Enabling At-Home Health Interventions with Personal Health Technologies

Richard Li

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

James Fogarty, Co-Chair

Shwetak Patel, Co-Chair

Sean A. Munson

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Enabling At-Home Health Interventions with Personal Health Technologies

Richard Li

Co-Chairs of the Supervisory Committee:
James Fogarty
Shwetak Patel
Computer Science & Engineering

While the growing ubiquity of personal health technologies holds promise for supporting healthier lifestyles, many of these technologies currently focus on data collection, with limited support for translating that data into meaningful action. In this dissertation, I will illustrate how designing these technologies for the challenges of at-home, longitudinal use can enable novel health interventions. First, I will introduce Beacon, a system for self-measurement of cognitive function, and how it can be used to support patients in adjusting their medication dosage. Then, I will demonstrate how ExerciseRx, an interactive movement feedback system, promotes exercise in patients and enables them to monitor their progression in movement quality. Finally, I will show how PARX builds upon ExerciseRx and closes the gap between patients and healthcare providers by enabling providers to generate and tailor exercise plans to individual needs. I will conclude by discussing on-going work in translating each of these projects into the real world as well as overarching lessons learned from developing and deploying personal home health technologies.

Contents

Acknowledgments

Humans are hard, but they also make so many things better. People always say the best of academia is the work that you get to do. I'd argue, more importantly, it's the people that you do the work with. Before I get to the work (the dissertation research), I first want to acknowledge all the people that have supported me along the way to this point.

First, I want to thank my masters co-advisors at Georgia Tech: Thad Starner and Gregory Abowd. Thad, your enthusiasm for science and hacking is infectious. I admire how you are so generous with your time, including with so many undergrads, and you will always be my role model in hard work and persistence. Gregory, your ability to ask the hard questions inspires me. However, it's not only your bold and fearless leadership that is inspirational, but also (and maybe even more so) your ability to love the people around you so much. Thank you both for putting me on this path.

Next, I want to acknowledge my co-advisors, James Fogarty and Shwetak Patel, and my unofficial third co-advisor, Sean Munson. James, your attention to detail (all while grumbling about life) has helped me attain levels of rigor in research I didn't know I was capable of, and I will always be so thankful for that. Shwetak, you were always an absolutely endless source of creativity and positivity, and you provided so much safe space and freedom to explore. I will never take that for granted. Sean, your feedback was always thoughtful yet practical, your ability to be efficient with time will always remind me that work-life balance is possible. I hope I get there one day.

I also want to give a quick thanks to Jake Wobbrock for not only being on my committee, but also creating many publicly available resources (like "Catchy Titles Are Good: But Avoid Being Cute") that saved me from writer's block so many times.

My dissertation research sits at the intersection of computing and health, and none of this work would have been possible if not for my UW Med collaborators: Cindy Lin and George Ioannou. Cindy, your fearless leadership and ability to get things done will always inspire me. George, your enthusiasm for trying things in the name of science will always remind me to have fun.

I have been very fortunate to receive so much incredible mentorship, and I hope I was able to pay it forward some in mentoring some wonderful undergrad research assistants. Annalice Ni and Eddy Zhou, it felt like we navigated being new to UW and also survived the pandemic together. I hope you're doing well out there. Millicent Li, everything will be fine; also, 🐈. Toma Itagaki, I'm proud of you for following your heart. Nancy Liu, even though you weren't my mentee, you were a big part of my PhD, so I'm sticking you here. Jason Miller and Jonathan Shu, I hope you both had fun doing all kinds of research. I'm so proud of all of you.

The best part of having two advisors is having two sets of labmates and friends. I'll start with the Fogies. Ravi Karkar, you were the person I needed most as a first year PhD student. Underneath the wit and snark, you're down to earth, calm and collected, and you get a big share of James's credit in helping me grow as a researcher. Annie Ross, Jessie Schroeder, Jina Suh, and Alex Okeson, you set the lab culture and showed us how to deal with James. Jesse Martinez and Mingyuan Zhong, we're somehow the only ones left from our generation? I think we turned out alright. Yasaman Sefidgar and Anant Mittal, I'm glad you joined the Fogies and provided wisdom as post-COVID senior PhD students. Shaan Chopra, Tae Jones, and Aaleyah Lewis, I'm so excited for the new energy you all brought to the group.

The Ubicomp lab is a special place to be because it's a large space with access to a dedicated fabrication space and it's always loud from being jampacked with people talking. Even though we did not overlap, I'm grateful to Gabe Cohn, Sidhant Gupta, and Mayank Goel for reaching out and being supportive from where ever they were. Then there were the senior students that showed us the ropes: Alex Mariakakis, Edward Wang, Eric Whitmire, Farshid Parizi, CJ Park, Xin Liu, Manuja Sharma, and Matt Whitehill. And then there's us, the cursed "unprecedented" generation: Joe Breda, Alvin Cao, Ishan Chatterjee, Jason Hoffman, Shirley Xue, Girish Narayanswamy, Jacob Peplinski, and Anand Waghmare. Despite the pandemic and the recession, I'm so glad I got to do this with you all. The next generation helped a lot with rebuilding the lab culture after the pandemic: Jerry Cao, Zachary Englhardt, and Zhihan Zhang. I'm excited to see the great things you all will accomplish. Vidya Srinivas, Poojita Garg, Jiexin Ding, you too. I also wanted to acknowledge Parker Ruth and Libby Lavitt for being instrumental in getting the ExerciseRx project off the ground.

I'm also lucky to have the most incredibly reliable friends to not talk research with. Akshay Chandrasekhar, we've come a long way from Southside, and I'm glad we still stay in touch. Jerry Lin and Ravi Konjeti, you guys are always there when I need it, and we need to plan our next trip soon! Raj Ammanabrolu: 🐕 🐕 🐕. Thanks to all of you for sticking around.

Thanks to my family for always being there, even after I moved away from you all for the first time: Dad (*"Die"*), Mom (*"Abba"*), and Sarah (*"Gui"*).

Finally, thanks to my incredible partner MeiXing Dong for everything. I've cherished having your company all this time. Thanks for being so patient with my crazy work hours and all the ups and downs of PhD student life. You've helped me grow while letting me figure it out in my own way, and I really appreciate that. I'm grateful that you were a huge part of this PhD journey with me, and I cannot wait to see what adventures we go on next. <3 !!!

Chapter 1

Introduction

A century ago, most doctor visits in the United States were house calls; doctors would visit patients in their homes and treat them there. The main concern of the time was acute diseases, such as tuberculosis or polio. However, through public health interventions such as vaccines and sanitation, deaths from infections dramatically reduced. Doctor attention then shifted to chronic diseases, such as diseases or neurodegenerative conditions. Through this transition, doctors also became more specialized and began using more specialized equipment. The rise of specialization began to reverse the relationship between doctor and patients; patients increasingly began to visit doctors in specialized clinics and hospitals, which housed specialized equipment. By the 2000s, this model had become the norm. However, this transition also puts increased pressure on the healthcare system. Specifically, chronic conditions require continuous monitoring and long-term management, which are poorly supported by hospital-based healthcare models [37, 4]. This specific flaw was particularly exacerbated by the COVID-19 pandemic, when patient need for sustained treatment overwhelmed hospitals [42]. Projections show that these conditions will only continue to worsen; chronic conditions are increasing, and an aging population will face multi-morbidities [124].

Many research and commercial efforts have developed personal health technologies in support of patients with chronic health conditions. However, despite the plethora of data generated by these technologies, patients still struggle to derive value from these technologies as shown by their abandonment of these technologies [38, 39]. Patients have cited that they find it challenging to align the data produced by a given piece of health technology and their own health goals [24]. Patients have also described that they struggle to collaborate with their healthcare providers around these technologies [25]. There is thus an opportunity to design personal health technologies that are precise to a condition and can support multiple stakeholder relationships.

In my research, I aim to support individuals with chronic conditions through personal health technologies that deliver new models of care for patients and their providers and caregivers. Towards this end, I discuss the development of three systems that demonstrate this goal: (1) Beacon, a system to support chronic liver disease patients in self-monitoring their cognitive function; (2) ExerciseRx, a system to provide patients with chronic conditions with feedback on their movement while exercising; and (3) PARX, a system to enable healthcare providers in generating and tailoring exercise instructions to individual patient needs. Each of these systems is deployed in longitudinal use by the target patient population to understand their lived experience in integrating such technologies into their everyday life.

1.1 Thesis Statement

My dissertation research demonstrates the following thesis statement:

Designing and deploying novel personal health technologies for the challenges of at-home longitudinal use can enable and inform new data and models of care for patients, caregivers, and providers.

1.2 Thesis Overview

My dissertation mainly makes contributions in the form of three artifacts, while secondarily contributing design implications based on empirical findings.

In Chapter 2, I review the related work in health and human-computer interaction and ubiquitous computing and then present the medical context for the projects in my dissertation. I first discuss the space of health research in human-computer interaction and ubiquitous computing and how I position the contributions of my work relative to this space. I next describe related work specific to the challenges of my dissertation projects, including prior work deployed in home settings and prior work that has studied the role of multiple stakeholders in patient care. I then provide the medical context for Beacon by introducing the chronic liver disease patient population: discussing what makes their condition difficult to diagnose and support, and describing the current standard of diagnosis and care. I finally provide the medical context for ExerciseRx by reviewing the literature on physical activity, particularly as it pertains to two patient populations: individuals with cerebral palsy and individuals with bladder cancer.

In Chapter 3, I introduce Beacon, a system that enables chronic liver disease patients to self-administer criticial flicker frequency (CFF) tests to monitor their cognitive function. As part of my dissertation's demonstration of my thesis statement, in this chapter I show:

1. Technical contributions required in iterating 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. Validation of the CFF measurements produced by Beacon in a clinical setting with chronic liver disease patients.

3. Evidence of the feasibility of patients in taking measurements longitudinally in home

settings.

4. Qualitative patient experiences in using Beacon as part of daily life, and findings from interviews with patients and hepatologists on the overall efficacy of using Beacon in clinical and at-home practice.

In Chapter 4, I introduce ExerciseRx, a system that provides patients with feedback on their movements while exercising and enables patients to track their progress in completing exercise routines. As part of my dissertation's demonstration of my thesis statement, in this chapter I show:

1. Formative work in needs-finding based in interviews with patients from the cerebral palsy and bladder cancer populations to understand the role of exercise in their lives.
2. Technical contributions required in iterating on ExerciseRx's sensing, machine learning, and app front-end, motivated by the need to enable at-home patient deployments and scale across multiple patient populations.
3. Evidence of the feasibility of patients in sustained use of the ExerciseRx app to facilitate completing exercise sessions.
4. Qualitative patient experiences in using ExerciseRx as part of daily life.

In Chapter 5, I introduce PARX, a system that supports healthcare providers in generating and tailoring an exercise plan to individual patients using natural language interaction. This work is motivated by my experience in ExerciseRx and identifying a significant gap in the relationship between patient and provider; this project fills this gap, making scaling ExerciseRx up more viable. As part of my dissertation's demonstration of my thesis statement, in this chapter I show:

1. Insights from collecting a dataset of fine-grained text instructions paired with photographic exercise instructions.
2. Technical contributions in designing an agentic workflow for enabling generating and editing human motion animations using natural language queries.
3. Findings from a technology probe with healthcare providers interested in prescribing exercise plans for their patients and feedback from their experience using the system.

In Chapter 6, I reflect on my work conducting translational, multidisciplinary research, discussing themes that span all three projects. These themes include: (1) the circular (i.e., "chicken and egg") problem in investigating novel data sources, and motivating patients in collecting data of unknown value; (2) the contrast between developing systems general-to-specific versus specific-to-general; (3) the nature of collaborations between engineering and medicine, and the different forms of research contribution between these fields; and (4) challenges in translation, particularly in health technologies.

Chapter 2

Background

2.1 Health in HCI and UbiComp

My work draws on and contributes to the literature in using technology to move clinical and health measures into the home, including the benefits and design challenges. Here I first provide background information on health research in the broader HCI and UbiComp literature. I then review the related work in moving health technologies into home settings and the implications for stakeholders in this process.

2.1.1 Background

Health research in HCI and UbiComp includes many styles of work, spanning participatory design through building technical systems. Work in the latter category includes designing, building, and evaluating systems for advancing healthcare. Such systems have allowed researchers to investigate areas such as new sensing and measurement techniques [11, 2, 122], actuation [134, 89, 31], and health tracking [59, 60, 65]. The work I will present in this dissertation falls under sensing and health tracking techniques. Other work in these areas include systems that have been built to explore tracking vital signs [40, 105, 48, 131, 92, 47], monitoring chronic conditions [69, 32, 68, 101, 59], and screening for acute conditions [94, 127, 123, 28, 85]. This dissertation presents contributions specifically in deploying systems that leverage sensing into patient homes in support of new models of care for patients and their providers and caregivers.

2.1.2 Personal Health Technologies in Home Settings

As healthcare technology evolves, the ability to monitor health conditions outside of traditional clinical environments has become increasingly feasible through the availability of smartphones [67] and wearables [113]. In many cases, this involves the adaptation of well-established measures, such as blood pressure [62, 129], body weight [61], core temperature [128, 16], and spirometry [71, 43], to be taken outside of the traditional clinical environment. This process can afford several benefits, including mitigating reliability confounds and increased measurement resolution in timing.

Taking measures outside of the clinical environment can mitigate potential influences, such as anxiety, that affect reliability of health measures 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 measures taken in a clinical setting [129, 22, 84, 114, 45].

Adapting measures to be taken outside of the clinical setting can have another benefit in the potential for increased time resolution of measures, 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 [87, 56, 99].

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 [126]. Cordeiro et al. similarly found that the effort required for food journaling was a barrier to consistent logging [26]. 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 [76]. Kay et al. studied perception of the bathroom weight scale and enumerated design recommendations for measure presentations that help people understand the daily fluctuations and uncertainty around weight measure [61]. Kendall et al. also studied people's reactions to frequent blood pressure measures [62]. The work in this dissertation contributes additional evidence in the functional and perceptual value of longitudinal usage of health technologies in home settings.

2.1.3 Multi-Stakeholder Health Research

Management of a patient's health condition involves multiple stakeholders. 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 [14], parents [98, 23], adult children [50], 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 [13], and Bascom et al. similarly presented techniques for reducing the implicit biases of healthcare providers when encountering patients [8]. Seo et al. examined techniques specifically for supporting communication between child patients and their providers [108].

Other systems have been designed to instead facilitate interaction between patient and caregiver. Mynatt et al. designed digital family portraits [91] to remotely give caregivers awareness of their parents. Khan et al. described the development of a personal health application to help patients and their caregivers manage medications [63]. Hong et al. deployed diary probes to help family members of adolescents with chronic conditions understand their experiences [49].

In this dissertation, I discuss the role of stakeholders in each project. Clinician engagement in designing the systems naturally leads to surfacing the role of a clinician in using these systems. The deployment aspect of these projects has also helped me develop an understanding of patient lived experiences in using these systems in daily life, including how they may or may not discuss using the system with other people in their life.

2.2 Medical Background

Here I provide the medical background and context for the work presented in this dissertation.

2.2.1 Background on Chronic Liver Disease

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 [90]. Over 80% of cirrhotic patients develop neurocognitive impairments known as hepatic encephalopathy (HE) [119]. Impairments can fluctuate over time, ranging from minimal hepatic encephalopathy (MHE) to overt hepatic encephalopathy (OHE), affecting nearly all aspects of life. OHE is end-stage, often leading to coma or death, while MHE, results in slight confusion or disruptions to fine motor skills. More importantly, if detected promptly, MHE can be effectively controlled with medication and thus prevent the condition from progressing to OHE.

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 [102, 117], and commercially-available systems such as the Lafayette Flicker Fusion System have been used in studies for clinical validation of this measure [130]. Although initially proposed as a clinical screening test, Karkar et al. contributed a reframing of the CFF measure from a clinical screening test to a self-administered self-tracking measure using an initial Beacon prototype introduced in [58].

My work extends Karkar et al.'s reframing by validating Beacon in a clinical setting, showing feasibility of patient longitudinally self-monitoring, demonstrating stability in longitudinal measurements, and confirming with patients and providers the potential utility of Beacon in real-world settings.

2.2.2 Background on Exercise for Chronic Conditions

Physical activity is not only important for a healthy lifestyle and general wellness, it can also lower risk factors and improve management of many chronic conditions [77, 79]. In this section, I will provide background on the utility of exercise for two specific chronic conditions: cerebral palsy and bladder cancer.

Exercise in Cerebral Palsy

Cerebral palsy (CP) is a neurological disorder that appears in infancy or early childhood as a result of damage or abnormalities in the brain that permanently affects body movement and muscle coordination. The motor skills of individuals with CP are characterized using the Gross Motor Function Classification System (GMFCS), which categorizes people into five different levels [96]. CP is a non-reversible condition, and in general, children will not improve their GMFCS level after the age of 5 years old [19]. Although prior work has studied the benefits of physical activity in helping manage the implications of CP [118, 82], physical therapy alone does not increase physical activity in the long term. Further work on motivation and behavioral interventions is needed for sustained physical activity and lasting impact [15]. People with CP face a range of significant barriers to getting the physical activity they need. In addition to the significant financial cost of physical therapy (PT) sessions, particularly with specialists that tailor their sessions for people with CP [64], access to PT is generally limited due to factors such as transportation and insurance coverage [33]. Even for individuals who are able to receive PT, they struggle to stay motivated to practice their prescribed exercises at home [116, 97]. While PT solutions have been developed for at-home use by providing video libraries and text instructions, they are generally not applicable or adaptable to this patient population.

Exercise in Bladder Cancer Surgery Patients

Bladder cancer patients that require surgery for removal of the entire bladder (i.e., radical cystectomy) will first undergo a round of chemotherapy to shrink the area of cancer coverage. These typically older patients experience significant morbidity due to the frailty and deconditioning caused by chemotherapy [81]. Physical fitness has been determined to be a primary metric of care for this population [70, 88]. Prehabilitation [17] and rehabilitation [100] programs have shown great potential in improving physical fitness and thus accelerating recovery after surgery. Furthermore, digital health technologies that deliver prehabilitation and rehabilitation interventions, primarily based on data produced by wearable devices, have demonstrated success in causing behavior changes that lead to increased physical activity [53, 12]. Many of these interventions, however, are based on general exercise and physical activity metrics and not tailored to the specific needs of bladder cancer patients.

Chapter 3

Beacon: At-Home, Patient Self-Monitoring of Cognitive Function

Chronic liver disease can lead to neurological conditions, called hepatic encephalopathy (HE) that result in coma or death. Although early detection can allow for effective intervention, screening for these conditions is infrequent and unstandardized. Prior work has shown that the critical flicker frequency (CFF) is sensitive to even the earliest stages of HE (i.e., minimal hepatic encephalopathy) [102, 117], and identified CFF as an excellent candidate for timely screening of HE [58]. I led work on the Beacon project to realize the vision of supporting individuals with chronic liver disease in self-monitoring their cognitive conditions in home settings.

In this chapter, I discuss the technical challenges in iterating on Beacon's hardware and software to (1) be appropriate for use by members of the target population, and (2) be appropriate for use in home settings. I then show how these iterations enabled our team to (1) validate the CFF measurements produced by Beacon in a clinical setting with chronic liver disease patients, and (2) demonstrate the feasibility of patients in taking at-home longitudinal measurements. I then report on patient experiences in using Beacon as part of daily life, and findings from interviews with patients and hepatologists on how Beacon might integrated into clinical and at-home practice. I finally describe ongoing work in improving the experience of taking measurements using Beacon (i.e., via accelerating the process and reducing the mental burden required by the process), and expanding the scope of Beacon beyond the chronic liver disease population (e.g., towards detecting early onset of Alzheimer's disease).

As part of my dissertation's demonstration of my thesis, in this chapter:

1. Technical contributions required in iterating 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. Validation of the CFF measurements produced by Beacon in a clinical setting with chronic liver disease patients.
3. Evidence of the feasibility of patients in taking measurements longitudinally in home settings.
4. Qualitative patient experiences in using Beacon as part of daily life, and findings from interviews with patients and hepatologists on the overall efficacy of using Beacon in clinical and at-home practice.

This research was done in collaboration with research coordinators Isabella Bueno, Michael Yacoub, Sabrina Omer, and James Kashima; undergraduate research assistant Jonathan Shu; Human Centered Design & Engineering professor Sean Munson; Computer Science & Engineering professors Ravi Karkar and James Fogarty; and hepatologists Philip Vutien and George Ioannou. This line of research has led to a poster at AASLD Liver Meeting 2020 [120], a workshop paper at CHI WISH 2023 [74], a 2023 journal publication in the American Journal of Gastroenterology [121], and a full paper at CHI 2025 [75]. This work has also been presented at the hepatology grand rounds at the Veterans Affairs Puget Sound Medical Center and the Computer Science & Engineering Industry Affiliates Day. The commercialization potential of this work has also been shown through our participation in the I-Corps Program and Innovation Gap Fund.

