research question was addressed through interviews with older adults to investigate their expectations when using voice assistants for health information seeking. Our interview data suggested three key themes. First, participants in our study expected to receive personalized health information from the device when it responded with health recommendations. Second, participants expected that the voice assistants would provide detailed responses and not merely narrate back information from health information websites or provide generic information. Third, participants expected guidance on how to best scaffold conversational interactions in that the voice assistant, similar to human conversation, would probe or follow-up for clarification to help improve the conversation flow.

*4.6.1 Personalized Health Information Seeking.* Personalized or custom health information was a frequent expectation. Eight participants indicated how they wanted the voice assistant to incorporate personalized content that would include 1) individual-specific symptom information, 2) health suggestions based on demographics, and 3) information about their healthcare networks and providers. For example P12 said, "*I probably wouldn't ask it 'cause it's too generic an answer. I would want to dig deeper into what causes it. 'Cause maybe my migraine headache would be different from some- body else's.*" She recognized that symptoms could be different depending on the individual and wanted the voice assistant to be able to offer personalized information. Such customization may require the voice assistant to have access to one's short or long-term medical history, or engage the user in a conversation before providing health-related advice.

Participants also recognized that such advice may need to be customized based on demographics like age or gender. P5 said, "*...I think that older people, like myself* *(laughs), tend to become more vulnerable, especially with the flu, you know? ... So I think the older you are, the more serious it is when you get ill...if you are a senior citizen, it is recommended that you go and see a health professional in any case.*" He describes how the voice assistant would be more likely to suggest seeing a medical professional if the assistant knew the person was older. Other demographic information that could be useful for health information seeking includes disability or income information as these could affect what medication a device could recommend.

Additionally, participants wanted the device to provide custom information based on who was in their healthcare network. For example, when the voice assistant was unable to provide an answer to a query, P34 said she wanted the device to "*ask my doctor...Maybe they could suggest a doctor, or someone who could help you deal with it.*" To accomplish this, the voice assistant would need to know the primary care physician of its users, their location, and and health insurance network information to suggest a doctor. While access to this information can be useful to for more personalized experience with a voice assistant, there are design implications such as when a device is used by more than one person. Moreover, there are ethical implications of a device, or company who creates the device, accessing such information. We discuss these implications further in the discussion section.

*4.6.2 Expectations for Detailed Responses.* Some participants also wanted the voice assistant to engage in conversation that would lead to recommendation beyond web searches. For example, P9 said, "*I was a little disappointed in that response. I was expecting at some point for it to say, have you sought medical help? Or have you talked to your doctor, your health provider about this? You could take an over-the-counter pain relief.*" Similar to P5's desire to recommend a doctor, P9 wanted the voice assistant to suggest health resources other than websites, like speaking to a medical professional if it recognized that the person repeatedly asked questions about the same content or that it could not provide a response.

*4.6.3 Conversational Scaffolding.* Participants often indicated how they wanted the voice assistant to be more human-like (n=19), with most requests being for a conversational interaction. For example, when P20 was asked how she expected the voice assistant to respond to health-related queries, she said, "*I was just trying to put it in terms of, well if I were speaking with a person what would I expect a person to respond back*". Similarly, P8 said, "*...if we started a dialogue that it sort of began to zero in...so far this doesn't lead the conversation and maybe that's not its role. Its role is to just do a specific task for you. I sit here with a bit of an expectation and it should participate in the conversation ... just much like what you and I are doing. You're asking me a question and I'm responding and then you're asking me a further question to that or reverse, out of it. Its interactiveness in terms of allowing me to look at alternatives or explore more detail, I don't know that I feel that.*" Here, P8 describes how engaging in a series of questions and answers would allow him to better scope and refine his queries to get to more detailed information. In some instances, the voice assistant did interact with participants asking if they wanted more information about a particular topic, but it was unclear what triggered such an interaction to begin.

