

1

2

3

4

A Voice App Design for Heart Failure Self-Management: A Pilot Study

6

7

8

9 Antonia Barbaric MASc^{1,2,6}, Cosmin Munteanu PhD³, Heather J. Ross MD MHSc^{4,5}, and Joseph
10 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 **Abstract**

28 There is a growing interest to investigate the feasibility of using voice user interfaces as a
29 platform for digital therapeutics in chronic disease management. While mostly deployed as
30 smartphone applications, some demographics struggle when using touch screens and often
31 cannot complete tasks independently. This research aimed to evaluate how heart failure patients
32 interacted with a voice app version of an already existing digital therapeutic, *Medly*, using a
33 mixed-methods concurrent triangulation approach. The objective was to determine the
34 acceptability and feasibility of the voice app by better understanding who this platform is best
35 suited for. Quantitative data included engagement levels and accuracy rates. Participants (n=20)
36 used the voice app over a four week period and completed questionnaires and semi-structured
37 interviews relating to acceptability, ease of use, and workload. The average engagement level
38 was 73%, with a 14% decline between week one and four. The difference in engagement levels
39 between the oldest and youngest demographic was the most significant, 84% and 43%
40 respectively. The *Medly* voice app had an overall accuracy rate of 97.8% and was successful in
41 sending data to the clinic. Users were accepting of the technology (ranking it in the 80th
42 percentile) and felt it did not require a lot of work (2.1 on a 7-point Likert scale). However, 13%
43 of users were less inclined to use the voice app at the end of the study. The following themes and
44 subthemes emerged: (1) feasibility of clinical integration: user adaptation to voice app's
45 conversational style, device unreliability, and (2) voice app acceptability: good device
46 integration within household, users blamed themselves for voice app problems, and voice app
47 missing desirable user features. The voice app proved to be most beneficial to those who: are

48 older, have flexible schedules, are confident with using technology, and are experiencing other
49 medical conditions.

50 **Introduction**

51 **Background**

52 Chronic diseases are the leading cause of death and disability worldwide, with over 41 million
53 people dying each year due to these diseases (1). Cardiovascular diseases, such as heart attacks
54 and high blood pressure are responsible for most chronic disease deaths (17.9 million people)
55 (1). Patient self-care is considered to be essential in the prevention and management of chronic
56 diseases (2) as studies have shown the benefits of this approach, which include improved health
57 outcomes, decreased clinic visits, and decreased health costs (3). Mobile health, also referred to
58 as mHealth, is a type of digital health technology that involves the use of mobile devices for
59 medical and public health practice (4), and enables the integration of self-care support into a
60 patient's routine (5). While mHealth apps are one of the most popular tools for helping patients
61 with chronic conditions manage their health at home (6), there is research to suggest voice apps
62 are an emerging platform that will create alternative interaction models that some patient
63 demographics may find more accessible.

64 Voice user interfaces (VUIs) are becoming more prevalent in the healthcare field for a variety of
65 different purposes. With VUIs the user is able to interact with a computing system using only
66 speech, with voice apps being one example of this technology. The primary advantage of
67 implementing VUIs in any environment is simplicity, since it does not require the user to interact
68 with a hand-held technology, as we are typically accustomed to. So far, VUIs have been used to
69 help those who have speech or hearing difficulty, to improve patient engagement, as well as
70 aging in place (7). This technology has also been used in the clinical setting by supporting

71 physician note transcription and the patient registration process (7). Devices that offer VUI
72 capability are also continuing to gain popularity in consumer households and are becoming more
73 integrated into our daily lifestyles due to their convenience, ease of use, and affordability. As a
74 result, there is growing interest to investigate the feasibility of using smart speakers to improve
75 patient engagement, with a specific focus on chronic disease management.

76 **Heart Failure**

77 Previous research has begun to investigate the potential for patients to manage their heart failure
78 (HF) using a voice app (publication pending). HF is a condition that develops after the heart
79 muscle becomes damaged or weak due to cardiovascular diseases, such as heart attacks and high
80 blood pressure. When the heart muscle becomes damaged or weakens, it is unable to pump
81 enough blood to meet the body's needs for blood and oxygen (8). *Medly* is an evidence-based,
82 HF self-management program that has been developed by the University Health Network (UHN)
83 and is implemented as part of the standard of care at UHN's Ted Rogers Center of Excellence for
84 Heart Failure clinic (9). This program is currently available to patients through a smartphone app
85 and enables them to log clinically relevant physiological measurements and symptoms, which is
86 then used in the *Medly* algorithm to generate an automated self-care message. A voice app
87 version of *Medly* has already been built as part of previous work, and a usability study has been
88 performed with the voice app at UHN's Heart Failure Clinic.

89 **Objectives**

90 The results from a previous usability study show promise that a voice app for chronic disease
91 management, such as HF, is feasible to deploy and acceptable by patients (publication

92 forthcoming). With these findings, we sought to perform a more in-depth clinical evaluation by
93 using *Medly* as a case study, with 20 HF participants. The goal of this study is to determine if
94 voice apps can be a practical alternative of enabling patients to receive a digital therapeutic. The
95 following research question will be investigated:

96 *What is the acceptability of a voice application for patients, through the use of a smart*
97 *speaker, for a home chronic disease management platform?*

98 Through this research we hope to uncover whether deploying a voice app version of *Medly* adds
99 any benefit to the current model of interaction and care.

