

1

2

3

4 **Design, Development, and Usability**

5 **Evaluation of a Voice App Experience**

6 **for Heart Failure Management**

7

8

9 Antonia Barbaric MASC^{1,2,6}, Cosmin Munteau PhD³, Heather J. Ross MD MHSc^{4,5}, Kuo-Cheng

10 Tong², and Joseph A. Cafazzo PhD PEng^{1,2,6}

11

12

13

14 ¹ Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

15 ² Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON,

16 Canada

17 ³ Department of Computer Science, University of Toronto, Toronto, ON, Canada

18 ⁴ Ted Rogers Centre for Heart Research, Peter Munk Cardiac Centre, University Health Network,

19 Toronto, ON, Canada

20 ⁵ Department of Medicine, University of Toronto, Toronto, ON, Canada

21 ⁶ Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University

22 of Toronto, Toronto, ON, Canada

23

24

25 Corresponding author

26 Email: antonia.barbaric@mail.utoronto.ca (AB)

27

28 Abstract

29 The use of digital therapeutics (DTx) in the prevention and management of medical conditions
30 has increased through the years with an estimated 44 million people using one as part of their
31 treatment plan in 2021, nearly double the amount from last year. DTx are commonly accessed
32 through smartphone apps, but offering these treatments through an alternative input can improve
33 the accessibility of these interventions. Voice apps are an emerging technology in the digital
34 health field, and may be an appropriate alternative platform for some patients. This research
35 aimed to identify the acceptability and feasibility of offering a voice app as an alternative input
36 for a chronic disease self-management program. The objective of this project was to design,
37 develop, and evaluate a voice app of an already existing smartphone-based heart failure self-
38 management program, *Medly*, to be used as a case study. A voice app version of *Medly* was
39 designed and developed through a user-centered design process. We conducted a usability study
40 and semi-structured interviews with representative end users (n=8) at the Peter Munk Cardiac
41 Clinic in Toronto General Hospital to better understand the user experience. A *Medly* voice app
42 prototype was built using a software development kit in tandem with a cloud computing
43 platform. Three out of the eight participants were successful in completing the usability session,
44 while the rest of the participants were not due to various errors. Almost all (7 out of the 8)
45 participants were satisfied with the voice app and felt confident using it. Half of the participants
46 were unsure about using the voice app in the future, though. With these findings, design changes
47 were made to better improve the user experience. With rapid advancements in voice user
48 interfaces, we believe this technology will play an integral role when providing access to DTx
49 for chronic disease management.

50 Introduction

51 Background

52 The prevalence of heart failure (HF) continues to be on the rise as more people are surviving
53 cardiovascular disease (1). Cardiovascular diseases can cause the heart muscle to become
54 damaged and weak, leading to the development of HF. HF occurs when the pumping action of
55 the heart muscle is not strong enough to meet the needs of the body, or when the heart muscle
56 does not relax properly to accommodate blood flow back into the heart. When this occurs, fluid
57 can build up in the lungs and other parts of the body, such as the ankles, creating congestion in
58 the lungs and results in a lack of oxygen being delivered to the rest of the body (2). As of 2017, it
59 is estimated that 64.3 million people are living with HF globally (3) and many countries are
60 reporting a steady increase in this condition's prevalence.

61 HF not only creates a burden on healthcare resources and expenditures (1), but also on the
62 patient's well-being if not cared for properly. HF limits a patient's capacity to live well, either
63 through physical, psychological, or social means (4). Patient self-management plays an integral
64 role in the treatment of HF, with studies reporting improved health outcomes, decreased clinic
65 visits, and decreased health costs (5). Mobile health, also referred to as mHealth, is a type of
66 digital health technology that involves the use of mobile devices (smartphone, patient monitoring
67 device, wireless devices, etc.) for medical and public health practice (6), and enables the
68 integration of self-care support into a patient's daily routine (7). Smartphone applications (apps)
69 remain to be one of the most popular tools for helping patients who are diagnosed with chronic
70 conditions manage their health at home (8).

71 *Medly* is an evidence-based, HF self-management program that has been developed by the
72 University Health Network (UHN) and is implemented as part of the standard of care at UHN's
73 Peter Munk Cardiac Centre (PMCC) (9). This program is available to patients as a prescribed
74 digital therapeutic, cleared by Health Canada in 2020. *Medly* consists of a smartphone app to
75 enable patients to log clinically relevant physiological measurements and symptoms daily, which
76 is then used in the *Medly* algorithm to generate an automated self-care message. The care team is
77 able to review the patient's data daily, and can view current trends and historic data. If the
78 algorithm detects the patient is deteriorating, the care team receives alerts via email and on the
79 dashboard, and is able to contact the patient to advise. Previous studies evaluating *Medly* have
80 proven that this program can reduce health service utilization and improve clinical, quality of
81 life, and patient self-care outcomes (9).

82 With these encouraging results, there is a desire to improve the accessibility of the *Medly*
83 program. More specifically, there is evidence in the literature that suggests older adults, as well
84 as those with cognitive and physical impairments have difficulties when using touch screen
85 devices, such as the smartphone (10–13), and often require additional assistance from their
86 caregivers, making it difficult to complete tasks independently (14–16). Similarly, a study
87 performed by Ware et al. showcased that adherence rates for *Medly* over a 12 month period were
88 highest in older age groups and progressively lower in younger age groups (17). This data
89 suggests that younger adults may require a more convenient way of interacting with self-
90 management apps that require consistent interaction. While smartphone design guidelines for
91 these specific demographics exist (12), there is potential for other platforms, such as voice user
92 interfaces (VUIs), to increase the uptake of this program and to create a better user experience.

93 Voice apps are an emerging technology in the healthcare field, ranging from integration with in-
94 clinic registration processes to helping people manage their chronic illness and live
95 independently in their homes. There is growing interest in VUIs such as Google Home and
96 Amazon Alexa due to their simple set-up, ease of use, and low cost when compared to current
97 interactive voice response technologies already on the market. Voice apps enable ubiquitous
98 connectivity by allowing the user to access services using only their voice, making for a more
99 convenient experience. Research relating to voice apps for chronic illness are still in the
100 development and piloting phases and have limited efficacy in testing to support final outcomes
101 and conclusions. As described by Sezgin et al., most voice apps currently being developed
102 provide information and assistance services, which includes general educational content and
103 guidance, as well as reminders and tracking (18). There is limited research showcasing voice
104 apps as a tool to provide more personalized, user specific support, creating an opportunity to
105 further investigate using this technology to provide healthcare services and support.