## 5 DISCUSSION

In this paper, we contribute to prior work that primarily describes *perceptions* of voice assistants for health information seeking. We achieve this by contributing self-reported (survey) and in-situ (interview) empirical data showing *behaviors and expectations* of how older adults use voice assistants for health information seeking. To summarize, we first conducted a survey to understand how older adults use voice assistants for health search compared to other sources (RQ1). Our findings align with prior work showing that people prefer to seek subjective health information and advice from other people. Our data reinforces the need for better systems to support high-quality, subjective health information seeking when access to a medical professional is limited, further motivating the need to study behaviors with voice assistants. Research has suggested that older adults prescribe human-like qualities to voice assistants and enjoy the conversational interaction style (e.g., [47]). Further, voice assistants are accessible sources of information seeking in the home with little digital skill required. Yet, survey responses showed how older adult participants did not use voice assistants for health information seeking as often as other technologies. This echoes recent interview data on older adults' perceptions of a voice assistant probe [41], often using them as a last resort or to "test" the device's responses compared to other sources. Next, we analyzed older adult health query data to answer how they engage in conversational health information seeking (RQ2). Our findings show that subjective and multi-part queries were difficult for the voice assistant to parse, particularly for behavior-driven health scenarios. We also observed how participants struggled with reformulating queries for a better outcome across scenarios. Lastly, we interviewed older adults to understand how voice assistants align with their expectations of health information seeking (RQ3). We find that participants expected personalized responses that were contextualized their current and prior health needs, and that they expected better guidance for how to get to an on-topic health response.

In this discussion, we argue that the computing research community needs to rethink how "intelligent" these devices can be in a health information seeking context. There are risks when use devices without human feedback such as health information from medical professionals and health advice from family and friends. Our findings suggest that voice assistants are more "intelligent" in symptom-driven (fact seeking) scenarios than the behavior-driven (advice seeking) scenarios. The latter requires taking a more personalized approach and being able to engage with users in a way that closely resembles human-human interaction, of which is difficult for the devices to accurately respond. Below, we discuss ethical implications of designing for personalized health contexts and design implications for intelligent voice assistants within and beyond health information seeking.

### 5.1 Ethical Implications of Designing for Personalized Health

We observed that participants wanted to engage in a more personalized health search experience with the voice assistant. Although current voice assistants are capable of responding to basic health questions, many participants wanted the voice assistant to know about their health and medical history, including physician information and medical records, and possess the agency to make recommendations based on this information. These individual desires can incompatible with ethical and privacy implications of privacy-sensitive users or companies having access to this data. Recent work shows one primary concern of using voice assistants is privacy [17, 60], yet health-related customization introduces complex implications that are not as prevalent with general search and recommendations. Although organizational access to a user's search history is currently available and used to make inferences about one's life (e.g. demographics, preferred brands), it does not yet make use of actual medical and health history. On one hand, this can be used to recommend relevant content. Yet, the potential for information misuse by the organization or third party who stores or owns this information can have great consequences. This scenario is a reality for Amazon, who partnered with Britain's National Health Service to provide health information to users via the Amazon Echo [5]. Other consequences include data misuse to increase prices for healthcare needs or predatory advertising.

One common strategy for managing private information across contexts is to **create fine-grained privacy controls** such as those found on platforms like Facebook (e.g., friend lists) and Instagram (e.g., 'Close Friends' in Instagram Stories). These controls allow users to select when to share content with a smaller audience in their network. Similar information groups could be configured by voice assistant users such that certain information about their search behavior or demographics can be seen by audiences such as advertisers, physicians, or friends. In a health context, a user could grant their physician or other professional (e.g., a therapist) access to their search behavior to better guide their experience during in-person healthcare appointments. Additionally, such information collected during healthcare appointments could also be particularly helpful for people who have difficulty recalling health-related information in between their appointments. These cross-channel health conversations could also be useful when the flow of information is reversed - conversations with medical professionals being accessible when older adults use voice assistants in their homes. Thus, voice assistants could be used to improve the continuity of care by facilitating information exchange between patients and providers. In this way, voice assistants can be positioned to supplement rather than replace interactions with medical professionals. Yet, with any verbal form of information input, privacy and security risks include overhearing private health content, if using a mobile voice assistant rather than an in-home voice assistant.