100 **Methods**

101 **Participant Recruitment**

102 HF patients were asked to participate in this study and interact with the *Medly* voice app while in
103 their homes for a four week period. Participants were considered eligible if they had been
104 diagnosed with HF by a physician at UHN's HF clinic, and were prescribed the *Medly* program.
105 Participants were also required to speak and read English adequately to understand the voice
106 prompts in the *Medly* application. The *Medly* nurse coordinator first provided a brief overview of
107 the research study to interested patients prior to introducing them to the study coordinator. If they
108 agreed to participate, informed written consent was obtained by the study coordinator prior to
109 onboarding.

110 A total of 20 participants were recruited for this study, based on findings from the literature
111 which suggested that a sample size between 10 and 30 users (10) is appropriate to use for pilot

112 studies. To help mitigate potential bias, we aimed to recruit both participants who were ‘new’ to
113 *Medly* (less than 2 month since being onboarded to the program), and also those who were
114 ‘existing’ (more than 2 months since being onboarded) *Medly* patients. In the end we recruited a
115 total of 7 new and 13 existing *Medly* patients.

116 Recruited participants were required to perform a double-entry of their *Medly* measurements for
117 the four week duration, more specifically they were asked to first input their *Medly*
118 measurements on the smartphone app prior to interacting with the voice app. Each participant
119 received a gift card to compensate them for their time participating in the study. Ethics approval
120 was obtained from the University Health Network (UHN) Research Ethics Board (20-6095).

121 **Study Outcome Measures**

122 The evaluation of the *Medly* voice app was influenced by two frameworks: Proctor et al.’s
123 Implementation Outcomes (11) and Unified Theory of Acceptance and Use of Technology 2
124 (UTAUT2) (12). The outcomes listed below were selected from both frameworks to help guide
125 the quantitative and qualitative data that was gathered to determine the acceptability of the *Medly*
126 voice app.

127 **Table 1. Definitions of the selected outcomes chosen from frameworks to identify the**
128 **acceptability of using the *Medly* voice app.**

<i>Proctor et al.’s Implementation Outcomes</i>

Acceptability	The perception among patients that the <i>Medly</i> voice app is agreeable or satisfactory.
Feasibility	The extent to which the <i>Medly</i> voice app can be successfully used by patients.
<i>Unified Theory of Acceptance and Use of Technology 2 (UTAUT2)</i>	
Effort expectancy	The degree of ease associated with using the <i>Medly</i> voice app.

129 Questionnaires and semi-structured interview questions were influenced by the System Usability
130 Scale (SUS) (13) and NASA Task Load Index (TLX) (14) standardized assessment tools.
131 Quantitative data was also gathered through semi-structured interviews by asking participants
132 how often the voice app misheard their measurements, how many times they were required to
133 correct wrongly recorded data, and how many times they missed inputting their measurements
134 and why (engagement levels). Accuracy rates were calculated by comparing the measurements
135 inputted on the smartphone app versus the voice app.

136 **Data Collection**

137 The study coordinator performed an onboarding session over the phone with each participant to
138 help them set-up and access the *Medly* voice app, and provided them with an instructions manual

139 (Fig S1, Multimedia Appendix 1). Participants were also required to answer a baseline
140 questionnaire (Table S1, Multimedia Appendix 2) to help the study coordinator understand their
141 comfort levels with using technology. Participants were made aware that they needed to perform
142 a double entry of their *Medly* measurements for the four week duration and were told to prioritize
143 the *Medly* smartphone app, namely to input measurements on the phone first, and to only follow
144 guidance from the smartphone app.

145 A one week check-in was scheduled with all participants over the phone. During this check-in,
146 the study coordinator asked each participant about their experience so far and collected the
147 quantitative data described above. Following the week one check-in, participants were emailed a
148 questionnaire at the end of week two. The questions were influenced by the frameworks and
149 standardized assessment tools mentioned earlier in order to gauge the participant's thoughts and
150 opinions of using the voice app thus far. As part of each participant's offboarding, the
151 questionnaire was sent out again (to see if thoughts or opinions changed), and a semi-structured
152 interview took place with the study coordinator over the phone.

153 **Study Analysis**

154 Descriptive statistics for the standardized questionnaire responses were performed and recorded
155 using Microsoft Excel. Graphical representations of engagement levels were also created using
156 Excel. The responses from the System Usability Scale questionnaire were analyzed as per
157 standard protocol (13) and averages were calculated for the NASA-TLX and UTAUT2
158 questionnaires, both overall and question specific. Data was categorized in different ways using
159 various characteristics and then analyzed to identify any trends or commonalities.

160 The quantitative data was then triangulated with the qualitative data findings, namely from the
161 semi-structured interviews. Interview transcripts were analyzed by the study coordinator (AB)
162 using an deductive approach. Specifically, the transcripts were analyzed under the guidance of
163 Proctor et al.'s Implementation Outcomes framework, with a focus on the acceptability and
164 feasibility constructs. Sub-themes were then identified to better describe the study findings. The
165 transcripts and coding was organized using Microsoft Word.