My roles in this project include: (1) leading the design, development, and iteration of the hardware/software platform used in [120, 121, 74, 75]; (2) contributing to the data analysis in [121], (3) leading the design and administration of the studies in [75]; (4) leading the data analysis in [75]; (5) seeking and pursuing opportunities for translation and adoption of the Beacon project in real world clinical and home settings.

This work was supported in part by the National Institutes of Health under awards R21DK117431 and R01LM012810, by the University of Washington Population Health Initiative, and by Google.

3.1 Vision & Design Goals

Although the guidelines for diagnosing overt hepatic encephalopathy (OHE) are well-defined, treatment of OHE is difficult. Conversely, treatment of minimal hepatic encephalopathy (MHE) can be very effective, but screening is inadequate, even "uncommon" [6]. 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 [5]. Our vision is thus 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

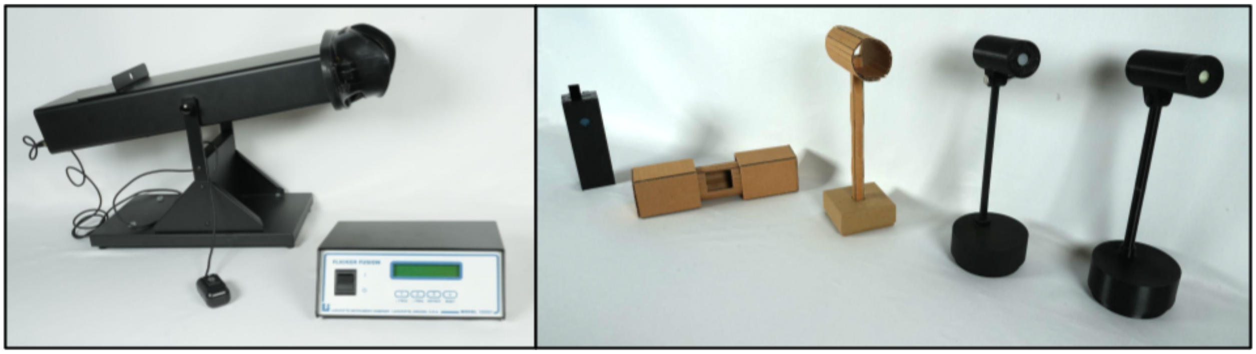


Figure 3.1: **Left:** The Lafayette Flicker Fusion device, a commercially-available device for administering CFF tests. **Right:** Explorations and iterations in Beacon's form, from the early prototype developed by Karkar et al. [58] (left-most), through early low-fidelity explorations of potential forms, to development and refinement of Beacon's current form (right-most).

choices to control MHE, through decisions to promote safety (e.g., not driving), and through more prompt treatment.

This vision helped define our design goals: we sought to create a portable, inexpensive, and self-administrable device for measuring the CFF threshold. We selected the Lafayette FFS, a medical gold standard device, as the reference device for us to model and compare performance against while developing Beacon (shown in the left panel of Figure 3.1). Karkar et al. implemented the first four iterations of the Beacon device (shown in the right panel of Figure 3.1), and I led the design and implementation of the final version. In particular, significant effort was put towards improving the ability to produce many copies of the device and deploy these devices at scale.

3.2 System

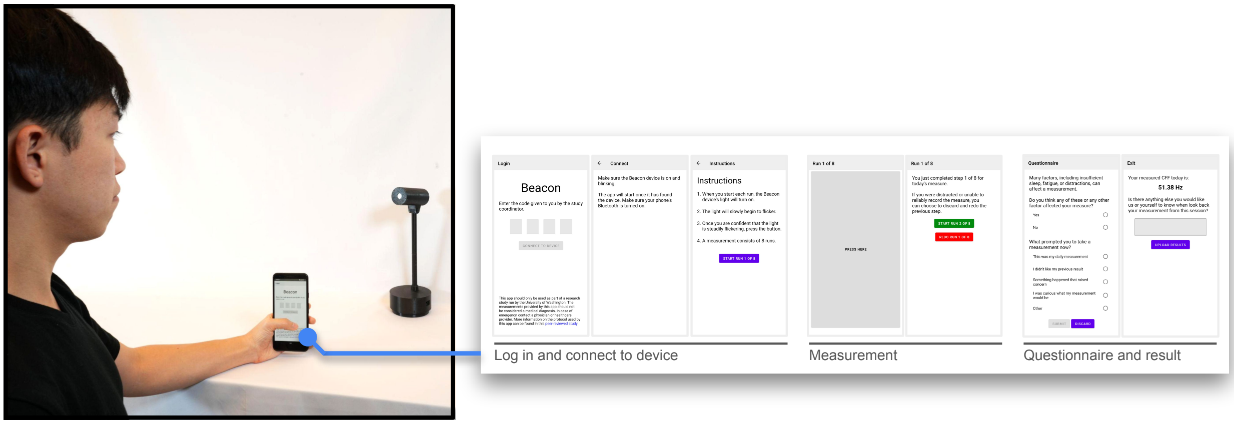
The Beacon system consists of three components: the physical Beacon device, an accompanying cross-platform smartphone app, and a cloud server. The physical Beacon device is for precisely rendering the flickering light stimulus. The smartphone app controls the Beacon device over Bluetooth Low Energy (BLE) and provides a interface for patients to interact with the device. The smartphone app retrieves protocol specifications and saves measurement data to the cloud server over WiFi.

3.2.1 Beacon Device

The physical Beacon device adopts the appearance of a desktop lamp. Most of the components are 3D printed, with the exception of a metal pipe providing structural integrity in the neck section, a metal screw for the hinge, and a piece of translucent acrylic acting as the optical diffuser. Internally, the device's electronics comprise of an ATmega 32u4 microcontroller with an nRF52 co-processor for BLE, a $10,000\ mAh$ lithium-polymer (LiPo) battery with power switch, and an LED placed behind the optical diffuser. The device's behavior is dictated by a simple state machine: if no phone is connected, render an extremely slow flicker at $1\ Hz$ to remind patients to use the device or turn it off. If a phone is connected, turn the light off until a command is received dictating what frequency to flicker at. All flickering patterns are driven using timer interrupts for high precision and are generated with 50% duty cycle square waves.

3.2.2 Beacon App

The current exploratory data collection Beacon app is implemented using React Native to allow for cross-platform use. The app provides three essential features: (1) logging in, which pulls the measurement protocol details and target device from the server and connects to that device; (2) guiding the patient through the measurement process; and (3) giving the patient a chance to reflect on the process before showing them their measurement. After completing all three steps of the app, patients are required to entirely close the app to intentionally and clearly mark the end of the session. We designed the app to be simple, maintaining a linear flow to mitigate issues with patients getting lost while using the app.



Enter the code given to you by the study coordinator.

This app should only be used as part of a research study run by the University of Washington. The measurements provided by this app should not be considered a medical diagnosis. In case of emergency, contact a physician or healthcare provider. More information on the protocol used by this app can be found in this peer-reviewed study.

Make sure the Beacon device is on and blinking.

The app will start once it has found the device. Make sure your phone's Bluetooth is turned on.

Instructions

1. When you start each run, the Beacon device's light will turn on.
2. The light will slowly begin to flicker.
3. Once you are confident that the light is steadily flickering, press the button.
4. A measurement consists of 8 runs.

You just completed step 1 of 8 for today's measure.

If you were distracted or unable to reliably record the measure, you can choose to discard and redo the previous step.

Many factors, including insufficient sleep, fatigue, or distractions, can affect a measurement.

Do you think any of these or any other factor affected your measure?

- Yes ○
- No ○

What prompted you to take a measurement now?

- This was my daily measurement ○
- I didn't like my previous result ○
- Something happened that raised concern ○
- I was curious what my measurement would be ○
- Other ○

Your measured CFF today is:

51.38 Hz

Is there anything else you would like us or yourself to know when look back your measurement from this session?

Figure 3.2: 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.

3.2.3 Cloud Server

The cloud server infrastructure is implemented using PHP and vanilla Javascript. It is used to provide the app with session details, including which measurement protocol and which Beacon device to use, as well as to facilitate storage of measurements as they are completed. We also built a dashboard for the server to enable research coordinators to configure protocol details as well as remotely monitor measurements in real time.

3.2.4 Measurement Protocols

The literature describes a number of ways to measure CFF, including the method of limits (MOL) and 2-alternative forced choice (FC) [36]. We describe the procedures for these protocols here as they are the basis for the rest of this chapter.

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).

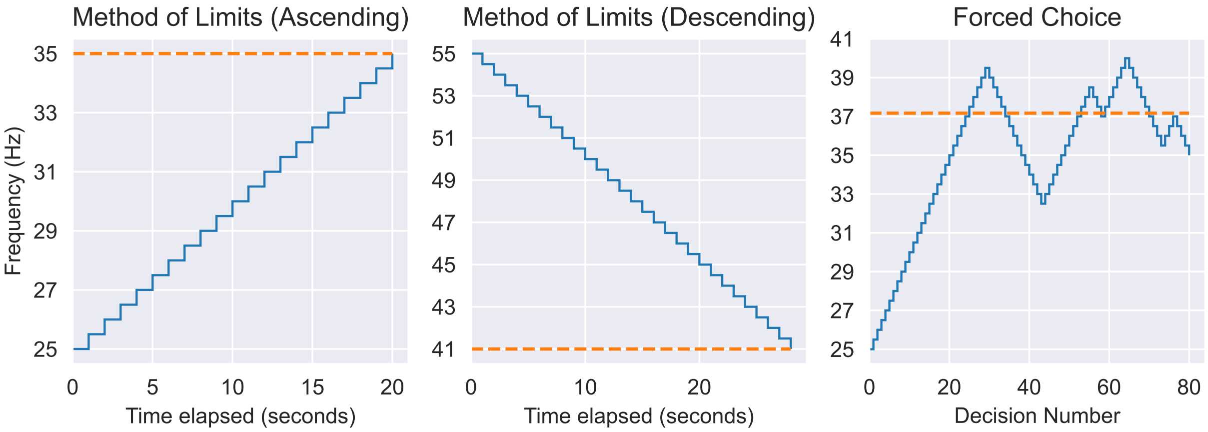


Figure 3.3: Example trials for three different measurement protocols for obtaining a CFF measure.

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.

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 3.3. 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 3.3 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 3.3, 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. [36], 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 [3, 66, 103, 110, 117]. 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.

3.3 Clinical Validation

The primary goal for this study was to obtain medical evidence for Beacon by validating it with a chronic liver disease population. However, we also wanted to begin to gather indicators that Beacon could be used in at-home settings. We quickly learned that as part of their visit to the clinic, patients have very limited time and capacity (both mental and physical) to engage with our study. As a result, our study design process required us to optimize for both duration – by enforcing a strict time limit – and make considerations for physical and mental load. Nonetheless, through our evolving study protocols, we were able to accumulate encouraging evidence supporting the potential at-home use of Beacon. In Phase 1, we validated Beacon's CFF measurements against a gold standard device, the Lafayette Flicker Fusion System. In Phase 2, we showed the efficacy of a

Figure 3.4: The study room, in the hepatology clinic at the UW Medical Center, in which we conducted our clinical validation studies.

new CFF measurement protocol, the forced choice algorithm, to mitigate response bias. In Phase 3, we demonstrated that test-retest measurements over time are reliable. In addition to answering these research questions, we wanted to gauge Beacon's usability with the patient population. We thus report on usability findings from patients across all phases of the study.

3.3.1 Participants

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. We excluded patients with a history of neurocognitive disorders, seizure disorders or epilepsy, severe migraines or photosensitivity, or ophthalmologic diseases. Patients were compensated with a $50 gift card, along with the cost of parking. For each of the three phases of the study, we recruited 37 participants, 35 participants, and 81 participants, respectively. In total, 153 participants were involved in the clinical validation study (mean age=53.4 years, standard deviation=13.8).

3.3.2 Apparatus

The study was conducted in the same building as the hepatology clinic. Consented patients were brought to the study room (see Figure 3.4), where the patient and research coordinator sat next to each other. Due to COVID-19 policies enacted during Phase 2, a plexiglass divider was installed between the patient and research coordinator (see Figure 3.4). The room was prepared with a Beacon device, smartphone preloaded with the Beacon app, and a Lafayette Flicker-Fusion System.

3.3.3 Procedure

Phase 1

We wanted to first establish a baseline relationship between all of our proposed tasks. We therefore asked patients to complete the MOL-A and MOL-D tasks using both the Beacon and Lafayette devices. The order of the tasks to be completed was randomized by the research coordinator.

Phase 2

Over the course of Phase 1, we became concerned about potential issues with response bias affecting the CFF measurements. These biases could be both unintentional (e.g., lapse in focus) or intentional (e.g., impatience and wanting to finish the study quickly). Such issues could escalate if measurements were taken at home, without being overseen by a research coordinator. As a result, we wanted to investigate the efficacy of employing a forced alternative choice (FC) protocol, in which patients decide which stimulus out of two options is flickering. To ensure that the study remained under one hour, we decided to remove both the Lafayette condition as well as the MOL-A condition. Phase 1 already provided encouraging evidence that Beacon's measurements correlated well with Lafayette's measurements. We also follow precedent in the literature that defaults to using only MOL-D when the researchers choose not to employ a full a MOL protocol [109, 57, 7]. The study procedure for phase 2 in total thus included the MOL-D and FC tasks using Beacon.

Phase 3

We finally sought to investigate the consistency of Beacon's CFF measurements across repeated tests. This finding would be critical for informing the potential design of a longitudinal study. In particular, we needed to demonstrate that Beacon's CFF measures were consistent in the short-term to help us feel confident in determining fluctuations over the long-term to be meaningful. This

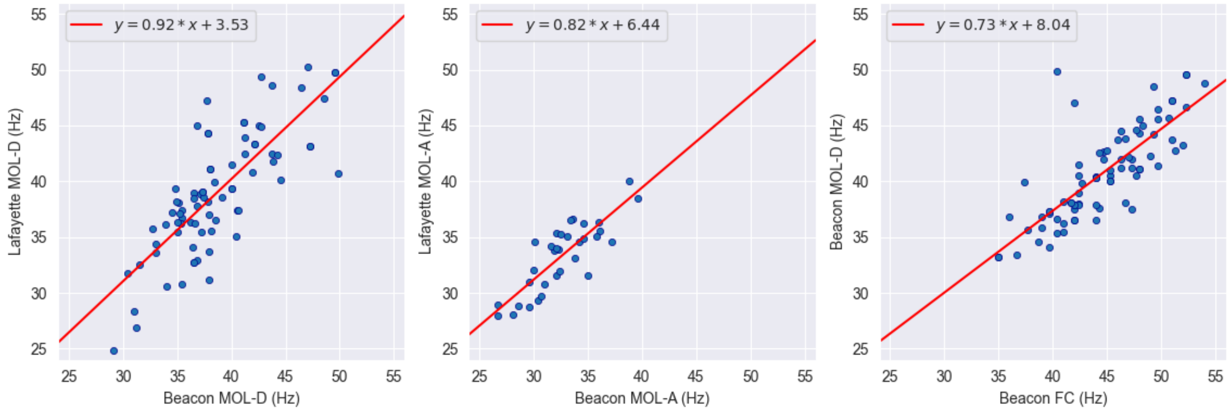


Figure 3.5: Regression analyses to compare the CFF measurements produced by the Lafayette Flicker Fusion device and those produced by the Beacon device.

phase is also particularly important since Beacon does not have a viewing chamber like the Lafayette. The viewing chamber is designed to increase consistency across measurements by reducing external visual noise and supporting alignment of the eyes to the light source. Patients in this phase of the study completed each of the following twice: (1) the MOL-D test with Beacon, (2) the MOL-D test with Lafayette, and (3) the FC test with Beacon. The order of the tasks was randomized, but with the limitation that the same task would not be performed immediately after each other. All 6 tasks were completed in the same one-hour study session.

3.3.4 Results

Phase 1

To compare the CFF measurements produced by the Beacon and Lafayette devices, we calculate the intraclass correlation between measurements taken with each protocol from each device. With the MOL-A protocol, CFF measurements from Beacon (M: 32.55 Hz, SD: 3.12) and Lafayette (M: 33.29 Hz, SD: 3.09) correlated with a coefficient $r = 0.84$. With the MOL-D protocol, CFF measurements from Beacon (M: 39.20 Hz, SD: 4.64) and Lafayette (M: 39.30 Hz, SD: 5.43)

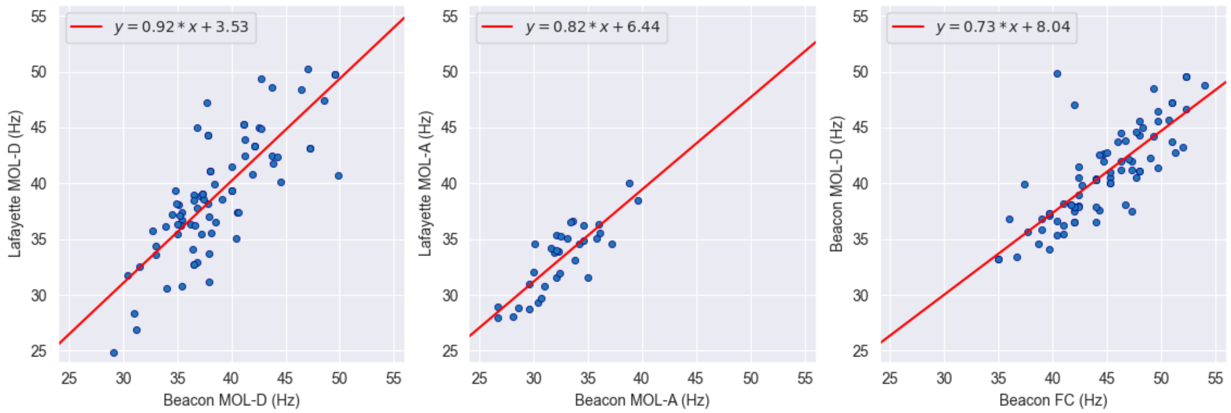


Figure 3.6: Regression analyses to compare different CFF measurement protocols.

correlated with a coefficient $r = 0.78$. In general, we found measurements from Beacon to correlate well with those obtained using Lafayette. This initial phase of the study helped us develop confidence in the Beacon system, especially in how well its measurements corresponded to those obtained with the Lafayette.

Phase 2

Because the Lafayette device does not support the FC protocol, we validated Beacon FC measurements by comparing them to Beacon MOL-D measurements, which were shown in Phase 1 to correlate well with Lafayette MOL-D measurements. To do so, we calculate the intraclass correlation between the measurements from Beacon FC and Beacon MOL-D. The CFF measurements from the FC (M: $44.60\ Hz$, SD: 4.47) and MOL-D (M: $40.74\ Hz$, SD: 4.15) protocols correlated with a coefficient $r = 0.79$. Due to the nature of the FC protocol allowing users more time to inspect the presented light stimuli, we expected the measurements it produces to be higher on average.

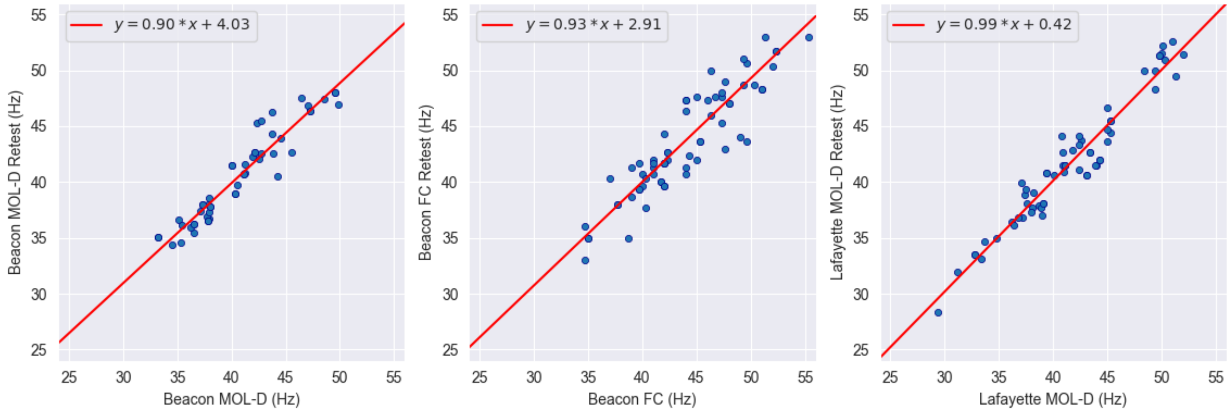


Figure 3.7: Regression analyses to demonstrate the test-retest reliability of CFF measurements.