Additionally, we observed that participants preferred and expected a conversational interaction style when engaging in health information seeking. We urge designers to **leverage conversational interaction preferences** when designing personalized

[5]https://www.nytimes.com/2019/07/10/world/europe/alexa-nhs-amazon-privacy.html

voice-based health experiences. Voice assistants could ask users if they have any known medical conditions to customize the results given. In our study, we observed how doing so would have prevented the assistant from responding with a medication that could not be taken by someone with kidney disease. If unsure how to respond to a query, a voice assistant can also offer to connect to the user's preferred and appropriate medical professional. Doing so would turn the output from passive ("I'm not sure how to help with that") to active ("I cannot answer this question, so I will connect you to your primary care physician"). We note that engaging in such behaviors when voice assistants are still in their infancy may burden medical professionals with an overwhelming amount of patient calls. On the other hand, voice assistants could be a welcome alternative to help medical professionals handle more complex queries if connecting to middle-skilled medical professionals with appropriate knowledge and availability.

## 5.2 Designing Intelligent Conversational Agents

*5.2.1 Learning from Human Conversation Expectations.* Analysis of interview and query reformulation data show how a better workflow is needed for unfavorable voice assistant responses (e.g., generic answer, can't answer, answer of topic). Subjective and multi-part queries were hard for the voice assistant to parse. These queries are reflective of human-human communication. Findings show how older adult participants did not necessarily adjust their queries as though they were searching with a computer or adjust their communication to reflect human-machine conversational exchanges [18]. This suggests older adults want to engage with voice assistants in a human-like way. However, this can be a hindrance for users seeking on-topic information if voice assistants are unable to adjust to human conversational styles.

*5.2.2 Guiding with Improved Query Feedback.* Further, interview and query data showed how most participants were unsure of how to proceed when needing to rephrase their query to address their original request or for additional information, yet increased system transparency could help to address this challenge. Researchers have noted the need for intelligent systems to help users to uncertainty (e.g., alt-text for people with vision impairments [34], algorithms in autonomous vehicles (i.e. [9, 22]). We extend this work by highlighting the importance of designing for transparency and mitigating uncertainty with voice assistants. For instance, participants were unsure if a "failed" request was due to an error on their part because it was unclear whether the voice assistant correctly interpreted their original request. To address this, we propose designers mimic uncertainty reduction strategies in visual search engines which display the request at the top of the screen, and **repeat the users' request verbally**. This intentional design feature would provide the necessary feedback for the user to understand whether the system correctly interpreted the request. In other instances of unfavorable responses, particularly for high-risk health topics, the voice assistant could automatically connect the user with a medical professional.

In a health context, improved feedback and transparency mechanisms could increase trust in digital information sources and mitigate any user or system uncertainty, which can lead to negative, and even harmful, health outcomes [11]. In contrast to younger age groups, research continues to show how older people have less trust in online sources for health information seeking because of information overload [55]. Yet, prior work shows how using online information to supplement medical visits and information from medical professionals when searching visually is beneficial [38, 61]. Challenges arise when people exhibit ineffective search and evaluation skills, leading medical professionals having to "educate" patients on effective search strategies [38, 52]. Yet, this is a challenge as medical professionals have limited interaction time with patients and advice would not be customized to individual search needs. As such, machine learning offers the potential to improve visual and voice health search practices.

