Deploying and Examining Beacon for At-Home Patient Self-Monitoring with Critical Flicker Frequency

Richard Li,
University of Washington, Seattle, WA, USA

Philip Vutien,
UW Medicine and VA Puget Sound, Seattle, WA, USA

Sabrina Omer,
UW Medicine, Seattle, WA, USA

Michael Yacoub,
UW Medicine, Seattle, WA, USA

George Ioannou,
UW Medicine and VA Puget Sound, Seattle, WA, USA

Ravi Karkar,
University of Massachusetts Amherst, Amherst, MA, USA

Sean A. Munson,
University of Washington, Seattle, WA, USA

James Fogarty
University of Washington, Seattle, WA, USA

Abstract

Chronic liver disease can lead to neurological conditions that result in coma or death. Although early detection can allow for intervention, testing is infrequent and unstandardized. Beacon is a device for at-home patient self-measurement of cognitive function via critical flicker frequency, which is the frequency at which a flickering light appears steady to an observer. This paper presents our efforts in iterating on Beacon's hardware and software to enable at-home use, then reports on an at-home deployment with 21 patients taking measurements over 6 weeks. We found that measurements were stable despite being taken at different times and in different environments. Finally, through interviews with 15 patients and 5 hepatologists, we report on participant experiences with Beacon, preferences around how CFF data should be presented, and the role of caregivers in helping patients manage their condition. Informed by our experiences with Beacon, we further discuss design implications for home health devices.

lichard@cs.washington.edu .

1 INTRODUCTION

4.5 million adults in the United States are diagnosed with chronic liver disease [22]. Chronic liver disease over time leads to cirrhosis, an end-stage condition in which scarring occurs in the liver. Reduced liver function due to cirrhosis results in accumulations of neurotoxic substances that induce a spectrum of neurological impairments called *hepatic encephalopathy* (HE). These impairments fluctuate in severity from *minimal hepatic encephalopathy* (MHE) through *overt hepatic encephalopathy* (OHE). In its most advanced form, OHE leads to gross disorientation, coma, or death. In contrast, MHE manifests in subtle cognitive impairments. These cognitive impairments can cause driving accidents [7] or serious falls [46], contributing to individual and societal impacts of HE. Undetected and untreated, MHE may also progress into OHE. On the other hand, early detection can support use of medication to effectively control and treat MHE. Early detection of MHE requires frequent administration of tests, but MHE testing in clinical settings remains "uncommon" [2] and undetected MHE is a "hidden epidemic" [58]. Barriers to MHE testing include: (1) the time it would add to a clinical visit; (2) tests that are difficult, expensive, and require trained personnel to administer; and (3) a lack of standardization [3].

Prior work has identified *critical flicker frequency* (CFF) as an excellent candidate for timely screening of MHE [31]. The CFF measure is the minimum frequency at which an individual perceives a flickering light as fused. Despite its potential, available solutions for measuring CFF are generally expensive and complicated to use, while also not designed to be self-administered.[1] Our vision is to enable more widespread and frequent usage of CFF measurements, both for clinical screening and for at-home self-monitoring, through a portable, low-cost, and self-administered system. By enabling self-measurement in a home setting, we envision cirrhotic patients being more frequently screened for the onset of MHE without adding to the burden of healthcare providers. CFF self-monitoring could inform cirrhotic patients, their caregivers, and their collaborations with healthcare providers. Timely detection of MHE could in turn support cirrhotic patients in better managing their condition: through lifestyle choices to control MHE, through decisions to promote safety (e.g., not driving), and through more prompt treatment.

In this paper, we first introduce iterative changes to the Beacon system, depicted in Figure 1, necessary to enable an at-home deployment. Building upon the initial prototype described by Karkar et al. [31], we report changes across the entire system: (1) updates to the physical device, (2) transformation of its form factor into a common piece of furniture (i.e., a lamp); (3) a redesign of the smartphone app to be cross-platform; and (4) a new cloud server with dashboard for remote monitoring of devices. We then present findings from an at-home

---

[1]The most commonly used research devices are the Lafayette Flicker Fusion System, the Hepatonorm Analyzer, and the Schuhfried Flicker Fusion Analyzer. Each cost approximately $3000 USD. They involve large and bulky form factors and have not been designed for self-measurement or self-monitoring.

deployment with 21 patients each using Beacon over a 6-week period, demonstrating robustness of their CFF measurements with Beacon despite being self-administered and taken in a variety of environments at different times of day. As part of this, we evaluated two different measurement protocols, finding that one requires less time to complete (mean: 2.78 minutes) but has greater variance in measurement (standard deviation: 1.09) while the other requires more time to complete (mean: 6.37 minutes) but has significantly less variance in measurement (standard deviation: 0.61). Finally, we share further findings from interviews with 15 patients and 5 hepatologists, gathering feedback on their experiences using Beacon and exploring how they might want to engage with such measurements in managing their condition. The primary course of action for patients experiencing HE is to adjust their medication (i.e., lactulose) dosage in response to progression of their condition. We learned that this responsibility often falls on the caregivers of patients, leading us to identify opportunities for Beacon to support management of their condition through facilitating communication between patient, caregiver, and healthcare provider.

In summary, this paper makes the following contributions:

1. We contribute an iteration on Beacon's form, electronics, and software, motivated by the need to scale from a proof-of-concept research prototype to enable clinical validation and at-home patient data collection.

2. Through a 6-week at-home study with 21 patients, we demonstrate the feasibility of patients using Beacon to collect a novel CFF measure in at-home settings, and we demonstrate the robustness of Beacon CFF measures to environmental factors.

3. Based on interviews with 15 patients and 5 hepatologists, we contribute considerations and opportunities for deploying Beacon in practice: the task load and temporal burden of taking Beacon measurements, the presentation and actionability of Beacon measurements, and the role of caregivers and providers in patient management of their chronic condition. We further discuss how these factors might apply for others working on home health devices.

2 RELATED WORK

Our research draws on and contributes to literature in using technology to move clinical and health measures into the home, including the benefits and design challenges associated with home measurement. We review the related work in moving health monitoring measures into home settings and the implications for stakeholders in this process. We then introduce the specific domain area of our work: patients with chronic liver disease.

2.1 Health Monitoring in Home Settings

As healthcare technology evolves, the ability to monitor health conditions outside of traditional clinical environments has become increasingly feasible [54]. In many cases, this transition involves the adaptation of well-established measures, such as blood pressure [34, 62], body weight [33], core temperature [12, 61], and spirometry [23, 37], to be taken outside of the traditional clinical environment. This mitigates potential influences, such

as anxiety, that affect reliability of health measurements taken in these settings. In the context of blood pressure, this phenomenon is referred to as *white coat hypertension*, in which patients exhibit a higher blood pressure due to being in a clinical setting, or *masked hypertension*, in which patients exhibit a seemingly normal blood pressure in a clinical setting but have elevated blood pressure in daily living. Studies have shown that home blood pressure monitoring has higher sensitivity for detecting hypertension than measurements taken in a clinical setting [14, 24, 40, 55, 62]. Another benefit of adapting measures to be taken outside of the clinical setting is the potential for increased time resolution of measurements, enabling faster detection of the onset of a change from status quo. In the context of blood glucose monitoring, the increasing availability of portable glucose monitors and continuous glucose monitors has enabled people diagnosed with diabetes to act on more real-time changes in glucose levels, such as eating a snack if their glucose level is low [29, 41, 44].

Beyond the functional benefits of health monitoring outside of the clinical setting, researchers have also investigated patient perception of these technologies, including factors such as the burden required and the intelligibility of the outputs. For example, Xu et al. found that unexpected events, temporally adjacent events, and fatigue were barriers to executing plans for physical activity [60]. Cordeiro et al. similarly found that the effort required for food journaling was a barrier to consistent logging [17]. On the other hand, Lim et al. found that exposing the certainty of a system, such as the confidence region in a location-tracking system, can improve perception of the system [39]. Kay et al. studied perception of the bathroom weight scale and enumerated design recommendations for measurement presentations that help people understand the daily fluctuations and uncertainty around weight measurements [33]. Kendall et al. also studied people's reactions to frequent blood pressure measurements [34]. In this work, we investigate how members of the chronic liver disease patient population perceive a novel CFF measurement device in terms of burden and intelligibility through a 6-week at-home deployment.

2.2 Multi-Stakeholder Health Research

Management of a patient's health condition is a multi-stakeholder operation. In the clinical setting, healthcare providers, nurses, and pharmacists work with patients to educate them and help them manage their condition. In a home setting, patients interact with an existing social infrastructure, such as caregivers [11], parents [15, 43], adult children [28], or other people in their life who they engage with regarding their health. The design of a health monitoring device, particularly those intended for home use, involves considering who might be best positioned to use, prescribe, or administer the device and who might be best positioned to act on its output. For example, some systems are designed to facilitate interaction between patient and healthcare provider. Berry et al. presented techniques for supporting communication about personal values between people with multiple chronic conditions and their providers [10], and Bascom et al. similarly presented techniques for reducing the implicit biases of healthcare providers when encountering patients [9]. Seo et al. examined techniques specifically for supporting communication between child patients and their providers. Alternatively, other systems were designed to facilitate interaction between patient and caregiver. Khan et al. described the development of a personal health

application to help patients and their caregivers manage medications [35]. Hong et al. deployed diary probes to help family members of adolescents with chronic conditions understand their experiences [27]. In our work, we seek to understand different stakeholders around chronic liver disease patients and how they might use Beacon and act on the measurements it produces.

2.3 Chronic Liver Disease and Critical Flicker Frequency

Cirrhosis, or scarring of the liver, is a severe outcome of liver disease, responsible for over 48,000 deaths in the United States in 2021, the 9th-leading cause of death [42]. Over 80% of cirrhotic patients develop neurocognitive impairments known as hepatic encephalopathy (HE) [58]. Impairments can fluctuate over time, ranging from minimal hepatic encephalopathy (MHE) to overt hepatic encephalopathy (OHE), affecting nearly all aspects of life.

The critical flicker frequency (CFF) test assesses neurophysiological state by measuring the minimum frequency at which a flickering light appears fused to an individual. CFF has been shown to have potential in diagnosing MHE [47, 57], and commercially-available systems such as the Lafayette Flicker Fusion System have been used in studies for clinical validation of this measure [63]. Although initially proposed as a clinical screening test, Karkar et al. contributed a reframing of the CFF measurement from a clinical screening test to a self-administered self-tracking measure using an initial Beacon prototype introduced in [31]. Then, Vutien et al. used an improved version of the device to clinically validate Beacon across 153 chronic liver disease patients [59]. Building upon prior contributions of Karkar et al. [31] and Vutien et al. [59], this paper will:

1. Briefly describe additional improvements made to the Beacon device to enable at-home deployments (Section 4).

2. Present the protocol for the at-home deployment and follow-up patient and hepatologist interviews (Section 5).

3. Report on findings from both at-home deployment and patient and hepatologist interviews (Section 6).

4. Discuss design implications and considerations based on lessons learned in our research (Section 7).

3   MOTIVATION AND CONTEXT

In this section, we first present the need for more consistent screening of HE as motivation for our work in this paper. We then provide definitions of the CFF protocols used in our work.