Phase 3

To evaluate the reliability of CFF measurements taken over time, we calculate the intraclass correlation between measurements obtained from the test and retest trials. Using Beacon, the MOL-D protocol produced a first (M: 40.76 Hz, SD: 4.38) and second (M: 40.58 Hz, SD: 4.12) measurement that correlated with a coefficient $r = 0.95$. Using Beacon, the FC protocol produced a first (M: 44.71 Hz, SD: 5.73) and second (M: 44.38 Hz, SD: 5.76) measurement that correlated with a coefficient $r = 0.92$. Using Lafayette, the MOL-D protocol produced a first (M: 41.62 Hz, SD: 5.40) and second (M: 41.74 Hz, SD: 5.54) measurement that correlated with a coefficient $r = 0.97$. Consistent with Phase 2's results, the CFF measurements produced using the FC protocol were still higher than those produced using the MOL-D protocol on either device. However, all three protocols demonstrated strong consistency across multiple measurements, serving as evidence for the potential efficacy of Beacon in longitudinal use.

Cross-Phase Results

We report on two sets of results that span patients across all three phases of the clinical study.

CFF and Liver History. We compared CFF measurements taken using Beacon to patient liver

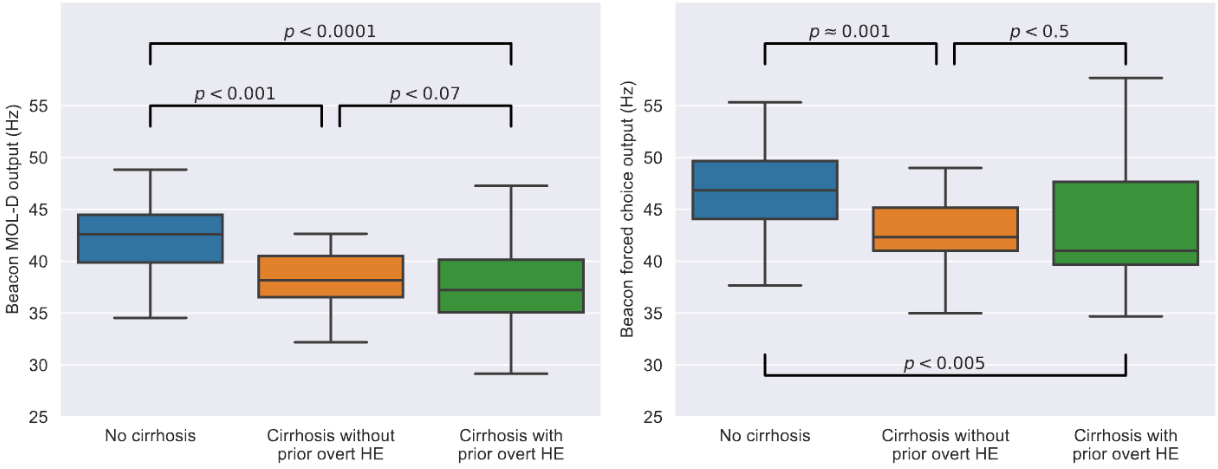


Figure 3.8: Comparison of patients without cirrhosis, with cirrhosis but without prior overt HE, and with cirrhosis and with prior overt HE, based on thresholds derived by MOL-D and FC.

condition history. Using MOL-D, 44 patients without a history of cirrhosis had a mean CFF measurement of $42.0\ Hz$, significantly higher than 53 patients with a history of cirrhosis without prior overt HE (mean: $38.1\ Hz$, $t(43) = 442$, $p < .001$) and 56 patients with a history of cirrhosis with prior overt HE (mean: $37.4\ Hz$, $t(43) = 302, p < .001$). Using FC, patients without a history of cirrhosis had a mean CFF measurement of $46.8\ Hz$, significantly higher than patients with a history of cirrhosis without prior overt HE (mean: $43.0\ Hz$, $t(43) = 402, p \approx .001$) and patients with a history of cirrhosis with prior overt HE (mean: $43.26\ Hz$, $t(43) = 333, p < .005$).

Usability Feedback. In addition to the individual tasks performed in each phase of this study, patients were asked to respond to the prompt "X was easy to use", where X was each task the patient completed, on a scale from 1 to 7. Since the number of patients that used Beacon with MOL-D alone was different from the number of patients that used Beacon with both MOL-D and FC, we present two sets of results. Phase 1 patients ($N = 37$) found Beacon with the MOL-D protocol significantly easier to use than the Lafayette device with the MOL-D protocol ($t(36) = 262$, $p < 0.001$). Phase 2 and Phase 3 patients ($N = 119$) found Beacon with MOL-D and FC protocols

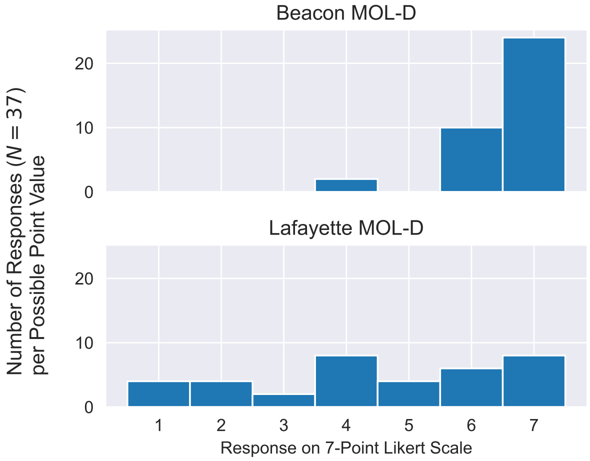


Figure 3.9: Distribution of patient responses to how easy each condition was to use on a 7-point Likert scale. Phase 1 patients ($N = 37$) used Beacon with the MOL-D protocol, finding it significantly easier to use than the Lafayette device with the MOL-D protocol.

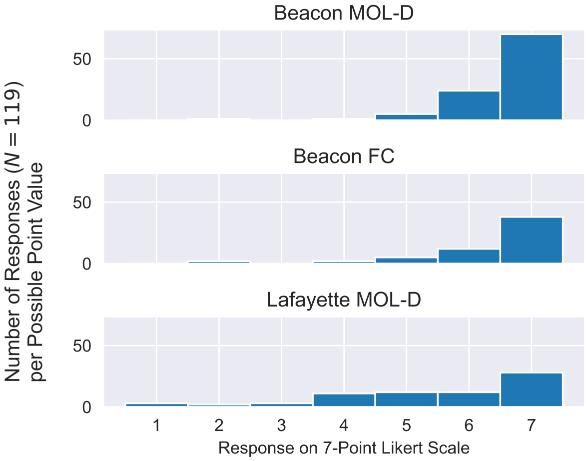


Figure 3.10: Distribution of patient responses to how easy each condition was to use on a 7-point Likert scale. Phase 2 and Phase 3 patients ($N = 119$) used Beacon with MOL-D and FC protocols, finding both significantly easier to use than the Lafayette device with the MOL-D protocol.

significantly easier to use than the Lafayette device with the MOL-D protocol ($t(118) = 5020$, $p < 0.001$, $t(118) = 2770, p < 0.001$).

Although we did not formally interview patients, many alluded to the usability of both devices while completing the study tasks. For example, patients expressed discomfort when using the Lafayette due to the viewing chamber. The most commonly reported issue was the lack of compatibility between the viewing chamber and patient eyeglasses, namely the viewing chamber causing their eyeglasses to develop condensation. Patients also commented on the claustrophobic nature of putting their face in the viewing chamber, and further cited that their eyes struggled to adjust to two light sources. One patient asked to withdraw from the study due to this last reason. In contrast, and consistent with the quantitative feedback, patients did not raise any significant challenges in using Beacon.

3.4 At-Home Deployment

At this stage, we had confirmed that the Beacon device was indeed measuring the CFF phenomenon. We then wanted to see if patients could measure their CFF on their own, in home settings.

3.4.1 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=52.6, standard deviation=11.9). Twelve of these patients had previously participated in the clinical validation 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; we include reflections of these participants on the contrast according to their experiences with at-home measurement. The remaining 9 were patients drawn from a similar population but had never previously encountered Beacon or any other device for CFF measurement.

3.4.2 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, 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

Beacon At-Home Study Instruction Booklet

Thank you for agreeing to participate in the at-home phase of the Beacon study.

At any time during the study, if you have any questions or concerns, please contact [redacted]. In case of emergency, please contact your health provider.

1. Study Kit

Shipment

The study kit should have been shipped directly to you with this instruction booklet inside.

Please keep all packaging to safely return all equipment at the end of the study.

At the end of the study, simply place the included prepaid UPS return label over the previous sticker.

Page 2

Figure 3.11: 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.

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).

3.4.3 Procedure

Depending on proximity of a patient to our institution, we either 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

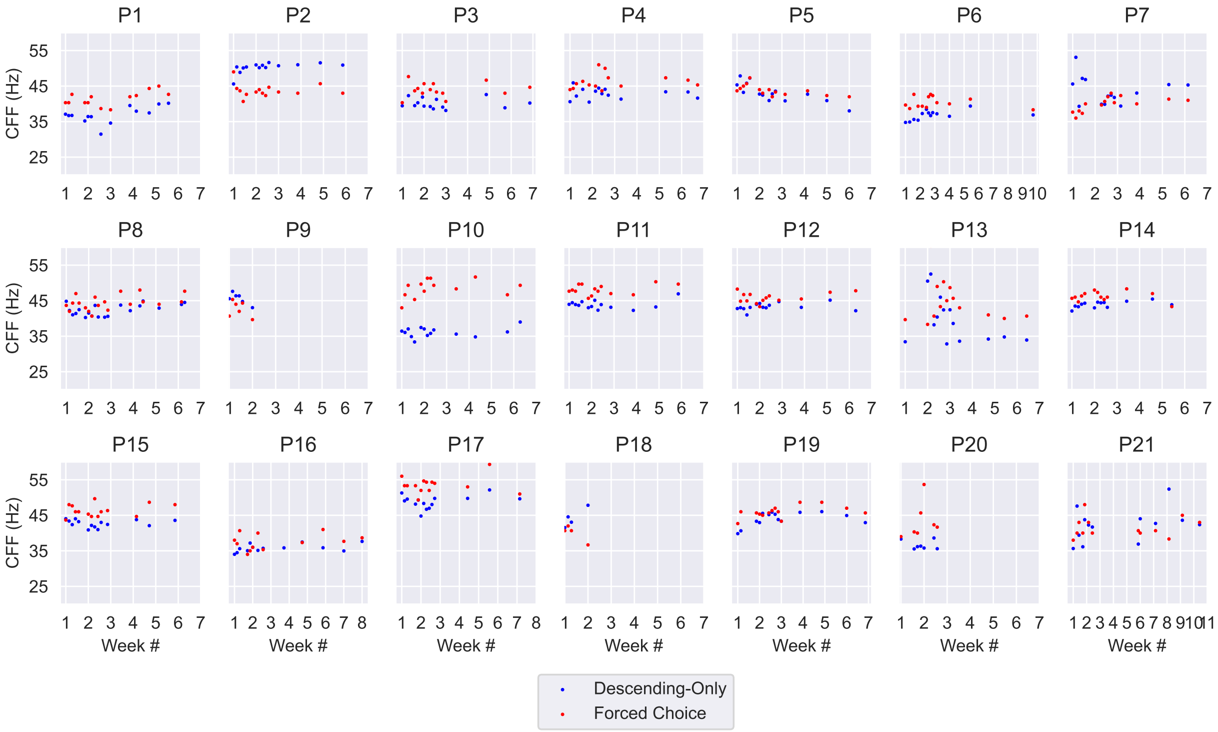


Figure 3.12: A plot of each CFF measure taken by 21 patients using Beacon at-home for 6 weeks with MOL-D and FC protocols.

patient took two CFF measurements, using the MOL-D and FC protocols in a randomized order as prescribed by the calendar.

3.4.4 Results

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, 8 participated

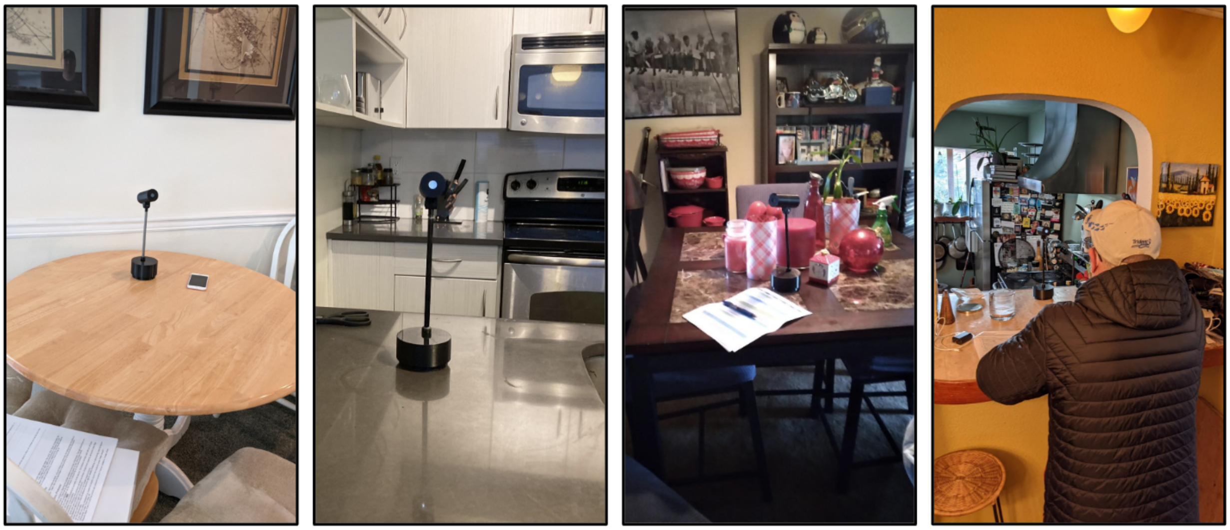


Figure 3.13: Participant-provided photos of environments where they used Beacon to take CFF measurements in their own homes. Participants described keeping Beacon out in their home throughout the study, including because its visibility served as a reminder to take measurements. Participant responses thus validate our design of Beacon to be small, portable, and attractive.

in exit interviews. Of these, 6 expressed they preferred Beacon over the Lafayette device (shown in Figure 3.5) 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). 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 design principles 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, 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.

Characterization of Longitudinal CFF Measurements

We report on at-home CFF measurements obtained from 21 patients. 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 the previous clinical validation, even while gathered in naturalistic conditions.

3.4.5 Patient and Hepatologists Insights

Patient and Caregiver Motivation to Taking Measurements

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 one 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.

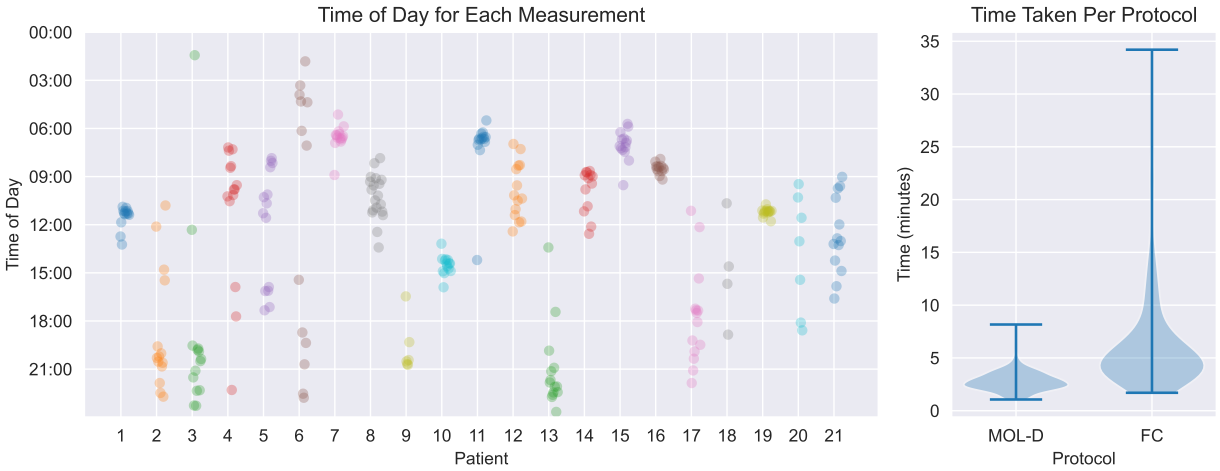


Figure 3.14: Overview of the time of day for each patient measurement and the time required with each protocol.
Left: Time of day patients took measurements. Patients were encouraged to set a time for taking measurements each day, but also demonstrated variation due to schedule disturbances. CFF measures appeared stable even in the presence of 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.

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).

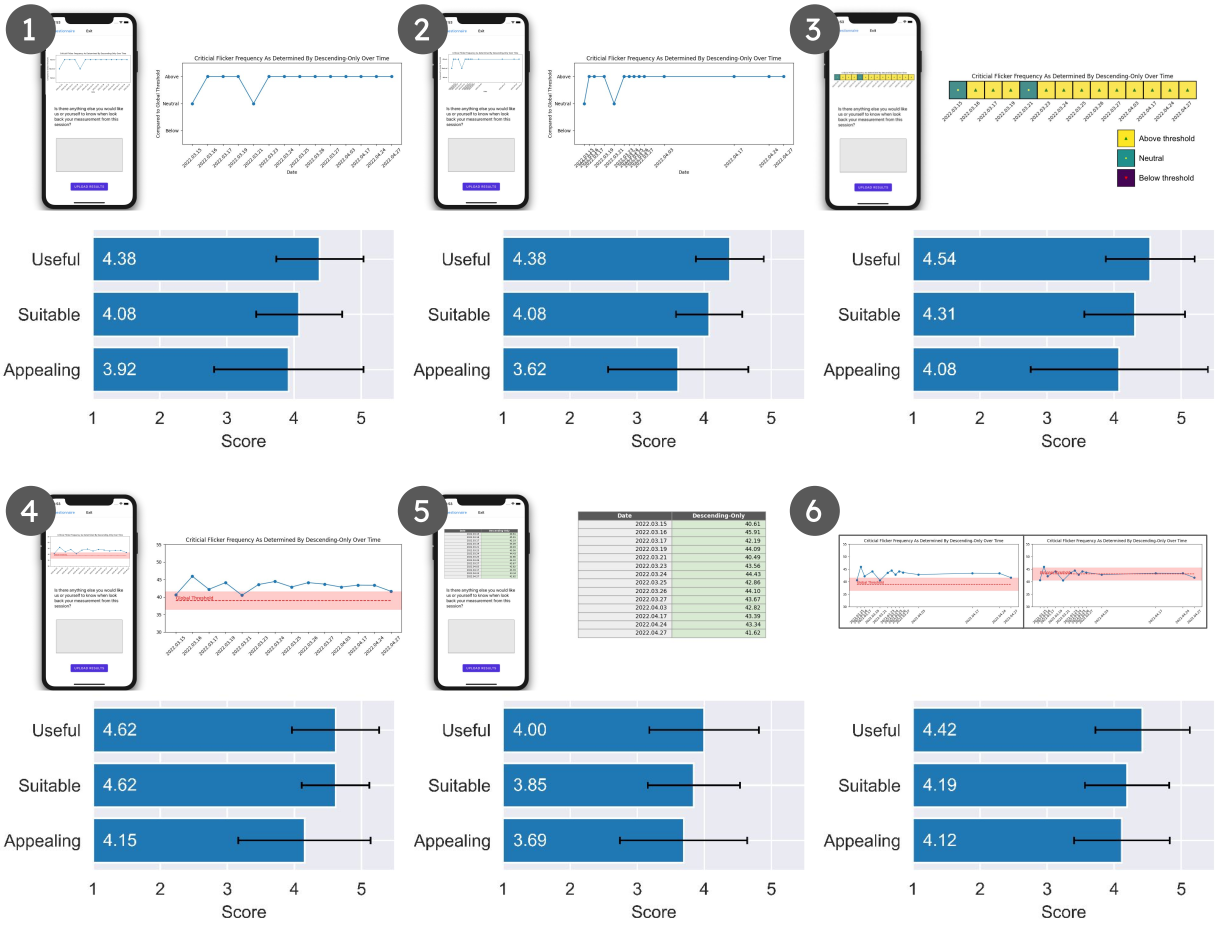
We first investigated the temporal burden of a CFF measurement by examining the time required for each protocol. The right panel of Figure 3.14 shows the distribution of time taken per CFF measurement by protocol. The mean was 2.78 minutes using MOL-D versus 6.37 minutes using FC, but visual inspection 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 3.14 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 experience 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 [18]. SUS scores above 84.1 correspond to a letter grade of A+, indicating extremely high usability [107]. 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 [18]. 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.

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 3.15, 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



Date	Descending-Only
2022.03.15	40.61
2022.03.16	45.91
2022.03.17	42.19
2022.03.19	44.09
2022.03.21	40.49
2022.03.23	43.56
2022.03.24	44.43
2022.03.25	42.86
2022.03.26	44.10
2022.03.27	43.67
2022.04.03	42.82
2022.04.17	43.39
2022.04.24	43.34
2022.04.27	41.62

Figure 3.15: 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.

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 [62]).

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 3.15, 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 [60] 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.

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 by disallowing patients to inflate their measure, 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 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.