106 Objectives

107 Guided by the findings from the literature review and coupled with a user-centered design
108 process (UCD), we sought to design, develop, and evaluate a voice app for an existing
109 smartphone-based HF self-management program, *Medly*, to determine if a voice app version
110 adds benefit to *Medly*'s current model of interaction and care. The *Medly* voice app was
111 evaluated in a usability study, and the findings from this evaluation helped inform the final
112 design and development of the *Medly* voice app, as presented in this paper.

113 Methods

114 Overview

115 The *Medly* voice app was created in two main phases: 1) design and development, and 2)
116 usability evaluation of the preliminary design. A prototype version of the *Medly* voice app was
117 created and used in a usability study. Findings from the usability evaluation helped inform any
118 redesign and redevelopment work needed to create a voice app that better met the needs of HF
119 patients.

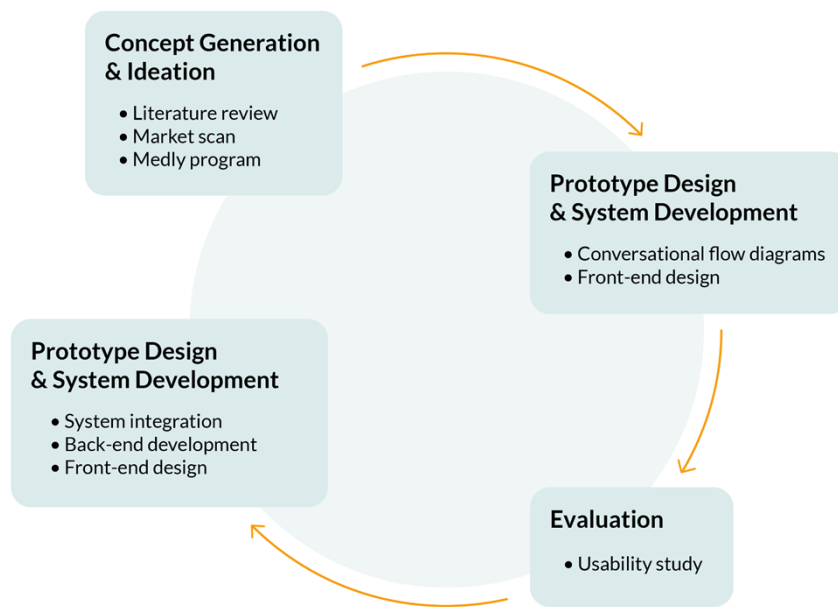
120 The *Medly* voice app was designed and developed for deployment on VUIs and was built using
121 an external cloud computing service. The usability study was performed under the UHN REB
122 19-5051.2 and collected both quantitative and qualitative data regarding the user's experience
123 when interacting with the *Medly* voice app prototype. The findings from this usability study
124 influenced any redesign and/or redevelopment work that occurred after the usability evaluation.

125 Design Process

126 A UCD process was followed when the *Medly* voice app was created to ensure that end user
127 needs were included in all phases of the design project. This framework is an iterative process
128 that allows design teams to create technologies that users are able and willing to use, rather than
129 expecting them to change their behaviors in order to accommodate the technology (19). The first
130 step in this process is concept generation and ideation, which involves identifying end user needs
131 to better understand the technology's intended use. Following this investigation prototypes are
132 developed, based on the design requirements created from the user needs assessment, and

133 evaluated through usability testing. Multiple cycles of prototyping and user testing are performed
134 to improve the ease of use and adoption of the technology.

135 For this research, a literature review was performed during the first phase of the UCD process to
136 help guide the voice app design by better understanding the demographic that would use this
137 technology and the environment in which it would be used. Keywords such as, “chronic illness”,
138 “self-management”, and “Amazon Alexa” or “Google Home” were used in library database
139 searches in order to find relevant research related to this project. A market scan was also
140 performed in addition to the literature review and specifically searched for voice apps related to
141 healthcare management. The ideation of the *Medly* voice app was also influenced by the *Medly*
142 program requirements. Using the information gathered from the first phase, a prototype of the
143 *Medly* voice app was developed and deployed on a VUI. A usability study with the voice app
144 was then performed with representative end users in the Ted Rogers Centre of Excellence in
145 Heart Function clinic at the PMCC in Toronto General Hospital. Based on the usability study
146 results, the prototype was further developed and redesigned to better meet end user needs, and is
147 expected to be evaluated in a future clinical study. Fig 1 illustrates the process that was followed
148 for this project.



149

150 **Fig 1. Design process was guided by UCD process; details describing what was done during each phase are**
 151 **included above**

152 **Design Specifications of Medly *Voice***

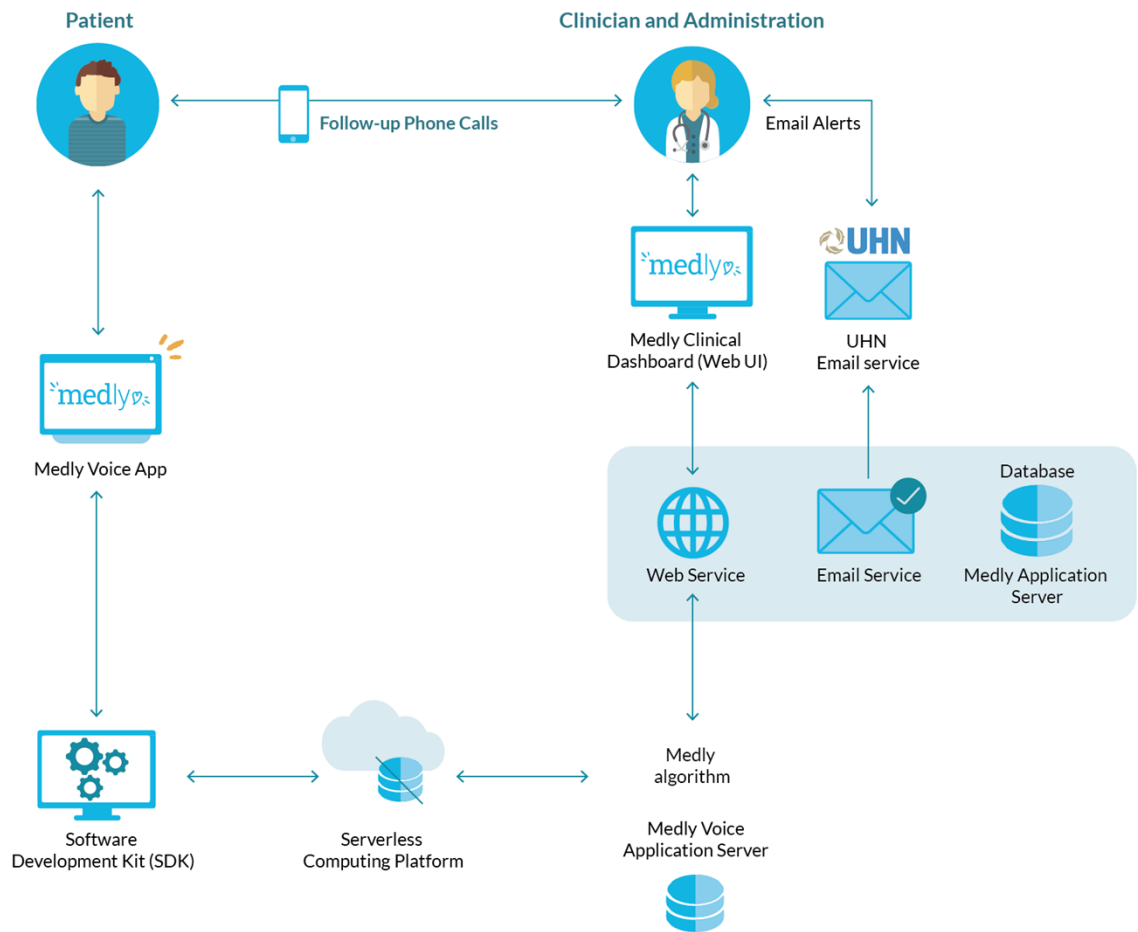
153 Design requirements were created prior to developing the *Medly* voice app to help guide the
 154 development process and define the voice app’s functionality. The main requirements
 155 implemented for the voice app are presented in Table S1 in Multimedia Appendix 1, which
 156 summarize the main objectives of the program. The voice app was designed to have the same
 157 functionality as the mobile app which includes: 1) asking the user to measure their weight, blood
 158 pressure, and heart rate; 2) saving and storing those values in the *Medly* clinical dashboard; 3)
 159 asking the user a series of yes/no questions relating to HF symptoms; 4) processing the data
 160 using the *Medly* algorithm; and 5) outputting a message to the user based on the algorithm result.
 161 These requirements helped create an app that was appropriate for voice interaction and are based
 162 on research as well as guidelines related to VUI design (20,21). Conversational flow diagrams