166 **Results**

167 **Characteristics of Study Participants**

168 A total of 20 patients were recruited for the study, with a fairly even split among genders
169 (females: 9/20, 45%, male: 11/20, 55%) and an average age of 57.8 (SD 13.1) years. All patients
170 who were recruited were required to be enrolled in the *Medly* program, with a mix between those
171 who have just recently (defined as less than 2 months since the time of recruitment) been
172 onboarded to the program (7/20 patients, 35%) and those who have been enrolled in the program
173 for longer (13/20 patients, 65%). Other patient characteristics were also collected for the
174 purposes of this study, such as: comfort levels with technology and whether or not they have
175 used a smart speaker before through a baseline questionnaire (18 out of the 20 participants
176 returned this questionnaire). The statistics for each of these characteristics can be seen in Table
177 2.

178 **Table 2. Patient characteristics used to categorize and sort data in the study.**

Characteristics	Statistics
Age (years), mean (SD)	57.8 (13.1)
Age (years; categorical), n (%)	
20 or younger	0 (0.0)
21 - 40	2 (10.0)
41 - 60	7 (35.0)
61 - 80	11 (55.0)
81 or older	0 (0.0)
Sex, n (%)	

Female	7 (35.0)
Male	13 (55.0)
Interacted with smart speaker before, n (%)	
Yes	7 (35.0)
No	11 (55.0)
Comfort level with technology, n(%)	
Very uncomfortable	1 (5.0)
Somewhat uncomfortable	0 (0.0)
Neutral	6 (30.0)

Somewhat comfortable	2 (10.0)
Very comfortable	9 (45.0)

179 **Quantitative Data**

180 **Engagement Levels and Accuracy Rates**

181 The overall engagement level for the entire study population during the four week period was
 182 73%, with noticeable drops in engagement as the weeks progressed (Table 3) and an overall
 183 decline of 14% when comparing week one and four averages.

184 **Table 3. Overall and weekly average engagement levels over the four week study duration.**

Overall engagement level/SD (%)		73.4 / 9.5	
<i>Week #</i>	<i>Average (%)</i>	<i>Average Days Missed</i>	<i>SD (%)</i>
1	80.7	1.4	11.3
2	75.0	1.8	5.8

3	70.7	2.0	7.9
4	67.1	2.3	8.1

185 In addition to calculating the overall engagement levels, patient characteristics from Table 3
186 were also used to group the study population and compare the results among sub-groups to
187 identify any noticeable trends. These results can be seen in Table S1, Multimedia Appendix 3.

188 In summary, there were no noticeable trends when comparing voice app engagement levels
189 among new and existing *Medly* patients, with an ~1% difference overall. Similar to the findings
190 related to the entire study population, engagement levels were lower in week four when
191 compared to week one for both of these groups.

192 In contrast, engagement levels when compared to the different age groups showcased more
193 obvious trends. Overall, average engagement levels increased as the age groups increased, with
194 the oldest demographic (61-80 year olds) having the best engagement level of 84.1%, almost
195 double the overall engagement level when compared to the youngest age group in the study.
196 Those in the 61-80 age group were also the most consistent throughout the four week duration
197 and had the smallest difference between the different engagement week averages.

198 A similar trend was observed when comparing participants based on their reported comfort
199 levels with technology. Those who were very confident consistently used the technology more
200 through the four weeks than those who reported less confidence (13.6% overall difference).

201 There were also consistently higher engagement levels with the group that had never interacted
202 with smart speakers before when compared to those who have (7.6% difference). Both groups
203 steadily declined in engagement as the weeks progressed, with similar overall differences
204 between week one and four averages.

205 Over the four week duration (28 days), nine entries (out of 411) were incorrect measurements
206 that were submitted using the *Medly* voice app, indicating an overall accuracy rate of 97.8%. The
207 errors varied between weight and blood pressure. A small subset of participants (four) were not
208 able to successfully submit their correct readings which led to the nine errors that were recorded.

209 **Acceptability of the *Medly* Voice App**

210 Findings from the SUS questionnaire paired with the findings from the semi-structured
211 interviews were used to better understand the acceptability of using a voice app version of the
212 *Medly* program.

213 The responses from the SUS questionnaire from week two resulted in an overall average score of
214 69 (out of 100), ranking it in the 53rd percentile based on previous studies. In contrast, the
215 average score from week four results is 77 (out of 100), ranking it in the 80th percentile based on
216 previous studies. This data indicates an overall increase in the level of satisfaction of using the
217 *Medly* voice app (by 27%) from the study population. The difference in averages for each
218 individual question between weeks two and four was also calculated, with the last question in the
219 survey having the biggest difference of 13%. Participants felt that as time went on they needed to
220 learn more things about the voice app to successfully interact with it (consistent with NASA

221 TLX cognitive load results). Response distributions for week two and four results were fairly
222 similar for all of the questions (Fig S1, Multimedia Appendix 4).