3.1 Context in Chronic Liver Disease

Although the guidelines for diagnosing OHE are generally well-defined, treatment of OHE is difficult. Our work instead focuses on improving screening for MHE, when treatment is easier and more effective than for OHE. Studies have demonstrated that, if administered in the early stages of MHE, lactulose and rifaximin can significantly improve the quality of

life of patients [5, 45, 53]. In particular, studies observed improvement in patient driving ability as well as significantly reduced episodes of recurring HE in patients with a history of HE [64]. On the other hand, if not controlled, MHE can develop into OHE, resulting in severe consequences such as coma or death. We seek to improve detection of early MHE by making testing more accessible and enabling patients to take more frequent tests, both in clinical and home settings. Upon detection of MHE, all parties involved are better informed: the patient can make applicable lifestyle changes, such as adjusting their diet [6], making sure to be careful when climbing stairs [46], and not operating heavy machinery [7]; caregivers can pay more attention to their patients; and providers can prescribe relevant drugs, such as lactulose or rifaximin.

Cirrhotic patients are recommended frequent screening for MHE [58]. However, studies have shown that they are not adequately screened; screening is even considered to be "uncommon" [3]. This is largely due to the amount of time available tests require; tests that are difficult, expensive and require trained personnel; and a lack of standardization [2]. Although patients with more severe levels of HE will outwardly exhibit symptoms such as tremors, disorientation, or slurred speech, patients with MHE do not present any clinical symptoms. In fact, MHE was historically referred to as subclinical hepatic encephalopathy [48]. Instead, MHE must be screened for using psychometric tests (i.e., tests measuring mental capabilities and behavioral style) or neurophysiological tests (i.e., tests measuring function of the nervous system). Psychometric tests such as the Stroop test (e.g., implemented as the EncephalApp [4]) or the psychometric hepatic encephalopathy score (PHES) [21] have been previously employed to screen for MHE. However, their results are strongly affected by effort, training, age, interaction with the test administrator, literacy, numeracy, and education level [30, 52]. Most critically, psychometric tests are subject to learning effects, in which repeated use can make people better at performing the test and thus undermine its sensitivity. These properties limit psychometric testing as a candidate for repeated use, leaving a gap for a test that can be longitudinally self-administered.

3.2 Critical Flicker Frequency Protocols

The literature describes a number of ways to measure CFF, including the method of limits (MOL) and 2-alternative forced choice (FC) [19]. We describe the procedure for these protocols here as they are the basis for the rest of the paper.

The MOL measurement protocol includes two components: an ascending option, and a descending option. In the ascending component, referred to as MOL-A, the device presents a light flickering at a frequency visible to any person with a typical visual system (e.g., 25 *Hz*). The frequency of the light is steadily increased (e.g., at 0.5 *Hz* per 0.1 *sec*). A person is asked to indicate when the light has begun to flicker too quickly for them to perceive it as flickering. This perception of only a solid light source is when the light source has "fused". An example of this procedure is shown in the left panel of Figure 2. The descending option, MOL-D, is simply the opposite of MOL-A. The device presents a light flickering at a frequency too fast for a person to perceive (e.g., 55 *Hz*). The frequency is steadily decreased (e.g., at 0.5 *Hz* per 0.1 *sec*) and the person is asked to indicate when the light begins to appear to flicker. The center panel of Figure 2 shows an example of the MOL-D process.

A person's CFF as determined by the MOL is then typically calculated as the average of frequencies determined by the MOL-A and MOL-D. However, some of our experiments also consider the MOL-A and MOL-D measures separately.

In the FC protocol, as shown in the right panel of Figure 2, a person is presented with two different stimulus options, and they must decide which one appears to flicker and which one appears to be fused. One option, which will always appear fused, is a flickering light at a frequency significantly higher than human perception (e.g., 120 *Hz*). A high frequency is used here instead of a constant light to avoid luminance artifacts that could bias a person's decision. The other option, which the person is asked to identify, is flickering at the current variable frequency. Options are presented one after another, in random order, and the person can revisit either option. The person is asked to select which option appears to be flickering. Following the procedure described by Eisen-Enosh et al. [19], if they select the correct option three times, the frequency of the target option is increased by 2 *Hz*. If they select an incorrect option, the variable frequency immediately decrements by 2 *Hz*. The protocol ends after 8 turns (i.e., changes in direction of the variable frequency). The frequencies of the final 6 turns are then averaged to obtain a CFF value.

Multiple studies have shown that a CFF of 39 *Hz* is a strong indicator of MHE [1, 36, 49, 51, 57]. The MOL protocol is the most often used protocol for measuring CFF due to its simplicity, and indeed is used in all of these cited studies. However, we hypothesize that MOL might suffer from different types of response biases. For example, due to its time-sensitive nature, it is easy for a person to miss the onset of the stimulus transitioning from flickering to fused (or vice versa), such as during a blink. Furthermore, a person's response strategy might change over time; because the transition from flickering to fused (or vice versa) is not immediate, what people consider flickering might be inconsistent. On the other hand, we conjecture that the FC protocol can mitigate some of these issues by requiring a person to make a decision between options. As a result, it might produce measurements that are closer to a person's discrimination threshold, possibly higher than the accepted 39 *Hz*. However, such a protocol is also less frequently used in the literature because of how much longer it takes. In our work, towards our goal of increasing the availability of CFF measurements to screen for MHE, we investigate these different protocols in terms of both their quantitative output (i.e., the consistency of the measurements themselves) as well as patient subjective attitudes towards them. Understanding these properties will help inform future iterations of Beacon and increase its appropriateness for routine use.

4 SYSTEM

We identified several areas of improvement in the systems described by Karkar et al. [31] and Vutien et al. [59]. Our implementation, as shown in Figure 3, preserves the same functional properties and specifications (e.g., brightness and resolution) in order to maintain the same degree of validation obtained in Vutien et al.'s [59] clinical studies. We then improve Beacon across several areas: physical form factor, internals and assembly, an app for serving an interface, and a dashboard for monitoring a set of remote devices.

Internals.

To drive an accurate and precise flicker stimulus, we use the ATMega32u4's Timer 1 functionality, which is an interrupt based on the 8 MHz crystal oscillator and thus extremely stable with aging and temperature. To enable wireless communication between the Beacon device and an accompanying smartphone, we selected the nRF52 as a co-processor for supporting Bluetooth Low Energy functionality. Finally, in order for patients to not have to charge Beacon's battery for the duration of a deployment, we selected a 10,000 mAh lithium-polymer battery, which sustains Beacon for 3-months of daily measurements.

Assembly.

Although previous Beacon devices were functional, we observed that the assembly process required considerable manual effort and could also be error-prone. For example, placing the optical diffuser required carefully gluing it into place, and a shaky hand resulted in leaving streaks of glue on the diffuser. To make the assembly process more robust, we designed Beacon such that all of its components could be 3D printed and then snapped or screwed together. By mitigating the error-prone steps of assembly, we were able to produce more devices and with more consistency than before. As part of this design process, we also implemented quality control measures as standard procedure to objectively verify that the functional properties of each new device meet the desired specifications (i.e., brightness and flicker rate).

App for serving an interface.

The phone app primarily acts as an interface for a person to interact with the physical Beacon device. In that role, the app renders the interface elements necessary to administer a given measurement protocol. In addition to facilitating a measurement, the app also allows patients to redo or reflect on their measurement. We designed the app to be simple, maintaining a linear flow to mitigate issues with patients getting lost while using the app. In anticipation of deployment across different phones, we implemented the app using React Native for cross-platform support.

Dashboard for monitoring a set of remote devices.

Although keeping all measurements on a local device was sufficient for the in-lab and clinical testing conducted by Karkar et al. [31] and Vutien et al. [59], our goal of at-home deployment required the ability to remotely monitor devices and store measurements. For this at-home deployment, we set up a cloud server and implemented a dashboard interface to allow research coordinators to create new patient accounts and remotely monitor device interaction and measurements in real time.

Physical form factor.

In anticipation of deploying Beacon devices in homes, we also sought to improve the form of the device. Prior work has found medical devices can carry a stigma that deters their use [38, 65]. With the same underlying principle as keeping the app simple, we also decided to keep the form simple by modeling it off of a common piece of furniture: a lamp. This design decision had the additional benefit of introducing a hinge to the light source, allowing

patients to adjust the device such that they could look straight into the light instead of moving their own head or body.

Usability Adjustments.

In the course of at-home deployments described in this paper (Section 5), we found that patients often forgot to turn off the device after use, draining the internal battery and resulting in the burdensome and disruptive need to exchange devices, including an extra round of shipping back and forth. As a result, we introduced a slow flicker (at 1 *Hz*) for the Beacon device to indicate when it is not connected to a phone. Implementing this feature prompted participants to turn off their Beacon devices and entirely eliminated issues with patients leaving the device on.

5 METHODS

We designed and conducted studies towards two objectives. First, we sought to demonstrate the feasibility of patients taking CFF measurements in home settings. Second, we sought to understand how patients and hepatologists might interpret and use those home measurements in care and practice. In this section, we describe our methods in a pair of studies studies: a 6-week at-home deployment of Beacon with 21 chronic liver disease patients, and qualitative interviews with 15 patients and 5 hepatologists.

5.1 At-Home Deployment

Through the at-home deployment, we seek to (1) establish feasibility of people using Beacon to collect CFF measures and (2) provide the first characterization of CFF measurements taken over a longitudinal period of time.

5.1.1 Recruitment. We recruited patients from the hepatology clinic at the University of Washington Medical Center. Candidates were pre-screened by a hepatologist on our research team to determine whether their participation was appropriate. For safety reasons, we excluded patients with a history of neurocognitive disorders, seizure disorders or epilepsy, severe migraines or photosensitivity, or ophthalmologic diseases. Patients that met our exclusion criteria were approached about participating in a research study related to MHE and cirrhosis of the liver. Because adherence was not a primary outcome measure, we compensated participants in a tiered structure: $30 for week 1 (at least 5 measurements completed); $40 for week 2 (at least 5 measurements completed); $20 per week for weeks 3 – 6 (at least 1 measurement per week); and $20 for a 1-hour exit interview.

5.1.2 Participants. We report findings from 21 patients who participated in our deployment of Beacon for at-home self-measurement. According to patient medical records, 16 had a biological sex of male and 5 had a biological sex of female. Patients were between the ages of 22 and 70 (mean=52.65, standard deviation=11.96). Twelve of these patients had previously participated in Vutien et al.'s clinical study [59] and agreed to be contacted for participation in future research. We highlight this because those 12 participants had previously encountered both Beacon and the Lafayette Flicker Fusion System, a commercially available CFF device not designed for at-home use. Section 6 includes some

of these participants reflecting on that contrast according to their experiences with at-home measurement. The remaining 9 were patients drawn from a similar population as those examined by Vutien et al. [59], but had never previously encountered Beacon or any other device for CFF measurement.