163 were created by also using VUI design guidelines, and each scenario was tested on a VUI
164 following its implementation (22,23).

165 Development Specifications of *Medly Voice*

166 *Medly Voice* App Architecture (Fig 2)

167 The architecture of the *Medly Voice* program consists of various components required to receive,
168 store, and send data. The *Medly* voice app serves as the primary point of contact for the user and
169 was built using a software development kit. The voice input of physiological measures is
170 captured by the *Medly* voice app and is processed through an external serverless computing
171 platform into algorithm input data. This data is transmitted to the UHN *Medly Voice* application
172 server. This server receives the input data from Amazon Web Services, pulls relevant patient
173 parameter thresholds from the *Medly Voice* application server, and evaluates the combined data
174 using the *Medly* algorithm to generate a patient assessment. The patient data and resulting
175 assessment are sent to the UHN *Medly* Application server, where it is stored in a database and
176 available to view on a dashboard by the *Medly* care team.



177

178 **Fig 2. Software architecture diagram of the Medly voice app system. The participant interacts with the Medly**
 179 **voice app through the smart speaker, which is connected to an external cloud computing platform. This data**
 180 **is then sent to the Medly voice application**

181 Software Platform

182 Voice apps consist of both a VUI and application logic. When the user speaks to the VUI and
 183 invokes the voice app, a request is created and processed in the context of the voice app's
 184 interaction model using tools such as machine learning, natural language understanding, and
 185 automatic speech recognition. Once the interaction model has processed the speech, a request is
 186 sent to the application logic, which provides a response back to the user. For the purposes of this

187 project, a Hosted Alexa Skill built using the Amazon Alexa Skills Kit (24) was used to
188 implement the *Medly* voice app's logic, but other frameworks could have been used as well.

189 A voice interaction model consists of the following three inputs: 1) intents, 2) custom slot types,
190 and 3) sample utterances. An intent represents an action that will fulfill a user's request and uses
191 slots to store key information extracted from the user's request. Sample utterances are specific
192 words and phrases that the user may say to invoke an intent. Developers are encouraged to
193 include multiple sample utterances per intent to increase the likelihood of the interaction model
194 entering into that intent when appropriate. Machine learning and natural language understanding
195 are also used to help train the interaction model in identifying which intent to enter based on the
196 phrase, even if it does not exactly match the sample utterances listed.

197 Verification and Validation

198 The verification and validation of the *Medly* voice app were ongoing processes that were
199 completed at various steps in the design process. The verification process involved both the
200 voice app code (25), as well as the *Medly* algorithm. The validation process consisted of
201 evaluating the front-end of the voice app, which is what the user is expected to interact with.

202 Verification was done on the *Medly* voice app to ensure that it met the design requirements and
203 performed as expected. The code implemented to help guide the user through the conversation
204 was verified using a demonstration method. The demonstration method is one of the four
205 fundamental approaches used in design verification, and can be appropriate for designs that
206 consist of software development. Using the demonstration method, different scenarios were
207 created with various inputs in order to ensure that the produced results were as expected. The
208 *Medly* algorithm was verified using unit testing to ensure that the algorithm outputs the expected

209 response. Various combinations of weight, blood pressure, heart rate, and symptom responses
210 were designed using a test *Medly* patient account to determine if the correct self-care messages
211 were being produced.

212 Different conversational flows were brainstormed and used to help validate the voice app. These
213 conversations were then tested using an online voice app simulator, which responds with voice
214 and text and also includes the JSON request to help fix errors. In addition to validating the *Medly*
215 voice app by testing different conversational flows, the usability study helped determine the
216 different conversations/phrases being used that were originally unaccounted for. These findings
217 were then used to help inform the re-development work which included modifying the
218 interaction model.

219 Usability Study

220 Overview

221 The purpose of the usability study was to investigate the potential of voice communication
222 devices as an alternative healthcare delivery platform to smartphone apps for providing patient
223 self-care. Representative end-users were recruited and asked to explore the intuitiveness of the
224 *Medly* voice app. Through the evaluation we sought to identify user preferences and expectations
225 through the following: a usability session, the standardized System Usability Scale (SUS)
226 questionnaire (Multimedia Appendix 3), and a qualitative semi-structured interview session. The
227 findings from this study helped identify the acceptability and ease of use of the voice app and
228 also informed the service design that would be most appropriate to implement for a voice app
229 offering a self-management program.