3.5 Discussion

3.6 Summary

My research aims to deliver personalized health interventions through longitudinal monitoring. With Beacon, I have examined how a measure of cognitive function, CFF, can be used for longitudinal monitoring towards delivering more timely health interventions. Although CFF has been previously described in a clinical context (i.e., administered by a provider, decisions made on a single point of measure), there was no certainty that previous definitions would translate to an at-home, longitudinal setting. Our efforts in adapting Beacon to home settings therefore involved a level of risk that the measure would be utterly useless on a personal level (i.e., as opposed to on a population level). Indeed, even now that we have demonstrated feasibility and stability in the measures, much work remains to demonstrate sufficient sensitivity to lead to improved outcomes (i.e., the signals fluctuate enough for patients to act upon it).

This project represents 8-years of sustained collaboration between researchers in engineering and medicine. My particular role in this project has led to some unique observations and insights about collaborative, multidisciplinary, and translational research. As a human-computer interaction and ubiquitous computing researcher, we are often motivated by research problems in health because of its importance and potential for impact. A single research paper, however, often does not help it live up to its potential; significant additional effort is required for impact. We must work closely with researchers in medicine to precisely contextualize the work. However, researchers in medicine also have their own values and goals in research; a research contribution in medicine is different from what constitutes a contribution in human-computer interaction. My experience with Beacon has thus involved navigating the publication process in multiple communities, thus far perfectly alternating between an HCI venue and a medical venue; I believe that this level of close engagement is setting us on the right track towards making a positive impact with Beacon.

Chapter 4

ExerciseRx: Promoting At-Home Exercise for Mitigating Chronic Conditions

Exercise is important not only for general well-being, but also for preventing and mitigating the effects of many chronic conditions and musculoskeletal disorders. The current healthcare system focuses on medication prescription as the basis of managing many chronic medical conditions. However, lifestyle factors including exercise can also play a large role in treatment. For people with chronic conditions that lead to declining function, regular exercise can be used for rehabilitation to help maintain their functional abilities. Exercise can also be used for preventive purposes (i.e., “prehabilitation”) as greater physical fitness prior to an operation results in better postoperative outcomes. Despite these proven benefits of exercise, however, it is underutilized in clinical care. I led work on the ExerciseRx project to realize the vision of helping promote exercise in individuals with chronic diseases.

In this chapter, I will present our work on promoting exercise in two different patient populations: teenagers with cerebral palsy and older adults undergoing bladder surgery. First, I will motivate this work based on findings from a series of formative interviews with patients from these populations and healthcare providers that work with these populations. Then, I will discuss the technical details of implementing the ExerciseRx app. Finally, I will show results from deploying the ExerciseRx app in longitudinal, at-home settings with participants from these two patient populations.

I demonstrate this project's relevance to my dissertation's thesis in this chapter through:

1. Formative work in needs-finding based in interviews with patients from the cerebral palsy and bladder cancer populations to understand the role of exercise in their lives.
2. Technical contributions required in iterating on ExerciseRx's sensing, machine learning, and app front-end, motivated by the need to enable at-home patient deployments and scale across multiple patient populations.
3. Evidence of the feasibility of patients in sustained use of the ExerciseRx app to facilitate completing exercise sessions.
4. Qualitative patient experiences in using ExerciseRx as part of daily life.

This research was done in collaboration with research coordinators David Bridges, Leah Cantor, Theresa Kehne; undergraduate research assistant Hana Smahi; software engineers Otari Ioseliani and Kristen Gustafson; designer Aileen Moroney; Human Centered Design & Engineering professor Sean Munson; Computer Science & Engineering professor Shwetak Patel; and sports medicine professor Cindy Lin. Additional collaboration involved Seattle Children's Hospital via research collaborator Deborah Grenard and physical therapist Connie Leibow; and Fred Hutch Cancer Center via research coordinator Ellie Brewer, oncologist Sarah Psutka, and urologist Hanna Hunter. This line of research has led to a poster at the APTA Pediatrics Annual Conference [72], an abstract in the Journal of Investigative Medicine [86], an abstract in the Journal of Clinical Oncology [54], and a manuscript in Urologic Oncology [136]. This work has also been presented at the cerebral palsy grand rounds at Seattle Children's Hospital and the Computer Science & Engineering Industry Affiliates Day. The commercialization potential of this work has also been shown through our participation in the UW CoMotion Innovation Gap Fund program.

My roles in this project include: (1) leading the design, development, and iterations on the software platform used in [72, 86, 54, 136], including coordinating with engineers and designers; (2) contributing to the design and administration of the studies, including coordinating with research coordinators; (3) seeking and pursuing opportunities for translation and adoption of the ExerciseRx project in real world clinical and home settings.

This work was supported in part by the Bladder Cancer Advocacy Network, the National Multiple Sclerosis Society, and by Google.

4.1 Formative Interviews

We conducted a series of formative semi-structured interviews with 5 teenagers with cerebral palsy and 9 bladder cancer survivors who previously underwent chemotherapy and surgery. The goals of the interviews were to: (1) understand current practices around exercise, including barriers and facilitators to engaging in physical activity in home settings; (2) their relationship with their healthcare provider, if applicable; (3) their relationship with technology, and to understand what features of digital health tools might be helpful in supporting physical activity.

We found multiple themes common to interview responses provided by participants from both populations. All participants acknowledged that exercise plays a valuable role in their health: flexibility exercises for cerebral palsy patients helps them with maintaining range of motion, and strengthening exercises for bladder cancer patients helps with postoperative outcomes. However, many participants sought greater support from caregivers and healthcare providers. Multiple cerebral palsy participants said that additional family support could be helpful in staying motivated, particularly when they did not feel motivated by their physical therapist. Similarly, bladder cancer participants reported feeling that healthcare providers did not prioritize discussions about physical activity after treatment, instead focusing on oncological outcomes. These participants suggested that they would have liked their providers to ask if they were able to move around regularly or how fatigued they felt.

In addition to the role of health, participants also offered suggestions for features in a digital health tool for facilitating exercise sessions. Participants most prominently stressed the importance of being able to monitor and reflect on their progress over time. They compared it to existing app designs such as the progress bar in Duolingo or the rewards systems often used in video games. While some participants also potentially liked the idea of being able to compare their progress to that of others, they were concerned about some of the side effects that may come from a feature like this, such as toxic comments or deceptive artificial progress. In addition to tracking progress, participants also suggested that they would like to see instructional content tailored to them, such that the "teacher" in the video looked and moved like them, to build empathetic engagement.

We also found themes in which participants across populations disagreed on. The greatest difference was around the set of exercises prescribed; while bladder cancer patients were specifically concerned about the utility of the exercises prescribed, cerebral palsy patients were instead more interested in having a wide variety of exercises to choose from. While this difference may be attributed to the difference in age ranges, it may also be a result of how long each patient population is intended to use the ExerciseRx app; bladder cancer patients might want the most effective set of exercises to complete in their prehabilitation process, while cerebral palsy patients might want to have many different choices to choose from throughout the course of their lifetime. Aside from the exercise prescription, patients also disagreed on the frequency of notifications and reminders to exercise. While cerebral palsy patients were optimistic about receiving notifications frequently, bladder cancer patients were more protective of their attention

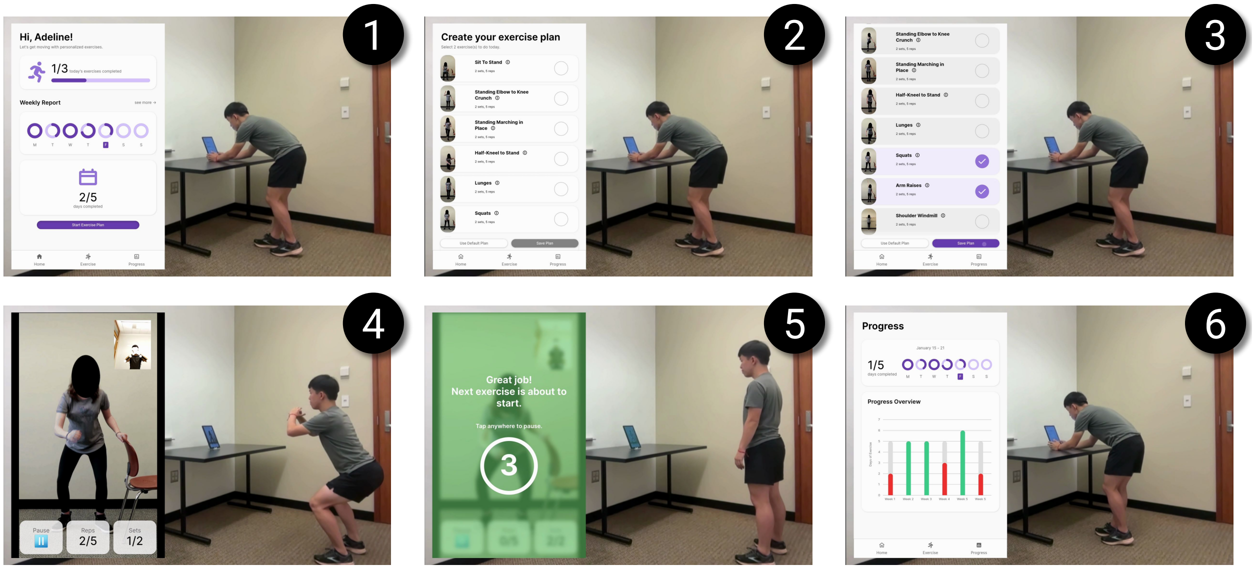


Figure 4.1: Walk through of the ExerciseRx app interface. (1) Home screen showing progress on daily, weekly, and monthly scales. (2) and (3) Exercise selection from a physical therapist prescribed list of choices. (4) Exercise screen showing an instructional video, a selfie-view camera feed, and real-time repetition and set counting. (5) Breaks are inserted between sets of exercise and can be requested on-demand through the Pause button. (6) Detailed progress is displayed after each workout, and can also be accessed from the home screen.

span and thus conservative about notifications, with many specifying they would want to receive "a few notifications a week at most".

4.2 System

The app was designed to be simple while facilitating three primary functions: (1) support healthcare providers to prescribe an exercise routine for patients, (2) provide real-time exercise feedback through repetition counting, and (3) enable users to reflect on progress made towards their exercise goals. The app was implemented using JavaScript as a Progressive Web Application to enable convenient cross-platform testing and so that it can be installed onto the home screen of devices, appearing alongside native apps. The app uploaded usage statistics and participant adherence to their prescribed exercises to a remote server, which our research coordinators could monitor in real-time from a dashboard.

To support real-time feedback through automatic repetition counting, we implement a computer vision pipeline that first extracts body landmark coordinates from the camera video feed, then processes those coordinates to detect when exercise repetitions are completed. Specifically, we employed the BlazePose [10] machine learning model which returns 2D coordinates for 33 body landmarks. This model is capable of running in real-time on-device, maintaining user privacy. We then consider each exercise as a state machine. We train a separate support vector machine for each exercise. The support vector machine takes as input the body landmarks for a single video frame and outputs the current state in the exercise's state machine. Completing a cycle of the entire state machine results in incrementing the repetition count, which is displayed as feedback.

The app is designed to support dynamically updating the set of available exercises. Specific lists of exercises are made available to providers of different specialties (i.e., flexibility exercises for cerebral palsy providers), and providers prescribe sets of exercises for patients through a dashboard by selecting them from the list of available exercises. Patients then receive their exercise prescription (i.e., consisting of exercises with an instructional video and a machine learning model for recognizing repetitions of that specific exercise) on their local copy of the ExerciseRx app.

Table 4.1: Summary of study participants in the at-home study. AHP stands for at-home participant. Grade refers to year in school.

ID	Gender	Age	Grade	GMFCS Level	Prescribed Exercises
AHP1	Male	16	10th	II	Arm Raises, Marches, Sit-to-Stands, Squats
AHP2	Female	13	8th	II	Lunges, Marches, Squats
AHP3	Male	15	9th	III	Arm Raises, Sit-to-Stands, Marches
AHP4	Male	13	8th	II	Arm Raises, Lunges, Sit-to-Stand
AHP5	Female	15	9th	II	Arm Raises, Lunges, Sit-to-Stands

4.3 Validation in Cerebral Palsy

4.3.1 Participants

We recruited participants through the providers on our team reaching out to their local networks via word of mouth, online communities, and mailing lists. Participants were required to be between the age of 13 and 17 years old, have a cerebral palsy diagnosis, and be classified with a GMFCS level of II or III, meaning that they have some motor impairment but can walk with or without an assistive device. Participants were also required to be English speakers and have regular access to an internet connection. In total, we recruited 6 participants, with 4 male and 2 female, for this study. The mean age was 14, with the youngest at age 13 and the oldest at age 16. Participants were compensated for completing each stage of the study. Participants were paid $15 for enrolling in the study and completing the onboarding process, $60 for completing the 5-weeks of exercises, $25 for completing an exit interview. In this section, we will refer to these patient participants as AHP#, for “at-home participant”, where # is their participant ID.

4.3.2 Apparatus

We loaned a Samsung Galaxy Tab A7 tablet with a 10.7 inch screen running the Android operating system to each participant. The prototype app described in Section 4.2 was preloaded onto each tablet. We also included a small foldable tablet stand to help participants stand the tablet up while performing their exercises.

We selected 8 exercises to support in this app: (1) arm raises, (2) half-kneel to stand, (3) lunges, (4) shoulder windmill, (5) sit-to-stand, (6) standing elbow to knee crunch, (7) standing marching in place, (8) squats. These exercises are dynamic, in contrast to stationary exercises such as stretches or balance exercises, so that they can be visually tracked with a camera. They were also selected as exercises relevant to people with cerebral palsy with GMFCS levels II and III. We recorded custom instructional videos for these 8 exercises featuring an adolescent volunteer with cerebral palsy.

4.3.3 Procedure

Participation in the study involved around 7 weeks of app engagement. After participants consented to participate in the study, two members of our study team traveled to the participant's home to facilitate the onboarding process. At this meeting, the pediatric physical therapist from our research team prescribed 3 to 5 exercises with corresponding repetition and set goals for each exercise. The physical therapist made sure that the selected exercises could be safely performed by the participant, and that the participant was interested in doing the exercise. The prescription was made by first providing instructions in writing to the participant, then configuring the app through an online secure HIPAA compliant dashboard. After creating the prescription, the study team representatives introduced the participant to the tablet and app, walking the participant through how to use it. In particular, they made sure the participant could navigate the app on their own and complete the exercises safely while ensuring the machine learning models correctly counted each repetition of exercise. Finally, the study team representatives helped the participant find a place in their home that they could consistently do their exercises, helped the participant determine a time to do their exercises, and configured notification reminders, if requested. After onboarding, participants were asked to exercise 3 to 5 days for five weeks. Each week, participants were asked to complete a brief online survey. At the end of the five weeks, participants were invited to participate in a remote exit interview to discuss their experience using the app and their engagement in physical activity as supported by the app. This study procedure was approved by our Institutional Review Board.

4.3.4 Results

Participant Adherence to Exercise Prescription

According to our usage logs, participants completed an average of 3.87 exercise sessions per week. AHP5 completed the most sessions, with an average of 4.6 sessions per week, and AHP4 completed the least, with an average of 3.0 sessions per week. Through the exit interviews, 5 out of 6 participants self-reported that the app increased their exercise levels compared to before using the app.

Feedback on Instructional Video

Participants reported in general finding the instructional videos more motivating than the paper handouts that PT's have provided them in the past. For example, AHP2 explained how *"it [the instructional video] was nice for showing me how to do the exercise... [unlike] the paper instructions that didn't really give me motivation to do it, and it was kind of annoying cause I would have to try and find the paper, and probably lost the paper several times."* AHP3, however, commented that they thought that the model in the instructional video *"kind of didn't feel representative to me... like, way more able-bodied, and she made it look easy while you were having a really hard time."* Although we intentionally filmed the instructional videos with a model from the target population (i.e., adolescents with CP), because of the variations within the CP condition, this participant suggested that the video was insufficient for making them feel engaged.

Progress Tracking Features

Some participants did not notice the progress tracking features of the app. AHP6 commented that *"It would be cool if on the app you could see how many days you had done [exercises]"*, even though it was displayed on the home screen. There was a consensus that greater integration of the progress tracking elements inline with the exercising experience could increase its visibility and enhance motivation. AHP3 suggested that *"They [Duolingo] have this way of sort of incentivizing streaks and giving people awards for certain number of streaks... It's pretty smart, so I could see how that might be really motivating if it was in this app."*

Pose Tracking and Repetition Counting Experience

Participants found it useful to be able to see themselves move and connect their movements with the on-screen visual feedback. However, AHP3's parents commented that they got a *"... a little obsessed with watching it [the selfie camera]"*, ultimately affecting their posture and form during exercise. AHP6 cited how they found less utility for the instructional videos after they had learned how to perform the exercises, and suggested *"What if those videos [the instructional videos] are smaller and the video of you is bigger?"*

The repetition counting feature was well-received, with participants noting that it reduced the burden of keeping track of repetitions. AHP1 described *"It was pretty helpful that it [the app] showed like the progress of the sets."* However, some participants found the feature to have issues with reliability. AHP6 shared that *"It seems like it didn't catch every time I was doing it. So you had to do extra reps to get the number... and that made it a little hard to finish sessions."*

Preference for Personal Device

Participants appreciated the convenience of exercising at home. AHP1 shared that *"the option of being able to work out at home was a good thing... it's a lot quicker than moving to the gym."* However, multiple participants also expressed a preference for using the app on their personal devices rather than a provided tablet, as this would increase autonomy and ease of use. AHP6 commented that *"I feel like the biggest constraining factor was the device, just because I feel like, if it was my personal device, and I was more in charge, and I would have more reason to be motivated to do it."*

4.4 Validation in Bladder Surgery

In this subsection, I will present preliminary results of the ongoing "Get Moving Trial" [54]. In this study, we investigated the hypothesis that increased physical activity leading up to a surgical operation will improve post surgical outcomes. We examined physical activity in the form of walking and running (i.e., via step count metrics) and through a home exercise plan (i.e., via the previously described app). In this chapter, I will only discuss the home exercise portion of the study since that was my primary responsibility. Preparation for this two-phase randomized control trial began in May 2023, and patient enrollment began in October 2023.

4.4.1 Procedure

Patients are randomized upon enrollment in a 1:1 ratio to either the *"(P)REHAB"* or standard-of-care arms. Patients in the (P)REHAB arm are prescribed a home exercise program via the ExerciseRx app and asked to complete exercise sessions (designed to last approximately 20 to 30 minutes) four times a week. They are asked to begin completing exercise sessions during neoadjuvant chemotherapy, which typically begins 4 to 6 weeks prior to surgery, through 90 days after surgery. Patients in the standard-of-care arm are provided printed guideline-based recommendations for physical activity as is typically given during perioperative care.

4.4.2 Apparatus

Patients in the (P)REHAB arm of the study will be onboarded into the ExerciseRx system. This technology onboarding process involves creating an account for the patient, having a rehabilitation doctor prescribe a specific set of exercises, having the patient download the ExerciseRx app onto their personal smartphone, and logging in. Patients are provided a small foldable stand to prop up their smartphone while completing exercises.

We selected 14 unique exercises to support in this study: (1) arm pull back, (2) back leg lift, (3) elbow to knee crunch - left and right, (4) forward arm circles, (5) seated front arm raise, (6) marching, (7) chair march - with and without resistance, (8) seated leg arc, (9) shoulder squeeze - with and without resistance, (10) arm pull apart - with and without resistance, (11) side leg lift - left and right, with and without resistance, (12) sit to stand - assisted and not, (13) squat, (14) assisted lunge - left and right. Many of these exercises also involve variations. The most common variation is when an exercise uses only one side of the body, and the mirrored version of the exercise is also explicitly implemented. Other exercises were presented with option to make them easier (i.e., with chair assistance) or more challenging (i.e., with resistance bands). We recorded custom instructional videos for these 16 exercises. Some featured older adults (i.e., in the approximate age range of the target patient population), but due to time limitations, others featured young adults on our team.

4.4.3 Participants

We recruited participants through bladder cancer patients recommended for surgical operations at Fred Hutch Cancer Center. The target enrollment for powered analysis is 102 participants. As of June 2025, 418 patients have been screened, 48 patients have been enrolled as participants, and 16 have completed the study.

4.4.4 Results

We report adherence in terms of completed sets of exercises, as given by:

$$\frac{completed\_sets}{prescribed\_sets * 4 * num\_weeks} \tag{4.1}$$

Here, $completed\_sets$ is the number of completed sets of exercises over the entire duration of the study; $prescribed\_sets$ is the number of sets to be completed in a single exercise session as prescribed by the provider during onboarding; 4 is the number of exercise sessions per week recommended to patients; num_weeks is the total number of weeks the patient participated in the study. Out of the 16 patients that have completed the study thus far, 9 were in the (P)REHAB arm. In these 9 patients, the mean adherence to their exercise prescription is 42.8%. Although this adherence rate may seem low, prior work in exercise monitoring has shown adherence rates of 20% to 50% in groups with general musculoskeletal conditions [112, 9]. We therefore consider this encouraging evidence for the feasibility of bladder cancer patients engaging with the ExerciseRx app.