*5.2.3 Learning Scaffolding Practices Over Time.* Interview data showed how older adults wanted the voice assistant to **guide them on how to rephrase their query**. While shorter responses can lessen information overload present in visual forms of search, brief responses when the voice assistant is unsure how to respond can be frustrating. Limited feedback is a known issue with voice assistants, with proposals to "differentiate between lack of device knowledge vs. speech recognition error due to not understanding the user or due to background noise" [48]. More complex approaches could leverage machine learning techniques for the system to learn query best practices for favorable/on-topic queries across users over time. The voice assistant could then make suggestions for improving a series of queries. A simpler approach could allow voice assistants to give feedback based on query features. Although not statistically significant, we observed how participants were either interrupted when making longer queries or the assistant was unable to respond to multi-part queries. If using query complexity to offer recommendations, the voice assistant could prompt the user to parse a multi-part request into shorter requests. Or, after responding to one part of the query, the voice assistant could offer query guidance for improving responses to subsequent parts of a query. Combined with verbally repeating the request, these design features could improve system transparency of voice assistants within and beyond health information seeking contexts.

*5.2.4 Improving User Agency.* Lastly, voice assistants can better communicate the possibility of receiving inaccurate information or misinformation in health contexts as suggested for automatically generated alt-text in visual online communities [34] and recent approaches to algorithmic uncertainty estimation in neural networks [5]. We found that people relied on information about sources, lists, and level of detail to evaluate the quality of the voice assistant's responses. These are features that research has shown to be important with health information seeking on a computer or mobile phone [59], but there remains little to no research for how to provide audio-based cues of information quality beyond those that mimic visual cues. Therein lies an opportunity to consider cues such as tone, gender, and age to communicate information quality with voice interfaces.

Further, since participants preferred certain, recognizable sources (e.g. Mayo Clinic, New England Journal of Medicine), it could be helpful for them to **choose their preferred sources** from a list of verified health organizations during setup procedures of a voice assistant. This could, in part, lead to filter bubbles where only certain content is made visible when diversity of content could be useful, particularly for health information [51, 54]. Yet, if the sources are verified, neutral, science-driven organizations rather than partisan or biased subjective sources, choosing from a limited list should not be cause for concern. Another approach is to **offer multiple results**, similar to searching using a visual browser and allow people to select or compare from a list. While this could be more cognitively demanding, it places the user in control and empowers them to be involved in health decision making. Inspired by the complexity of screen readers for blind computer users, Vtyurina et al., (2019) have explored how voice assistants can provide more detail, but some participants described how being required to do so by recalling specific commands was confusing [69]. Voice assistants could leverage this approach of giving users control of expanding query results, but with improved conversational flow that does not require command recall. In some instances during the interviews, we observed how the voice assistant would ask participants, 'Would you like to hear more?', but this was not a consistent response. We urge developers and designers to consider settings for users to enable this follow-up question after any informational voice query, or for specific topics such as health information seeking.

### 5.3 Limitations

We acknowledge that our study was in a quasi-naturalistic setting in that interviews did not take place in participants' homes and were scenario-based. Our attempts to minimize the Hawthorne effect and bias due to the researcher's presence included starting with an open-ended and non-health related query, and asking follow-up questions after all queries were completed. Additionally, participants in our survey and interview study had relatively high computer and health self-efficacies and were primarily residents of a white, high SES community. These demographic characteristics likely shape how the participants in our study access and subsequently use technology for personal use. For example, the survey was completed online, which means respondents likely have higher rates of internet access and use than the broader older adult population. However, because our study goals focus on health and technology use, we acknowledge that people with internet access are more likely to use the internet for health information seeking. Also, interview participants had a range of experiences with voice assistants, which may have affected their search and reformulation behaviors during the scenarios. Regarding analysis, much of the query analysis was descriptive rather than inferential because of the small sample size and we encourage researchers to replicate our findings with larger groups of older and younger adults. Lastly, although we do not focus on hearing loss in this paper, we recognize that one limitation of using voice assistants with an aging population is a perceived lack of accessibility for people with hearing loss. However, recent work shows that deaf and hard-of-hearing communities also engage with voice-based interfaces and there are opportunities for design [12, 23].