5.1.3 Apparatus. Each patient was provided a study package consisting of the Beacon device, a phone (Android or iPhone) with the Beacon app, a printed instruction booklet, and a printed calendar with the study schedule. Items were transported in custom packaging, shown in Figure 4, designed to keep the devices safe. After the patient provided consent, the research coordinator called the patient to go through the process of setting up the device in their own home. This included helping familiarize the patient with each item in the packaging, helping the patient connect the provided phone to the their home WiFi, and walking them through each measurement protocol. Finally, the research coordinator worked with the patient to identify a time in their daily routine to regularly take measurements. The recommendation was to take measurements in the morning after grogginess had subsided (e.g., after breakfast or coffee).

5.1.4 Procedure. Depending on proximity of a patient to our institution, we either hand-delivered or mailed the study package. After the initial setup call with the research coordinator, patients used the Beacon device for 6 weeks according to the provided calendar, which prescribed 5 measurements in each of week 1 and week 2 then 1 measurement in each of week 3 to week 6. Due to a miscommunication, P1 was asked to take 5 measurements only in week 1, then 1 measurement in each of week 2 to week 6. We arrived at this procedure because our target duration for the at-home study was one month. Although we were curious about day-to-day fluctuations of the CFF measurements, we also did not want to overwhelm or overly burden patients as part of the study. In each of these sessions, a patient took two CFF measurements, using the MOL-D and FC protocols in a randomized order as prescribed by the calendar.

5.1.5 Analysis. Server logs were analyzed to extract CFF measures, time taken per CFF measure, adherence to prescribed study regimen, and number of decisions made in the forced choice protocol.

5.2 Patient and Hepatologist Interviews

Through interviews with both patients and hepatologists, we seek to understand how Beacon might be used in clinical and self-tracking practice and potential pathways to achieve that goal.

5.2.1 Participants. We report findings from interviews with 15 patients and 5 hepatologists. Patient participants were recruited from participants in the at-home deployment. Hepatologist participants were recruited through word-of-mouth from the healthcare providers on our team.

5.2.2 Procedure. We conducted semi-structured interviews with participants over Zoom. Interviews with patients involved understanding their experiences with setting up the

device at home, how measurement fit into their routine, any unexpected issues, and preferences between the MOL-D and FC protocols. We explored patient impressions around interpretation of CFF measures by presenting visualizations of their own measurements and asking them to reflect on that data. Interviews with providers involved understanding their role in providing care to patients. We also presented examples of patient data and gathered provider impressions of how they might like to see and use this data.

5.2.3   Analysis and Reporting. We recorded all interviews and used Zoom to generate transcripts. We had developed our interview protocols with specific questions in mind, and we coded transcripts deductively according to these areas. For patients, this included motivations for using Beacon, challenges they encountered in use of Beacon, experiences with the different protocols, understanding of the CFF measurements, and preferences for receiving CFF measurements. For hepatologists, we coded for when they would seek to have a patient take CFF measurements, their expectations of patient adherence to a measurement protocol, how they anticipated using the measurements in the care they provided, and implementation details (e.g., distribution of Beacon to patients). We also remained open to insights that we had not anticipated, conducting some open coding and identifying an additional theme around the role of caregivers in supporting the measurement process. In reporting on our results, we begin by presenting patient perspectives on their experiences of using Beacon in a home setting over a longitudinal period. We then present the patient and hepatologist interviews together to build an understanding of how they anticipate measurements from Beacon integrating into their clinical and self-monitoring practices.

Throughout this paper, we use P## to attribute quotes from the patient interviews according to Table 1 and use H## to attribute quotes from the hepatologist interviews according to Table 2.

6   RESULTS

We present findings from the at-home deployment and ensuing interviews with patients and hepatologists. Through the at-home deployment, we first establish the feasibility of patients self-administering CFF measurements using Beacon over a longitudinal period of time and report on their experiences using Beacon (Section 6.1.1). We then present characterizations of the longitudinal CFF measurements and discuss how our design decisions supported this robustness (Section 6.1.2). Next, we sought to understand the barriers and facilitators to the use of Beacon in clinical and self-tracking practice. Taking the patient and hepatologist interviews together, we present insights into the motivations of our patient population (Section 6.2.1), burden of taking a measurement with Beacon (Section 6.2.2), preferences around presentation of measurement data (Section 6.2.3), and the proposed actionability of Beacon measurements (Section 6.2.4).

6.1   At-Home Deployment

6.1.1   Feasibility and Experience of Taking Longitudinal CFF
Measurements. Patients generally took all 6 weeks of requested measurements, showing the feasibility of chronic liver disease patients taking consistent CFF measurements using

Beacon. P9 withdrew after week 1, saying they were too busy to continue. P18 also withdrew after week 1, due to a liver transplant.

We first wanted to assess preferences about the usability and form of the device. Of the 12 at-home deployment patients who were previously exposed to the Lafayette device in Vutien et al.'s clinical study, 8 participated in exit interviews. Of these, 6 expressed they preferred Beacon over the Lafayette device (shown in Figure 5) and 2 were neutral. Preference for Beacon was attributed to usability issues with Lafayette, such as "*[having] a problem keeping both eyes open and using it so I shut one*" (P1) and "*[feeling] claustrophobic and had an anxiety attack*" (P5). Form was also cited as a concern. For example, P2 said "*I liked the compactness of it [Beacon] better… Other one [Lafayette] looked like a dinosaur.*" and P10 said "*it [Beacon] was relaxed compared to the big machine [Lafayette]*". On the other hand, P3 had "*no preference between Lafayette and Beacon*" and said that they would be willing to use the Lafayette at home if needed. Patients thus generally preferred Beacon's usability and form, but might accept a less usable or aesthetic device if medically necessary.

We then sought to understand usage patterns of patients. We found that previously-discussed form factor preferences also had implications for where patients used Beacon in their homes. Participants were instructed to place Beacon in a location with minimal visual and auditory distractions. All participants used the device in an open area of their home (e.g., the kitchen counter, the dining room table), some examples are shown in Figure 7. Although some simply decided on a location because "*that's where the available space was*" (P7), others used its location as a strategic mechanism, as with P8 saying "*it helped remind me to do it.*" Four patients reported keeping Beacon in its box or another safe location, while 11 patients left the device out in the same spot in their home for the duration of the study. Patients who put the device away described doing so when they had company (e.g., P8) or in concern for its robustness (e.g., in the presence of young children (P10)). These considerations validate our decisions from Section 4 around physical form factor, including prioritizing the need to keep home devices small and portable.

Of the 8 at-home exit interview participants who were previously exposed to the Lafayette device in Vutien et al.'s clinical study, 4 expressed a strong preference for taking measurements at home, while the other 4 were neutral. Patients said they preferred at-home measurements because "*[there was] less pressure to get it right*" (P5) and "*[you are] more relaxed since you are in familiar conditions*" (P10). P7 also cited flexibility as an advantage: "*Could do it when I wanted to.*" Although P3 did not express a strong preference for at-home measurements, they acknowledged that clinic measurements could cause "*anxiety or stage fright*", but said they did not personally experience those symptoms. P4 simply said the experience of taking measurements was "*pretty much the same in clinic and at home.*" These responses suggest that although some people might value taking measurements at home, the environment was not a critical issue for every patient.

6.1.2 Characterization of Longitudinal CFF Measurements.

We report on at-home CFF measurements obtained from 21 patients, depicted in Figure 6. Visual inspection indicates stability in the measurements. Indeed, the standard deviation of CFF within-patient measured by MOL-D ranged from 0.61 to 5.28, and the standard deviation of CFF measured

by FC ranged from 0.90 to 3.63. Although MOL-D produced the smallest standard deviation in a single patient, the distribution of standard deviations obtained by FC was much smaller ($t(21) = 88$, $p < 0.05$). This comparison indicates that, over extended time and within a patient, FC produces relatively more consistent CFF measures than MOL-D. We observe that FC measurements have a tendency to be higher than MOL-D measurements taken in the same session, as also noted in clinical validation. Data collected in this at-home study is therefore consistent with prior clinical validation conducted by Vutien et al. [59], even while gathered in naturalistic conditions.

6.2   Patient and Hepatologists Insights

6.2.1   Patient and Caregiver Motivation to Taking Measurements.
In Section 6.1.1, we showed that chronic liver disease patients can feasibly take consistent CFF measurements using Beacon. However, hepatologists pointed out that motivation may be a barrier for many people, given their experiences recommending established medical devices: "*In the world of primary care, like, I would prescribe so many glucose monitors. But if people don't use it, it doesn't do anything. So you have to think about it from the user interface standpoint. It's like, why would I want to use it? And if I use it, how is it going to benefit me? …again, who are the people who actually going home and actually check their blood pressure or blood sugar?*" (H3). Hepatologists further remarked that motivation may be a barrier in this population: "*There are people who are gonna be good about checking their blood pressure, and would do this [Beacon], and I would say that maybe that's 20%. The vast majority probably would be like, maybe spotty. And then, like 20% would never do it, just because they're just not that motivated in their health to do it, and don't even take their medications daily, much less check something daily*" (H3). Indeed, we found that the highest yield in recruitment was from patients trying to get on a transplant waitlist. Being motivated in seeking help and demonstrating the ability to follow the required procedures to get a transplant suggests that many of these patients are people that are reasonably proactive about engaging with their health. Patient interviews confirmed this, with 14 of 15 interviewed patients reporting that they were primarily motivated to participate in the deployment and study due to curiosity about how the technology might be helpful for understanding their own condition and for the potential to advance science, while 1 patient said that they were participating solely for the compensation. P2 said "*any little bit that I could do to help other people would be beneficial,*" while P5 shared that they "*needed a liver transplant… which was scary… any information or anything that could identify stages of disease… would help.*" This motivation was also evident when patients were asked if they would be willing to using the device after the study, with 10 patients responding they were interested in continuing to use the device, "*even if not being paid*" (P1).

H5 also suggested that caregivers might play a large role in the motivation of patients using Beacon: "*so either the patient himself or herself is very motivated, and will do this app on their own, or they have a very motivated caregiver.*" Of 15 interviewed patients, 3 reported relying on a caregiver (i.e., someone who helps manage their condition on a daily basis), while an additional 2 cited needing a caregiver in the recent past. Caregivers that also participated in patient interviews reported relying on recognizing subtle behavioral cues to determine changes in patient condition. Hepatologist interviews additionally revealed

that, in appointments with patients, they will separately instruct caregivers on what kinds of behaviors to look for. P21's caregiver, who also participated in his interview, described it as: "*he doesn't remember the time of day. Sometimes he gets a little attitude… I can tell when he's a little more irate. He's not very receptive to whatever I ask him to do.*" Caregivers suggested in interviews that they may be separately motivated to use Beacon to more objectively and accurately track a patient's condition as part of their caregiving responsibilities.