4.5 Discussion

In this chapter, I presented studies with two different patient populations. I will now discuss commonalities and differences in qualitative patient feedback between the two populations.

4.5.1 Anticipated Usage Durations

The intended use of the ExerciseRx app varies between the two populations. Although bladder cancer is a chronic condition, surgery *can* completely cure it; in this case, the ExerciseRx app is intended to be used leading up to surgery and for limited time after surgery in order to improve surgical outcomes. On the other hand, cerebral palsy is a lifelong condition, thus the ExerciseRx in this case is intended to be used for the patient's entire life as a means of supporting efforts to maintain functional abilities (i.e., completing activities of daily living). Perception of the app is understandably different given this context. For example, bladder cancer patients seemed to be more realistic but also more motivated; they were cautious about receiving too many notifications, but they also cited wanting control of their exercise schedules, and wanting to know that they were doing the most effective set of exercises to improve their post-operative outcomes. On the other hand, cerebral palsy patients seemed overly positive about notifications, perhaps because many of them did not own a personal smartphone or tablet.

The research nature of this work likely also distorts these findings. While the design of the randomized control trial with bladder cancer patients allows us to follow these patients through their entire oncological process, the cerebral palsy study gives us only a brief segment of their lifetime with the condition. We expect the results of the cerebral palsy patients to be optimistic and that realistically, engagement would diminish over time as the novelty wears off. However, we selected this age range to study because it is the time in these young peoples' lives that they need to develop healthy habits, including around exercise.

4.5.2 Stakeholder Involvement

The patient journey involves a number of other stakeholders, including healthcare providers and caregivers. We examine distinctions and commonalities in the clinical ecosystem and caregiver relationships around the two different patient populations studied, and discuss implications for designing digital health tools around these stakeholders.

Although the diagnosis for cerebral palsy is typically provided by a neurologist, most care is provided by to specialists such as physical therapists. In the cerebral palsy study, we thus primarily collaborated with pediatric physical therapists that specialized or had experience with patients with cerebral palsy. Bladder cancer patients are similarly diagnosed by a urologist, operated on by an oncologist, and their physical condition is managed by a rehabilitation doctor. However, in the context of our study examining the period of time between chemotherapy and surgery, all of these clinical stakeholders are actively involved in the care of the patient. We thus had to collaborate with all of these healthcare providers in order to effectively coordinate on patient status and progress, such as if surgery dates had to be moved. In order to facilitate these collaborations, we iterated on the provider dashboard so that it did not only support core functionality (e.g., prescribe exercises and view patient progress), but also enable new means of cross-provider collaboration (e.g., leaving provider notes for other providers to see).

On the other hand, the caregiver relationships of the two populations on the other hand is very different. All of the participants in the cerebral palsy study were teenagers that lived at home with their families, thus under the supervision of their parents. The participants in the bladder cancer study however were older, and thus were cared for by a spouse or family but remained independent adults. This distinction in caregiving dynamics likely affected motivation and engagement. For example, parents of patients in the cerebral palsy study commented on how they reminded their children about doing their exercises. Video recordings of the exercise sessions confirmed that some parents got involved in their child's exercise sessions by watching and giving feedback. We did not find any analogous behavior in the bladder cancer patient videos or in the exit interviews. Systems that use pose estimation should be designed to handle scenarios where multiple people are visible, especially if targeting children or adolescents. Furthermore, systems can be designed to support and facilitate interaction between a patient and their caregivers, such as by presenting progress data that is understandable by all stakeholders.

4.5.3 Metrics of Success

There are multiple ways to evaluate the success of this work. Physical therapists and sports medicine doctors will often say that *"any movement is helpful"*, and that the ExerciseRx app's ability to increase exercise by any amount makes it an effective system. We however seek to define more objective metrics of success. For bladder cancer patients, demonstrating that usage of the app leads to increased strength and thus improves post-surgical outcomes is a clear metric of success. As part of motivating patients to continue engagement, systems could be designed to monitor progress over time and also provide feedback on movements. Towards this goal, we conducted an experiment in which the physical therapists from the cerebral palsy study annotated a subset of the video recordings of home exercise sessions across different standard movement quality metrics such as range of motion and stability. While we did not see any trends in these movement quality metrics over the course of the 5-week study period, the physical therapists commented qualitatively that there were noticeable improvements. This suggests there is potential for improvement in movement quality metrics that go beyond current clinical standards, such as the novel biomechanical signatures found in Ruth et al.'s work in discerning different neuromuscular diseases [104].

4.6 Summary

In building and studying systems that deliver personalized health interventions, ExerciseRx poses a unique opportunity for simultaneously providing the intervention (i.e., through the exercise instructions) and also passively monitoring patient movement quality for improvement or declining function. In this chapter, I presented our work in promoting exercise for patients of two specific chronic conditions: cerebral palsy and bladder cancer. Through longitudinal, at-home deployments in these populations, and through engagement with clinicians outside of strictly academic research settings, we have demonstrated potential for the clinical utility of this system. Our latest efforts in integrating our system into the EPIC electronic health record system represent encouraging progress in making real-world impact with ExerciseRx.

Chapter 5

PARX

In the previous chapter, I discussed the importance of exercise and the potential for exercise monitoring and feedback systems to improve adherence to exercise plans. We also found that to be effective, such systems should be tailored to the needs of specific populations and specific individuals. However, the process of tailoring is time-consuming and resource-intensive. For example, for each of the patient populations we worked with in the previous chapter, we had to: (1) record custom instruction videos reflecting the movement profile of the target population, and (2) train custom machine learning models to recognize each exercise, which was time consuming and resource intensive. I will now show how generative AI can support healthcare providers in generating and customizing exercise plans for patients using natural language prompting.

In this chapter, I discuss the process of collecting and curating a dataset of exercise videos paired with text descriptions, and show the distribution of this dataset contrasts with that of existing text-motion datasets. I then discuss the design of a transformer encoder to generate exercise instruction videos given a text description, which is used in an agentic workflow context. I finally present preliminary feedback from healthcare providers using this tool.

I demonstrate this project's relevance to my dissertation's thesis in this chapter through:

1. Insights from collecting a dataset of fine-grained text instructions paired with photographic exercise instructions.
2. Technical contributions in designing an agentic workflow for enabling generating and editing human motion animations using natural language queries.
3. Findings from a technology probe with healthcare providers interested in prescribing exercise plans for their patients and feedback from their experience using the system.

My roles in this project include: (1) designing, developing, and evaluating the machine learning models; (2) designing and administering the studies. Undergraduate research assistant Jason Miller helped with the model training process. This line of research is in the peer-review process.

5.1 Motivation and Vision

In this section, we first present a concrete example scenario to motivate the vision for our work. We then use this vision to contextualize our design decisions and system architecture in developing and evaluating PARX.

5.1.1 Motivating Scenario

Consider the following scenario: a physical therapist meets with a patient seeking to rehabilitate from surgery on the left leg. The physical therapist decides that *"leg raises"* are an appropriate exercise for this patient to practice at home to increase range of motion. The most common and accessible solution currently is for the physical therapist to search the Internet (e.g., via Google or YouTube) for instructions on how to perform this exercise (Figure 5.1, left panel, step 1). The physical therapist might share the instructions as is, or they might also make modifications to

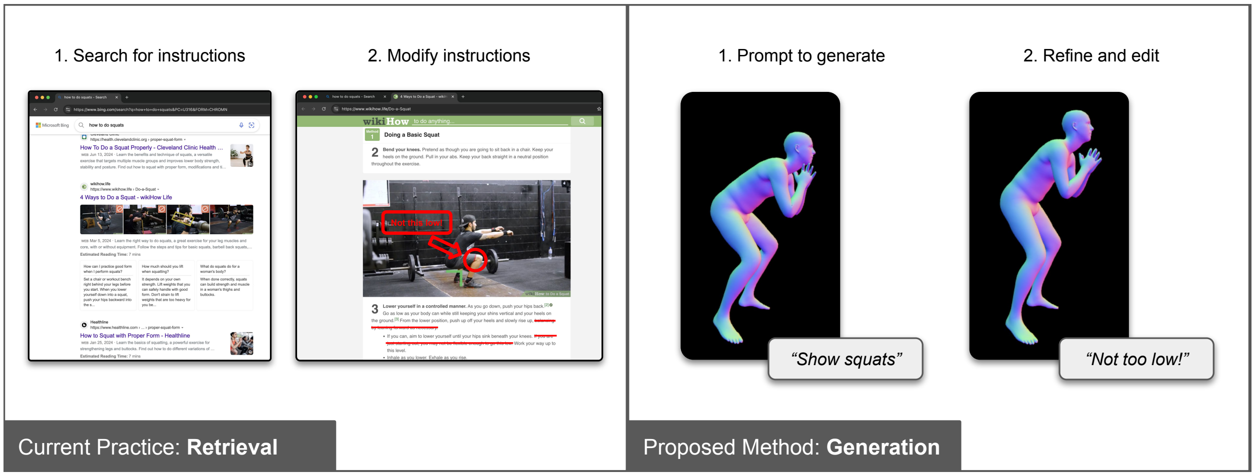


Figure 5.1: High-level illustrations of exercise prescription workflows. **Left:** Current practice involves a retrieval workflow, requiring providers to seek exercise instructions that they modify manually. **Right:** Our proposed workflow takes a generative approach, allowing providers to synthesize exercise instructions through text prompts.

those instructions in order to tailor it to the patient's specific needs. For instance, an example video that the physical therapist found shows a person raising their leg 90 degrees; the physical therapist may warn the patient against raising their leg all the way to 90 degrees, instead asking the patient to start by raising it to 45 degrees. The physical therapist may reflect this modification in different ways. On text instructions, they may scratch out or write over the text. On photographic instructions, they may annotate the photo with labels or text (Figure 5.1, left panel, step 2). On video instructions, there is no realistic way to modify any videos they might have found. They then provide these modified instructions to the patient, who follows them to complete exercises in a home setting. In this scenario, the physical therapist must complete this manual process, including the modification steps, within a limited window of time (e.g., appointments are often only 30 minutes). Our work seeks to facilitate this process.

5.1.2 Vision and Implementation

To address the needs previously described in the scenario, we introduce our technical decisions in the context of Xia et al.'s design space for motion-learning systems [125], which includes the follow dimensions:

- **Customizable target motion:** Currently, physical therapists work with existing instructions found on the Internet, or with instructions in a fixed exercise library. Some systems such as VoLearn [125] support manual customization of specific parameters such as speed or amplitude of motion through interface elements such as sliders. However, such designs must make a compromise between the number of options presented and the expressivity of the system. We investigate the potential for open-vocabulary interaction with a pose model to enable greater customization of target motions.
- **Availability of devices:** Our system needs to run on-device for both the provider and for the patient. On the provider side, we support natural language interaction in real-time on a commercial grade laptop by using quantized large language models. On the patient side, we demonstrate that our approach of using the generated synthetic data can be used to bootstrap smaller exercise tracking models across different sensing modalities, giving patients the flexibility to choose which sensing modality they are comfortable with (e.g., electing to use a wearable instead of a camera for privacy reasons).

In Section 5.2, we will describe the implementation of PARX, focusing on the design and evaluation of the text-to-motion pipeline. Then, in Section 5.3, we will report on semi-structured interviews conducted with healthcare providers, using PARX as a technology probe for understanding how they might interact with it in clinical practice.

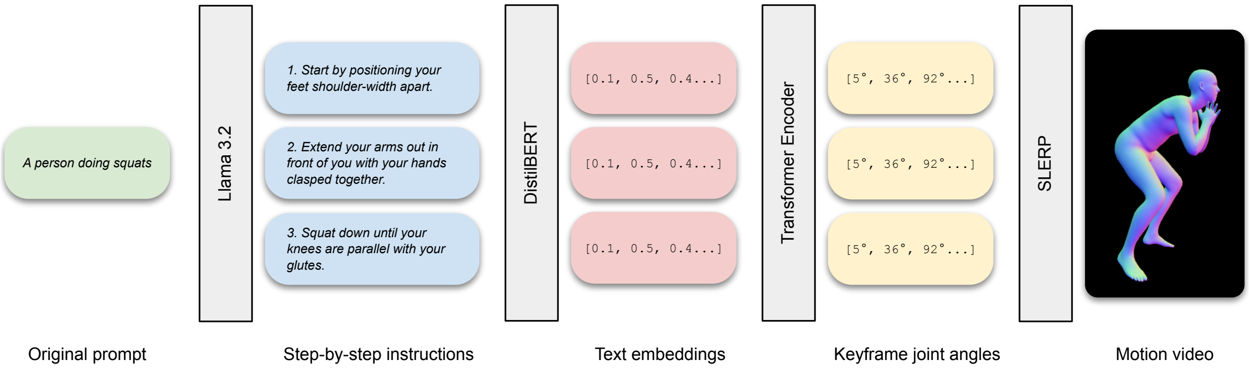


Figure 5.2: PARX's text-to-motion pipeline is designed to generate keyframe poses that are sequenced together to form a video. This design exploits the procedural nature of exercises, which are often described using step-by-step instructions.

5.2 System

This section introduces PARX, a tool that supports healthcare providers in designing exercise prescriptions. We first discuss the design and implementation of the text-to-motion pipeline used to drive the app (Section 5.2.1). We then evaluate this text-to-motion pipeline by comparing its performance on a dataset we curated for exercise instructions with the performance of other models in the literature(Section 5.2.2). We finally discuss how the text-to-motion pipeline is incorporated into the overall PARX app (Section 5.2.3).

5.2.1 Text-to-Motion Implementation

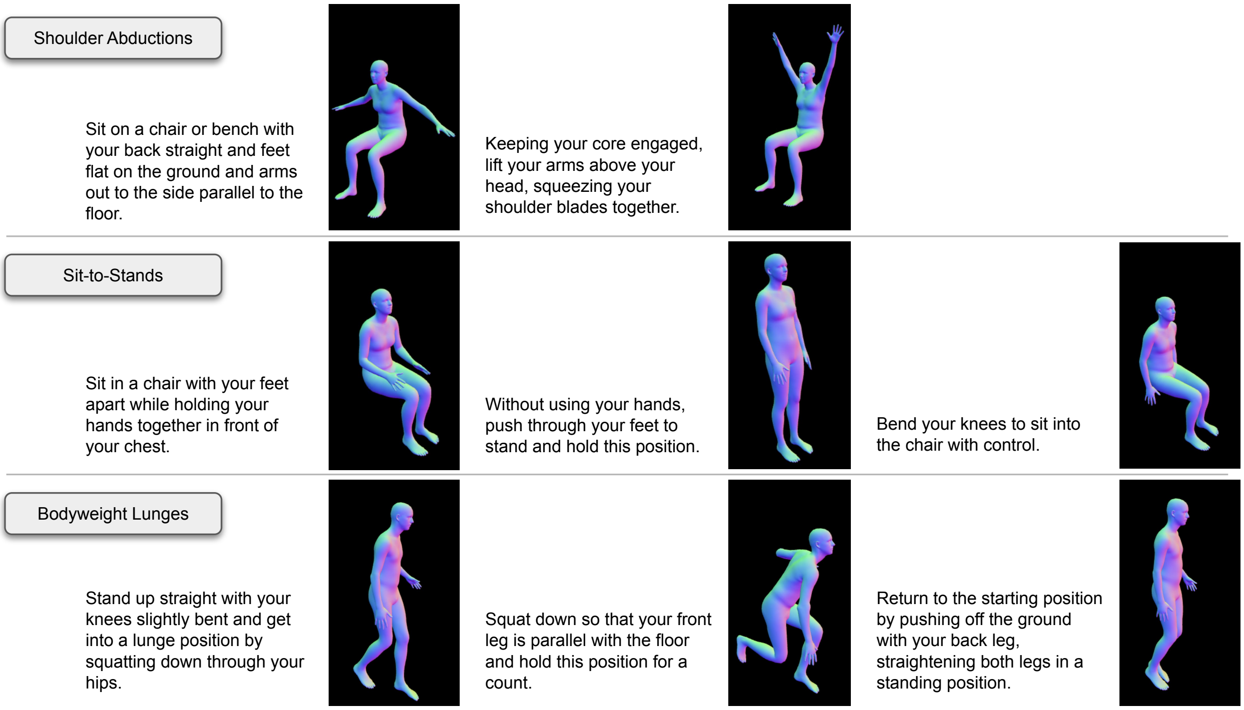
Decomposing Text Prompts into Step-by-Step Instructions

Prior work has found that motion generation can be more fine-grained by turning high-level prompts into more detailed descriptions with greater spatial, temporal, or spatiotemporal fidelity [51, 133]. We similarly decompose high-level instructions (e.g., *"Show a person doing squats."*) into a series of step-by-step instructions (e.g., *"1. Hold your arms straight out in front of you. 2. Lower your pelvis toward the ground until your knees reach 90 degrees..."*). Concretely, we employ the 3-billion parameter version of Llama 3.2 [35] with instruction finetuning provided by Ollama [95] to facilitate this process through the following prompt:

For the exercise ”<EXERCISE>”, please a provide a series of step-by-step instructions, focusing on the physical aspects of the movement. Give your response in JSON format, with the step number as the key and the instruction as the value. Do not include any other keys. Do not provide any other text. Do not include steps that just instruct the reader to repeat. Here is an example:

```
[
    {”0”: ”Stand with feet shoulder-width apart”},
    {”1”: ”Lower body down into squat, keeping weight in heels and knees tracking over toes”},
    {”2”: ”Pause briefly at bottom of squat with weight in heels and knees bent at 90 degrees”},
    {”3”: ”Push through heels to return to standing, maintaining control throughout movement”}
]
```

We found that Llama 3.2 was able to consistently adhere to the JSON format, and that the description provided in each step was largely accurate. We hypothesize that Llama 3.2 was able to successfully perform this task because it was trained on many webpages on the Internet containing step-by-step instructions on how to perform exercises. The response is then parsed as a JSON object, and each item in the list separately provided to the following step.



Shoulder Abductions

Sit on a chair or bench with your back straight and feet flat on the ground and arms out to the side parallel to the floor.

Keeping your core engaged, lift your arms above your head, squeezing your shoulder blades together.

Sit-to-Stands

Sit in a chair with your feet apart while holding your hands together in front of your chest.

Without using your hands, push through your feet to stand and hold this position.

Bend your knees to sit into the chair with control.

Bodyweight Lunges

Stand up straight with your knees slightly bent and get into a lunge position by squatting down through your hips.

Squat down so that your front leg is parallel with the floor and hold this position for a count.

Return to the starting position by pushing off the ground with your back leg, straightening both legs in a standing position.

Figure 5.3: Example pose sequences from the text-to-motion generation pipeline. The detailed instructions are generated by prompting the LLM to decompose the given exercise into steps. The text-to-pose model is run for each instruction step, and the LLM produces different numbers of steps for each exercise, resulting in different numbers of keyframe poses. These keyframes are sequenced together to produce a motion video.

Text-to-Pose Generation and Editing Models

We first preprocess the text in each instruction step by tokenizing and embedding it using DistilBERT [106]. These embeddings are then positionally encoded and passed into an encoder model to produce a corresponding pose (represented as 24 3D joint angles in SMPL format [80]). Two separate transformer models are used: the *PoseGenerate* model and the *PoseEdit* model. For the first step in each sequence of instructions, the text embeddings are passed into PoseGenerate's multi-head attention mechanism as the query, key, and value. For subsequent steps in each sequence of instructions, the text embeddings are passed into PoseEdit's multi-head attention mechanism as the key and value; the query is the pose generated from the previous step. This design allows the system to produce temporally consistent poses by being aware of the previous frame. We initially tried to train a single model for both of these tasks by providing a generic pose (i.e., a T-pose) as the source pose in for first step instructions. However, consistent with Delmas et al.'s [30] findings, the model learned to ignore the source pose because it is too infrequently used. Although the editing model was initially designed to support generation of temporally consistent poses, we also found that it enables prompted editing of poses, such as in the right panel of Figure 5.1.

Sequencing Pose Keyframes into a Motion Video

After each instruction step is processed through the transformer models, we treat each resulting pose as a keyframe. The SMPL format represents poses with a sequence of joint angles, so we apply a spherical linear interpolation (SLERP) [111] to each joint of each pose to sequence the keyframes into a motion video. The quaternion interpolation process produces constant-speed motion along a unit radius, creating a smooth upsampling effect. We empirically found a constant target frame rate that produced visually adequate results. Finally, the pose keyframes are rendered into meshes by running a forward pass of the SMPL neutral model [80], producing a motion video as seen in the right most element of Figure 5.2.

5.2.2 Text-to-Motion Evaluation

In this section, we present an evaluation of the text-to-motion pipeline by comparing its performance against other models from the literature. The primary goal for this evaluation is to characterize our pipeline and give readers a reference point for what study participants experienced.