## 6 CONCLUSION

This work provides empirical evidence of how one group of users, older adults, engage in health information seeking using voice assistants. Through scenario-based interviews, we highlight how their expectations are met and not met. We describe challenges people had formulating their initial queries and engaging in an iterative search process when reformulating queries. As such, we reflect and provide actionable recommendations about ethics of personalization in health contexts and designing better interactive voice systems. These design recommendations could also support other groups of people who face challenges using voice assistants such as people with speech or cognitive impairments, or those with limited trust in technological systems such as low income and minoritized populations. Future work could engage the research and design community in developing better solutions to these challenges and evaluating them with users.

## ACKNOWLEDGMENTS

We would like to thank Madel Leal, Desiree McLain, and Ruchi Ookalkar for their work on this project. We would also like to thank Dr. Aqueasha Martin-Hammond, Dr. Gabriela Marcu, and Dr. Tiffany Veinot for their feedback on drafts of this paper.

## 7 REFERENCES

## REFERENCES

[1] 2017. *Nearly half of Americans use digital voice assistants, mostly on their smartphones*. Technical Report. Pew Research Center. https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

[2] 2019. *NHS health information available through Amazon's Alexa*. Technical Report. Department of Health and Social Care. https://www.gov.uk/government/news/nhs-health-information-available-through-amazon-s-alexa

[3] Ali Abdolrahmani and Ravi Kuber. 2016. Should I Trust It When I Cannot See It? Credibility Assessment for Blind Web Users. *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '16* (2016), 191–199. https://doi.org/10.1145/2982142.2982173

[4] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. Siri Talks at You: An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 249–258.

[5] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. 2019. Deep Evidential Regression. *arXiv preprint arXiv:1910.02600* (2019).

[6] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 17.

[7] Francisco G Barbeite and Elizabeth M Weiss. 2004. Computer self-efficacy and anxiety scales for an Internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior* 20, 1 (2004), 1–15.

[8] Heather Becker, Alexa Stuifbergen, Hyun Soo Oh, and Sharon Hall. 1993. Self-rated abilities for health practices: A health self-efficacy measure. *Health Values: The Journal of Health Behavior, Education & Promotion* (1993).

[9] Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the driver–automation interaction: An approach using automation uncertainty. *Human factors* 55, 6 (2013), 1130–1141.

[10] Ronald W. Berkowsky and Sara J. Czaja. 2018. Challenges associated with online health information seeking among older adults. In *Aging, Technology and Health*. Elsevier, 31–48. https://doi.org/10.1016/B978-0-12-811272-4.00002-6

[11] Timothy W Bickmore, Ha Trinh, Stefan Olafsson, Teresa K O'Leary, Reza Asadi, Nathaniel M Rickles, and Ricardo Cruz. 2018. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research* 20, 9 (2018), e11510. https://doi.org/10.2196/11510

[12] Johnna Blair and Saeed Abdullah. 2019. Understanding the Needs and Challenges of Using Conversational Agents for Deaf Older Adults. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 161–165.

[13] Robin Brewer, Raymundo Cornejo-García, Tedmond Schwaba, Darren Gergle, and Anne Marie Piper. 2016. Exploring Traditional Phones as an E-Mail Interface for Older Adults. *ACM Transactions on Accessible Computing* 8, 2 (jan 2016), 1–20. https://doi.org/10.1145/2839303

[14] Clare Carroll, Catherine Chiodo, Adena Xin Lin, Meg Nidever, and Jayanth Prathipati. 2017. Robin: enabling independence for individuals with cognitive disabilities using voice assistive technology. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 46–53.

[15] Mr SHOMIR CHAUDHURI, Mr THAI LE, Ms CATHY WHITE, Hilaire Thompson, and George Demiris. 2013. Examining health information–seeking behaviors of older adults. *Computers, informatics, nursing: CIN* 31, 11 (2013), 547.