6.2.2 Task Load and Temporal Burden of Taking Measurements.

Overall adherence to the measurement schedule was good. The lowest standard deviation of time range was 13 minutes (P18), and the largest standard deviation of times was 8 hours (P6).

We first investigated the temporal burden of a CFF measurement by examining the time required for each protocol. The right panel of Figure 8 shows the distribution of time taken per CFF measurement by protocol. The mean was 2.78 minutes using MOL-D versus 6.37 minutes using FC, but visual inspection found the mean of FC to be misleading due to outliers. We therefore note the median was 2.69 minutes using MOL-D versus 4.84 minutes using FC, and the right panel of Figure 8 shows the underlying distribution as a violin plot. All patients also got faster over time with both protocols, with none getting slower. Each patient's final MOL-D measurement on average took 2.23 minutes less than their first MOL-D measurement (median: −1.25 minutes), and their final FC measurement on average took 4.87 minutes less than their first FC measurement (median: −3.47 minutes). P15 most reduced time for their MOL-D measurements, decreasing by 11 minutes. P11 most reduced time for their FC measurements, decreasing by 12 minutes. These improvements show that practice allows patients to get much faster at taking measurements. The differences in time taken to complete each measurement, across protocol and across patients, primarily reveals the variability between patients in the amount of time they spend between trials or decisions. In the context of our work in designing a system for long-term, regular use, timing is an important factor in informing which protocol to prescribe or in the design of a novel protocol. This consideration is also critical from a patient perspective, in terms of long-term engagement and the likelihood of being incorporated into a routine.

We also gathered quantitative survey instruments intended to gauge subjective experience using three measurement protocols. We presented modified NASA Task Load Index (TLX), System Usability Scale (SUS), and User Burden Scale (UBS) survey instruments on scales from 1 to 5 regarding each of the two measurement protocols. Questions on the SUS instrument are presented with alternating positive and negative tone, so each individual response must be adjusted and scaled before summing to obtain a final score [13]. SUS scores above 84.1 correspond to a letter grade of A+, indicating extremely high usability [50]. Final adjusted mean SUS scores for MOL-D and FC were 88.50 (SD: 1.00) and 86.00 (SD: 0.71), a letter grade of A+ for usability for both protocols [13]. NASA TLX and User Burden Scale responses were consistent with this score, generally indicating low load and low burden. Although there were no significant differences in responses in comparing MOL-D versus FC protocols, 9 patients qualitatively commented on FC's interactive nature, in contrast to simply waiting for MOL-D. P3 explained "*I liked the forced choice better… it was a little more interactive as opposed to the sitting and waiting for [MOL-D]*". P8

compared FC to a fun game: "*I liked the forced choice… That was fun… like a game.*" The overall consistent positive response to Beacon is encouraging. The further positive responses to FC, despite it requiring slightly more time to take a measurement, are then especially encouraging when considering potential response biases associated with MOL protocols.

6.2.3   Presentation of Measurements. The app used in the at-home deployment simply displayed the CFF measurement at the end of the measurement process as a single number, showing no previous data. Towards our goal for personalizing monitoring of CFF, wherein an individual's CFF measurements dipping below a personal or population-based baseline could indicate a worsening condition and motivate intervention, we developed 6 potential visualizations of CFF history. Shown in Figure 9, these were intended to investigate patient preferences in the presentation of their data. Visualizations were designed to display the data at varying degrees of abstraction:

- **Vis-1**. A line chart of binned CFF measures over time. The X-axis is the time of each measurement, equally spaced regardless of the time between measurements. The Y-axis is the binned CFF: measures greater than 2.5 *Hz* above 39 *Hz* were considered "above average", measures less than 2.5 *Hz* below 39 *Hz* were considered "below average", and measures in between were considered "neutral".

- **Vis-2**. This visualization is the same as Vis-1, except that the spacing of measurements on the X-axis are proportional to how far apart they are in time. This change introduces more information than the previous, perhaps helping patients recall or associate measures with other events in their life.

- **Vis-3**. Also similar to Vis-1, except that each bin is encoded using a color and icon combination instead of a line chart. This change gets at a different aesthetic.

- **Vis-4**. A line chart of raw CFF measures over time. The X-axis consists of equally-spaced measurement times, and the Y-axis is the raw CFF measure. A dashed line is drawn to indicate the standard 39 *Hz* threshold, and a highlighted area is used to visually suggest the bins (i.e., "above", "below", "neutral") described in prior visualizations. This visualization further increases the resolution of data available to the patient.

- **Vis-5**. A table with two columns: Date and Measure. Each row consists of a measurement date and result. This visualization provides the greatest resolution of data to the patient by exposing the raw numbers themselves.

- **Vis-6**. Two side-by-side line charts of raw CFF measures over time. The X-axes consists of proportionally-spaced timestamps, and the Y-axis is the raw CFF measure. On the left line chart, the highlighted region centers on 39 *Hz*. On the right line chart, the highlighted region centers on the patient's own mean CFF. This visualization is used to demonstrate the difference between comparing one's measures against an absolute threshold versus comparing one's measures against themselves (i.e., internal consistency). This visualization might be useful in the future scenario where we determine that a stable measure is more

appropriate than comparisons made against an absolute threshold (e.g., as with blood pressure [34]).

As part of the exit interview, patients were shown each of these visualizations, one at a time, populated with their own data as collected during the at-home study. The interviewer explained each visualization as described above. Each patient was asked to provide a score on a Likert scale from 1 (strongly disagree) through 5 (strongly agree) for the prompts "This is appealing to me", "This seems suitable", and "This seems easy to use". These scores are depicted below the corresponding visualization in Figure 9, with Vis-4 having the highest combined score. However, when asked to rank order the visualizations, participants ranked Vis-3 (the colored boxes) as the highest-ranked option because "*I actually like it better cause its colored… It's not that I can't understand the other ones, this is just more appealing*" (P8). P5 also commented that it might make "*decision making to visit doctor [easier]*" because the data was simple to understand. Although the colorful aesthetic was an appealing factor, many patients also mentioned that being able to see the raw numbers was important, and the most commonly suggested improvement was for Vis-3 (the colored boxes) to also show the CFF number. Consistent with Kay et al.'s findings in [32] that there is "*a need to support… connecting high-level summaries to the low-level data,*" this result shows that patients not only want high level abstractions of their data, but also to understand the derivation of those abstractions. This feedback suggests patients have a strong interest in engaging with their CFF measurements.

We also investigated how hepatologists might want to be presented with patient data. Although they saw the potential benefits of at-home longitudinal monitoring, they were concerned about the potential burden posed by the additional responsibility of keeping track of another source of data. H2 explained: "*I'm just saying I don't need every single patient on my panel who has cirrhosis sending me an email like once a week with this [CFF measurements], because so much of that, then, would be almost like data overload. You risk the human error of not identifying the actual sort of significant results within that [CFF measurements]. But I would definitely want [to see] some [data].*" Hepatologists did offer possible strategies for mitigating this burden while still making the most of the potential benefit of longitudinal CFF monitoring, such as by suggesting they would adjust how frequently they would like to see measurements from patients depending on the severity and progression of their condition. Although they previously commented they would not want to see measurements once a week from stable patients, H2 continued "*…maybe for that patient that I'm very worried about, I'd have them do it 3 times a day for those first 3 weeks, you know, and then like, decrease the frequency over time.*" We were encouraged by hepatologists suggesting scenarios in which using Beacon for multiple measurements within a day could be useful as part of their ability to assess a patient's holistic condition, validating our hypothesis that at-home, longitudinal CFF measurements are valuable.

6.2.4 Actionability of Measurements. As part of the at-home deployment, patients were intentionally not informed of what the CFF measurements meant. We made this decision for two reasons: (1) to mitigate response bias, and (2) because we were not confident that we could ethically provide a precise, meaningful, or actionable interpretation to patients. However, in the patient interviews, we explained the high-level concept of what

CFF measurements can indicate with respect to their condition, and we were transparent about current challenges around uncertainty in interpretation of measurements. Nonetheless, patients immediately found potential utility in having the measurements. P5 suggested that "*[Beacon would] make it easy to decide to visit the doctor.*" P10 commented that "*If I had this [Beacon] prior to my surgery it would have been so much easier to anticipate and handle episodes of HE.*" P15 observed that his stable data "*affirms what the doctors have been telling me about being consistent with my treatment and being compliant with what I'm supposed to be doing, medications, dietary restrictions, watching what I do, don't tire myself out, staying the healthiest I can be,*" further suggesting that deviations in his data would prompt him to adjust choices such as medication and dietary restrictions.

Hepatologists explained that the primary action taken by patients experiencing episodes of HE is to adjust their lactulose dosage. H4 suggested that "*if they're consistently above that red bar [39 Hz threshold], I would say you need to take more lactulose… I think it's helpful in that regard because people tend to under judge how bad their control is, or how good their control is. They they think they're fine, but then you see their family members going: Oh, no, they're not under control at all. So this [Beacon] just gives it a an objective test that they can do at home.*" However, the side effects of lactulose are not pleasant and thus a significant deterrent to patients complying with using it on a regular basis. H4 went on to explain that caregivers are heavily relied upon to help prompt and remind patients to take their medication. Indeed, P17's caregiver, who also participated in his interview, offered that "*he just really can't stand it cause it just gives him really explosive diarrhea. He won't die from that [lactulose], but it's not pleasant. That's why I think as caregivers, we're like, just take more, it's fine. As the patient, he's like, this is horrible, and I don't want it. So I think that's where the battle comes in. That's why, when he is off and I tell him he needs this, and he doesn't want to take it, I think, having a number that says, or an agreement that says, if you're above a certain level, then you do need to take additional lactulose, could be really helpful.*" Beyond lactulose, the second item that can be acted on at home is the medication rifaximin. H3 explained that "*rifaximin is essentially our second line treatment, which is highly effective and well-tolerated, but quite expensive. So often there is hesitance to start patients on that medication due to cost. In that way, it [Beacon] is almost a tool for patient advocacy by providing an objective measurement showing that despite an appropriate dosing of lactulose, the patient still has this change in their flicker frequency, the insurance company should like pay for this patient to have rifaximin because it's gonna prevent them from being hospitalized.*" We find that all of these stakeholders agree on the potential utility of Beacon's CFF measurements for helping patients and their caregivers regulate medication usage and manage their condition.

In addition to the vital role caregivers play in helping patients manage their condition, pharmacists can potentially also play a role in their care. H2 suggested that "*you could also imagine this being a pharmacist-driven initiative. So let's say they [patients] send all of their values and they're increased from their baseline. You could have your clinic pharmacist call them, do a medication reconciliation. Try to get rid of sedating medications, modify lactulose and rifaximin prescriptions as needed. I highlight that because physician specialists [such as hepatologists] are a very expensive, very limited commodity, whereas we have many excellent ancillary professionals, such as pharamacist PharmD's who, if it*

is something that is based on a protocol, it is within their expertise. They would have both the bandwidth, the time, and the training to process that data. That could actually be very interesting if you were able to have, like a pharmacist-driven protocol to optimize hepatic encephalopathy medications, using the flicker frequency as sort of the data to drive that." This suggestion presents another opportunity for us to investigate in the process of translating Beacon into practice. Nonetheless, all interviewed hepatologists agreed that, if their patients are taking measurements related to their liver health, then the hepatologist would like to stay updated. H5 said "*You still need a provider to educate and to see them. It's a matter of teaching them, having the provider saying this [CFF measurement] is what's normal. If you're abnormal, go talk to your provider.*" We are encouraged that Beacon was well-received by interviewees and for the potential actionability it can drive for helping patients, caregivers, and healthcare providers manage their patient's condition.