223 Average SUS scores were also calculated based on the different patient characteristics (age,
224 *Medly* status, comfort levels, and interaction with smart speaker familiarity). Overall the scores
225 were similar in range for all of the characteristics. The largest range in data however, was
226 identified in the age groups, with the oldest (61-80 years) demographic providing the lowest
227 score (72 out 100), ranking it in the 62nd percentile, while the middle-age demographic provided
228 an average score of 87.5 out of 100, ranking it in the 96th percentile. The average score from the
229 youngest demographic was 77.5, ranking it in the 80th percentile.

230 **Feasibility of *Medly* Voice**

231 The NASA-TLX was used in this study to better assess the perceived workload when using the
232 *Medly* voice app by the study participants. A 4% increase was seen in average scores between
233 week two and four results, indicating a slightly higher workload. While the averages for each of
234 the questions were fairly low, questions relating to: 1) success rates, 2) how hard they needed to
235 work to accomplish the task, and 3) feeling of discouragement, irritation and stress scored the
236 worse when compared to the rest of the questions. These results can be seen in Fig S2,
237 Multimedia Appendix 4. Participants also felt less successful with using the *Medly* voice app at
238 the end of the study than they did at the end of week two (22% difference in results).

239 When analyzing the scores based on the different age groups, it was found that the youngest
240 demographic felt they needed to work the most (highest average of 2.67) when compared to the
241 middle-age (average of 1.61) and oldest demographics (average of 2.12). It was also found that

242 those who were new to *Medly* specifically felt more rushed when using the voice app and less
243 successful when inputting their measurements, when compared to those who have been on the
244 *Medly* program for a longer time (~15% difference in scores for each question). The difference in
245 average scores for those who described themselves as less confident when using technology
246 consistently gave poorer scores for each of the questions, indicating they had a more difficult
247 time than those who described themselves as confident (Table S1, Multimedia Appendix 4).

248 In summary, the youngest age group felt they needed to work the most, the study population
249 collectively felt they needed to provide slightly more effort as time went on, and those who were
250 less familiar with technology had more difficulty using the voice app when compared to those
251 who were more confident.

252 The UTAUT2 questionnaire was used to better understand participants' thoughts regarding
253 facilitating conditions, effort expectancy, habit, and behavioral intention when it comes to using
254 the voice app. The biggest difference between week two and four results was with whether they
255 would use the *Medly* voice app in the future, with a 13% decline in the average score. The oldest
256 demographic was the least keen on using it in the future, while the middle-aged demographic
257 was the most interested in future use. When asked if the voice app became a habit, those who
258 used the technology before agreed more than those who did not have experience using the device
259 (19% difference in responses).

260 Overall, all participants felt the voice app required low effort to use, and that it was easy for them
261 to operate. They were less certain with whether or not using the voice app had become a habit for
262 them (this can be supported with engagement levels), and were least certain about whether they
263 would use the voice app in the future, as seen in Table S2, Multimedia Appendix 4.

264 **Qualitative Data**

265 Interview themes were classified using Proctor et al.'s Implementation Outcomes, specifically
266 focusing on the *feasibility* and *acceptability* constructs to better understand which patient
267 demographics would benefit from using the voice app. The themes: (1) feasibility of clinical
268 integration and (2) voice app acceptability are presented below, each with their own set of
269 accompanying sub-themes.

270 **Feasibility of Clinical Integration**

271 The feasibility of providing patients with the option of accessing DTx through voice apps in the
272 clinic first involves understanding how patients interact with the technology, as well as the
273 technology's reliability when being used. Subthemes for this section include (a) users adapted to
274 voice app's conversational style, and (b) device unreliability.

275 **Users Adapting to Voice App's Conversational Style**

276 Most participants found the device set-up and instructions fairly straightforward, but at times
277 struggled to successfully log their measurements on the *Medly* voice app. When participants
278 struggled, they adjusted the way they spoke instead of continuing in their natural manner in
279 hopes that the voice app would understand them better.

280 "I learned how to get into her rhythm as opposed to her getting into my rhythm."

281 [Participant 04]

282 Specific strategies were employed to change their speaking style and most often involved
283 modifying the volume, tone, pace, and style they spoke at. Different strategies seemed to work
284 better for different participants, specifically with the pace at which they spoke at.

285 “Now I just say 116.4 pounds (faster) and there's absolutely no issues with her now.”
286 [Participant 12]

287 “Of course I would either make sure to be speaking directly at it or elevate my voice or
288 something like that.” [Participant 15]

289 “I want to record one hundred, but it's very typical to say 'a hundred' and not 'one
290 hundred', but I notice it doesn't pick up on that.” [Participant 17]

291 Once participants changed their conversational tone when speaking to the voice app, they began
292 to notice difficulties in the interaction since it no longer felt like a natural conversation.

293 “It’s like when you talk to someone foreign or you know from another country or another
294 language and you try to say a few words for them to understand it.” [Participant 12]

295 “I try to, like, separate each word, almost like I had to speak robotic.” [Participant 18]

296 “I have to be serious, slow and sure of how I say the numbers.” [Participant 17]

297 Another interaction strategy employed by most participants involved using the device’s
298 touchscreen capabilities. In most cases this alternative input was the favorable approach over
299 using voice since it was simpler to use and most importantly, faster.