[16] Angela Chen. 2019. Amazon's Alexa now handles patient health information. https://www.theverge.com/2019/4/4/18295260/amazon-hipaa-alexa-echo-patient-health-information-privacy-voice-assistant

[17] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. What can i help you with?: infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 43.

[18] Aurelie Dommes, Aline Chevalier, and Sarah Lia. 2011. The role of cognitive flexibility and vocabulary abilities of younger and older users in searching for information on the web. *Applied Cognitive Psychology* 25, 5 (2011), 717–726.

[19] Andrew J. Flanagin and Miriam J. Metzger. 2007. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society* 9, 2 (apr 2007), 319–342. https://doi.org/10.1177/1461444807075015

[20] Ido Guy. 2016. Searching by talking: Analysis of voice queries on mobile web search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 35–44.

[21] Ahmed Hassan Awadallah, Ranjitha Gurunath Kulkarni, Umut Ozertem, and Rosie Jones. 2015. Characterizing and predicting voice query reformulation. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 543–552.

[22] Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. 2013. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications*. ACM, 210–217.

[23] Dhruv Jain, Angela Lin, Rose Guttman, Marcus Amalachandran, Aileen Zeng, Leah Findlater, and Jon Froehlich. 2019. Exploring sound awareness in the home for people who are deaf or hard of hearing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[24] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How do users respond to voice input errors? Lexical and phonetic query reformulation in voice search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 143–152.

[25] Jiepu Jiang and Chaoqun Ni. 2016. What affects word changes in query reformulation during a task-based search session?. In *Proceedings of the 2016 ACM on conference on human information interaction and retrieval*. 111–120.

[26] Shaheen Kanthawala, Amber Vermeesch, Barbara Given, and Jina Huh. 2016. Answers to health questions: internet search results versus online health community responses. *Journal of medical Internet research* 18, 4 (2016), e95.

[27] Johannes Kiesel, Kevin Lang, Henning Wachsmuth, Eva Hornecker, and Benno Stein. 2020. Investigating Expectations for Voice-Based and Conversational Argument Search on the Web. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) *(CHIIR '20)*. Association for Computing Machinery, New York, NY, USA, 53–62. https://doi.org/10.1145/3343413.3377978

[28] Erin Klawitter and Eszter Hargittai. 2018. Shortcuts to well being? Evaluating the credibility of online health information through multiple complementary heuristics. *Journal of Broadcasting & Electronic Media* 62, 2 (2018), 251–268.

[29] Erin Klawitter and Eszter Hargittai. 2018. "I Went Home to Google": How Users Assess the Credibility of Online Health Information. In *EHealth : Current Evidence, Promises, Perils, and Future Directions*, Timothy M Hale, Wen-Ying Sylvia Chou, and Shelia R Cotten (Eds.). Emerald Publishing Limited, 11–41.

[30] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.

[31] Mengyang Liu, Yiqun Liu, Jiaxin Mao, Cheng Luo, and Shaoping Ma. 2018. Towards designing better session search evaluation metrics. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1121–1124.

[32] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[33] Widad Machmouchi, Ahmed Hassan Awadallah, Imed Zitouni, and Georg Buscher. 2017. Beyond Success Rate: Utility as a Search Quality Metric for Online Experiments. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 757–765.

[34] Haley MacLeod, Cynthia L. Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (2017), 5988–5999. https://doi.org/10.1145/3025453.3025814

[35] Elizabeth Manafo and Sharon Wong. 2012. Exploring older adults' health information seeking behaviors. *Journal of nutrition education and behavior* 44, 1 (2012), 85–89.

[36] Jiaxin Mao, Yiqun Liu, Ke Zhou, Jian-Yun Nie, Jingtao Song, Min Zhang, Shaoping Ma, Jiashen Sun, and Hengliang Luo. 2016. When does Relevance Mean Usefulness and User Satisfaction in Web Search?. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 463–472.