Dataset and Procedure

We first describe an internal exercise instruction dataset curated for research which we will refer to as *ExerciseInstruct*. The dataset comprises of sets of instructions for 1171 exercises; they may not necessarily be unique exercises, but they will present different ways of explaining the exercise. Each set of instructions contains 2 to 6 steps, and each step comprises of a text description and still photo illustration pair. In total, the dataset contains 4612 text and image pairs. We preprocess the images by fitting a SMPL model to them using Goel et al.'s HMR 2.0 model [44], now forming text and pose pairs. We then randomly split the dataset into train (70%), test (15%), and validation (15%) subsets.

For training and evaluation, we use a combination of publicly available datasets and our exercise instruction dataset. For the text-to-pose task, we pretrain PoseGenerate using the PoseScript dataset [29], which pairs over 6000 poses from the AMASS dataset [83] with human-annotated text descriptions. We then finetune this model using the first step of instructions (e.g., such as what follows in "Begin this exercise by...") from the training subset of ExerciseInstruct. We finally evaluate our model on the testing subset of the exercise instruction dataset. For the pose editing

Model	FID↓	R@1↑	R@2↑	R@3↑	MPJRE↓	MPJPE↓	Time↓
T2M-GPT [132]	0.046	11.47	23.68	54.75	23.12	3.998	1.699
FineMoGen [133]	0.051	11.50	23.69	54.84	22.51	3.556	4.891
Ours (pretrained)	0.045	10.67	21.34	53.36	26.94	3.999	0.097
Ours (finetuned)	0.027	13.34	32.03	56.04	16.63	3.152	0.097

Table 5.1: Characterization of our system's generative performance through comparison with baseline methods on ExerciseInstruct. We report on our system's performance when pretrained on the AMASS dataset [83], and then finetuned on a split of ExerciseInstruct. **MPJRE** is reported in degrees, **MPJPE** is reported in centimeters, and **Time** is reported in seconds.

task, we pretrain PoseEdit using the PoseFix dataset [30], which provides several thousand pairs of poses and corresponding text description. We then finetune this model using each post-initial step of instructions from the training subset of ExerciseInstruct. The pose pair for the pose editing task is formed by sequentially adjacent poses in our dataset (e.g., the source pose is from step 1 in the instructions, and the target pose is from step 2).

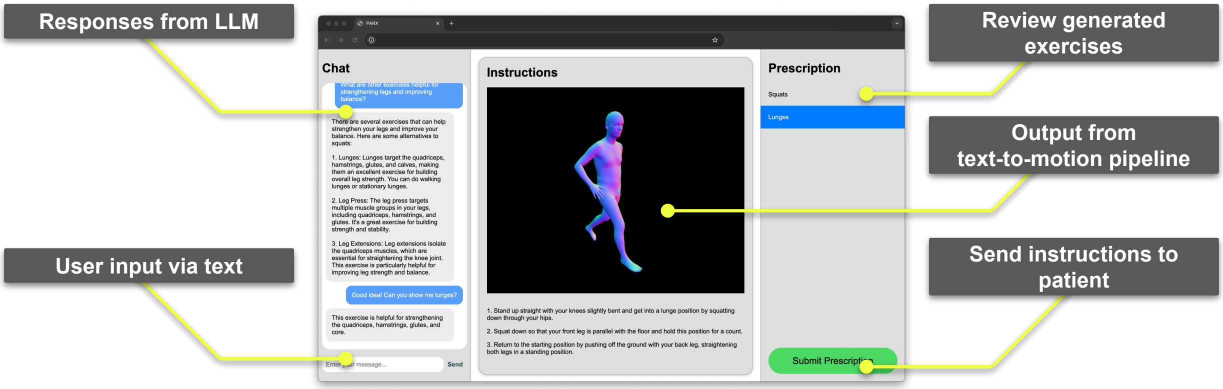
Results

We evaluate our text-to-motion generation pipeline using a relevant subset of the metrics defined by Guo et al. [46]. Since exercises are very precise motions, and because healthcare providers likely have a specific motion in mind when prompting PARX, we elected to exclude the following metrics:

- **Diversity:** Variance of motion across categories.
- **Multimodality:** Variance of motion within categories.

We use the remaining metrics from Guo et al. [46] and also metrics from pose estimation [52]:

- **Frechet Inception Distance (FID):** Comparison of distribution as a measure of realism.
- **Top-K Recall:** Alignment of generated motion with text prompt.



What are other exercises helpful for strengthening legs and improving balance?

There are several exercises that can help strengthen your legs and improve your balance. Here are some alternatives to squats:

1. Lunges: Lunges target the quadriceps, hamstrings, glutes, and calves, making them an excellent exercise for building overall leg strength. You can do walking lunges or stationary lunges.

2. Leg Press: The leg press targets multiple muscle groups in your legs, including quadriceps, hamstrings, and glutes. It's a great exercise for building strength and stability.

3. Leg Extensions: Leg extensions isolate the quadriceps muscles, which are essential for straightening the knee joint. This exercise is particularly helpful for improving leg strength and balance.

Good idea! Can you show me lunges?

This exercise is helpful for strengthening the quadriceps, hamstrings, glutes, and core.

1. Stand up straight with your knees slightly bent and get into a lunge position by squatting down through your hips.

2. Squat down so that your front leg is parallel with the floor and hold this position for a count.

3. Return to the starting position by pushing off the ground with your back leg, straightening both legs in a standing position.

Figure 5.4: The PARX app enables healthcare providers to generate exercise instructions by presenting a user interface for interacting the the text-to-motion pipeline.

- **Mean per joint rotation error (MPJRE):** Difference in angle across all joints.
- **Mean per joint position error (MPJPE):** Difference in distance across all joints.

We also evaluate the runtime required to generate a single pose to gauge the level of interactivity supported by our system. The times reported here are the average times of 100 rounds of generation on a 14-inch MacBook Pro with M3 Max processor and 36 GB of RAM.

As seen in Table 5.1, our pretrained model performs comparably to that of two baseline models from the literature, T2M-GPT [132] and FineMoGen [133]. Due to our pipeline's simplicity, finetuning is exceptionally beneficial, leading to strong gains across all metrics.

5.2.3 The PARX App

To enable interaction with the text-to-motion pipeline, we built the PARX app, as shown in Figure 5.4. The provider provides input via the text box, which is sent to a router LLM with tool-calling capabilities; we again use Llama 3.2 served via Ollama. When prompted with a query, in addition to providing a text response, the router LLM can decide to call one of two tools:

PoseGenerate or PoseEdit. The tools run the corresponding pipeline as described previously. When the PoseGenerate tool is called, a new exercise entry is created on the right-most panel of Figure 5.4. When the PoseEdit tool is called, the router LLM must decide which frame of the motion video to edit. By using this tool-calling LLM architecture, PARX enables a seamless experience: (1) the provider can ask questions, such as suggestions for what exercise might be helpful for a given condition, and receive a text reply; (2) the provider can then ask to see what one of those suggestions looks like, and the router LLM is capable of rendering the motion video.

5.3 PARX Provider Interviews

We conducted semi-structured interviews with healthcare providers, using the PARX system as a technology probe [55], with three goals: (1) to understand their specific needs and desires in the context of exercise prescription, (2) to gather data about how they might interact with a system such as PARX, and (3) to inspire them to consider how generative AI might be used in their practice.

5.3.1 Methods

Participants

Providers were recruited through word-of-mouth from the clinicians on our team. Participants included healthcare providers that personally prescribe exercise plans (e.g., physical therapists) and providers that recommend seeing a physical therapist or suggest increasing physical activity levels (e.g., psychiatrists). Further background information is shown in Table 5.2.

P#	Exp.	Job Title	Specialty	Relevance to Exercise
P1	2	Rheumatologist	Axial Spondyloarthritis	Recommend physical therapy
P2	5	Physiatrist	Cancer Rehabilitation	Prescribe exercise routines
P3	4	Physiatrist	Cancer Rehabilitation	Prescribe exercise routines
P4	26	Physical Therapist	Pediatric Cerebral Palsy	Prescribe exercise routines
P5	24	Physical Therapist	Manual Therapy	Prescribe exercise routines
P6	14	Physiatrist	Sports Medicine	Prescribe exercise routines
P7	19	Physical Therapist	N/A	Prescribe exercise routines
P8	37	Psychiatrist	Aging	Suggest physical activity

Table 5.2: Healthcare provider participant background information. **P#** refers to participant number. **Exp.** is short for experience and refers to years of experience in their specialty. Physiatrists are also known as physical medicine & rehabilitation doctors.

Apparatus

Participants were provided a 14-inch MacBook with M3 processor and 36 GB of RAM with the PARX app preloaded. The interface was opened before the laptop was given to the participant.

Procedure

Participants were first engaged in a series of background questions to understand their background in using generative AI as well as how their medical specialty relates to exercise. We then introduced them to the PARX system, explaining the motivation for developing the system, and walking through the process of generating a prescription through prompting the LLM. The participant was then asked to consider a patient for whom they have recently prescribed an exercise routine. They were then asked to use PARX to design and execute the routine that they recalled. Finally, we explored participant impressions of the system, seeking to understand their experience in interacting with the system and how they might incorporate such a system in their own practice. This protocol was approved our Institutional Review Board.

Analysis

The laptop screen was recorded to capture how the participant interacted with the system, and audio was recorded for any questions or comments the participant may have provided. We developed interview protocols with specific questions in mind, and we coded transcripts of the recordings deductively according to these areas. We then associated usage interactions and patterns with these codes to substantiate these themes and inform the design of future systems.

5.3.2 Results

Overall Participant Impressions

PARX was generally well-received by provider participants. On the mHealth App Usability Questionnaire (MAUQ) [135], participants gave an average score of 5.3 out of 7, indicating good overall usability. The ease of use component score was 5.9, the system information arrangement component score was 4.4, and the usefulness score was 5.4. System information arrangement being the lowest score is understandable, given that we did not spend significant time on polishing the interface. However, we are encouraged that participants found the system easy to use and also useful. Provider participants also qualitatively added that they were excited to see and use it in clinical practice.

Comparing a Generative Approach to Current Methods

Participants shared current methods for finding and distributing exercise instructions to their patients. Among these methods, 1 participant primarily relied on pre-printed brochures for common exercises, 5 participants used general-purpose search engines such as Google to find instructions, and 2 participants had access to a specialized exercise library to find instructions. They further shared that distribution of instructions was often through printed sheets of paper, but sometimes also attached to the follow-up instructions section of an electronic health record-generated visit summary. Two participants further described how they had configured macros in their electronic health record systems to automatically populate the follow-up instructions section with commonly prescribed exercises.

With this context for current methods, the most commonly recurring theme in discussion was the issue around the names of exercises. Participants described how current retrieval-based systems, including general-purpose search engines as well as exercise library systems, require knowing the specific, often colloquial, names of exercises. P4 explained: *"So that's the hard part... in my mind I know what something is called, but then you might search that in the search bar, and nothing comes up... You have to think about the whole spectrum of a layman's term name of the exercise... someone else would call this exercise a clamshell, but I would call this exercise a side-lying hip external rotation."* All participants agreed that a generative approach that flexibly captures intent would be extremely compelling. A few (3) participants experimented with how PARX would respond if prompted with different names for the same exercise. It was successful in generating a "clamshell" (another name for "side-lying hip external rotation") and a "bridge lift" (another name for "hip raise"), but not in generating a "double chin" (another name for "chin tucks"), for which it generated a crunches-like movement on the ground that seemed to move the chin. Participants also responded positively regarding the pose editing functionality, suggesting that it would be helpful for tailoring an exercise for a patient's condition or supporting a patient's exercise progression as their condition changes. The customization process might involve composing multiple movements and forming a new exercise. P2 described a recent experience that could have benefited from this functionality: *"[A patient] had breast cancer, so their right shoulder had pain and deconditioning... the goal was for them to do arm circles... I had them start with standing next to a wall, with their arm at neutral, and slowly raise their arm against the wall... maybe there's not a name for it."* PARX was able to successfully edit a previously generated video of "arm circles" to match this modification.

However, its inability to generate props (i.e., a wall) meant that the resulting video did not fully reflect P2's request.

In addition to personalizing content for patients, participants also expressed interest in customizing PARX for themselves as providers. Participants shared that because there is not a well-defined set of exercises for a given condition or diagnosis. In practice, individual providers will gravitate towards their personal favorite exercises. P5 shared that they *"...have like common prescription sets I like... I usually do these 4 for this functional impairment, and I do these 3 for this impairment... I'm sure there's literature out there on alternatives, but this is what I do..."* Participants added that they appreciated how PARX supports them in expressing themselves, and that they would like to be able to save and reuse exercises that they designed.

Some participants suggested that although having personal preferences is natural and a reality when there are so many options, it may be helpful for the system to also be able to provide suggestions. This functionality may also be helpful for providers that are not physical therapists (e.g., rheumatologists such as P1) to feel more confident in exercises they would recommend or prescribe. Although we did not explicitly build the functionality to suggest exercises into PARX, the router LLM was able to offer some suggestions when prompted; additional prompt-tuning could increase the system's efficacy in this ability. Participants also suggested that it would be helpful to see what exercises other providers were prescribing.

Interacting with the System

As described previously, participants considered the ability to express ideas in free-form text to be an advantage of PARX. However, a few participants identified opportunities for traditional interface elements to complement the text interface. For instance, P3 suggested that checkboxes could be used to enable or disable an exercise modification described with text, or sliders could be used for more precise input. Participant desire to customize PARX for themselves also motivated suggestions for stars to mark a generated exercise as a "favorite".

In response to the animation being a 3D mesh, participants found it powerful to be able to watch the animation from multiple views. They cited that many videos found on the Internet will present different views serially for the patient's benefit (e.g., a video might show a person doing squats from the front first, and then from the side, to demonstrate the angle of the knees), but found the 3D mesh more flexible and powerful. P6 specifically mentioned that the combination of animation and corresponding text instructions would be an inclusive way to help patients with different learning strategies. Some participants also found the ability to view the 3D model from different angles helpful for debugging their exercise design. However, a few participants found the "OrbitControls" camera controller used in our implementation to be challenging to use. They suggested that the camera could be constrained to certain angle ranges; for example, they suggested that users would never need to look underneath the 3D model.

Integration into Clinical Practice

When prompted about how PARX might be integrated into provider clinical practice, the most common concern raised was how much time it would take to use the system. While most participants found the time it took for results to be generated (on average, 5.6 seconds to run all of the LLMs and text-to-pose pipelines end-to-end) to be acceptable, others expressed some concern. P1 shared their perspective: *"I guess three to five seconds is fine... it felt slower in the moment... you know, like when you're wrapped up in clinic, five seconds can feel like five minutes."* On the other hand, some participants suggested that including the patient in the generation process (i.e., having the patient watch the provider iterate on the exercise design) could help alleviate some of the time pressure while also helping the patient understand the reasoning for the exercise. P4 thought that: *"I think it's nice to do it [use PARX], while you're working with the patient... and then the patient can make a connection with that avatar and imprint that in their brain, after they* *just practiced it."* Other participants cited that they currently often spend additional time after appointments putting together exercise prescriptions for their patients, and that they anticipated that PARX could accelerate that process. Finally, providers that would typically not prescribe exercises for patients but instead refer them to a physical therapist suggested that PARX could be helpful for lowering the barrier to prescribing exercises while their patient waits for a physical therapy appointment. Specifically, they thought that PARX could get the patient started with doing some exercises first, replacing brochures and pamphlets that they currently hand out, until they are able to meet with a physical therapist.

Most participants found the appearance of the 3D mesh (i.e., the SMPL neutral model) to be acceptable, specifically commenting that it conveyed movements clearly. Some participants even cited the neutrality of the model as a positive aspect, with no obvious gender or skin color associations. Other participants suggested that the ability to make the model look like the patient could be helpful, both in appearance (i.e., skin color, body shape) and in movement (i.e., natural tendencies due to a condition, such as limping). And yet other participants expressed that it might be helpful to make the mode look more different from the patient so as to help them not fixate on their condition. P4 suggested, due to their pediatric specialty, that cartoon characters could be helpful for engaging children in exercise.

Finally, participants were enthusiastic about the potential for using PARX for at-home patient monitoring. Their enthusiasm was primarily motivated by the ability to use objective adherence metrics to drive patient outcomes. For instance, participants suggested being able to use adherence metrics as a way to provide objective feedback to patients. They also commented that they might be able to use this feedback as a way to reflect on their exercise prescription, and could even make adjustments based on this feedback. Related to the earlier idea of seeing what other providers were prescribing, participants suggested patient adherence and engagement could be another metric to evaluate whether a given exercise or routine would be worth adopting.

5.4 Summary

While the previous two chapters introduced systems designed to be patient-facing, PARX is a provider-facing tool to accelerate the process of tailoring an exercise routine to an individual. The experience of designing a system for provider use is different from designing a system for patient use in a home setting. Providers often have an area of specialization. Designing a tool like PARX can both: (1) lower the barrier to entry for provider that do not specialize in physical therapy or rehabilitation, while (2) also making current workflows faster for providers that do specialize in those areas. Through the provider interviews, we found that these advantages were appreciated and that there is encouraging potential utility for a system like PARX, especially if used in tandem with ExerciseRx.

Chapter 6

Discussion and Conclusion

In this chapter, I will build on the contributions enumerated in this dissertation and present on-going and future work for each of the systems I have introduced. I will then discuss overarching observations and insights from designing, developing, and deploying these systems. Finally, I will end this dissertation with brief concluding remarks.

6.1 On-Going and Future Work

One of the most exciting aspects of the work presented in this dissertation is that they have all demonstrated potential for translation into real-world use. In seeking to make a positive impact with this work, all three of these projects are still in active progress. In this section, I will present some of the on-going extensions to these projects, as well as planned threads of future work.

6.1.1 Beacon

Through our at-home, longitudinal deployment, an apparent area of improvement is to reduce the burden of taking CFF measurements by making the process faster and easier to complete. I

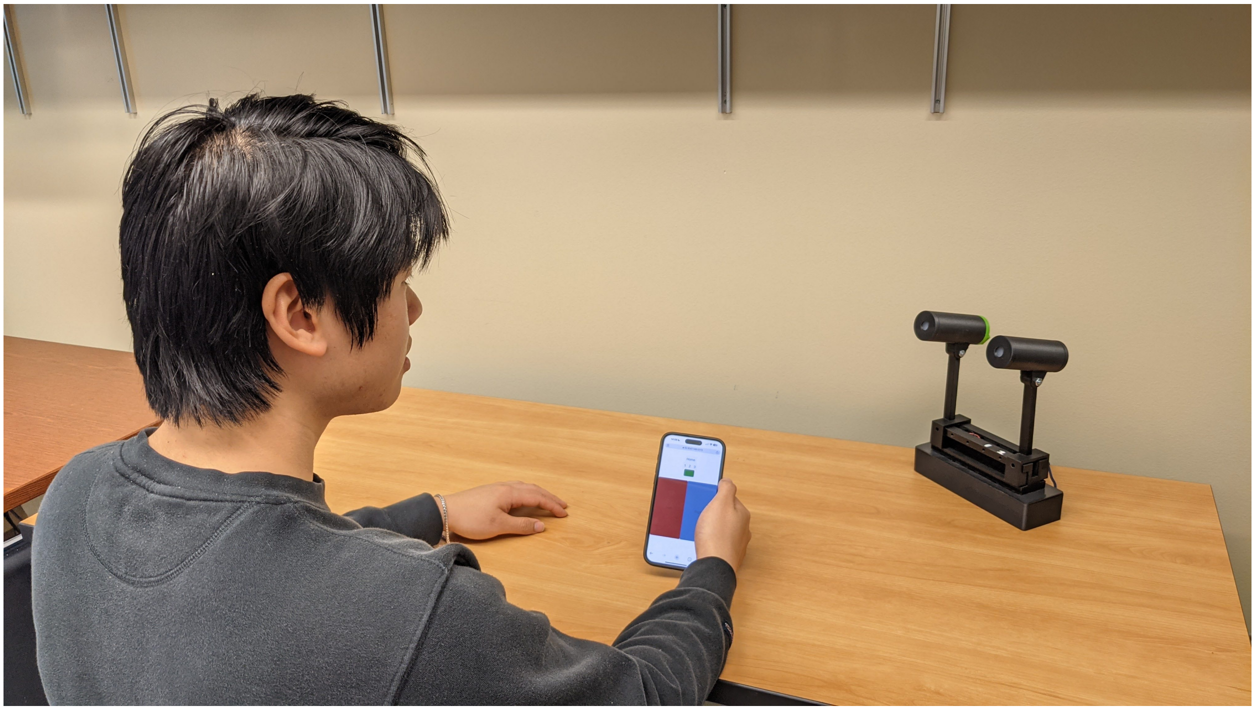


Figure 6.1: A prototype "two-headed" form factor that enables Beacon to present two different light stimuli simultaneously.

will first present an on-going investigation of a new form factor for Beacon to support multiple simultaneous light stimuli. I will then briefly discuss an area of exploration in using statistical and machine learning approaches to leverage information gained from measurements taken over time in order to reduce the amount of time required in a single session.