7 DISCUSSION

7.1 Designing Home Health Monitoring Systems

We discuss lessons we learned from our experiences with an at-home longitudinal deployment and how they can inform the design of future home health monitoring devices. First, we enumerate considerations towards designing for uptake (Section 7.1.1) of home health monitoring systems. We then present approaches for removing barriers to uptake by discussing strategies for reducing burden (Section 7.1.2) and surfacing actionability (Section 7.1.3).

7.1.1 Designing for Uptake. As reported in Section 6.2.1, hepatologists commented, based on their experiences in recommending established medical devices such as glucose monitors, that they were concerned that only proactive patients would be able to consistently complete measurements using Beacon. We therefore expect that the relatively high motivation levels of our patient participants represents an optimistic view of engagement with Beacon. From a research standpoint, this sampling was acceptable for demonstrating the feasibility of taking measurements. However, for researchers and designers interested in translating their work into practice, they should anticipate spending effort on outreach and education with patients. These types of efforts are consistent with the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework, which considers factors necessary to implement health and social care in practice [25]. Future work should investigate how other elements of this framework can be used to inform the design and implementation of systems and studies in a way that ultimately supports translation.

As part of validating the novel CFF measure, demonstrating the feasibility of patients taking measurements was the primary goal of our at-home deployment. In addition to feasibility, we were also interested in how consistently patients were able to take CFF measurements. To support this goal, as part of our protocol, we recommended that patients take Beacon measurements in the morning after being awake long enough to not be groggy. We made this broad recommendation because, due to the novelty of the CFF measure, we were unsure of exactly what factors could be the largest confounds. As we showed in Section 6.1.1, some patients were able to adhere to this recommendation, while others demonstrated

greater variability in their measurement timings. These participants cited barriers similar to the findings of Xu et al. [60] and Cordeiro et al. in adhering to physical activity and food journaling [17], respectively. For readers designing their own home health devices, we thus suggest limiting assumptions as to *when* measurements will be taken.

For many health measures, variability in timing also has implications in measurement fluctuations, such as fatigue or weight gain over the course of a day. Indeed, Kay et al. [33] made similar design recommendations around reflecting data uncertainty in a weight scale to account for natural fluctuations over the course of a day. Findings from our patient interviews supported the need for these design recommendations. Although patients were not aware of what the CFF measurements meant, they were curious about what was causing changes in their measurements. Observing a dip in their measures, P2 commented "*[it was] low probably because I was distracted or maybe I was tired.*" Because measurements were taken at home without the presence of research coordinator, patients also commented on the flexibility to try again. P4 shared that "*If [the measure was] low, I'd take a nap, do it again just to make sure I did it properly the last time.*" P10 similarly reported that "*[if] the score's low, let's do it 5 [more] times.*" To bridge this gap, home health monitoring systems should incorporate features that not only present measurements to patients but also help patients understand their results. Future work may also investigate strategies for adjusting or correcting obtained measurement values based on environmental factors.

7.1.2 Reducing Burden. Towards assessing the feasibility of taking consistent CFF measures, as described in Section 6.1.1, we investigated patient experience in the measurement process. Specifically, we were interested in understanding the burden involved and how it may deter patients from taking regular measures. We compared two measurement protocols with contrasting properties to understand their trade-offs: the MOL-D protocol, a quicker but less stable measure; and the FC protocol, a longer but more robust measure (as characterized in Section 6.1.2). We expected "burden" to align with the dimensions of the NASA-TLX assessment. However, we were surprised to find the difference in NASA-TLX scores across the two protocols was not significant, not even in temporal demand. Instead, patients explained how the interactivity of the FC protocol made it more engaging than the MOL-D protocol. Based on this finding, we recommend designing home health systems that are interactive and engaging in order to minimize the perceived burden of taking measurements. Future research can also investigate how measurements taken longitudinally can accelerate the measurement process based on already-collected longitudinal data.

7.1.3 Surfacing Actionability. Finally, a factor that we did not study as part of our at-home deployment is the actionability of Beacon's CFF measurements. With the limited evidence available prior to this study, we did not feel that it was ethical to provide guidance on how to interpret measurements, nor could we have anticipated how to provide guidance on the different ways people might consider acting on measurements. Because we did not explain to the patients what the CFF measurements meant until after the deployment, patients were not able to act upon their own CFF measurements. However, if they were informed of what Beacon's CFF measurements meant, it is possible that they would have taken measurements at a more consistent time or perhaps a more strategic time. For

example, a patient might choose to take a measurement before each time they drive a car, as a low measurement might suggest to them that they are not fit to drive. Similarly, a patient might choose to take a measurement only when they know they have time to spend in the restroom, in the case that Beacon's measurement suggests to them that they should take lactulose. Now that we have confirmed feasibility, future work should investigate more natural measurement habits and how patients might engage with their CFF data. More broadly, future work should study, once patients have access to a new data stream and are aware of how measures can inform treatment decisions, how patient behavior and everyday decisions might evolve.

7.2 The Role of Caregivers

Caregivers play an important role in the health of individuals with chronic conditions [43]. In many situations, the responsibilities of caregivers involve "*invisible work*" [28] in mitigating tensions between the medical needs of care recipients and their personal desires. For example, caregivers of diabetic patients often help with regulating sugar consumption, including limiting desired foods when necessary [15, 28]. Another example is in caregivers of patients with dementia who must balance between supporting the care recipient in performing activities of daily living and fully taking over these tasks [26, 28]. Consistent with prior studies [8, 16], our patient and hepatologist interviews discussed the unique role caregivers fill in the management of a liver disease patient's condition: adjusting medication (lactulose) dosage on a day-to-day basis. When possible and appropriate, healthcare providers will instruct caregivers in how to adjust medication dosage, and will depend on caregivers to understand the patient's changes in behavior and condition over time. However, the negative side effects of lactulose leads to tensions between the desire for the patient to feel better and the negative side effects of spending hours in the bathroom. Caregivers described needing to convince their patient that medication was a good idea, typically based on their own subjective observations of the patient's behavior. Future work should strive to continue to understand the lived experiences of people with chronic liver disease and their caregivers, and how their experiences with Beacon evolve in the face of clinical events. As suggested by caregivers and hepatologists in our interviews, and consistent with prior work in the design of tools for patients to share their experiences with caregivers [27], we anticipate that effective presentation of objective CFF measurements could be used to mediate some of these tensions.

7.3 Reflecting on Designing for Chronic Liver Disease Patients and Beyond

Although Karkar et al. [31] introduced the reframing of CFF from a clinical screening test to a self-administered measure for self-monitoring, their work was limited to testing on healthy participants. The goal for our work was to pursue evidence of Beacon's feasibility and appropriateness in home settings with our target population of patients with liver disease. We therefore began patient recruitment for this study by approaching patients in a hepatology clinic waiting area, mirroring the convenience sampling techniques commonly used by our colleagues in HCI. However, recruiting from a specialized clinic, such as the hepatology clinic, is different from a primary care clinic. Although liver disease is the 9th-leading cause of death in the United States, treatment often requires access to providers and equipment not available in a patient's own city or state. Many commute multiple hours and across

state boundaries to visit a clinic (e.g., our hepatology clinic serves our entire state and a cluster of 4 surrounding states). Because of the difficulty of getting to the clinic, patients are often scheduled for a full day of back-to-back tests and exams, making it infeasible to recruit patients on the day of a clinic visit. Ultimately, we arrived at a process in which we called patients before a clinic visit, checked for interest and eligibility, and explicitly added our study session to their clinical schedule. Through this process, we found that the highest yield in responses came from patients going through the process of qualifying for the liver transplant waitlist. This selection bias may have affected our findings, reflecting primarily the experiences of relatively motivated patients. Following our work on comparing the usability and burden of the two different measurement protocols as described in Section 6.2.2, better gauging how long a measurement takes (e.g., "a 5-minute test") can be used to inform the design of future research studies. The consideration that time must be built into a patient's schedule can be used by researchers to assess interest of potential participants. Future work can then seek to better gauge how interested less motivated individuals would be in using Beacon monitor their condition.

Although our work on Beacon thus far has revolved around the chronic liver disease patient population, the CFF measure has also been shown to relate to other chronic conditions, such as Alzheimer's disease [18], Lewy-body dementia [20], and multiple sclerosis [56]. Although there are opportunities for Beacon to be utilized by a broader range of patient populations, further research is necessary to determine the specific treatments that Beacon can effectively support in each of these conditions. Thus we suggest that future work can explore self-monitoring opportunities for other chronic conditions by adopting a similar approach to our work: investigating a novel measure, determining the treatments that the measure can be used to support, and understanding and engaging the stakeholders involved in using the measure.

8 CONCLUSION

We reported on our iterative design process to improve Beacon's physical device internals, assembly process, and form factor; our redesign of the phone app; and our development of a new cloud server with dashboard for remote monitoring of devices. We shared findings from a 6-week at-home deployment with 21 chronic liver disease patients, demonstrating feasibility of taking consistent CFF measurements and enabling a characterization of longitudinal CFF measurements We also reported on insights from interviews with 15 patients and 5 hepatologists, revealing the role of caregivers in patient motivation to take measurements, the burden of taking measurements, preferences around the presentation of measurements, and the actionability of measurements. Finally, we discussed considerations for designing home health systems informed by our experiences and how they might apply broadly.

This paper therefore represents important progress toward our vision to enable more widespread and frequent usage of CFF measurements for both clinical screening and for at-home self-monitoring in supporting chronic liver disease patients, their caregivers, and their healthcare providers in better managing their condition. Our demonstration of the feasibility of taking CFF measurements using Beacon and our exploration of the clinical relevance of

these measurements provides a basis for future work exploring real-world deployments in which patients interpret and use measurements as part of their related decision-making.