230 Participant Recruitment

231 Eight participants were recruited for this study, and all of them use *Medly* as part of their
232 standard of care. All eight participants recruited were adults living with chronic HF and could
233 read, write, and understand English. *Medly* patients were asked to participate so that unique
234 insights and themes could be identified since they already have experience with the program.
235 Participants were recruited through the *Medly* nurse coordinator, who provided a brief overview
236 of the research study and then asked if they were interested in participating before introducing
237 them to the study coordinator. Most usability studies require 5-8 participants in order to receive
238 response saturation (26,27) and as a result we chose to recruit eight participants to help identify
239 the potential this platform may have as a healthcare delivery technology in the future, by
240 identifying major themes emerging from the study.

241 Usability Study Design

242 Participants who volunteered attended a single, one hour session with the study coordinator after
243 their scheduled appointment and were asked to complete four main tasks: 1) inputting
244 measurements on *Medly* smartphone app, 2) inputting measurements on *Medly* voice app, 3)
245 completing the SUS questionnaire, and 4) engaging in a semi-structured interview with the study
246 coordinator. The first task asked the participant to interact with the *Medly* app on a phone
247 provided by the study coordinator. While the participant was completing this task, the study
248 coordinator recorded observational notes. Following this, the participants were then asked to
249 verbally interact with the *Medly* voice app using a VUI, which was provided by the study
250 coordinator. The user was also provided with an instruction card with steps to follow that
251 included suggestive phrasing to help guide them through the conversation. During this time, the

252 facilitator also recorded observational notes. Participants then completed a standardized, post-
253 study SUS questionnaire regarding their experience with the voice app. Once participants
254 completed the questionnaire, the study coordinator led a semi-structured interview with the
255 patient which included questions pertaining to their perceptions of the usability of the *Medly*
256 voice app, and specific issues or areas for improvement. The results of this research highlighted
257 key findings regarding the acceptability of integrating voice communication platforms into
258 patient-self-care tools, and informed decisions for future design iterations.

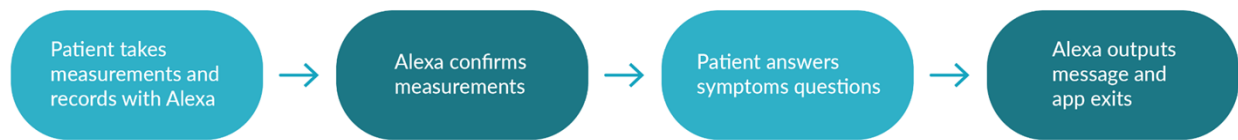
259 Results

260 Preliminary Design of Medly Voice App

261 A voice app prototype of *Medly* was created for the purpose of it being used during a usability
262 study in order to identify problems in the design, uncover any opportunities for improvement,
263 and to learn more about the users' behaviors and preferences when interacting with the app. An
264 instructions card was created and used to help guide the user through the *Medly* voice app
265 interaction (Figure S1, Multimedia Appendix 1). A preliminary design of the card was used by
266 participants in the study so that feedback could be collected about the design to make changes, if
267 required.

268 A high level overview of the conversation was first mapped out and created (Fig 3). Using this,
269 the full conversation was designed and is described in more detail in Figure S2 in Multimedia
270 Appendix 1. The first half of the voice app interaction consisted of the user measuring and
271 recording their weight, blood pressure, and heart rate with non-lyrical music playing in the
272 background in between prompts. Playing music in the background not only created a more

273 pleasant user experience, but was also a strategy used to ensure the app did not time out (28).
274 After the user inputted their heart rate, the voice app reiterated the measurements it captured to
275 the user and gave them an opportunity to correct any wrong measurements. Once the user
276 recorded their measurements, the app began to ask them a series of yes/no questions of HF-
277 related symptoms. Once complete, the app outputted a message and exited. The *Medly* algorithm
278 was not implemented for the prototype because it was not needed to accomplish the purpose of
279 the usability study.



280

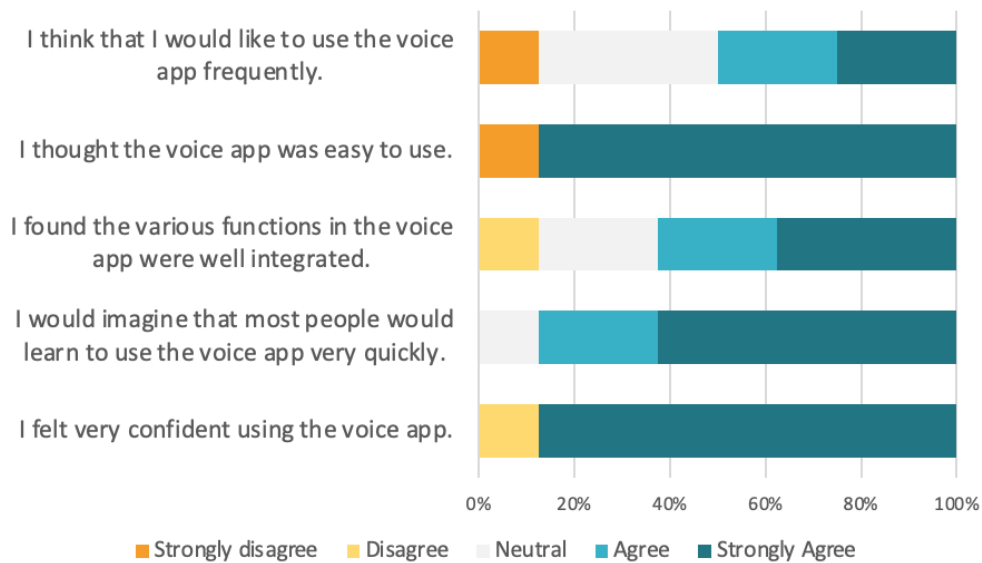
281 **Fig 3. A high level description of the conversational flow for the Medly voice app. ‘Milestones’ are outlined**
282 **and signifies an important event.**

283 An interaction model was then created to implement the conversational flow described above
284 and consisted of intents, training phrases, and slot types. An intent was created for each piece of
285 required information: weight, blood pressure, heart rate, and symptom questions. Within each
286 intent, training phrases were added and consisted of predictable utterances the user may say to
287 record their responses. Within each training phrase, the most important pieces of data were
288 identified and labeled as a slot type (weight, blood pressure, heart rate categorized as ‘numbers’,
289 with symptom responses labeled as ‘yes’ or ‘no’). The slot types were used to help extract the
290 measurements needed to send to the *Medly* voice server. Measurements were extracted from the
291 interaction model and the application logic was used to help direct the flow of conversation.
292 Once the required measurements were captured, a POST request was made to the *Medly* server,
293 and a self-care message sent back to the voice app to relay to the user.