[37] Lisa M Soederberg Miller and Robert A Bell. 2012. Online health information seeking: the influence of age, information trustworthiness, and search challenges. *Journal of aging and health* 24, 3 (2012), 525–541.

[38] Janet M Morahan-Martin. 2004. How internet users find, evaluate, and use online health information: a cross-cultural review. *CyberPsychology & Behavior* 7, 5 (2004), 497–510.

[39] Meredith Ringel Morris. 2012. Web on the wall: insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. ACM, 95–104.

[40] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.

[41] Phani Nallam, Siddhant Bhandari, Jamie Sanders, and Aqueasha Martin-Hammond. 2020. A Question of Access: Exploring the Perceived Benefits and Barriers of Intelligent Voice Assistants for Improving Access to Consumer Health Resources Among Low-Income Older Adults. *Gerontology and Geriatric Medicine* 6 (2020), 2333721420985975.

[42] Cameron D Norman and Harvey A Skinner. 2006. eHEALS: the eHealth literacy scale. *Journal of medical Internet research* 8, 4 (2006), e27.

[43] NPR. 2019. *The Smart Audio Report*. Technical Report. NPR and Edison Research. 1–36 pages. https://www.nationalpublicmedia.com/wp-content/uploads/2019/06/The{_}Smart{_}Audio{_}Report{_}Spring{_}2019.pdf

[44] Kenneth Olmstead. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones.

[45] Adam Palanica, Anirudh Thommandram, Andrew Lee, Michael Li, and Yan Fossat. 2019. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. *npj Digital Medicine* 2, 1 (2019), 55.

[46] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[47] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[48] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[49] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 459.

[50] Chen Qu, Liu Yang, W. Bruce Croft, Johanne R. Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and Characterizing User Intent in Information-Seeking Conversations. In *The 41st International ACM SIGIR Conference on Research amp; Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 989–992. https://doi.org/10.1145/3209978.3210124

[51] Stephen A Rains and Erin K Ruppel. 2016. Channel complementarity theory and the health information-seeking process: Further investigating the implications of source characteristic complementarity. *Communication Research* 43, 2 (2016), 232–252.

[52] Cong Ren, Zhaohua Deng, Ziying Hong, and Wei Zhang. 2019. Health information in the digital age: an empirical study of the perceived benefits and costs of seeking and using health information from online sources. *Health Information & Libraries Journal* 36, 2 (2019), 153–167.

[53] Soo Young Rieh et al. 2006. Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management* 42, 3 (2006), 751–768.

[54] Erin K Ruppel and Stephen A Rains. 2012. Information sources and the health information-seeking process: An application and extension of channel complementarity theory. *Communication Monographs* 79, 3 (2012), 385–405.

[55] Laura Sbaffi and Jennifer Rowley. 2017. Trust and credibility in web-based health information: a review and agenda for future research. *Journal of medical Internet research* 19, 6 (2017), e218.

[56] Johan Schalkwyk, Doug Beeferman, Françoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Kamvar, and Brian Strope. 2010. "your word is my command": Google search by voice: A case study. In *Advances in speech recognition*. Springer, 61–90.

[57] Stephan Schlögl, Gérard Chollet, Markus Garschall, Manfred Tscheligi, and Grégory Legouverneur. 2013. Exploring voice user interfaces for seniors. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 52.

[58] Milad Shokouhi, Umut Ozertem, and Nick Craswell. 2016. Did you say U2 or YouTube? Inferring implicit transcripts from voice search logs. In *Proceedings of the 25th International Conference on World Wide Web*. 1215–1224.

[59] Elizabeth Sillence, Pam Briggs, Peter Harris, and Lesley Fishwick. 2007. Going online for health advice: Changes in usage and trust practices over the last five years. *Interacting with computers* 19, 3 (2007), 397–406.

[60] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–23.