300 “I got into a routine which allowed me to go through it as quickly as possible, and that
301 routine would be that I would speak the results for weight, blood pressure and heart rate,
302 and then I would interact directly on the touch screen for symptoms so we didn't have to
303 wait for her. So yes, every time I use the touch screen it works fine and the fact that I
304 could use a touch screen and it would work even though she hadn't finished speaking is a
305 big plus for me.” [Participant 15]

306 Interactions were found to be most successful when participants did not multitask on other items.

307 “You can multitask if you really want to, but that's what I think mistakes can be made
308 easier.” [Participant 02]

309 “I knew the questions that were going to be asked after a while, but I still listened. Only
310 because you know I'd rather do it right than wrong if I can.” [Participant 08]

311 Despite the learning curve experienced by most participants, the mitigation strategies described
312 above support the feasibility of deploying a voice app, such as *Medly*, in the clinic due to the
313 perseverance displayed by these participants to make the interaction easier for themselves as
314 time went on.

315 **Device Unreliability**

316 Almost all participants experienced some level of difficulty when they interacted with the voice
317 app. Sometimes the voice app froze and the session ended abruptly, and other times it would not
318 provide the user with an opportunity to correct any of the wrong measurements.

319 “You can go back and correct it, right, but sometimes it gives you a little bit of a hassle
320 so I have to start over.” [Participant 02]

321 “Then she just shut down ... When she couldn't get the measurements or something, she
322 would just turn off.” [Participant 04]

323 Participants also described instances where the voice app was not able to correctly pick up the
324 information they were saying, making them feel frustrated, annoyed, panicked, and discouraged
325 to the point where they no longer wanted to use the device that day.

326 “Yeah, I'd wake up in a great mood and oftentimes it was so frustrating that it made me
327 cranky afterwards. Yeah, it really switched my mood. One time she repeated it to me and
328 I thought she got it alright and then she repeated it and said that I fainted and I had not
329 fainted, so I panicked.” [Participant 18]

330 When the voice app was not able to pick up the correct measurements, participants often felt the
331 need to speak louder. This was considered to be specifically problematic in situations where a
332 participant may not be feeling well and does not have the ability to project their voice. As
333 explained by one participant, with the smartphone they are able to share information without
334 needing to exert a lot of energy:

335 “I would never want it to not be on my phone when I go into the hospital and I have a
336 hard time talking. If my blood pressure is through the roof or it's way too low from
337 retaining water, it's so hard to speak and I love that I could just throw my phone at the
338 doctor and be like “look, this was [my data] two days ago.” ... I really like that.”
339 [Participant 18]

340 Although the voice app seems feasible to deploy from a patient interaction perspective, users
341 also experienced difficulties when interacting with the device for various tech-related reasons.
342 Understanding the causes and frequencies of these malfunctions will help identify when and
343 where it is appropriate to use voice apps like *Medly*.

344 **Voice App Acceptability**

345 This theme describes the extent to which the study participants found the *Medly* voice app
346 satisfactory. This level of acceptability includes not only the participant's thoughts, but also
347 other factors that may influence their experience as described by the following subthemes: (1)
348 device integrated well within household and user lives, (2) users blamed themselves when
349 problems arose with the voice app, (3) voice app missing specific features desired by users.

350 **Device Integration in Household**

351 In addition to using the device to access the *Medly* voice app, many participants also found they
352 used it for other things during their time on the study. Over the four weeks some participants
353 described the device as a companion, with one participant noting:

354 “She became like a buddy. I know it’s little quirks, specifically when it makes mistakes...
355 I would say for people that live on their own or whatever it can become like a friend,
356 right?” [Participant 08]

357 Some participants also described their experience as “pleasant” when interacting with the device,
358 and others specifically feeling the need to use manners and to be polite while conversing with it:

359 “And I’ve gotten along with Alexa just fine. It was so cute. I was inputting on *Medly* and
360 I did it with Alexa at the same time and at the end I said ‘Alexa, thank you’ and she said
361 ‘you bet’... One night I said, ‘oh Alexa goodnight’ and she said ‘night night, sleep
362 well’.” [Participant 08]

363 Not only did the device become a companion for the user, but for other family members and
364 friends as well:

365 “She did give my granddaughter a knock knock joke the other night. [The grandkids]
366 have fun with her by asking what the weather is or something like that.” [Participant 10]

367 This interaction is an example of how easily the device can fit in and become integrated within a
368 space in the household. While in common areas, users have noted using the device for other
369 activities, such as:

370 “I let it play music for me or I ask what’s the weather like today and I do the CTV News
371 first thing in the morning, so yeah, I think it’s a great thing.” [Participant 02]

372 Having the device in common spaces also served as a reminder for some participants who have
373 difficulty remembering to perform their *Medly* measurements. Others also mentioned that
374 because the device was sitting in a common space, they would be more inclined to use *Medly* on
375 it:

376 “Seeing the monitor right there on the counter I feel like it definitely encourages and
377 motivates me and is a visual reminder as opposed to the app on the phone to actually do
378 it.” [Participant 11]

379 “At first I thought it would be my phone. But probably you know, now it's Alexa. She sits
380 right there, so probably Alexa.” [Participant 02]

381 Some participants also situated the device in other places in their house, such as the bedroom. In
382 these cases they also found the set-up to be useful:

383 “I use it at night time when I’m going to bed like you know, relaxing music.” [Participant
384 06]

