Bridging a Gap Between Data Science Research and Health DIY Movement

Abstract
Healthcare initiatives have accelerated development of self-management tools for disease populations and are opening new areas of study in their domains. We believe that inclusive collaborations between academic researchers and DIY health communities hold tremendous potential for improving research and development for all parties involved. Our research group envisions health hackers leveraging our existing work in predictive analytics and modeling to advance their community, while promoting an open-source environment for the sharing of ideas, methods, data, and results. We hope to engage with the HCI research community in exploring new ways to bring these similarly interested groups together.

Author Keywords
DIY, diabetes, health self-management, open-source, data science

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction
User-driven design of software, hardware, and everything in between has been studied and discussed under the monikers of hacker, maker, and Do-It-
Yourself (DIY) culture. These movements have been incredibly successful in bringing users together to develop and share products that they themselves seek. This trend swept into the mobile health field (mHealth) with an early call to action of makers and hackers by Estrin and Sim [1]. Carrera and Dalton describe Do-It-Yourself (DIY) Healthcare as a movement to help “consumers...monitor and manage their health, and guide their healthcare consumption on their own, as well as with the involvement of providers” [2]. Estrin and Sim outlined a growing problem with mHealth initiatives, describing them as “a patchwork of incompatible applications (“apps”) serving narrow, albeit valuable, needs,” and proposed an open architecture for these applications that could be brought to consumers through public-private partnerships.

While interoperability is improving through software like Apple Health, DIY groups have made additional strides in accessing and integrating medical device data that is typically locked behind private manufacturer protocols. Individuals have been hacking code and repurposing hardware for decades, but the application to mHealth is a newer concept with less precedence and higher stakes with respect to the consequences of malfunction. Thus, data-liberators like Nightscout, a group dedicated to getting blood glucose readings off of continuous glucose monitors (CGM) and into the cloud for remote real-time viewing, were in uncharted regulatory territory until the United States Food and Drug Administration (FDA) published guidance for Medical Device Data Systems in 2015 that established systems for active monitoring to be exempt from regulatory control [3], [4].

While Nightscout may have the most users (estimates from 2015 suggest well over 2000 individual users, and over 15,000 members of the CGM In The Cloud Facebook group), DIY healthcare in diabetes has gone in other exciting directions. Tidepool is attempting to lead a charge towards diabetes device data interoperability, which has the potential to catalyze improvements in medical devices and third party software for analytics [5]. Equally interesting is the OpenAPS (Open Artificial Pancreas System) project, which provides the technology for constructing a basic overnight closed-loop system that integrates CGM and insulin pump data and controls to automatically regulate blood glucose levels [6].

**Academic-DIY Health Partnerships**

Some DIY Health movements have evolved without any institutional impetus, but many groups have built strong ties with academic institutions. BlueStar and Tidepool, both startups that came largely out of the DIY diabetes culture, use their relationships with university researchers to coordinate studies that evaluate and promote their product [5], [7]. The Diabetes Informatics + Analytics Lab (DIAL) at University of California San Diego (UCSD) recruits academic researchers as well as hackers and makers to accelerate their research and development [8].

**An opportunity for academics developing advanced data science methods**

Our research group is motivated by the same goal as the diabetes DIY movement: to unlock the potential of self-management data to improve quality of life and clinical outcomes. Our focus is on using advanced data science methods to provide personalized glucose forecasts and meaningful retrospective analyses for
improved self-management and clinical care [9]. We also pursue an understanding of how diabetics reason about their self-monitoring data to improve the relevance of our data-driven support efforts [10]. So far, we have centered our research on Type 2 Diabetes, but our versatile modeling techniques make Type 1 Diabetes a logical next step. We believe that stronger connections between academic data science researchers and DIY health communities can be mutually beneficial and lead to new advances in useful and usable technologies for diabetes self-management. In this workshop, we would like to connect with other researchers interested in studying and supporting diabetes DIY communities and discuss potential directions for such partnerships.