In addition to iterating on Beacon, we are also making the most of the momentum gained from our AJG and CHI publications, and are actively expanding the scope of Beacon. I will briefly report on engagements with third parties interested in using Beacon, and one instance where we have begun supporting a "deployment" of Beacon in an external clinic. I will also outline an on-going pilot study in dementia patients.

Evaluating a "Two-Headed" Beacon

Undergraduate research assistant Jonathan Shu led an initial investigation into a new form factor for Beacon involving presenting two light stimuli simultaneously. The hypothesis is that this will make the FC protocol faster to complete. Jonathan presented this work at the 2025 UW Undergraduate Research Symposium.

We conducted a pilot study with 10 healthy participants. Participants completed two rounds of 6 conditions, in counterbalanced order, with a 5-minute break between rounds:

1. Two-headed FC with heads at 13 cm apart (greater than average IPD)
2. Two-headed FC with heads at 8 cm apart (approximately average IPD)
3. Two-headed FC with heads at 4 cm apart (smaller than average IPD)
4. Single-headed MOL-D
5. Single-headed MOL-A
6. Single-headed FC

Participants also completed a NASA-TLX survey after each condition.

CFF Measurements. Measurements obtained in two-headed conditions were on average higher than all single-headed conditions. While this finding could be perceived as getting closer to a person's "true CFF", it may also be confounded by differences in CFF depending on viewing angle. Biologically, human CFF will be higher at their peripheral vision. Future work will need to investigate ways to mitigate this confound. Nonetheless, it is noteworthy that test-retest was most consistent specifically with the two-headed FC condition with heads at 13 cm apart.

Timing. All two-headed conditions were completed significantly more quickly than the single-headed FC condition. TODO: timing for MOL-D and MOL-A? TODO: get means for timing?

Task load. Participant responses to the NASA-TLX revealed that the single-headed FC required significantly ($p < 0.01$) greater mental demand and effort than single-headed MOL-D, single-headed MOL-A, and specifically the two-headed FC with heads at 13 cm apart. The mean score for mental demand and effort for single-headed FC was higher than the other two-headed conditions, but without significance. Participants reported finding all conditions to be similarly (not) frustrating.

Reducing Measurement Burden

Using the longitudinal dataset collected in our at-home deployment, we are interested in developing and validating algorithms to address two needs: (1) “What frequencies to sample during FC measurement?”, and (2) “When should a person take a measurement?”. Our preliminary literature search shows possible approaches in level set estimation [73] and in active learning [21]. We anticipate that frequency sampling informed by an individual’s recent measurements can preserve reliability while reducing FC median measurement time to less than 2 minutes. Analyses of trends can then also inform when a measurement should be taken, offering further potential for reducing burden.

External Users of Beacon

Since our American Journal of Gastroenterology (AJG) publication [121], we have had multiple inbound requests to use Beacon in their own studies and clinics. Jasmohan Bajaj is a professor at at Virginia Commonwealth University and a leader in hepatic encephalopathy (i.e., as editor-in-chief of the AJG) We have already shared a Beacon device with his team in a study to compare Beacon’s CFF measurements to that of other experimental conditions, including CFF via the Hepatonorm and a series of psychometric tests (i.e., the Stroop test). Gerald Kircheis is another expert in hepatic encephalopathy and was previously affiliated with the company that developed the Hepatonorm

Analyzer, another device for administering CFF measurements. He has also reached out regarding evaluating Beacon in his own research and clinic. Towards supporting external researchers, we have iterated on the Beacon device (i.e., introduced a charging port) and plan to also make additional refinements to the software (i.e., develop a more robust provider dashboard).

Pilot Study in Dementia Patients

We have also begun to investigate the potential utility of Beacon in populations outside of chronic liver disease patients. Informed by results showing differences in CFF for patients with Alzheimer's disease (e.g., [27, 1]), we have successfully begun including Beacon in a clinical study being conducted by Debby Tsuang, a dementia expert at the VA and UW, to examine trends in CFF measurements in patients who have early-stage dementia or are being evaluated for dementia. In this study, Beacon measures are being collected alongside other clinical measures of cognitive function at two different sessions 6-months apart. This collaboration is timely given that a number of pharmacotherapies (e.g., Kisunla) for early Alzheimer's disease have been recently approved by the FDA. Pending multiple grants, we will be able to launch an at-home deployment with this patient population similar to our deployment with chronic liver disease patients.

6.1.2 ExerciseRx

In this section, I will describe on-going and future work for the ExerciseRx project from technical, programmatic, and clinical efficacy standpoints.

Holistic Understanding of Physical Activity

ExerciseRx began as an app for monitoring step count, and delivered an intervention in which users that reached their target step count for a given week would have their goal increased by

15% the following week. In Chapter 4, I reported specifically on the technical implementation for facilitating home exercise sessions and patient experiences with using this functionality in home settings. Both of these features however capture only a portion of a person's physical activity levels. It is important to note that in the spirit of promoting exercise, it is most effective for systems to monitor and reward all forms of physical activity. While this work began as a way to leverage commodity smartphones to monitor physical activity, future work might explore the use of wearable devices for passively tracking exercises. In this regard, we have begun to deploy Fitbits as part of the bladder cancer study as a way to increase the accuracy of tracked step count as opposed to via smartphones. We will explore and investigate implementations towards enabling tracking exercises using wearables in the future.

Pilot Study in Axial Spondyloarthritis

We have begun to collaborate with Rachael Stovall, acting assistant professor in the Division of Rheumatology at the UW Medicine on a pilot study to investigate the efficacy of ExerciseRx in a rheumatology setting. This context specifically is interesting because rheumatologists work with conditions related to inflammation in the bones and joints, but they themselves are not experts in physical rehabilitation or sports medicine, instead referring a patient to a physical therapist as needed. However, because there is rarely a physical therapist immediately available, patient condition is often left untreated and sometimes to deteriorate while waiting for the next available physical therapy appointment. Our hypothesis is that ExerciseRx can be used to fill this gap and help patients start to practice specific helpful movements before receiving direct care from a physical therapist. Towards this end, we have thus begun a pilot study in evaluating the accuracy of ExerciseRx's machine learning models in a small population of rheumatology patients (i.e., those with axial spondyloarthritis). Multiple grants have been submitted to continue this line of work, particularly through a longitudinal deployment to study the utility of the ExerciseRx platform in bridging the gap between rheumatology and physical therapy.

Integration into Electronic Health Record Platforms

Towards adoption of ExerciseRx in clinical practice, we have begun to collaborate with the IT department at UW Medicine in an effort to integrate our physical activity tracking into their EPIC electronic health record platform, enabling healthcare providers to view physical activity progress alongside all of their other clinical measurements such as vital signs. The ExerciseRx provider dashboard is currently integrated into EPIC as an embedded <iframe>, enabling providers to view data through an isolated window. We are now working on deeper integration by appending ExerciseRx's data directly into patient health records. In success, this coupling will enable to not only display data to providers, but also enable our algorithms to pull other aspects of a patient's health records, potentially to generate more personalized recommendations for exercise.

6.1.3 PARX

The PARX project is distinct from Beacon and ExerciseRx because it is provider-facing, as opposed to patient-facing. It is also relatively new, primarily completed in the last year. In this section, I will enumerate a couple of directions for the continued co-design of the PARX system between the technical improvements that can be made and the clinical opportunities and usage patterns of providers.

Adopting Advances in Artificial Intelligence

PARX currently uses two pre-trained large models (i.e., large language model and language embedding model) in tandem with a custom trained transformer encoder. This straight-forward system architecture is designed to support fast inference and meet the time constraint requirements of clinical use. Recent advances in machine learning research have made great progress in accelerating both the inference performance of transformer models [41] and also more powerful architectures such as diffusion-based models [78]. Future work includes investigating how some of these approaches to improving inference time could be applied to PARX.

While PARX is currently implemented as a proof-of-concept that mostly generates reasonable results, they have a lot of potential for improvement. In particular, the interpolation step for sequencing individual frames of poses into a motion does not return particularly naturalistic or realistic motions. Future work also includes investigating alternatives to the interpolation step, or potentially end-to-end modeling of text the text-to-motion pipeline with a single model. Generated movements with improved accuracy can also lead to greater utility as synthetic data for training exercise recognition models (i.e., the repetition counting models used in the ExerciseRx app).

Evaluation in Clinical Settings

Preliminary feedback from providers on the potential clinical utility of PARX has been encouraging. We are thus interested in studying how it might be used in practice. Towards this end, we are currently designing a study that mimics a medical appointment between patient and provider. In this study, the provider will be asked to generate a customized exercise routine for the specific patient. Demonstrating the efficacy of this workflow can greatly improve the potential for adoption and scaling of ExerciseRx in clinical practice.

6.2 Challenges and Lessons Learned

In this section, I will discuss some of the common challenges faced across both Beacon and ExerciseRx, and also report on some lessons learned in the process of resolving these challenges.

6.2.1 The Chicken or the Egg in Novel Health Data

In projects that involve developing a new system for producing an established health measure, demonstrating correspondence between the system and a gold standard device is sufficient for making the logical leap that the new system can be used to deliver the same interventions as that of the gold standard device. The projects presented in this dissertation however leverage novel health data to inform medical interventions. We thus needed to produce evidence (i.e., via data collection) to demonstrate that the novel health measure does carry utility. In order to ethically engage patients in collecting and contributing their own data towards building this evidence base, we needed to be transparent with patients that they are participating in a research study in which the hypothesis may fail. This seemingly lack of confidence however can be demotivating to a patient deciding on engaging in a research study in addition to managing their own health condition. We thus recommend that, for readers developing their own novel health data system, extra care should be taken in how research studies are advertised and introduced to patients. Dedicated research coordinators can be particularly helpful for facilitating this interaction.

6.2.2 Designing for Specificity versus Generality

In The Universal Design File [115], Story et al. present a guide to the concept of designing products and environments to be usable by the most number of people possible. This notion is nice in theory, because it is inclusive and removes the need to backtrack and redesign components, but it is challenging in practice: one needs the experience and foresight to design something that is truly completely generalizable. Consistent with Norman's argument that *"there is no such thing as the average person"* [93] and Buxton's claim that *"we must diverge towards a set of simpler, more specialized tools"* [20], my dissertation projects have followed the approach of designing for a specific group of people first before expanding its scope. In Beacon, we spent 8 years designing for and validating on chronic liver disease patients before we felt sufficiently confident to start exploring opportunities for it to be used in a different population (i.e., dementia). In ExerciseRx, we also began by designing for a specific population (i.e., teenagers with cerebral palsy) before iterating on the platform to expand its scope to be applicable for older bladder cancer patients. In both of these cases, the iterative steps to go from specific to general enable increasing scope while also developing confidence in the robustness of the core functionality.

6.2.3 Collaborations between Engineering and Medicine

The projects presented in this dissertation have been inherently multidisciplinary, with collaborators in Computer Science & Engineering, Human Centered Design & Engineering, and Medicine. Although we are unified by a common vision of doing good in the world, broadly in the scope of the respective projects, all of the stakeholders have their own individual goals and values. Most of these goals involve research, but healthcare providers for example may also have values in providing patient care. Even within research, what constitutes a research contribution is different across different communities. Multidisciplinary collaboration thus requires an understanding from all stakeholders that not all activities will directly lead to a research contribution in one's own research field. Establishing this understanding and being transparent about goals at the beginning of collaborations can help mitigate conflict that might arise later.

6.2.4 Challenges in Translation

Health research, particularly in computing, is motivated by its importance and potential impact on society. Frameworks such as the Bench-to-Bedside Translational Model [34] document possible routes to go from basic science to societal impact. However, we have found that our own experience is not precisely documented by any of the established frameworks. A main point of deviation is the iterative nature of the process. Many frameworks describe a linear process from basic science to translation in humans to translation in patients to implementation in practice. The HCI design process on the other hand explicitly includes refinement as a part of the process. We, as HCI researchers, thus understand that many rounds of "translation in humans" are necessary before a product can reach its full potential in real-world settings. Although neither Beacon nor ExerciseRx have made societal impact yet, multiple deployments in these projects have provided encouraging evidence to the potential efficacy of these systems in real-world practice.

6.3 Conclusion

In this dissertation, I presented my thesis statement:

Designing and deploying novel personal health technologies for the challenges of at-home longitudinal use can enable and inform new data and models of care for patients, caregivers, and providers.

I then introduced three projects to demonstrate this thesis statement: (1) In Beacon, I discussed how technical iterations to the system were necessary to validate the CFF measurements produced by Beacon in clinical and home settings with chronic liver disease patients. (2) In ExerciseRx, I described the development of an exercise monitoring and feedback system and findings from deploying this system into patient homes. (3) In PARX, I presented the design and implementation of a system that enables providers to generate and edit human motion animations in support of tailoring exercise instructions to individual needs. Finally, I described on-going and future work for each of these projects, and discussed some of the challenges and lessons learned common across all of these projects. This work has shown great promise in translation to clinical adoption and making real world impact, but more importantly, I believe that this work broadly represents exciting opportunities in the design of novel at-home health technologies. I hope readers can leverage some of the insights presented in this dissertation towards supporting a healthier future.

Bibliography

[1] Azar Abiyev, Funda Datlı Yakaryılmaz, and Zeynel Abidin Öztürk. “A new diagnostic approach in Alzheimer’s disease: The critical flicker fusion threshold”. In: *Dementia & Neuropsychologia* 16.1 (2022), pp. 89–96.

[2] Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. “Vid2doppler: Synthesizing doppler radar data from videos for training privacy-preserving activity recognition”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–10.

[3] Javier Ampuero, Macarena Simón, Carmina Montoliú, Rodrigo Jover, Miguel Ángel Serra, Juan Córdoba, and Manuel Romero-Gómez. “Minimal hepatic encephalopathy and critical flicker frequency are associated with survival of patients with cirrhosis”. In: *Gastroenterology* 149.6 (2015), pp. 1483–1489.

[4] John P Ansah and Chi-Tsun Chiu. “Projecting the chronic disease burden among the adult population in the United States using a multi-state population model”. In: *Frontiers in public health* 10 (2023), p. 1082183.

[5] Jasmohan S Bajaj. “Diagnosing minimal hepatic encephalopathy: from the ivory tower to the real world”. In: *Gastroenterology* 149.6 (2015), pp. 1330–1333.

[6] Jasmohan Singh Bajaj, Ashkan Etemadian, Muhammad Hafeezullah, and Kia Saeian. “Testing for minimal hepatic encephalopathy in the United States: an AASLD survey”. In: *Hepatology* 45.3 (2007), pp. 833–834.

[7] Michele Barone, Endrit Shahini, Andrea Iannone, Maria Teresa Viggiani, Valeria Corvace, Mariabeatrice Principi, and Alfredo Di Leo. “Critical flicker frequency test predicts overt hepatic encephalopathy and survival in patients with liver cirrhosis”. In: *Digestive and Liver Disease* 50.5 (2018), pp. 496–500.

[8] Emily Bascom, Reggie Casanova-Perez, Kelly Tobar, Manas Satish Bedmutha, Harshini Ramaswamy, Wanda Pratt, Janice Sabin, Brian Wood, Nadir Weibel, and Andrea Hartzler. “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*. 2024, pp. 1–12.

[9] S Frances Bassett. "The assessment of patient adherence to physiotherapy rehabilitation". In: *NZ J Physiother* 31.2 (2003), pp. 60–66.

[10] V Bazarevsky. "BlazePose: On-device Real-time Body Pose tracking". In: *arXiv preprint arXiv:2006.10204* (2020).

[11] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. "EarBit: using wearable sensors to detect eating episodes in unconstrained environments". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3 (2017), pp. 1–20.

[12] Muhammad S Beg, Arjun Gupta, Tyler Stewart, and Chad D Rethorst. "Promise of wearable physical activity monitors in oncology practice". In: *Journal of oncology practice* 13.2 (2017), pp. 82–89.

[13] Andrew BL Berry, Catherine Y Lim, Tad Hirsch, Andrea L Hartzler, Linda M Kiel, Zoë A Bermet, and James D Ralston. "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*. 2019, pp. 1–14.

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

[15] Manon Bloemen, Leontien Van Wely, Jurgen Mollema, Annet Dallmeijer, and Janke de Groot. "Evidence for increasing physical activity in children with physical disabilities: a systematic review". In: *Developmental Medicine & Child Neurology* 59.10 (2017), pp. 1004–1010.

[16] Joseph Breda, Mastafa Springston, Alex Mariakakis, and Shwetak Patel. "Feverphone: Accessible core-body temperature sensing for fever monitoring using commodity smartphones". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7.1 (2023), pp. 1–23.

[17] Logan G Briggs, Chanan Reitblat, Paul A Bain, Sara Parke, Ny-Ying Lam, Jonathan Wright, James WF Catto, Robert J Copeland, and Sarah P Psutka. "Prehabilitation exercise before urologic cancer surgery: a systematic and interdisciplinary review". In: *European Urology* 81.2 (2022), pp. 157–167.

[18] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[19] Andrea Burgess, Sarah Reedman, Mark D Chatfield, Robert S Ware, Leanne Sakzewski, and Roslyn N Boyd. "Development of gross motor capacity and mobility performance in children with cerebral palsy: a longitudinal study". In: *Developmental Medicine & Child Neurology* 64.5 (2022), pp. 578–585.

[20] William Buxton. "Less is more (more is less)". In: *The invisible future: the seamless integration of technology into everyday life*. 2001, pp. 145–179.

[21] Romain Camilleri, Andrew Wagenmaker, Jamie Morgenstern, Lalit Jain, and Kevin Jamieson. "Fair active learning in low-data regimes". In: *arXiv preprint arXiv:2312.08559* (2023).

[22] Francesco P Cappuccio, Sally M Kerry, Lindsay Forbes, and Anna Donald. "Blood pressure control by home monitoring: meta-analysis of randomised trials". In: *Bmj* 329.7458 (2004), p. 145.

[23] Yoon Jeong Cha, Alice Wou, Arpita Saxena, Joyce Lee, Mark W Newman, and Sun Young Park. "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*. 2023, pp. 1–15.

[24] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. "Understanding quantified-selfers' practices in collecting and exploring personal data". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, pp. 1143–1152.

[25] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A Kientz, and Sean A Munson. "Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data". In: *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 2016, pp. 770–786.

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

[27] Alice Cronin-Golomb, Suzanne Corkin, Joseph F Rizzo, Jennifer Cohen, John H Growdon, and Kathleen S Banks. "Visual dysfunction in Alzheimer's disease: relation to normal aging". In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 29.1 (1991), pp. 41–52.

[28] Lilian De Greef, Mayank Goel, Min Joon Seo, Eric C Larson, James W Stout, James A Taylor, and Shwetak N Patel. "Bilicam: using mobile phones to monitor newborn jaundice". In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2014, pp. 331–342.

[29] Ginger Delmas, Philippe Weinzaepfel, Thomas Lucas, Francesc Moreno-Noguer, and Grégory Rogez. "Posescript: 3d human poses from natural language". In: *European Conference on Computer Vision*. Springer. 2022, pp. 346–362.

[30] Ginger Delmas, Philippe Weinzaepfel, Francesc Moreno-Noguer, and Grégory Rogez. "Posefix: correcting 3D human poses with natural language". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 15018–15028.

[31] Artem Dementyev and Christian Holz. "DualBlink: a wearable device to continuously detect, track, and actuate blinking for alleviating dry eyes and computer vision syndrome". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.1 (2017), pp. 1–19.

[32] Steven Díaz, Jeannie B Stephenson, and Miguel A Labrador. "Use of wearable sensor technology in gait, balance, and range of motion analysis". In: *Applied Sciences* 10.1 (2019), p. 234.

[33] Helen Dolk, Sam Pattenden, and Ann Johnson. "Cerebral palsy, low birthweight and socio-economic deprivation: inequalities in a major cause of childhood disability". In: *Paediatric and perinatal epidemiology* 15.4 (2001), pp. 359–363.

[34] Brian C Drolet and Nancy M Lorenzi. "Translational research: understanding the continuum from bench to bedside". In: *Translational Research* 157.1 (2011), pp. 1–5.

[35] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. "The llama 3 herd of models". In: *arXiv preprint arXiv:2407.21783* (2024).

[36] Auria Eisen-Enosh, Nairouz Farah, Zvia Burgansky-Eliash, Uri Polat, and Yossi Mandel. "Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution". In: *Scientific reports* 7.1 (2017), p. 15621.

[37] James Elsey. "US Healthcare System Is in Crisis". In: *American College of Surgeons* (2025).

[38] Daniel A Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A Munson. "Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use". In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 1109–1113.

[39] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. "A lived informatics model of personal informatics". In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 2015, pp. 731–742.

[40] Andrea Ferlini, Alessandro Montanari, Chulhong Min, Hongwei Li, Ugo Sassi, and Fahim Kawsar. "In-ear PPG for vital signs". In: *IEEE Pervasive Computing* 21.1 (2021), pp. 65–74.

[41] Quentin Fournier, Gaétan Marceau Caron, and Daniel Aloise. "A practical survey on faster and lighter transformers". In: *ACM Computing Surveys* 55.14s (2023), pp. 1–40.

