Digital health technology data in biocomputing:
Research efforts and considerations for expanding access

Michelle Holko

Google, Google Public Sector
Washington, DC 20001, USA
Email: michelleholko@google.com

Chris Lunt

National Institutes of Health
Bethesda, MD 20892, USA
Email: chris.lunt@nih.gov

Jessilyn Dunn

Biomedical Engineering, Duke University
Durham, NC 27708, USA
Email: jessilyn.dunn@duke.edu

Data from digital health technologies (DHT), including wearable sensors like Apple Watch, Whoop, Oura Ring, and Fitbit, are increasingly being used in biomedical research. Research and development of DHT-related devices, platforms, and applications is happening rapidly and with significant private-sector involvement with new biotech companies and large tech companies (e.g. Google, Apple, Amazon, Uber) investing heavily in technologies to improve human health. Many academic institutions are building capabilities related to DHT research, often in cross-sector collaboration with technology companies and other organizations with the goal of generating clinically meaningful evidence to improve patient care, to identify users at an earlier stage of disease presentation, and to support health preservation and disease prevention. Large research consortia, cross-sector partnerships, and individual research labs are all represented in the current corpus of published studies. Some of the large research studies, like NIH's *All of Us* Research Program, make data sets from wearable sensors available to the research community, while the vast majority of data from wearable sensors and other DHTs are held by private sector organizations and are not readily available to the research community. As data are unlocked from the private sector and made available to the academic research community, there is an opportunity to develop innovative analytics and methods through expanded access. This Session solicited research results leveraging digital health technologies, including wearable sensor data, describing novel analytical methods, and issues related to diversity, equity, inclusion (DEI) of both the underlying research data sets and the community of researchers working in this area. We particularly encouraged submissions describing opportunities for expanding and democratizing academic research using data from wearable sensors and related digital health technologies.

Keywords: digital health technologies; wearables; sensors; waveform data; time-series data; algorithms.

1. Background

Use of digital health devices has grown; in 2016, only 12% of Americans were estimated to regularly use a wearable digital health device, but by 2020, the estimation jumped to 21% [1]. Digital Health Technologies (DHTs), including wearable sensors like smart watches, have the potential to inform us about our health. But there are gaps in who has access to data and devices, who is performing the research, and therefore who the new technologies are poised to help. Reviews of the current landscape of DHT research studies in the National Center for Biotechnology Information (NCBI)'s Clinical Trials database (clinicaltrials.gov), and of studies published by the top-20 funded private sector DHT companies, highlight several patterns and limitations:

1. **Small sample size:** Aside from a few large studies, most of the published clinical trials utilizing DHT have been relatively small, and are largely under-powered. "Nearly half the studies - 829, or 46.5% - had less than 100 enrollees. Only 8% had more than 1,000 [2]."

2. **Narrow Health Focus:** The majority of published DHT studies focus on cardiometabolic health and mental health/wellness, while relatively little published research examines critical healthcare burden diseases like stroke, chronic obstructive pulmonary disease (COPD), and diabetes [2].

3. **Narrow Population Focus:** Of studies published by the top 20 funded DHT private-sector companies, the majority (72%) include only healthy volunteers, rather than high-risk populations with comorbid conditions [3]. The breadth and diversity of the study population(s), including socioeconomic, healthcare status, and racial diversity, may be the most critical component of building AI-based DHT algorithms. This diversity is lacking in current published research, likely leading to biased results [4]. The "bring your own device" model has been used by many research studies, but this design may result in biased selection of participants, and therefore biased results [5].

4. **Limited Outcome Assessments:** Only 15% of published DHT studies measured clinical effectiveness, and only in relation to the patient outcomes and did not evaluate healthcare cost or access to care [6]. As healthcare cost and access are two of the most pressing needs in healthcare, it is important to expand research to examine these outcomes.

5. **Insufficient Reporting and Data Publishing:** Importantly, not only is reporting in clinicaltrials.gov not required for observational DHT trials, there is also no public database for DHT data and algorithms. This complicates the ability to understand the full range of DHT "real world evidence" (RWE)-based research, and undermines research reproducibility and

validation. The lack of a consensus DHT database also means that DHT data curation, feature (e.g., digital biomarker) discovery, and algorithm development is limited to those who have data, which is largely the private sector DHT companies. One attempt to develop standardized pipelines and data repositories for digital health data, the Digital Health Data Repository as part of the Digital Biomarker Discovery Pipeline [7], developed by co-organizer Jessilyn Dunn's lab, is still not fully funded.

6. **Bridging the Regulatory Gap and Moving to Clinical Implementation:** Despite tremendous progress in DHT research and development, there is still a lot of work to be done along the research → regulatory → clinical implementation continuum. The *All of Us* Research Program is uniquely situated within NIH to interact with FDA colleagues and assist in developing regulatory standards for this new and uncharted territory. There is also a relatively new FDA Center for Digital Health Excellence, led by Bakul Patel. The Digital Medicine Society is a professional organization that has been working across sectors with the community to support innovation and standardization, in part via the Digital Health Measurement Collaborative Community (DATAcc) [8] and the Digital Health Playbook [9]. There is also a Digital Health Consortium, housed within the Office of the National Coordinator, for senior leaders within the federal government to convene across the digital health continuum.