ACKNOWLEDGMENTS

REFERENCES

[1]. Ampuero Javier, Simón Macarena, Montoliú Carmina, Jover Rodrigo, Serra Miguel Ángel, Córdoba Juan, and Romero-Gómez Manuel. 2015. Minimal hepatic encephalopathy and critical flicker frequency are associated with survival of patients with cirrhosis. Gastroenterology 149, 6 (2015), 1483–1489. [PubMed: 26299413]

[2]. Bajaj Jasmohan S. 2015. Diagnosing minimal hepatic encephalopathy: from the ivory tower to the real world. Gastroenterology 149, 6 (2015), 1330–1333. [PubMed: 26416330]

[3]. Bajaj Jasmohan Singh, Etemadian Ashkan, Hafeezullah Muhammad, and Saeian Kia. 2007. Testing for minimal hepatic encephalopathy in the United States: an AASLD survey. Hepatology 45, 3 (2007), 833–834. [PubMed: 17326210]

[4]. Bajaj Jasmohan S, Heuman Douglas M, Sterling Richard K, Sanyal Arun J, Siddiqui Muhammad, Matherly Scott, Luketic Velimir, Stravitz R Todd, Fuchs Michael, Thacker Leroy R, et al. 2015. Validation of EncephalApp, smartphone-based Stroop test, for the diagnosis of covert hepatic encephalopathy. Clinical Gastroenterology and Hepatology 13, 10 (2015), 1828–1835. [PubMed: 24846278]

[5]. Bajaj Jasmohan S, Heuman Douglas M, Wade James B, Gibson Douglas P, Saeian Kia, Wegelin Jacob A, Hafeezullah Muhammad, Bell Debulon E, Sterling Richard K, Stravitz R Todd, et al. 2011. Rifaximin improves driving simulator performance in a randomized trial of patients with minimal hepatic encephalopathy. Gastroenterology 140, 2 (2011), 478–487. [PubMed: 20849805]

[6]. Bajaj Jasmohan S, Saeian Kia, Christensen Kenneth M, Hafeezullah Muhammad, Varma Rajiv R, Franco Jose, Pleuss Joan A, Krakower Glenn, Hoffmann Raymond G, and Binion David G. 2008. Probiotic yogurt for the treatment of minimal hepatic encephalopathy. Official journal of the American College of Gastroenterology| ACG 103, 7 (2008), 1707–1715.

[7]. Bajaj Jasmohan S, Saeian Kia, Schubert Christine M, Hafeezullah Muhammad, Franco Jose, Varma Rajiv R, Gibson Douglas P, Hoffmann Raymond G, Stravitz R Todd, Heuman Douglas M, et al. 2009. Minimal hepatic encephalopathy is associated with motor vehicle crashes: the reality beyond the driving test. Hepatology 50, 4 (2009), 1175–1183. [PubMed: 19670416]

[8]. Bajaj Jasmohan S, Wade James B, Gibson Douglas P, Heuman Douglas M, Thacker Leroy R, Sterling Richard K, Stravitz Todd R, Luketic Velimir, Fuchs Michael, White Melanie B, et al. 2011. The multi-dimensional burden of cirrhosis and hepatic encephalopathy on patients and caregivers. Official journal of the American College of Gastroenterology| ACG 106, 9 (2011), 1646–1653.

[9]. Bascom Emily, Casanova-Perez Reggie, Tobar Kelly, Bedmutha Manas Satish, Ramaswamy Harshini, Pratt Wanda, Sabin Janice, Wood Brian, Weibel Nadir, and Hartzler Andrea. 2024. Designing Communication Feedback Systems To Reduce Healthcare Providers' Implicit Biases In Patient Encounters. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–12.

[10]. Berry Andrew BL, Lim Catherine Y, Hirsch Tad, Hartzler Andrea L, Kiel Linda M, Bermet Zoë A, and Ralston James D. 2019. Supporting communication about values between people with multiple chronic conditions and their providers. In proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–14.

[11]. Bhat Karthik S, Hall Amanda K, Kuo Tiffany, and Kumar Neha. 2023. " We are half-doctors": Family Caregivers as Boundary Actors in Chronic Disease Management. Proceedings of the ACM on Human-Computer Interaction 7, CSCW1 (2023), 1–29.

[12]. Breda Joseph, Springston Mastafa, Mariakakis Alex, and Patel Shwetak. 2023. Feverphone: Accessible core-body temperature sensing for fever monitoring using commodity smartphones. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 7, 1 (2023), 1–23.

[13]. Brooke John et al. 1996. SUS-A quick and dirty usability scale. Usability evaluation in industry 189, 194 (1996), 4–7.

[14]. Cappuccio Francesco P, Kerry Sally M, Forbes Lindsay, and Donald Anna. 2004. Blood pressure control by home monitoring: meta-analysis of randomised trials. Bmj 329, 7458 (2004), 145. [PubMed: 15194600]

[15]. Cha Yoon Jeong, Wou Alice, Saxena Arpita, Lee Joyce, Newman Mark W, and Park Sun Young. 2023. It's like an educated guessing game: Parents' strategies for collaborative diabetes management with their children. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 1–15.

[16]. Chow Kenneth W, Ibrahim Brittney M, Yum Jung J, Dang An, Dang Long, Chen Kuan-Ting, Jackson Nicholas J, and Saab Sammy. 2023. Barriers to lactulose adherence in patients with cirrhosis and hepatic encephalopathy. Digestive Diseases and Sciences 68, 6 (2023), 2389–2397. [PubMed: 37119376]

[17]. Cordeiro Felicia, Epstein Daniel A, Thomaz Edison, Bales Elizabeth, Jagannathan Arvind K, Abowd Gregory D, and Fogarty James. 2015. Barriers and negative nudges: Exploring challenges in food journaling. In Proceedings of the 33rd annual ACM conference on human factors in computing systems. 1159–1162.

[18]. Curran Stephen and Wattis John. 2000. Critical flicker fusion threshold: a potentially useful measure for the early detection of Alzheimer's disease. Human Psychopharmacology: Clinical and Experimental 15, 2 (2000), 103–112. [PubMed: 12404339]

[19]. Eisen-Enosh Auria, Farah Nairouz, Burgansky-Eliash Zvia, Polat Uri, and Mandel Yossi. 2017. Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution. Scientific reports 7, 1 (2017), 15621. [PubMed: 29142231]

[20]. Erskine Daniel, Taylor John-Paul, Thomas Alan, Collerton Daniel, McKeith Ian, Khundakar Ahmad, Attems Johannes, and Morris Christopher. 2019. Pathological changes to the subcortical visual system and its relationship to visual hallucinations in dementia with Lewy bodies. Neuroscience Bulletin 35 (2019), 295–300. [PubMed: 30729454]

[21]. Ferenci Peter, Lockwood Alan, Mullen Kevin, Tarter Ralph, Weissenborn Karin, Blei Andres T, et al. 2002. Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology 35, 3 (2002), 716–721. [PubMed: 11870389]

[22]. National Center for Health Statistics et al. 2015. Summary health statistics: National health interview survey, 2018. Centers for Disease Control and Prevention (2015).

[23]. Goel Mayank, Saba Elliot, Stiber Maia, Whitmire Eric, Fromm Josh, Larson Eric C, Borriello Gaetano, and Patel Shwetak N. 2016. Spirocall: Measuring lung function over a phone call. In Proceedings of the 2016 CHI conference on human factors in computing systems. 5675–5685.

[24]. Green Beverly B, Anderson Melissa L, Cook Andrea J, Ehrlich Kelly, Hall Yoshio N, Hsu Clarissa, Joseph Dwayne, Klasnja Predrag, Margolis Karen L, McClure Jennifer B, et al. 2022. Clinic, home, and kiosk blood pressure measurements for diagnosing hypertension: a randomized diagnostic study. Journal of General Internal Medicine 37, 12 (2022), 2948–2956. [PubMed: 35239109]

[25]. Greenhalgh Trisha, Wherton Joseph, Papoutsi Chrysanthi, Lynch Jennifer, Hughes Gemma, Hinder Susan, Fahy Nick, Procter Rob, Shaw Sara, et al. 2017. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. Journal of medical Internet research 19, 11 (2017), e8775.

[26]. Guan Connie, Bouzida Anya, Oncy-Avila Ramzy M, Moharana Sanika, and Riek Laurel D. 2021. Taking an (embodied) cue from community health: Designing dementia caregiver support

technology to advance health equity. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–16.

[27]. Hong Matthew K, Lakshmi Udaya, Do Kimberly, Prahalad Sampath, Olson Thomas, Arriaga Rosa I, and Wilcox Lauren. 2020. Using diaries to probe the illness experiences of adolescent patients and parental caregivers. In Proceedings of the 2020 chi conference on human factors in computing systems. 1–16.

[28]. Hsu Long-Jing and Chung Chia-Fang. 2024. Dancing with the Roles: Towards Designing Technology that Supports the Multifaceted Roles of Caregivers for Older Adults. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–12.

[29]. Jiao Yuxin, Lin Rose, Hua Xinyang, Churilov Leonid, Gaca Michele J, James Steven, Clarke Philip M, O'Neal David, and Ekinci Elif I. 2022. A systematic review: Cost-effectiveness of continuous glucose monitoring compared to self-monitoring of blood glucose in type 1 diabetes. Endocrinology, diabetes & metabolism 5, 6 (2022), e369.

[30]. Kappus Matthew R and Bajaj Jasmohan S. 2012. Assessment of minimal hepatic encephalopathy (with emphasis on computerized psychometric tests). Clinics in liver disease 16, 1 (2012), 43–55. [PubMed: 22321464]

[31]. Karkar Ravi, Kocielnik Rafal, Zhang Xiaoyi, Zia Jasmine, Ioannou George N, Munson Sean A, and Fogarty James. 2018. Beacon: Designing a portable device for self-administering a measure of critical flicker frequency. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 3 (2018), 1–27.

[32]. Kay Matthew, Choe Eun Kyoung, Shepherd Jesse, Greenstein Benjamin, Watson Nathaniel, Consolvo Sunny, and Kientz Julie A. 2012. Lullaby: a capture & access system for understanding the sleep environment. In Proceedings of the 2012 ACM conference on ubiquitous computing. 226–234.

[33]. Kay Matthew, Morris Dan, Schraefel MC, and Kientz Julie A. 2013. There's no such thing as gaining a pound: Reconsidering the bathroom scale user interface. In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. 401–410.

[34]. Kendall Logan, Morris Dan, and Tan Desney. 2015. Blood pressure beyond the clinic: Rethinking a health metric for everyone. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 1679–1688.

[35]. Khan Danish U, Siek Katie A, Meyers Jane, Haverhals Leah M, Cali Steven, and Ross Stephen E. 2010. Designing a personal health application for older adults to manage medications. In Proceedings of the 1st ACM International Health Informatics Symposium. 849–858.

[36]. Kircheis Gerald, Wettstein Matthias, Timmermann Lars, Schnitzler Alfons, and Häussinger Dieter. 2002. Critical flicker frequency for quantification of low-grade hepatic encephalopathy. Hepatology 35, 2 (2002), 357–366. [PubMed: 11826409]

[37]. Larson Eric C, Goel Mayank, Boriello Gaetano, Heltshe Sonya, Rosenfeld Margaret, and Patel Shwetak N. 2012. SpiroSmart: using a microphone to measure lung function on a mobile phone. In Proceedings of the 2012 ACM Conference on ubiquitous computing. 280–289.

[38]. Li Chen, Lee Chang-Franw, and Xu Song. 2020. Stigma Threat in Design for Older Adults: Exploring Design Factors that Induce Stigma Perception. International Journal of Design 14, 1 (2020).