[61] Sharon Swee-Lin Tan and Nadee Goonawardene. 2017. Internet health information seeking and the patient-physician relationship: a systematic review. *Journal of medical Internet research* 19, 1 (2017), e9.

[62] SJ Tracy. 2013. Data analysis basics: A pragmatic iterative approach. *SJ Tracy, Qualitative Research Methods: Collective Evidence, Crafting Analysis, Communicating Impact* (2013), 183–202.

[63] Milka Trajkova and Aqueasha Martin-Hammond. 2020. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[64] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 325–328. https://doi.org/10.1145/3020165.3022144

[65] Gilad Tsur, Yuval Pinter, Idan Szpektor, and David Carmel. 2016. Identifying Web Queries with Question Intent. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 783–793. https://doi.org/10.1145/2872427.2883058

[66] Pertti Vakkari. 2020. The Usefulness of Search Results: A Systematization of Types and Predictors. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 243–252.

[67] Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. 2015. Social Media Platforms for Low-Income Blind People in India. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*. ACM Press, New York, New York, USA, 259–272. https://doi.org/10.1145/2700648.2809858

[68] Lakshmi Vikraman, W. Bruce Croft, and Brendan O'Connor. 2018. Exploring Diversification In Non-Factoid Question Answering. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval* (Tianjin, China) *(ICTIR '18)*. Association for Computing Machinery, New York, NY, USA, 223–226. https://doi.org/10.1145/3234944.3234973

[69] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. ACM, New York, NY, USA, 3590–3594. https://doi.org/10.1145/3308558.3314136

[70] Ryen W. White, Matthew Richardson, and Wen-tau Yih. 2015. Questions vs. Queries in Informational Search Tasks. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15 Companion)*. Association for Computing Machinery, New York, NY, USA, 135–136. https://doi.org/10.1145/2740908.2742769

[71] Jayme Wilder, Devin Nadar, Nitin Gujral, Benjamin Ortiz, Robert Stevens, Faye Holder-Niles, John Lee, and Jonathan Gaffin. 2019. Pediatrician Attitudes toward Digital Voice Assistant Technology Use in Clinical Practice. *Applied Clinical Informatics* 10, 02 (mar 2019), 286–294. https://doi.org/10.1055/s-0039-1687863

[72] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1083–1092.

[73] Maria Wolters, Kallirroi Georgila, Johanna D. Moore, Robert H. Logie, Sarah E. MacPherson, and Matthew Watson. 2009. Reducing working memory load in spoken dialogue systems. *Interacting with Computers* 21, 4 (aug 2009), 276–287. https://doi.org/10.1016/j.intcom.2009.05.009

[74] Maria Wolters, Kallirroi Georgila, Johanna D Moore, and Sarah E Macpherson. 2009. Being Old Doesn't Mean Acting Old : How Older Users Interact with Spoken Dialog Systems. 2, 1 (2009). https://doi.org/10.1145/1525840.1525842.http

[75] Bo Xie. 2009. Older adults' health information wants in the Internet age: Implications for patient–provider relationships. *Journal of health communication* 14, 6 (2009), 510–524.

[76] Zhaopeng Xing, Fei Yu, and Xiaojun Yuan. 2019. Research on voice search behavior in the last 10 years. *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 819–821.

[77] Zhaopeng Xing, Xiaojun Yuan, Dan Wu, Yeman Huang, and Javed Mostafa. 2020. Understanding voice search behavior: review and synthesis of research. In *International Conference on Human-Computer Interaction*. Springer, 305–320.

[78] Jeonghe Yi and Farzin Maghoul. 2011. Mobile Search Pattern Evolution: The Trend and the Impact of Voice Queries. In *Proceedings of the 20th International Conference Companion on World Wide Web*

(Hyderabad, India) *(WWW '11)*. Association for Computing Machinery, New York, NY, USA, 165–166. https://doi.org/10.1145/1963192.1963276