We envision two modes of engaging with diabetes hackers that call for new tools, environments, and technologies that can be used to facilitate effective collaborations: 1) bridge the gap between academic and DIY data scientists, and 2) connect members of DIY diabetes communities to data science methods.

**Bridging the gap between data scientists**
We wish to connect academic researchers with DIY scientists who are developing algorithms for personal and DIY-group-wide use in order to accelerate collective progress. Each participating entity has its own unique set of data access, computational resources, code repository, knowledge, skills, and experience, all of which can be leveraged in an open source platform to maximize efficiency, communication, and, ultimately, impactful results.

We have observed a need for centralization of experimental results and collective wisdom in the DIY diabetes modeling and analytics space that can help other researchers and DIY-ers understand and evaluate previous and ongoing efforts. To this end, we believe that mini write-ups akin to demonstrations posted in r-bloggers.com will be most useful for describing methods and results [11].

Concomitant with the desire to explain and share methods and results is an interest in promoting discussion. Members of DIY health communities like Nightscout and OpenAPS come from many different backgrounds and thus bring a variety of important perspectives. We want to facilitate mathematical problem solving and collective sense-making amongst a diverse group of novice, experienced, and expert members, and look to stackoverflow.com as a prime example of how a question and answer format can breed creative, on-record solutions [12]. Furthermore, an active group on stackoverflow.com could attract the attention of computational enthusiasts looking to work on exciting real-world problems.

We also recognize a need for efficient and secure data hosting. There are examples of academic institutions that take steps towards establishing infrastructures for secure data hosting and sharing. For example, data from UCSD’s Diabetes Management Integrated Technology Research Initiative pilot study is hosted publicly by Integrating Data for Analysis, Anonymization and Sharing (iDASH), a National Center for Biomedical Computing under the NIH Roadmap for Bioinformatics and Computational Biology; iDASH’s public data sharing implementation provides a potential model for secure diabetes data hosting [13].
Perhaps most obvious is a means of publishing, updating, and editing code for analytics, for which public GitHub repositories are typically favored. This can promote rapid evaluation of newly developed analytical tools, and would also enable users to analyze their data in novel ways.

**Connecting users to data science methods**
While patients are interested in maximizing the potential benefits of data they collect through self-monitoring to inform their actions, investigators are looking for data sets to support their method development and evaluation efforts. We see mutual benefit in connecting data collected by users (e.g. through Nightscout) with data science methods developed by academic institutions or in a collaborative open-analytics space to provide interesting experimental reports to patients. This allows researchers to evaluate and improve their methods and get real-time feedback from users. At the same time, users can gain new perspectives on their personal data, while supporting organizations in whose success they are invested. The developers of Tidepool have already mapped out this type of data sharing, but the data-information exchange we describe may also become a natural consequence of open-source collaboration among analytics developers who are already rooted in the diabetes data-liberation space.

**Benefits of academic-DIY Health partnerships**
Relationships with academic institutions have proven valuable for diabetes hackers to validate their products and become involved in studies. Groups like Nightscout may still need to perform studies of safety and efficacy of their product to ensure they remain in compliance with FDA regulations, and their large base of engaged users creates a cohort ripe for new studies that can, in turn, further drive group visibility and participation. Data sharing can be a simple yet effective way for both academics and DIY-ers to push the cutting edge. Most of all, collaboration between academic and DIY-researchers promotes a wider exchange of ideas and experiences, along with an exchange of computational machinery to bring their shared visions to fruition.

**Conclusion**
In this workshop, we hope to engage with other HCI researchers in exploring new ways for building stronger connections with the DIY diabetes community in a sustained, open partnership that works towards developing the next generation of analytical tools for diabetes data. Our goal is to identify ways to help diabetes hackers leverage our existing work in predictive analytics and mechanistic modeling, support ongoing computational efforts of DIY data scientists, and learn from the DIY analytics developers.

**References**