[39]. Lim Brian Y and Dey Anind K. 2009. Assessing demand for intelligibility in context-aware applications. In Proceedings of the 11th international conference on Ubiquitous computing. 195–204.

[40]. Mallick Salman, Kanthety Radhika, and Rahman Mahboob. 2009. Home blood pressure monitoring in clinical practice: a review. The American journal of medicine 122, 9 (2009), 803–810. [PubMed: 19699371]

[41]. McIntosh Brendan, Yu Changhua, Lal Avtar, Chelak Kristen, Cameron Chris, Singh Sumeet R, and Dahl Marshall. 2010. Efficacy of self-monitoring of blood glucose in patients with type 2 diabetes mellitus managed without insulin: a systematic review and meta-analysis. Open Medicine 4, 2 (2010), e102. [PubMed: 21686299]

[42]. Murphy Sherry L, Kochanek Kenneth D, Xu Jiaquan, and Arias Elizabeth. 2021. Mortality in the United States, 2020. (2021).

[43]. Pina Laura R, Sien Sang-Wha, Ward Teresa, Yip Jason C, Munson Sean A, Fogarty James, and Kientz Julie A. 2017. From personal informatics to family informatics: Understanding family practices around health monitoring. In Proceedings of the 2017 acm conference on computer supported cooperative work and social computing. 2300–2315.

[44]. Poolsup Nalinee, Suksomboon Naeti, and Kyaw Aye Mon. 2013. Systematic review and meta-analysis of the effectiveness of continuous glucose monitoring (CGM) on glucose control in diabetes. Diabetology & metabolic syndrome 5 (2013), 1–14. [PubMed: 23298687]

[45]. Prasad Srinivasa, Dhiman Radha K, Duseja Ajay, Chawla Yogesh K, Sharma Arpita, and Agarwal Ritesh. 2007. Lactulose improves cognitive functions and health-related quality of life in patients with cirrhosis who have minimal hepatic encephalopathy. Hepatology 45, 3 (2007), 549–559. [PubMed: 17326150]

[46]. Román Eva, Córdoba Joan, Torrens Maria, Torras Xavier, Villanueva Càndid, Vargas Víctor, Guarner Carlos, and Soriano Germán. 2011. Minimal hepatic encephalopathy is associated with falls. Official journal of the American College of Gastroenterology| ACG 106, 3 (2011), 476–482.

[47]. Romero-Gómez Manuel. 2007. Critical flicker frequency: it is time to break down barriers surrounding minimal hepatic encephalopathy. Journal of hepatology 47, 1 (2007), 10–11. [PubMed: 17507110]

[48]. Romero-Gómez Manuel, Boza Fernando, García-Valdecasas Maria Sierra, García Emilio, and Aguilar-Reina José. 2001. Subclinical hepatic encephalopathy predicts the development of overt hepatic encephalopathy. Official journal of the American College of Gastroenterology| ACG 96, 9 (2001), 2718–2723.

[49]. Romero-Gómez Manuel, Córdoba Juan, Jover Rodrigo, Del Olmo Juan A, Ramírez Marta, Rey Ramón, De Madaria Enrique, Montoliu Carmina, Nuñez David, Flavia Montse, et al. 2007. Value of the critical flicker frequency in patients with minimal hepatic encephalopathy. Hepatology 45, 4 (2007), 879–885. [PubMed: 17393525]

[50]. Sauro Jeff and Lewis James R. 2016. Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

[51]. Sharma Praveen, Sharma Barjesh Chander, and Sarin Shiv Kumar. 2010. Critical flicker frequency for diagnosis and assessment of recovery from minimal hepatic encephalopathy in patients with cirrhosis. Hepatobiliary & pancreatic diseases international: HBPD INT 9, 1 (2010), 27–32. [PubMed: 20133225]

[52]. Shawcross Debbie L, Dunk Arthur A, Jalan Rajiv, Kircheis Gerald, De Knegt Robert J, Laleman Wim, Ramage John K, Wedemeyer Heiner, and Morgan Ian EJ. 2016. How to diagnose and manage hepatic encephalopathy: a consensus statement on roles and responsibilities beyond the liver specialist. European journal of gastroenterology & hepatology 28, 2 (2016), 146–152. [PubMed: 26600154]

[53]. Sidhu Sandeep Singh, Goyal Omesh, Mishra Bholeshwar Prashad, Sood Ajit, Chhina Rajoo Singh, and Soni Ravinder Kumar. 2011. Rifaximin improves psychometric performance and health-related quality of life in patients with minimal hepatic encephalopathy (the RIME Trial). Official journal of the American College of Gastroenterology| ACG 106, 2 (2011), 307–316.

[54]. St Fleur Ruth Gaelle, St George Sara Mijares, Leite Rafael, Kobayashi Marissa, Agosto Yaray, and Jake-Schoffman Danielle E. 2021. Use of Fitbit devices in physical activity intervention studies across the life course: narrative review. JMIR mHealth and uHealth 9, 5 (2021), e23411. [PubMed: 34047705]

[55]. Stergiou George S and Bliziotis Ioannis A. 2011. Home blood pressure monitoring in the diagnosis and treatment of hypertension: a systematic review. American journal of hypertension 24, 2 (2011), 123–134. [PubMed: 20940712]

[56]. Titcombe AF and Willison RG. 1961. Flicker fusion in multiple sclerosis. Journal of Neurology, Neurosurgery, and Psychiatry 24, 3 (1961), 260. [PubMed: 13777050]

[57]. Torlot FJ, McPhail MJW, and Taylor-Robinson SD. 2013. Meta-analysis: the diagnostic accuracy of critical flicker frequency in minimal hepatic encephalopathy. Alimentary pharmacology & therapeutics 37, 5 (2013), 527–536. [PubMed: 23293917]

[58]. Vilstrup Hendrik, Amodio Piero, Bajaj Jasmohan, Cordoba Juan, Ferenci Peter, Mullen Kevin D, Weissenborn Karin, and Wong Philip. 2014. Hepatic encephalopathy in chronic liver disease:

2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. Hepatology 60, 2 (2014), 715–735. [PubMed: 25042402]

[59]. Vutien Philip, Li Richard, Karkar Ravi, Munson Sean A, Fogarty James, Walter Kara, Yacoub Michael, and Ioannou George N. 2022. Evaluating a Novel, Portable, Self-Administrable Device ("Beacon") That Measures Critical Flicker Frequency as a Test for Hepatic Encephalopathy. Official journal of the American College of Gastroenterology| ACG (2022), 10–14309.

[60]. Xu Kefan, Yan Xinghui, and Newman Mark W. 2022. Understanding people's experience for physical activity planning and exploring the impact of historical records on plan creation and execution. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1–15.

[61]. Xue Qiuyue Shirley, Liu Yujia, Breda Joseph, Springston Mastafa, Iyer Vikram, and Patel Shwetak. 2024. Thermal Earring: Low-power Wireless Earring for Longitudinal Earlobe Temperature Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 7, 4 (2024), 1–28.

[62]. Yarows Steven A, Julius Stevo, and Pickering Thomas G. 2000. Home blood pressure monitoring. Archives of Internal Medicine 160, 9 (2000), 1251–1257. [PubMed: 10809027]

[63]. Zacharias HD, Jackson CD, Morgan MY, and Olesen SS. 2017. stepwise diagnosis in covert hepatic encephalopathy-critical flicker frequency and MELD-score as a first-step approach. Replication and pitfalls. Alimentary Pharmacology and Therapeutics 45, 1 (2017), 187–189. [PubMed: 27910151]

[64]. Zhan Tianzuo and Stremmel Wolfgang. 2012. The diagnosis and treatment of minimal hepatic encephalopathy. Deutsches Ärzteblatt International 109, 10 (2012), 180. [PubMed: 22470407]

[65]. Ziefle Martina and Schaar Anne Kathrin. 2014. Technology acceptance by patients: empowerment and stigma. Handbook of smart homes, health care and well-being (2014), 1–10.

CCS CONCEPTS

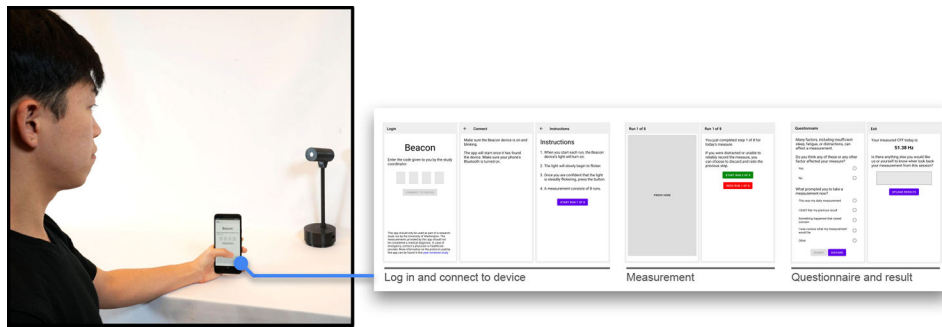Human-centered computing → Empirical studies in HCI.

Figure 1:
Beacon consists of a physical device and an accompanying app. The device is used to precisely render a flickering optical stimulus. The app is primarily used to control the device and facilitate measurements, but also provides basic functionality to support patients in selecting their desired measurement protocol and recording reflections on their measurement.
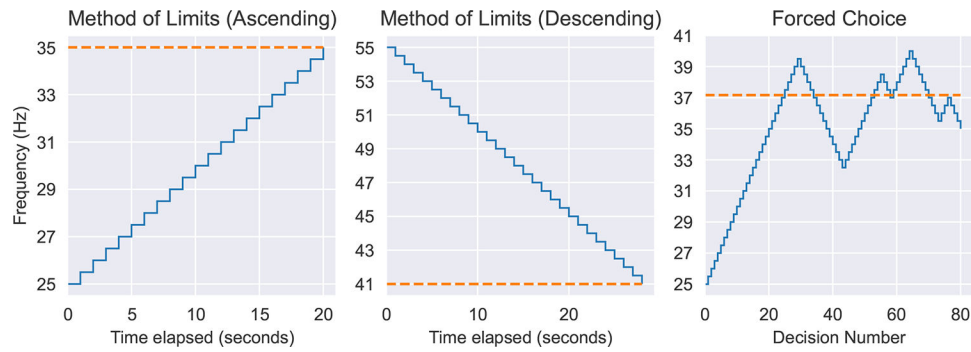
Figure 2:

Example trials for three different measurement protocols for obtaining a CFF measure.

Left: The MOL-A protocol renders a light stimulus that initially appears to flicker and steadily increases in frequency. At time $t = 20s$, the person indicates that it has begun to appear fused, and their CFF is determined to be 35 *Hz*.

Center: The MOL-D protocol renders a light stimulus that initially appears to be solid and steadily decreases in frequency. At time $t = 29s$, the person indicates that it has begun to appear to flicker, and their CFF is determined to be 41 *Hz*.