385 And even in some cases, more preferred when compared to the smartphone:

386 “I’m in my bedroom and I have a bathroom in the room, so when I go in the bathroom to
387 weigh myself, I do my blood pressure at the same time. So ideally that is where I talk to
388 [Alexa]... over the last week it's been working and I really like that because then I'm done
389 and then I can go right back to bed after I take my pills so it doesn't make my mind wake
390 up.” [Participant 03]

391 “I’m sort of having concussion symptoms and the phone makes me nauseous. So at the
392 moment, I prefer only having to do it with Alexa.” [Participant 14]

393 In addition to the benefits of the device integrating well within different spaces in the household,
394 there are also drawbacks that can exist when keeping the device in a public space. Most
395 participants noted the importance of having a quiet space to focus and successfully submit
396 readings:

397 “Honestly like I did it more often when I didn't have my son because everything here he
398 likes to speak over me... He would repeat ‘Alexa’ behind me.” [Participant 18]

399 “Like if my husband would walk into the kitchen as I was doing it, I would shoo him
400 away, literally.” [Participant 08]

401 **Users Blamed Themselves When Voice App Problems Arose**

402 Although frustration was experienced by some participants when the device abruptly stopped
403 working or incorrectly heard them, often times (especially in the first week) users felt like it was
404 their fault when a mistake was made:

405 “I wasn't annoyed by it. I just thought, oh, I'm not speaking clearly or loudly enough, or
406 you know.” [Participant 08]

407 “Well again, I go back to the learning curve in the first week. There was some frustration,
408 but you can't blame that on Alexa, that was all me.” [Participant 05]

409 These reflections indicate that users were generally understanding of the voice app and had some
410 patience when interacting with it.

411 **Missing, but Desired Voice App Features**

412 Participants shared some of the features they value in devices that can offer programs such as
413 *Medly* on. In particular, users would prefer to interact with a device that is fast and can quickly
414 record their data for the day. In some instances users compared the voice app to Bluetooth
415 capability, indicating the latter is a much faster and more simpler process:

416 “It's just really cumbersome, like the whole process. And I guess part of that is because
417 the [smartphone] app is so easy. And I think it could get even easier if I got the Bluetooth
418 blood pressure and scale.” [Participant 04]

419 “To me, honestly, because they want it in the morning, the smartphone is much faster.”
420 [Participant 09]

421 Most users also expressed concern about how they would use the voice app should they go on an
422 overnight trip. A device that is small enough in size to be portable when traveling was desired
423 and often mentioned.

424 “The only thing I don't like about it is it is big and bulky so it is not something I would be
425 too inclined to want to travel with. So yeah, so for me the mobility issue would be a bit of
426 a concern if I had to rely on it.” [Participant 13]

427 **Discussion**

428 **Principal Findings**

429 This manuscript presents the findings from a clinical pilot study for a voice app, designed for HF
430 patients, using a mixed methods approach. To our knowledge, this has been the first clinical pilot
431 evaluation of a voice app used for helping patients manage an advanced chronic condition at
432 home. So far, studies have only reported on accuracy and acceptability levels in a controlled lab
433 environment, however these findings are still consistent with the results presented in this paper
434 (15,16). This study indicates the level of acceptability and feasibility of a voice app for patient
435 self-management by measuring engagement levels. Our findings show that there was a 14%

436 decline in engagement levels between week one and week four levels. Despite engagement
437 decreasing as the weeks progressed, participants became more accepting of the technology as
438 time went on (higher SUS scores when compared to week two). The workload associated with
439 using the voice app was not seen as problematic based on NASA-TLX scores, although
440 participants reported needing to use a higher cognitive load in week four when compared to
441 week two data (4% increase). Regardless of these positive results however, there were 13% more
442 participants who stated they were less interested in using in the future when compared to week
443 two results.

444 Aside from understanding the acceptability and feasibility of using the voice app as an
445 alternative input for chronic disease management, we also sought to identify who this technology
446 would be best suited for. Similar to the findings presented by Ware et al., engagement levels
447 were highest in the older age group demographic, and progressively lower in younger age groups
448 (17). While the oldest group had the highest engagement levels, the middle-aged demographic
449 (41-60 year old's) had the highest SUS average score, indicating they were the most accepting of
450 the voice app.

451 One of the most common responses provided by participants during the interview was the notion
452 that the voice app takes a long time to complete, and in particular, takes longer than the *Medly*
453 smartphone app. Users would often describe being rushed out the door in the mornings, in which
454 case they appreciated being able to use the smartphone app to quickly input their measurements.
455 This type of lifestyle and response was observed less with the older demographics who generally
456 seemed to have more patience and understanding when interacting with the voice app. There
457 were also specific cases where the voice app actually proved to be more useful than the

458 smartphone. One participant was experiencing concussion-type symptoms, and as a result had
459 limited screen time, so the voice app worked well for them. Another participant often felt fatigue
460 as one of the side effects from their medications and experienced difficulties navigating the
461 *Medly* smartphone app in the mornings. In this case, they also appreciated how much easier it
462 was to perform the required tasks using the *Medly* voice app. Similar sentiments were echoed by
463 other participants who came to the realization that they can successfully record their readings
464 when speaking in a relaxed, non-strenuous manner. While this worked well for some, one
465 participant in a similar situation had a different experience, specifically because the voice app
466 was unable to decipher their speech when they were feeling unwell due to their weak and fragile
467 voice. As a result, further advancements are required to better recognize sound, specifically when
468 users are not able to exert large amounts of energy while speaking. Similar technical limitations
469 have also been outlined by other voice app studies (18).