294 Usability Study Findings

295 A usability study was performed with eight HF participants from UHN’s Ted Rogers Center of
296 Excellence for Heart Failure clinic. Three out of the eight participants were able to successfully
297 complete the usability session, while the other participants were unsuccessful due to the voice
298 app exiting because of various errors that occurred, mainly with the device unsuccessfully
299 understanding numerical values the user would say. Despite these results, all participants filled
300 out the SUS questionnaire to describe their experience with the voice app. The responses from
301 this questionnaire resulted in an average SUS score of 92 (out of 100), ranking the voice app in
302 the 98th percentile based on previous studies. Most participants were satisfied with the design and
303 development of the voice app. Participant responses for statements relating to the positive and
304 negative attributes of the voice app are shown in Figs 4a and 4b, respectively.

SUS Questions Relating to Positive Attributes

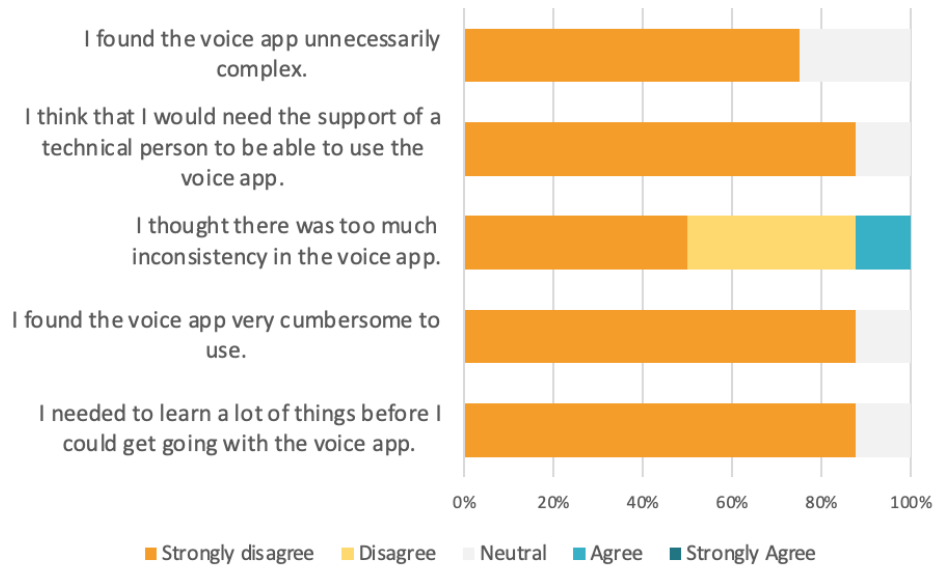


305

306

(a)

SUS Questions Relating to Negative Attributes



307

308

(b)

309

Fig 4. (a) Data showcasing participant results for the questions in the SUS relating to positive attributes of

310

the Medly voice app system (b) Data showcasing participant results for the questions in the SUS relating to

311

the negative attributes of the Medly voice app system.

312

Looking at the positive attribute statements, almost all participants (7 out of the 8) felt confident

313

using the app and thought it was easy to use. This is further supported with the responses to the

314

negative attributes statements, with the most popular response being ‘strongly disagree’. When

315

asked about whether they would use the voice app frequently however, there was a greater divide

316

in opinion, with the most popular response being neutral.

317

Overall, these findings show promise when determining user acceptability due to the overall

318

positive responses from the SUS questionnaire. These results also indicate that some minor

319

changes should occur in order to improve the patient experience. Participants who were unable to

320

complete the session had difficulties recording their blood pressure readings, either due to the

321

way they said the readings (saying each digit separately for systolic and diastolic), or if they took

322 a small pause between values, in which case the voice app incorrectly recorded the values. In
323 these cases the users were not able to correct the measurements. As a result, the blood pressure
324 part of the interaction model was further developed to mitigate the risk of incorrectly and/or
325 unsuccessfully capturing the user's data.

326

327 In addition to the SUS questionnaire results, observations were derived from the semi-structured
328 interviews that occurred with each user after their usability session. For example, an obvious
329 change in physical behaviour was observed with participants when interacting with the voice
330 app, with most users showcasing nervous and tense feelings during the session. Furthermore,
331 some voice app features helped the user navigate through their session making for a more
332 pleasant experience, such as: having the music play in the background so they didn't feel rushed
333 through the interaction, having the voice app reiterate measurements back to them to ensure the
334 device correctly recorded the values, as well as having the instructions card available with
335 suggested phrases to use. General frustrations were expressed by participants when using a
336 smartphone, and concerns about privacy were brought up by two (out of eight) participants.
337 These findings are described in more detail in Table S1 in Multimedia Appendix 2.

338 Prior to the second part of the development phase, a list of potential risks were identified that
339 relate to the user interacting with the voice app. The severity of each risk was assessed based on
340 its likelihood and consequence. Mitigation strategies for each risk were developed and
341 implemented in the voice app design during the redevelopment work. This data can be seen in
342 Table S2, Multimedia Appendix 2.