The above limitations don't begin to address potential bias in algorithm development due to a limited pool of researchers interacting with these data. The purpose of this Session is to provide a forum for current research, address issues related to Diversity, Equity and Inclusion (DEI) in terms of the types of research and the researchers engaged, and ultimately to energize non-commercial research in the area. Our motivating question is how can this community work together to create more equitable research in the digital health tech space to benefit the research community and resulting impact?

2. Relevance to biocomputing

Digital health technologies, including wearable sensors, lend themselves well to biomedical and computational biology research since they generate continuous or near-continuous data streams ripe for machine learning and artificial intelligence (ML/AI) research. Algorithms developed for detecting anomalies and other biomedically-related phenomena in wearable sensor data are increasingly being incorporated into research and moving into clinical practice and other health adjacent applications. In past years of this conference, there has been good representation of a variety of data types, including genomics, imaging and clinical data sets; there has been limited coverage of wearable sensors and digital health technologies research.

The topic is timely for PSB 2023 since not only is there a growing use of wearable sensors in research, but also because there are potential DEI issues for both research data sets and researchers working with these data. Searching PubMed for the keyword "wearable" (Figure 1) shows exponential growth in the number of publications, with 701 in 2021. "Digital health" shows a similar trend (graph not shown) with 1,306 publications in 2021. Some of the journals and conferences that generally cover DHT research include Nature Digital Medicine, Lancet Digital Health, AMIA, and IEEE Biomedical and Health Informatics (BHI). Many of the conferences are more focused on the clinical aspects and clinical trials, and not as much on the computational biology or biomedical research aspects of DHT data analysis and algorithm development. There have also been a few cross-sector seminars recently to explore regulatory and other issues related to digital health technologies research, including this one in early 2020:
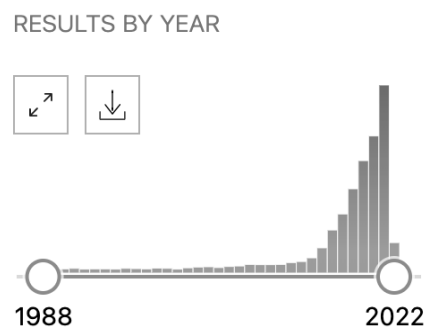
https://fnih.org/our-programs/biomarkers-consortium/digitalmonitoring

RESULTS BY YEAR

Fig. 1. Number of publications with "wearable" in PubMed from 1988-2021

This Session showcases recent research on digital health tech, DEI issues related to these data and research, and a discussion about what is needed to bridge these DEI gaps. The goal of this information sharing and discussion opportunity for participants and the community is to expand awareness and access to these data and tools, to enrich computational biology research, and bridge DEI gaps. The session also includes a range of voices from academia, government, and private sector. It's important to represent private sector voices in this discussion since much of the research is currently happening in tech companies developing digital health devices. Creating a forum for dialogue across sectors is important for bridging gaps in awareness and understanding, and encouraging more researchers to participate in developing computational methods and analysis of data from digital health tech.

The discussion will focus on key challenges facing the field, and participants are encouraged to contribute ideas to potential solutions and initiate lasting collaborations with researchers and

communities in this area (e.g., the Digital Medicine Society). Further, participants will be exposed to cutting edge tools in this space with brief demos on how to use them, including the Digital Biomarker Discovery Pipeline (DBDP.org) [10], the Digital Health Data Repository, the *All of Us* Researcher Workbench, and others. The Session will also provide an opportunity to discuss as a community what is needed to truly enable cross-sector and expanded research for digital health technologies.

3. Session overview

The organizers will introduce the session, followed by a keynote from Eric Perakslis, the Chief Science and Digital Officer at the Duke Clinical Research Institute. He brings to the discussion his wide range of experience working on collaborative efforts in data science that spanned medicine, policy, engineering, computer science, information technology, and security, all from positions in academia, private sector, and the government.

There will then be a series of brief talks from the authors of the papers that were accepted for inclusion in the proceedings, and a panel discussion to include voices from industry and government. A moderated Q&A discussion will conclude the session. The talks are original research for publication, are widely varied, and include 1) comparing two wearable devices to augmenting prediction of mild cognitive decline using not only MRI but also language markers from speech, 2) a computational method for image segmentation of medical images, and 3) how fitbit data in the *All of Us* cohort can be used to improve upon current methods of predicting quality of life post-surgery.

The panel discussion will feature speakers from industry, including Ed Ramos and Julia Moore Vogel from Scripps Digital Trials Center and Care Evolution, Aaron Coleman, founder and CEO of Fitabase, Bakul Patel, currently at Google Health but the founding director of the FDA's Center for Digital Health Excellence, and Joshua Stein, Founder and Chief Growth Officer at Fitbit.

For the moderated Q&A discussion session all speakers, session organizers, and session attendees are welcomed to participate. The speakers and organizers represent a diverse set of perspectives across research efforts and related DEI issues. For both the talks and the panel, diversity and inclusion across gender, race and other factors are incorporated into the Session organization.

References

1. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/

2. https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2725079
3. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.05081
4. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1377-7
5. https://preprints.jmir.org/preprint/29510/accepted
6. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.05081
7. https://github.com/DigitalBiomarkerDiscoveryPipeline/Digital_Health_Data_Repository
8. https://datacc.dimesociety.org/
9. https://playbook.dimesociety.org/
10. https://pubmed.ncbi.nlm.nih.gov/33948242/