470 The findings from this study also showcase how well integrated the device became in many
471 households and the potential benefits this may have for participants. Because of the device's
472 versatility, it quickly became a part of many users' daily routine, from listening to music to
473 asking for dinner recipes, and started turning into a companion. Not only did the device provide
474 social support, but it also served as a visual reminder to perform their *Medly* measurements. One
475 participant also noted they would be more inclined to use the *Medly* voice app simply because it
476 was in a common space they frequent in their house. Therefore, the natural integration of the
477 device into users' lives over the four weeks shows the possibility that it may make it more
478 convenient for some to perform their *Medly* measurements, and may encourage and motivate
479 others who often forget.

480 These findings help begin to uncover the “profile” of the patient demographic this technology
481 would be most suitable for. Data from the clinical pilot show that those who feel more confident
482 in using technology, have less busy schedules, as well as those who are older (60+ years) have an
483 easier time, are more successful and consistent when interacting with the voice app. Also, those
484 with multimorbidity can benefit from using this platform especially due to common side effects
485 they may experience from their conditions.

486 **Limitations**

487 Multiple limitations were identified over the course of the study and as a result should be
488 acknowledged to better understand the impact of the findings.

489 First, because there were numerous questionnaires and interviews, the study team was mindful of
490 the potential for social desirability bias (19). As a result, participants were encouraged to speak
491 honestly and were given the opportunity to disclose their thoughts through questionnaires instead
492 of over the phone. Second, our aim was to recruit 30 participants but only 20 patients were
493 onboarded. Therefore, most of the findings and results were interpreted in a qualitative manner
494 since they were not statistically powered. Third, specific study factors could have impacted the
495 participant’s thoughts, experiences, and feedback. Users were aware that the study duration was
496 only a four week period and as a result, may have had higher engagement levels than if they
497 were asked to use it for a longer period of time. Participants were also required to perform a
498 double entry of their measurements, which may have impacted some users’ routines had they
499 only been required to use the voice app. Fourth, because the inclusion criteria was general
500 enough to include any patient enrolled in the program, selection bias likely occurred during
501 recruitment. In this case, there may have been missed opportunities to include a greater variety of

502 demographics in the study, especially those who primarily spoke languages other than English.
503 Lastly, because most participants from this study never interacted with a smart speaker before,
504 their thoughts and feedback may be influenced by the fact that they were interacting with a novel
505 technology. As a result, their thoughts on the device itself could be reflected in their responses,
506 even though any voice user interface device could have been used for the study.

507 **Conclusions**

508 This study utilized a mixed methods approach to investigate the acceptability and feasibility of
509 deploying a voice app for DTx used in chronic disease management. Our findings were
510 consistent with previous research when it came to engagement levels, with the oldest age group
511 showcasing the best, most consistent results. We recommend this platform be offered to those
512 who: are older (60+ years), have less busy schedules, exhibit high confidence levels when using
513 technology, or experience symptoms (such as fatigue or headaches) from chronic conditions.
514 While the technology could benefit from some advancements, participants were successful in
515 finding ways to improve their conversational experience, proving that an app like this could be
516 feasible to deploy in the clinic for future use.

517 **Conflicts of Interest**

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

520 **Acknowledgments**

521 First, the authors wish to thank the patients who participated in this study. Thank you to the
522 *Medly* nurse coordinators: Mary O’Sullivan, Sarvatit Bhatt, Eva Pavic, Tina Carriere, and
523 Annabelle Fontanilla for helping with the recruitment process. We also wish to express our
524 gratitude to Quynh Pham and Patrick Ware in helping guide this research project’s methodology,
525 as well as Cait Nuun, Madison Taylor, and Denise Ng for their REB expertise and guidance.

526 **References**

- 527 1. Noncommunicable diseases [Internet]. World Health Organization. 2021. Available from:
528 <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
- 529 2. Dickson VV, Clark RA, Rabelo-Silva ER, Buck HG. Self-Care and Chronic Disease.
530 Nurs Res Pract [Internet]. 2013; Available from: <http://dx.doi.org/10.1155/2013/827409>
- 531 3. Seto E, Leonard KJ, Cafazzo JA, Masino C, Barnsley J, Ross HJ. Self-care and quality of
532 life of heart failure patients at a multidisciplinary heart function clinic. J Cardiovasc Nurs
533 [Internet]. 2011; Available from: <http://dx.doi.org/10.1097/JCN.0b013e31820612b8>
- 534 4. Ware P, Ross HJ, Cafazzo JA, Laporte A, Seto E. Implementation and evaluation of a
535 smartphone-based telemonitoring program for patients with heart failure: Mixed-methods study
536 protocol. J Med Internet Res [Internet]. 2018; Available from:
537 <http://dx.doi.org/10.2196/resprot.9911>
- 538 5. Irfan Khan A, Gill A, Cott C, Hans PK, Steele Gray C. mHealth Tools for the Self-
539 Management of Patients With Multimorbidity in Primary Care Settings: Pilot Study to Explore
540 User Experience. JMIR Mhealth Uhealth [Internet]. 2018 Aug 28;6(8):e171. Available from:
541 <http://dx.doi.org/10.2196/mhealth.8593>
- 542 6. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications
543 for health-promoting behavior among individuals with chronic medical conditions. Digit Health
544 [Internet]. 2019 Jan;5:2055207619882181. Available from:
545 <http://dx.doi.org/10.1177/2055207619882181>