[42] John Geyman. "COVID-19 has revealed America's broken health care system: what can we learn?" In: *International Journal of Health Services* 51.2 (2021), pp. 188–194.

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

[44] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. “Humans in 4D: Reconstructing and tracking humans with transformers”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 14783–14794.

[45] Beverly B Green, Melissa L Anderson, Andrea J Cook, Kelly Ehrlich, Yoshio N Hall, Clarissa Hsu, Dwayne Joseph, Predrag Klasnja, Karen L Margolis, Jennifer B McClure, et al. “Clinic, home, and kiosk blood pressure measurements for diagnosing hypertension: a randomized diagnostic study”. In: *Journal of General Internal Medicine* 37.12 (2022), pp. 2948–2956.

[46] Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. “Generating diverse and natural 3d human motions from text”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 5152–5161.

[47] Dana Habeeb, James Clawson, Arash Zakeresfahani, and Zebulon Holtz. “Investigating and Validating On-body Temperature Sensors for Personal Heat Exposure Tracking.” In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–14.

[48] Christian Holz and Edward J Wang. “Glabella: Continuously sensing blood pressure behavior using an unobtrusive wearable device”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3 (2017), pp. 1–23.

[49] Matthew K Hong, Udaya Lakshmi, Kimberly Do, Sampath Prahalad, Thomas Olson, Rosa I Arriaga, and Lauren Wilcox. “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*. 2020, pp. 1–16.

[50] Long-Jing Hsu and Chia-Fang Chung. “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*. 2024, pp. 1–12.

[51] Yiming Huang, Weilin Wan, Yue Yang, Chris Callison-Burch, Mark Yatskar, and Lingjie Liu. “Como: Controllable motion generation through language guided pose code editing”. In: *European Conference on Computer Vision*. Springer. 2025, pp. 180–196.

[52] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J Black, Otmar Hilliges, and Gerard Pons-Moll. “Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time”. In: *ACM Transactions on Graphics (TOG)* 37.6 (2018), pp. 1–15.

[53] Hanna Hunter, Nicole Bennington-McKay, Jessica Sher, Sarah P Psutka, and Cindy Lin. “Emerging role of mobile applications and wearable devices for prehabilitation in urologic oncology”. In: *European urology focus* 10.1 (2024), pp. 20–22.

[54] Hanna Hunter, Cindy Y Lin, Richard Li, Otari Ioseliani, Leah Cantor, Elena G Brewer, Samia Jannat, Karla Landis, David Bridges, Sean A Munson, and Sarah P Psutka. *The "Get Moving Trial": A phase I/II RCT of home-based (P) rehabilitation ((P) REHAB) with ExerciseRx in muscle-invasive bladder cancer (MIBC)—Study protocol for a randomized controlled trial.* 2024.

[55] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. "Technology probes: inspiring design for and with families". In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* 2003, pp. 17–24.

[56] Yuxin Jiao, Rose Lin, Xinyang Hua, Leonid Churilov, Michele J Gaca, Steven James, Philip M Clarke, David O'Neal, and Elif I Ekinci. "A systematic review: Cost-effectiveness of continuous glucose monitoring compared to self-monitoring of blood glucose in type 1 diabetes". In: *Endocrinology, diabetes & metabolism* 5.6 (2022), e369.

[57] Nina Kahlbrock, Markus Butz, Elisabeth S May, Meike Brenner, Gerald Kircheis, Dieter Häussinger, and Alfons Schnitzler. "Lowered frequency and impaired modulation of gamma band oscillations in a bimodal attention task are associated with reduced critical flicker frequency". In: *Neuroimage* 61.1 (2012), pp. 216–227.

[58] Ravi Karkar, Rafal Kocielnik, Xiaoyi Zhang, Jasmine Zia, George N Ioannou, Sean A Munson, and James Fogarty. "Beacon: Designing a portable device for self-administering a measure of critical flicker frequency". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (2018), pp. 1–27.

[59] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. "Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers". In: *Proceedings of the 2017 CHI conference on human factors in computing systems.* 2017, pp. 6850–6863.

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

[61] Matthew Kay, Dan Morris, MC Schraefel, and Julie A Kientz. "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.* 2013, pp. 401–410.

[62] Logan Kendall, Dan Morris, and Desney Tan. "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.* 2015, pp. 1679–1688.

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

[64] Seong Woo Kim, Ha Ra Jeon, Taemi Youk, and Jiyong Kim. "Cost of rehabilitation treatment of patients with cerebral palsy in Korea". In: *Annals of Rehabilitation Medicine* 42.5 (2018), p. 722.

[65] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. "OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3 (2017), pp. 1–28.

[66] Gerald Kircheis, Matthias Wettstein, Lars Timmermann, Alfons Schnitzler, and Dieter Häussinger. "Critical flicker frequency for quantification of low-grade hepatic encephalopathy". In: *Hepatology* 35.2 (2002), pp. 357–366.

[67] Predrag Klasnja and Wanda Pratt. "Healthcare in the pocket: mapping the space of mobile-phone health interventions". In: *Journal of biomedical informatics* 45.1 (2012), pp. 184–198.

[68] Jeffrey Konrad, Natasha Marrus, and Catherine E Lang. "A Feasibility Study of Bilateral Wrist Sensors for Measuring Motor Traits in Children With Autism". In: *Perceptual and motor skills* 129.6 (2022), pp. 1709–1735.

[69] Lampros C Kourtis, Oliver B Regele, Justin M Wright, and Graham B Jones. "Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity". In: *NPJ digital medicine* 2.1 (2019), p. 9.

[70] Naveen Krishnan, Benjamin Li, Bruce L Jacobs, Sapan N Ambani, Tudor Borza, Chang He, Brent K Hollenbeck, Todd Morgan, Khaled S Hafez, Alon Z Weizer, et al. "The fate of radical cystectomy patients after hospital discharge: understanding the black box of the pre-readmission interval". In: *European Urology Focus* 4.5 (2018), pp. 711–717.

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

[72] Connie Leibow, Richard Li, Deborah Grenard, Otari Ioseliani, Tiffany Li, Leah Cantor, David Bridges, Karla Landis, Sean A Munson, and Cindy Y Lin. *ExerciseRx-CP: At-Home Sensor-Driven Platform for Exercise in Adolescents with Cerebral Palsy.* 2024.

[73] Benjamin Letham, Phillip Guan, Chase Tymms, Eytan Bakshy, and Michael Shvartsman. "Look-Ahead Acquisition Functions for Bernoulli Level Set Estimation". In: *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*. Ed. by Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera. Vol. 151. Proceedings of Machine Learning Research. PMLR, 28–30 Mar 2022, pp. 8493–8513. URL: https://proceedings.mlr.press/v151/letham22a.html.

[74] Richard Li, Ravi Karkar, Philip Vutien, George N. Ioannou, Sean A Munson, and James Fogarty. "Developing and Validating Beacon: A Portable Device for Self-Administering a Measure of Critical Flicker Frequency". In: *2023 Workgroup on Interactive Systems in Healthcare*. 2023.

[75] Richard Li, Philip Vutien, Sabrina Omer, Michael Yacoub, George Ioannou, Ravi Karkar, Sean A Munson, and James Fogarty. "Deploying and Examining Beacon for At-Home Patient Self-Monitoring with Critical Flicker Frequency". In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 2025, pp. 1–17.

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

[77] Cindy Y Lin, Nicole L Gentile, Levi Bale, Melanie Rice, E Sally Lee, Lisa S Ray, and Marcia A Ciol. "Implementation of a physical activity vital sign in primary care: Associations between physical activity, demographic characteristics, and chronic disease burden". In: *Preventing Chronic Disease* 19 (2022), E33.

[78] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al. "Instaflow: One step is enough for high-quality diffusion-based text-to-image generation". In: *The Twelfth International Conference on Learning Representations*. 2023.

[79] Felipe Lobelo, Deborah Rohm Young, Robert Sallis, Michael D Garber, Sandra A Billinger, John Duperly, Adrian Hutber, Russell R Pate, Randal J Thomas, Michael E Widlansky, et al. "Routine assessment and promotion of physical activity in healthcare settings: a scientific statement from the American Heart Association". In: *Circulation* 137.18 (2018), e495–e522.

[80] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. "SMPL: A Skinned Multi-Person Linear Model". In: *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34.6 (Oct. 2015), 248:1–248:16.

[81] Lars Lund, Jacob Jacobsen, Peter Clark, Michael Borre, Mette Nørgaard, Northern Danish Cancer Quality Assessment Group, et al. "Impact of comorbidity on survival of invasive bladder cancer patients, 1996-2007: a Danish population-based cohort study". In: *Urology* 75.2 (2010), pp. 393–398.

[82] Carol Ann Maher, Monica Toohey, and Monika Ferguson. "Physical activity predicts quality of life and happiness in children and adolescents with cerebral palsy". In: *Disability and rehabilitation* 38.9 (2016), pp. 865–869.

[83] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. "AMASS: Archive of Motion Capture as Surface Shapes". In: *International Conference on Computer Vision*. Oct. 2019, pp. 5442–5451.

[84] Salman Mallick, Radhika Kanthety, and Mahboob Rahman. "Home blood pressure monitoring in clinical practice: a review". In: *The American journal of medicine* 122.9 (2009), pp. 803–810.

[85] Alex Mariakakis, Megan A Banks, Lauren Phillipi, Lei Yu, James Taylor, and Shwetak N Patel. "Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.2 (2017), pp. 1–26.

[86] Jackson McCue, Deborah Grenard, Tiffany Li, Chloe Reibold, Connie Leibow, and Cindy Y Lin. *Experience of Adolescents with Cerebral Palsy with the ExerciseRx App: Qualitative Feedback Regarding App Usability for Home Based Exercises*. 2024.

[87] Brendan McIntosh, Changhua Yu, Avtar Lal, Kristen Chelak, Chris Cameron, Sumeet R Singh, and Marshall Dahl. "Efficacy of self-monitoring of blood glucose in patients with type 2 diabetes mellitus managed without insulin: a systematic review and meta-analysis". In: *Open Medicine* 4.2 (2010), e102.

[88] Enrico Maria Minnella, Rashami Awasthi, Guillaume Bousquet-Dion, Vanessa Ferreira, Berson Austin, Christine Audi, Simon Tanguay, Armen Aprikian, Francesco Carli, and Wassim Kassouf. "Multimodal prehabilitation to enhance functional capacity following radical cystectomy: a randomized controlled trial". In: *European urology focus* 7.1 (2021), pp. 132–138.

[89] Pardis Miri, Mehul Arora, Aman Malhotra, Robert Flory, Stephanie Hu, Ashley Lowber, Ishan Goyal, Jacqueline Nguyen, John P Hegarty, Marlo D Kohn, et al. "FAR: End-to-End Vibrotactile Distributed System Designed to Facilitate Affect Regulation in Children Diagnosed with Autism Spectrum Disorder Through Slow Breathing". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–20.

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

[91] Elizabeth D Mynatt, Jim Rowan, Sarah Craighill, and Annie Jacobs. "Digital family portraits: supporting peace of mind for extended family members". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2001, pp. 333–340.

[92] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. "Contactless sleep apnea detection on smartphones". In: *Proceedings of the 13th annual international conference on mobile systems, applications, and services*. 2015, pp. 45–57.

[93] Donald Norman. *The Design of Everyday Things*. Hachette Book Group, 2013.

[94] Mikio Obuchi, Jeremy F Huckins, Weichen Wang, Alex Dasilva, Courtney Rogers, Eilis Murphy, Elin Hedlund, Paul Holtzheimer, Shayan Mirjafari, and Andrew Campbell. "Predicting brain functional connectivity using mobile sensing". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4.1 (2020), pp. 1–22.

[95] Ollama. *Ollama.* URL: https://ollama.com/.

[96] Robert Palisano, Peter Rosenbaum, Stephen Walter, Dianne Russell, Ellen Wood, and Barbara Galuppi. "Development and reliability of a system to classify gross motor function in children with cerebral palsy". In: *Developmental medicine & child neurology* 39.4 (1997), pp. 214–223.

[97] Una C Peplow and Christine Carpenter. "Perceptions of parents of children with cerebral palsy about the relevance of, and adherence to, exercise programs: a qualitative study". In: *Physical & occupational therapy in pediatrics* 33.3 (2013), pp. 285–299.

[98] Laura R Pina, Sang-Wha Sien, Teresa Ward, Jason C Yip, Sean A Munson, James Fogarty, and Julie A Kientz. "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*. 2017, pp. 2300–2315.

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

[100] Andrea Porserud, Amir Sherif, and Anna Tollbäck. "The effects of a physical exercise programme after radical cystectomy for urinary bladder cancer. A pilot randomized controlled trial". In: *Clinical rehabilitation* 28.5 (2014), pp. 451–459.

[101] James M Rehg, Agata Rozga, Gregory D Abowd, and Matthew S Goodwin. "Behavioral imaging and autism". In: *IEEE Pervasive Computing* 13.2 (2014), pp. 84–87.

[102] Manuel Romero-Gómez. "Critical flicker frequency: it is time to break down barriers surrounding minimal hepatic encephalopathy". In: *Journal of hepatology* 47.1 (2007), pp. 10–11.

[103] Manuel Romero-Gómez, Juan Córdoba, Rodrigo Jover, Juan A Del Olmo, Marta Ramírez, Ramón Rey, Enrique De Madaria, Carmina Montoliu, David Nuñez, Montse Flavia, et al. "Value of the critical flicker frequency in patients with minimal hepatic encephalopathy". In: *Hepatology* 45.4 (2007), pp. 879–885.

[104] Parker S Ruth, Scott D Uhlrich, Constance de Monts, Antoine Falisse, Julie Muccini, Sydney Covitz, Shelby Vogt-Domke, John Day, Tina Duong, and Scott L Delp. "Video-based biomechanical analysis captures disease-specific movement signatures of different neuromuscular diseases". In: *bioRxiv* (2024), pp. 2024–09.

[105] Justin Saluja, Joaquin Casanova, and Jenshan Lin. “A supervised machine learning algorithm for heart-rate detection using Doppler motion-sensing radar”. In: *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* 4.1 (2019), pp. 45–51.

[106] V Sanh. “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter”. In: *arXiv preprint arXiv:1910.01108* (2019).

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

[108] Woosuk Seo, Ayse G Buyuktur, Sanya Verma, Hyeryoung Kim, Sung Won Choi, Laura Sedig, and Sun Young Park. “Learning from healthcare providers’ strategies: Designing technology to support effective child patient-provider communication”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 2021, pp. 1–15.

[109] P Sharma, BC Sharma, V Puri, and SK Sarin. “Critical flicker frequency: diagnostic tool for minimal hepatic encephalopathy”. In: *Journal of hepatology* 47.1 (2007), pp. 67–73.

[110] Praveen Sharma, Barjesh Chander Sharma, and Shiv Kumar Sarin. “Critical flicker frequency for diagnosis and assessment of recovery from minimal hepatic encephalopathy in patients with cirrhosis.” In: *Hepatobiliary & pancreatic diseases international: HBPD INT* 9.1 (2010), pp. 27–32.

[111] Ken Shoemake. “Animating rotation with quaternion curves”. In: *Proceedings of the 12th annual conference on Computer graphics and interactive techniques.* 1985, pp. 245–254.

[112] Emily M Simek, Lucy McPhate, and Terry P Haines. “Adherence to and efficacy of home exercise programs to prevent falls: a systematic review and meta-analysis of the impact of exercise program characteristics”. In: *Preventive medicine* 55.4 (2012), pp. 262–275.

[113] Ruth Gaelle St Fleur, Sara Mijares St George, Rafael Leite, Marissa Kobayashi, Yaray Agosto, and Danielle E Jake-Schoffman. “Use of Fitbit devices in physical activity intervention studies across the life course: narrative review”. In: *JMIR mHealth and uHealth* 9.5 (2021), e23411.

[114] George S Stergiou and Ioannis A Bliziotis. “Home blood pressure monitoring in the diagnosis and treatment of hypertension: a systematic review”. In: *American journal of hypertension* 24.2 (2011), pp. 123–134.

[115] Molly Follette Story, James L Mueller, and Ronald L Mace. “The universal design file: Designing for people of all ages and abilities”. In: (1998).

[116] Nicholas F Taylor, Karen J Dodd, Helen McBurney, and H Kerr Graham. “Factors influencing adherence to a home-based strength-training programme for young people with cerebral palsy”. In: *Physiotherapy* 90.2 (2004), pp. 57–63.

[117] FJ Torlot, MJW McPhail, and SD Taylor-Robinson. “Meta-analysis: the diagnostic accuracy of critical flicker frequency in minimal hepatic encephalopathy”. In: *Alimentary pharmacology & therapeutics* 37.5 (2013), pp. 527–536.

[118] Olaf Verschuren, Mark D Peterson, Astrid CJ Balemans, and Edward A Hurvitz. “Exercise and physical activity recommendations for people with cerebral palsy”. In: *Developmental Medicine & Child Neurology* 58.8 (2016), pp. 798–808.

[119] Hendrik Vilstrup, Piero Amodio, Jasmohan Bajaj, Juan Cordoba, Peter Ferenci, Kevin D Mullen, Karin Weissenborn, and Philip Wong. “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”. In: *Hepatology* 60.2 (2014), pp. 715–735.

[120] Philip Vutien, Ravi Karkar, Richard Li, Kara Walter, Sean Munson, James Fogarty, and George Ioannou. “Evaluating a novel, portable, self-administered device (“flicker-app”) that measures critical flicker frequency as a test for hepatic encephalopathy in patients with cirrhosis”. In: *The Liver Meeting Digital Experience™*. AASLD. 2020.

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

[122] Anandghan Waghmare, Farshid Salemi Parizi, Jason Hoffman, Yuntao Wang, Matthew Thompson, and Shwetak Patel. “GlucoScreen: A Smartphone-based Readerless Glucose Test Strip for Prediabetes Screening”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7.1 (2023), pp. 1–20.

[123] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. “Tracking depression dynamics in college students using mobile phone and wearable sensing”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–26.

[124] Shin-Yi Wu and Anthony Green. “Projection of chronic illness prevalence and cost inflation”. In: *Santa Monica, CA: RAND Health* 18 (2000).

[125] Chengshuo Xia, Xinrui Fang, Riku Arakawa, and Yuta Sugiura. “VoLearn: A cross-modal operable motion-learning system combined with virtual avatar and auditory feedback”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.2 (2022), pp. 1–26.

[126] Kefan Xu, Xinghui Yan, and Mark W Newman. “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*. 2022, pp. 1–15.

[127] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. “GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.4 (2023), pp. 1–34.

[128] Qiuyue Shirley Xue, Yujia Liu, Joseph Breda, Mastafa Springston, Vikram Iyer, and Shwetak Patel. “Thermal Earring: Low-power Wireless Earring for Longitudinal Earlobe Temperature Sensing”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7.4 (2024), pp. 1–28.

[129] Steven A Yarows, Stevo Julius, and Thomas G Pickering. “Home blood pressure monitoring”. In: *Archives of Internal Medicine* 160.9 (2000), pp. 1251–1257.

[130] HD Zacharias, CD Jackson, MY Morgan, and SS Olesen. “stepwise diagnosis in covert hepatic encephalopathy-critical flicker frequency and MELD-score as a first-step approach. Replication and pitfalls”. In: *Alimentary Pharmacology and Therapeutics* 45.1 (2017), pp. 187–189.

[131] Youwei Zeng, Dan Wu, Jie Xiong, Jinyi Liu, Zhaopeng Liu, and Daqing Zhang. “MultiSense: Enabling multi-person respiration sensing with commodity wifi”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020), pp. 1–29.

[132] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Yong Zhang, Hongwei Zhao, Hongtao Lu, Xi Shen, and Ying Shan. “Generating human motion from textual descriptions with discrete representations”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, pp. 14730–14740.

[133] Mingyuan Zhang, Huirong Li, Zhongang Cai, Jiawei Ren, Lei Yang, and Ziwei Liu. “Finemogen: Fine-grained spatio-temporal motion generation and editing”. In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 13981–13992.

[134] Yiran Zhao, Yujie Tao, Grace Le, Rui Maki, Alexander Adams, Pedro Lopes, and Tanzeem Choudhury. “Affective Touch as Immediate and Passive Wearable Intervention”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.4 (2023), pp. 1–23.

[135] Leming Zhou, Jie Bao, I Made Agus Setiawan, Andi Saptono, Bambang Parmanto, et al. “The mHealth app usability questionnaire (MAUQ): development and validation study”. In: *JMIR mHealth and uHealth* 7.4 (2019), e11500.

[136] Alexander Zhu, J Sher, Richard Li, Otari Ioseliani, Leah Cantor, Elena G Brewer, Karla Landis, David Bridges, Sean A Munson, Hanna Hunter, Cindy Y Lin, and Sarah P Psutka. *What Motivates Bladder Cancer Patients to be Active? A Qualitative Study Assessing Attitudes towards Physical Activity and Digital Health Technologies*. 2025.