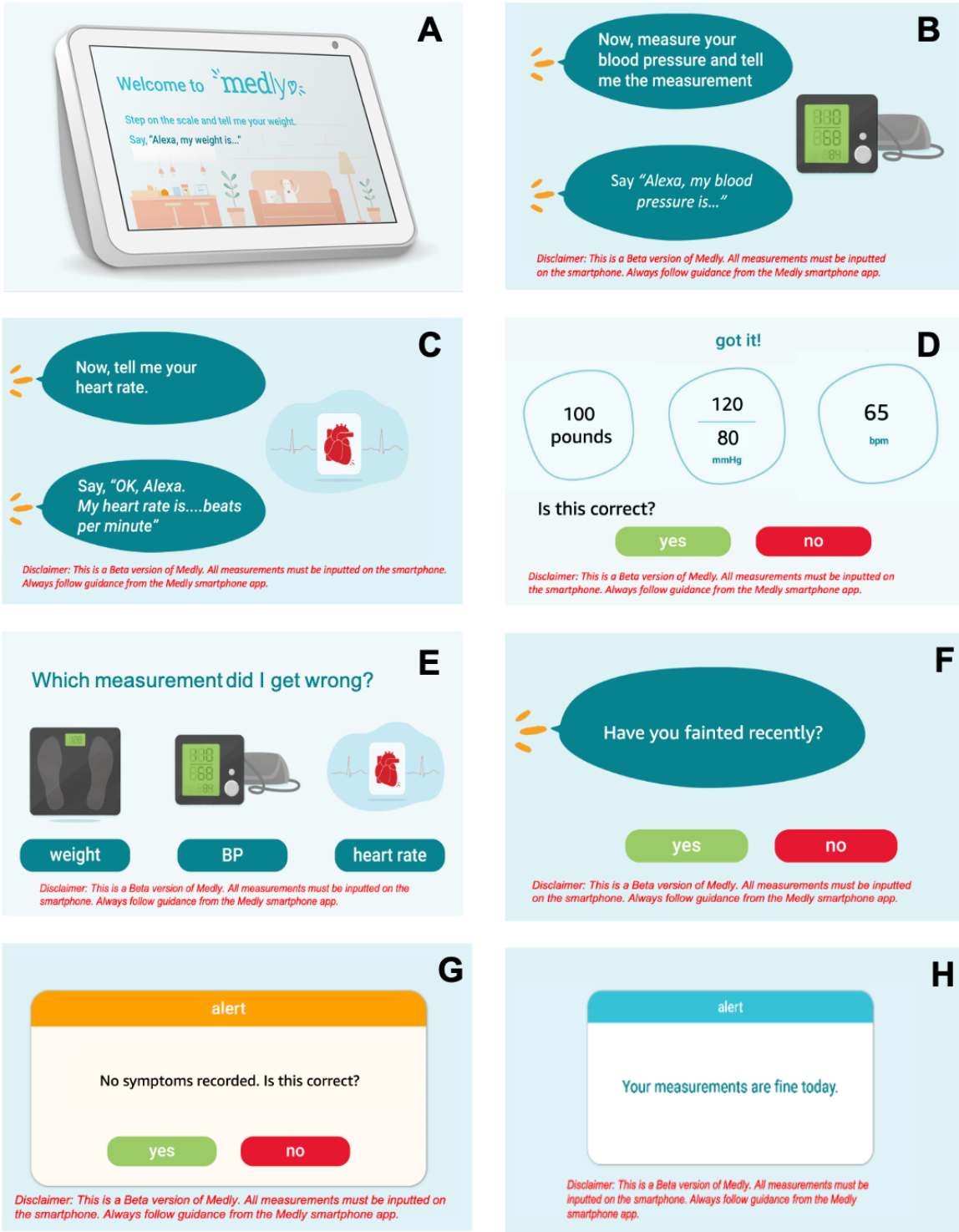
343 Redevelopment of *Medly* Voice App

344 The prototype that was built for the usability study was further developed to meet all of the
345 *Medly* voice app design requirements. The new features added were influenced by the findings
346 from the usability study and also involved changing VUIs to incorporate screen and touch screen
347 capabilities.

348 Screen Design

349 Having a screen display on the device meant that users were able to visually see instructions and
350 measurements, and had the option to respond using the touch screen for yes/no questions. Each
351 screen was designed to only include the necessary information the user needed to avoid any
352 confusion. A consistent theme, similar to the *Medly* smartphone app, was used for each screen
353 display by using the same color scheme, layout, font size and style. The written cues on the
354 screen aligned with the verbal prompts to avoid user confusion. Design guidelines were used for
355 font size and layout to ensure the prompts are presented at an appropriate reading size (29). The
356 screen design at various points of the conversation is showcased in Fig 5.

357 Incorporating a screen display for the *Medly* voice app would also allow users to see their
358 previous measurements and data trends, similar to what they're used to seeing when using the
359 smartphone app. While the implementation of seeing data history was not in scope for this
360 project, it is a feature that is common in most mHealth apps since it helps users understand their
361 health status better.



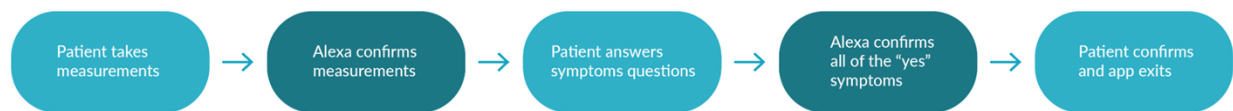
362

363 **Fig 5. Screen designs showcased through images A-D, representing different events that occur in the**
 364 **conversation. A disclaimer is said verbally at the start of skill interaction, and in between a written disclaimer**
 365 **is provided on the screen. (A) welcome screen that users see when they invoke the skill, (B) and (C)**

366 instructions to measure blood pressure and heart rate, respectively , (D) confirmation slide of the user's
367 weight, blood pressure, and heart rate, (E) what the user sees when they are correcting a measurement, (F)
368 an example of how the symptom questions appear on the screen, (G) the confirmation message for the
369 symptom responses, (H) an example of the output message provided by the Medly algorithm.

370 Conversational Design

371 The conversational script used in the preliminary design stayed the same during the re-
372 development work. Findings from the usability study helped inform which parts of the
373 conversation were most prone to error and needed to be changed. More training phrases and
374 symptom confirmations were new additions to the conversation and the summary of the
375 interaction is shown in Fig 6. A more detailed description of the new additions to the
376 conversation can be seen in Figure S3 in Multimedia Appendix 1. With these new changes, the
377 app only confirms the symptoms the user answered yes to for a more efficient process by giving
378 them an opportunity to correct any wrong responses recorded. Once all the responses have been
379 corrected, the data was sent to the *Medly* voice server. The voice server retrieved patient
380 thresholds, sent the daily data to the *Medly* research server and used the *Medly* algorithm to send
381 an output message back to the voice app, informing the patient on the status of their health.



382

383 **Fig 6. A high level description of the conversational flow for the final version of the Medly voice app. A new**
384 **milestone has been added to describe Medly voice app also confirming the symptom responses answered**
385 **'yes'**

386 Back End Development

387 The software architecture and *Medly* algorithm were implemented during the redevelopment
388 phase of the project. Each output message generated to the user was a personalized response
389 based on their measurements and pre-existing thresholds. In order to receive data from the voice
390 app and pull relevant patient thresholds, account linking was required to be set-up to connect the
391 patient's identity from the two different services. Personalized *Medly* accounts were created for
392 each study participant and connected to their respective Amazon account. The *Medly* algorithm
393 was built and deployed on the *Medly* voice server and pulls data from *Medly* application server
394 as well as the data sent from the VUI to generate an output message.

395 Discussion

396 Principal Findings

397 This manuscript describes the UCD approach that was used to develop a voice app experience
398 for HF management. Accessing chronic disease management programs through the use of voice
399 apps has the potential to increase these programs' uptake by making them more accessible. The
400 results from this usability study show promise in the possibility of offering these types of
401 programs as a voice app, and the development process proves the feasibility of deploying these
402 programs on VUIs.

403 VUIs are designed to be integrated within a household by having the ability to connect to
404 multiple devices throughout the home ("smart devices"), and also provide the ability to connect
405 to online services by just using voice. Their near ubiquitous nature makes it easy and convenient
406 to perform tasks through simple conversation and offers the flexibility to do other tasks

407 simultaneously. Patients with chronic illness experience a constant need to record and transmit
408 data to better track their health and VUIs have the potential to help them accomplish this
409 seamlessly, not only because it uses speech to communicate but also because it can help guide
410 the user through their tasks more explicitly than the smartphone.