- 546 7. Brownstein J, Lannon J, Lindenauer S. 37 Startups building voice applications for
547 healthcare [Internet]. MobiHealthNews. Available from:
548 <https://www.mobihealthnews.com/content/37-startups-building-voice-applications-healthcare>
- 549 8. Kannel WB, Belanger AJ. Epidemiology of heart failure. Am Heart J [Internet]. 1991
550 Mar;121(3 Pt 1):951–7. Available from: [http://dx.doi.org/10.1016/0002-8703\(91\)90225-7](http://dx.doi.org/10.1016/0002-8703(91)90225-7)
- 551 9. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart
552 failure telemonitoring program implemented as the standard of care in an outpatient heart
553 function clinic: Pretest-posttest pragmatic study. J Med Internet Res [Internet]. 2020; Available
554 from: <http://dx.doi.org/10.2196/16538>
- 555 10. Johanson GA, Brooks GP. Initial Scale Development: Sample Size for Pilot Studies.
556 Educ Psychol Meas [Internet]. 2010 Jun 1;70(3):394–400. Available from:
557 <https://doi.org/10.1177/0013164409355692>
- 558 11. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for
559 implementation research: Conceptual distinctions, measurement challenges, and research agenda.
560 Administration and Policy in Mental Health and Mental Health Services Research [Internet].
561 2011;38(2):65–76. Available from: <http://dx.doi.org/10.1007/s10488-010-0319-7>
- 562 12. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information
563 technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly:
564 Management Information Systems [Internet]. 2012; Available from:
565 <http://dx.doi.org/10.2307/41410412>

- 566 13. System Usability Scale (SUS) [Internet]. usability.gov. Available from:
567 <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- 568 14. Nasa. NASA Task Load Index. Human mental workload [Internet]. 2006;1(6):21–21.
569 Available from:
570 [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16243365)
571 [_uids=16243365](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16243365)
- 572 15. Jadczyk T, Kiwic O, Khandwalla RM, Grabowski K, Rudawski S, Magaczewski P, et al.
573 Feasibility of a voice-enabled automated platform for medical data collection: CardioCube. *Int J*
574 *Med Inform* [Internet]. 2019;129(March):388–93. Available from:
575 <http://dx.doi.org/10.1016/j.ijmedinf.2019.07.001>
- 576 16. Cheng A, Raghavaraju V, Kanugo J, Handrianto YP, Shang Y. Development and
577 evaluation of a healthy coping voice interface application using the Google home for elderly
578 patients with type 2 diabetes. In: 2018 15th IEEE Annual Consumer Communications &
579 Networking Conference (CCNC) [Internet]. IEEE; 2018. Available from:
580 <http://dx.doi.org/10.1109/ccnc.2018.8319283>
- 581 17. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient Adherence
582 to a Mobile Phone–Based Heart Failure Telemonitoring Program: A Longitudinal Mixed-
583 Methods Study (Preprint) [Internet]. Available from: <http://dx.doi.org/10.2196/preprints.13259>
- 584 18. Sezgin E, Noritz G, Elek A, Conkol K, Rust S, Bailey M, et al. Capturing at-home health
585 and care information for children with medical complexity using voice interactive technologies:

586 Multi-stakeholder viewpoint. J Med Internet Res [Internet]. 2020; Available from:

587 <http://dx.doi.org/10.2196/14202>

588 19. Grimm P. Social Desirability Bias. In: Wiley International Encyclopedia of Marketing

589 [Internet]. 2010. Available from: <http://dx.doi.org/10.1002/9781444316568.wiem02057>

590 **Abbreviations**

591 HF: heart failure

592 SUS: System Usability Scale

593 UHN: University Health Network

594 UTAUT2: Unified Theory of Acceptance and Use of Technology

595 VUI: voice user interface

596 **Supporting Information**

597 Multimedia Appendix 1 **Fig. S1 Instructions manual for *Medly* voice app.**

598 Multimedia Appendix 2 **Table S1. Baseline questionnaire for participants.**

599 Multimedia Appendix 3 **Table S1. Overall and weekly average engagement levels based on**
600 **various patient characteristics.**

601 Multimedia Appendix 4 **Fig S1. SUS score distributions.**

602 Multimedia Appendix 4 **Fig S2. NASA-TLX score distributions from week two and four**
603 **results (top and bottom, respectively).**

604 Multimedia Appendix 4 **Table S1. Average scores for each NASA-TLX question.**

605 Multimedia Appendix 4 **Table S2. Average scores for each of the constructs from the**
606 **UTAUT2 questionnaire.**