Right: The FC protocol involves the person deciding which one of two stimulus options is flickering: one always appears fused, and the other at a variable frequency as shown in this plot. Correct answers increase the frequency of the variable option (i.e., where the plot trends upward, such as the first 30 decisions). Incorrect answers decrease the frequency (i.e., where the plot trends downward, such as decisions 30 through 41). The person's CFF is determined to be 37.17 *Hz* based on the average of the frequencies at which the direction changes.

Figure 3:
Left: Explorations and iterations in Beacon's form, from the early prototype developed by Karkar et al. [31] (left-most), through early low-fidelity explorations of potential forms, to development and refinement of Beacon's current form (right-most).

Right: An exploded view of the main components in the current Beacon design. Note the threading designed into each 3D-printed component, allowing parts to be screwed together in assembly, reducing required time and potential error (e.g., glue damage).

Figure 4:
Beacon devices were shipped to patients in custom packaging. Patients were also provided a phone and charger to be used in the study. An instruction booklet and study calendar were included to help facilitate patient self-measurement.

Figure 5:
Of 21 participants in our at-home deployments, 12 had previously participated in Vutien et al.'s clinical study, in which they used the Lafayette Flicker Fusion device. Participant preference for Beacon was generally attributed to improved usability.
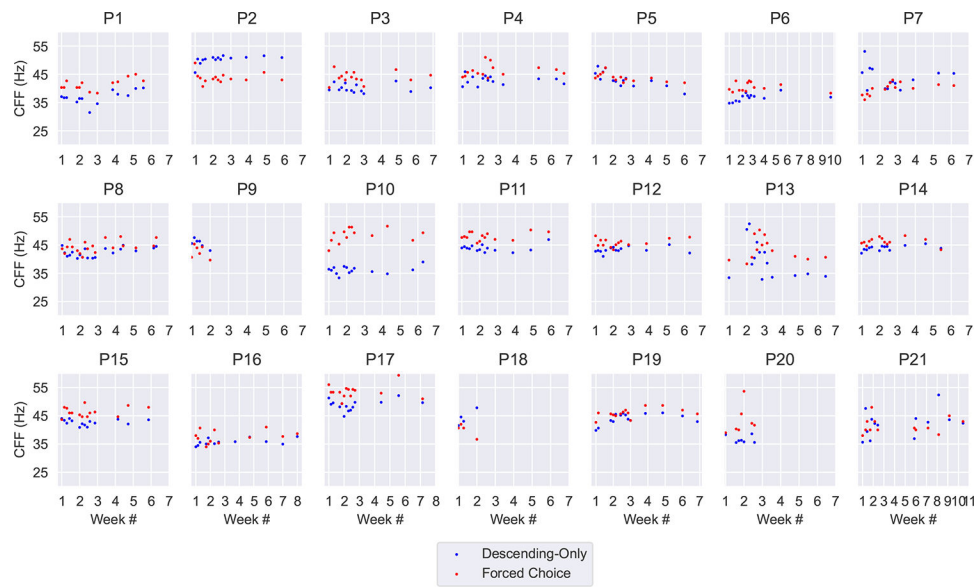
Figure 6:
A plot of each CFF measure taken by 21 patients using Beacon at-home for 6 weeks with MOL-D and FC protocols.
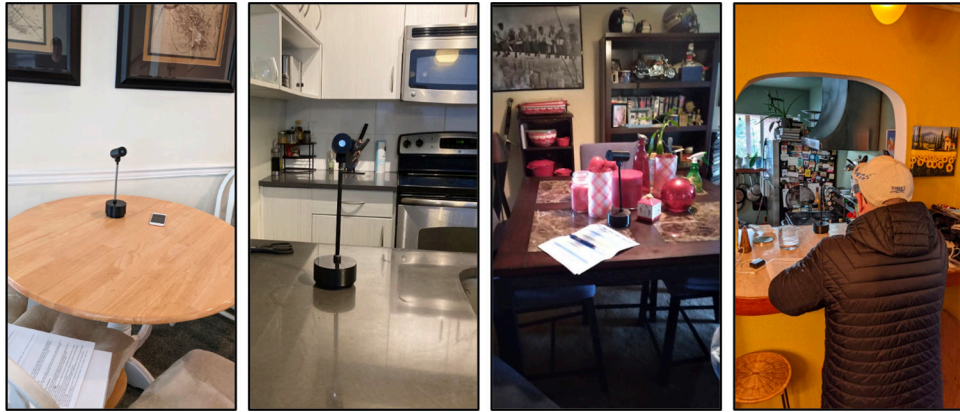
Figure 7:
Participant-provided photos of environments where they used Beacon to take CFF measurements in their own homes. Participants described keeping Beacon out in their home throughout the study, including because its visibility served as a reminder to take measurements. Participant responses thus validate our design of Beacon to be small, portable, and attractive.
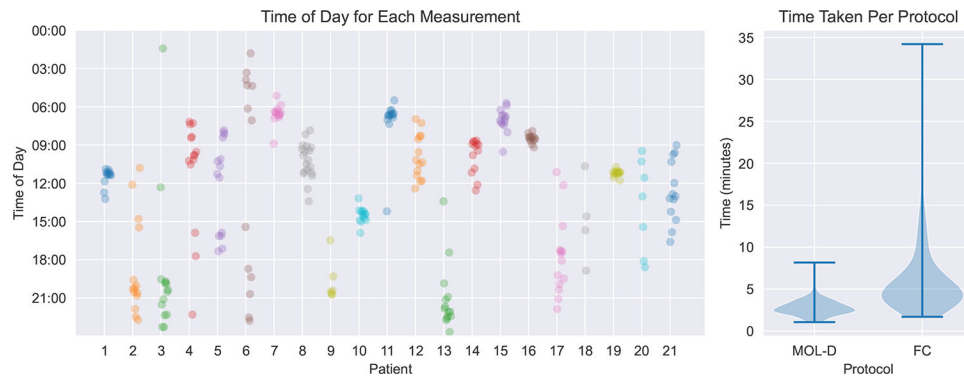
Figure 8:

Overview of the time of day for each patient measurement and the time required with each protocol.

Left: Time of day patients took measurements. Patients were encouraged to set a time for taking measurements each day, but also demonstrated variation due to schedule disturbances. CFF measures appeared stable even in the presence of this variation.

Right: Distribution of time taken per CFF measurement by protocol across all at-home patients. The filled area represents the probability density of the data, and the lines at the top and bottom indicate the maximum and minimum.
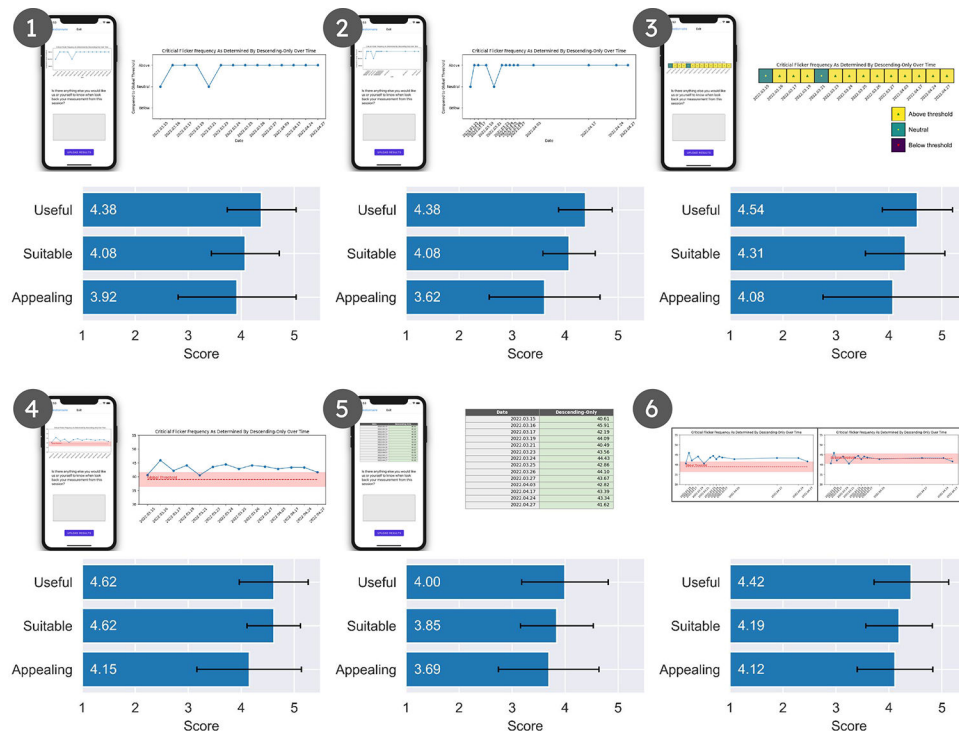
Figure 9:
Six different visualizations to depict CFF measurements over time, representing timestamps and measurements at different levels of resolution. Each visualization was presented to each patient with their own measurements. Patients were asked to score each visualization on whether it was useful, suitable, and appealing on a scale from 1 to 5. Means of these scores are shown as bar charts under the corresponding visualization, with standard deviation as error bars.

Table 1:

At-home study patient demographic information, including age, sex, occupation, and liver disease etiology. We also indicate whether a patient participated in Vutien et al.'s clinical study [59] and whether they participated in an exit interview.

Patient #	Clinical Study?	Exit Interview?	Age	Sex	Occupation	Liver Disease Etiology
P1	Yes	Yes	60–69	Male	Retired	Hepatitis C
P2	Yes	Yes	50–59	Male	Retail	Alcohol
P3	Yes	Yes	40–49	Male	Information Technology	Autoimmune liver disease
P4	Yes	Yes	20–29	Male	Information Technology	Wilsonation Tech
P5	Yes	Yes	50–59	Female	Retail	Alcohol
P6	Yes	No	50–59	Male	N/A	Autoimmune liver disease
P7	Yes	Yes	50–59	Male	Retired	Hepatitis C
P8	Yes	Yes	60–69	Female	Retired	Other
P9	Yes	No	30–39	Male	N/A	Alcohol
P10	Yes	Yes	60–69	Female	Retired	NAFLD/NASH
P11	Yes	No	60–69	Female	N/A	NAFLD/NASH
P12	Yes	No	50–59	Male	N/A	Alcohol
P13	No	Yes	50–59	Male	Unemployed	Hepatitis C
P14	No	Yes	30–39	Male	Unemployed	Hepatitis C
P15	No	Yes	60–69	Male	Retired	Alcohol
P16	No	Yes	40–49	Male	Accountant	Alcohol
P17	No	Yes	40–49	Male	Cook	Alcohol
P18	No	No	60–69	Female	N/A	Alcohol
P19	No	Yes	50–59	Male	Civil Service	NAFLD/NASH
P20	No	No	50–59	Male	N/A	TBD
P21	No	Yes	60–69	Male	Civil Service	TBD

Table 2:

Hepatologist interview participant background. Experience refers to years of experience as a hepatology specialist. Practice refers to the setting at which they practice.

Hepatologist #	Experience	Practice
H1	18	Public
H2	3	Public
H3	17	Public
H4	3	Public
H5	18	Private