411 Because this platform requires less technical knowledge than the smartphone (mHealth apps),
412 those who struggle to use technology have the opportunity to participate in the program. Based
413 on the usability study observations, most participants felt that it was easy to interact with the
414 voice app since it only required a conversation and no technical background. The platform's
415 multilingual capability offers another opportunity, which may make some more willing to log
416 their measurements. VUIs also allow users to initiate a call to anyone on their list through voice
417 command, which can be useful in cases where they need assistance but cannot reach their phone
418 to call for help. This scenario often occurs with older demographics who are also the most
419 common group to be diagnosed with chronic illnesses, and as a result are prescribed digital
420 therapeutics.

421 When inputting measurements on a mHealth app, it is typical for the user to see their data on the
422 screen prior to submitting. With a VUI that solely relies on verbal communication, visual
423 feedback is no longer a possibility. Study participants commented on the importance of knowing
424 which measurements were recorded on the voice app, and valued the ability to see historical data
425 and trends on the smartphone app. With these findings, we incorporated verbal feedback as a
426 necessary component in the conversational design to help with user confidence. Adding this step
427 in the conversation comes at the cost of prolonging the time it takes to complete the interaction,
428 but is not necessarily a drawback due to the lower focus required when compared to using a
429 smartphone. Advancements in VUI technology include the ability for the user to interrupt the

430 device when speaking, and an integrated screen display which can allow the users to view results
431 (as they are currently accustomed to), as well.

432 Although users only need to use speech when interacting with VUIs, the conversation style may
433 vary significantly when compared with human-to-human interaction. During the study, users
434 who had never used these types of devices before felt the need to raise their voice, in hopes that
435 it heard and understood them better. Expelling this energy repeatedly may make the user feel
436 tired, as one participant noted, especially since the demographic who uses these digital
437 therapeutics are older adults. Without knowing how the voice app will respond, a difference in
438 body language was also observed, namely users would tense up, felt the need to sit up straighter,
439 and focused more when interacting with the device. This behaviour is in stark contrast to how
440 users interacted with the smartphone, since the physical body language requires a much more
441 relaxed behaviour. Despite these differences, it is expected that user body language will shift
442 towards a more relaxed behaviour as they get more comfortable using voice apps.

443 It is also important to consider the potential barriers to entry that exist as a result of using this
444 technology. For example, most VUI devices require the user to have an account to interact with
445 the device, creating an accessibility issue. Another barrier to entry is the requirement to have a
446 constant, reliable internet connection to use the device. This is in contrast to the *Medly*
447 smartphone app which can be completed offline and is mobile. Not only can this create an equity
448 issue for those who do not have internet, but can also be problematic when internet issues occur
449 and power outages.

450 As advancements in VUIs progress, we believe they will play an integral role in providing access
451 to chronic disease management programs, by: 1) helping more patients complete their tasks

452 independently, 2) offering a more convenient experience to record relevant data, and 3) allowing
453 those with limited technical and English skills access to these programs.

454 Limitations

455 While the results from this research project show promise for future use, it is also important to
456 acknowledge the limitations associated with this work.

457 The main limitation for usability studies is the short interaction time. While the observations and
458 feedback from participants was helpful in identifying the potential voice apps may have for
459 chronic disease management, they only interacted with the technology for a short period of time.
460 Future work should include a study that requires the users to interact with the device for a longer
461 period of time to give them more experience with the voice app to reflect more grounded
462 feedback.

463 A sample size of eight participants was chosen since the usability study was only focused on
464 gathering user insights, and not statistically relevant data. According to the literature, the total
465 participants needed for a usability study is dependent on when saturation in results is achieved.
466 In most cases this occurs when the sample size is typically between 5-8 participants (30).
467 Therefore, it is recommended that multiple usability tests occur with fewer users (between 5-8
468 participants) and that changes be made between usability tests to mitigate challenges observed in
469 previous sessions. The research showcased in this paper underwent 1½ cycles of the UCD
470 process and requires further testing to identify whether the challenges uncovered in the usability
471 study are resolved with the latest iteration of design and development work. Due to the small
472 sample size, the usability study is also not statistically powered. In this case, the evaluation of the
473 results is mostly limited to qualitative data analysis. While a questionnaire was completed by

474 each user after the session, these results are only used to support the conclusions drawn from the
475 qualitative data.

476 Lastly, participants may have had social desirability bias when answering interview questions
477 and filling out the questionnaire, especially because they knew that the study coordinator also
478 had involvement in the design and development of the voice app (31). When things go wrong,
479 participants most often feel as though it is their fault for not understanding how to use the
480 technology, and as a result will not bring up comments about what they disliked about the
481 experience since they think that it is their fault.

482 Conclusion

483 This project involved the design, development, and usability evaluation of a voice app for HF
484 management. A UCD process was followed to systematically create a voice app that would meet
485 user needs and be easy to use. The usability study performed at the PMCC at the University
486 Health Network provided insightful user feedback about the voice app design, with the overall
487 response being positive with high user satisfaction. The findings from this usability study
488 impacted the re-development of the voice app, which will be used in a future clinical evaluation.
489 The findings from this research show promise in using VUIs to help with chronic disease
490 management.

491 Conflicts of Interest

492 JC and HR are part of the team that founded the *Medly* system under the intellectual property
493 policies of the UHN and may benefit from future commercialization of this technology.

494 Acknowledgments

495 The authors wish to thank Jacqueline Simpson for designing the *Medly* voice app screensavers
496 and figures presented in this paper. The authors are also grateful to all of the patients who gave
497 their time and participated in this study.

498 References

- 499 1. Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev.* 2017
500 Apr;3(1):7–11.
- 501 2. What is heart failure? [Internet]. www.heart.org. [cited 2022 Feb 17]. Available from:
502 <https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure>
- 503 3. Lippi G, Sanchis-Gomar F. Global epidemiology and future trends of heart failure. *AME*
504 *Med J.* 2020 Jun;5:15–15.
- 505 4. Megari K. Quality of Life in Chronic Disease Patients. *Health Psychol Res.* 2013 Sep
506 24;1(3):e27.
- 507 5. Seto E, Leonard KJ, Cafazzo JA, Masino C, Barnsley J, Ross HJ. Self-care and quality of
508 life of heart failure patients at a multidisciplinary heart function clinic. *J Cardiovasc Nurs.*
509 2011 Sep;26(5):377–85.
- 510 6. Chen X, Koskela M, Hyvääkkä J. Image based information access for mobile phones.
511 *Proceedings - International Workshop on Content-Based Multimedia Indexing.* 2010;3:130–
512 5.
- 513 7. Irfan Khan A, Gill A, Cott C, Hans PK, Steele Gray C. mHealth Tools for the Self-
514 Management of Patients With Multimorbidity in Primary Care Settings: Pilot Study to
515 Explore User Experience. *JMIR Mhealth Uhealth.* 2018 Aug 28;6(8):e171.
- 516 8. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for
517 health-promoting behavior among individuals with chronic medical conditions. *Digit*
518 *Health.* 2019 Jan;5:2055207619882181.
- 519 9. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a Heart Failure
520 Telemonitoring Program Implemented as the Standard of Care in an Outpatient Heart
521 Function Clinic: Pretest-Posttest Pragmatic Study. *J Med Internet Res.* 2020 Feb
522 8;22(2):e16538.
- 523 10. Elboim-Gabyzon M, Weiss PL, Danial-Saad A. Effect of Age on the Touchscreen
524 Manipulation Ability of Community-Dwelling Adults. *Int J Environ Res Public Health*
525 [Internet]. 2021 Feb 21;18(4). Available from: <http://dx.doi.org/10.3390/ijerph18042094>
- 526 11. Naumann AB, Wechsung I, Hurtienne J. Multimodal interaction: A suitable strategy for
527 including older users? *Interact Comput.* 2010 Sep 24;22(6):465–74.
- 528 12. Nunes F, Silva PA, Cevada J, Correia Barros A, Teixeira L. User interface design
529 guidelines for smartphone applications for people with Parkinson’s disease. *Univers access*
530 *inf soc.* 2016 Nov;15(4):659–79.
- 531 13. Part 6: Rehabilitation and Assistive Robotics. *Technol Disabil.* 2019 Jul 17;31(s1):147–54.

- 532 14. Williams P, Shekhar S. People with Learning Disabilities and Smartphones: Testing the
533 Usability of a Touch-Screen Interface. *Education Sciences*. 2019 Oct 27;9(4):263.
- 534 15. Patrick PA, Obermeyer I, Xenakis J, Crocitto D, O’Hara DM. Technology and social media
535 use by adult patients with intellectual and/or developmental disabilities. *Disabil Health J*.
536 2020 Jan;13(1):100840.
- 537 16. Kravitz N, Levanon Y, Cukierman-Yaffe T, Nota A, Kizony R, Rand D. Sensorimotor and
538 Cognitive Abilities Associated With Touchscreen Tablet App Performance to Support Self-
539 Management of Type 2 Diabetes. *Am J Occup Ther*. 2021 Jan;75(1):7501205080p1–9.
- 540 17. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient Adherence to a
541 Mobile Phone–Based Heart Failure Telemonitoring Program: A Longitudinal Mixed-
542 Methods Study (Preprint) [Internet]. Available from:
543 <http://dx.doi.org/10.2196/preprints.13259>
- 544 18. Sezgin E, Noritz G, Elek A, Conkol K, Rust S, Bailey M, et al. Capturing At-Home Health
545 and Care Information for Children With Medical Complexity Using Voice Interactive
546 Technologies: Multi-Stakeholder Viewpoint. *J Med Internet Res*. 2020 Feb
547 13;22(2):e14202.
- 548 19. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth
549 consumer apps: the case for user-centered design. *Biomed Instrum Technol*. 2012
550 Autumn;Suppl:49–56.
- 551 20. What is the Alexa Skills Kit? [Internet]. alexa developer documentation. Available from:
552 [https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-](https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html)
553 [kit.html](https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html)
- 554 21. Murad C, Munteanu C, Clark L, Cowan BR. Design guidelines for hands-free speech
555 interaction. In: *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International*
556 *Conference on Human-Computer Interaction with Mobile Devices and Services,*
557 *Conference Proceedings Adjunct* [Internet]. 2018. Available from:
558 <http://dx.doi.org/10.1145/3236112.3236149>
- 559 22. Build Your Design Artifact [Internet]. alexa developer documentation. Available from:
560 <https://developer.amazon.com/en-US/docs/alexa/alexa-design/build-artifact.html>
- 561 23. Chakraborty P. Creating voice interaction flows [Internet]. UX Collective. 2018 [cited 2021
562 Jun 28]. Available from: [https://uxdesign.cc/ux-design-for-voice-interfaces-part-ii-](https://uxdesign.cc/ux-design-for-voice-interfaces-part-ii-3b0056020cd3)
563 [3b0056020cd3](https://uxdesign.cc/ux-design-for-voice-interfaces-part-ii-3b0056020cd3)
- 564 24. Alexa Skillts Kit [Internet]. Available from: [https://developer.amazon.com/en-](https://developer.amazon.com/en-US/alexa/alexa-skills-kit)
565 [US/alexa/alexa-skills-kit](https://developer.amazon.com/en-US/alexa/alexa-skills-kit)
- 566 25. AWS Lambda [Internet]. AWS. Available from: <https://aws.amazon.com/lambda/>
- 567 26. Nielsen J, Landauer TK. Mathematical model of the finding of usability problems. In:

- 568 Conference on Human Factors in Computing Systems - Proceedings [Internet]. 1993.
569 Available from: <http://dx.doi.org/10.1145/169059.169166>
- 570 27. Virzi RA. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?
571 Hum Factors. 1992 Aug 1;34(4):457–68.
- 572 28. Voice Design for Alexa Experiences [Internet]. alexa developer documentation. Available
573 from: <https://developer.amazon.com/en-US/docs/alexa/alexa-design/available.html>
- 574 29. Alexa Design System for Alexa Presentation Language (APL) [Internet]. alexa developer
575 language. Available from: [https://developer.amazon.com/en-US/docs/alexa/alexa-](https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-alexa-packages-overview.html)
576 [presentation-language/apl-alexa-packages-overview.html](https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-alexa-packages-overview.html)
- 577 30. Why you only need to test with 5 users [Internet]. [cited 2021 Jun 28]. Available from:
578 <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users>
- 579 31. Grimm P. Social Desirability Bias. In: Wiley International Encyclopedia of Marketing
580 [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2010. Available from:
581 <http://doi.wiley.com/10.1002/9781444316568.wiem02057>

582 Abbreviations

583 HF: heart failure

584 PMCC: Peter Munk Cardiac Center

585 SUS: System Usability Scale

586 UCD: user-centered design process

587 UHN: University Health Network

588 VUI: voice user interface

589 Supporting Information

590 Multimedia Appendix 1 Table. S1

591 Multimedia Appendix 1 Fig. S1

592 Multimedia Appendix 1 Fig. S2

593

594 Multimedia Appendix 2 Table. S1

595 Multimedia Appendix 2 Table. S2

596

597 Multimedia Appendix 3 Fig. S1