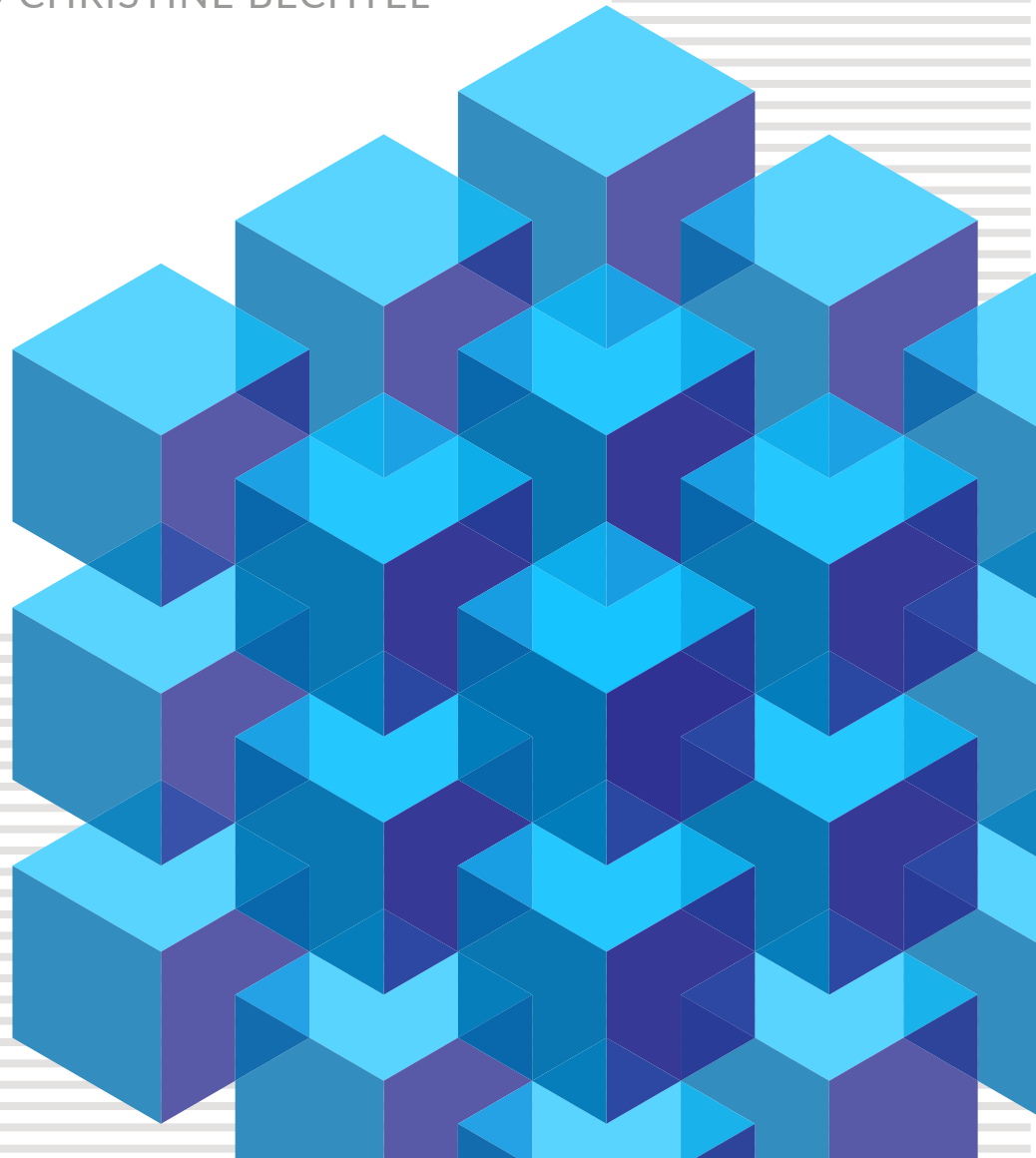# Leveraging Data-Driven Patient Participation to Accelerate Medical Research

RACHEL TUNIS AND CHRISTINE BECHTEL

ABOUT THE MILKEN INSTITUTE

The Milken Institute is a nonprofit, nonpartisan think tank.

For the past three decades, the Milken Institute has served as a catalyst for practical, scalable solutions to global challenges by connecting human, financial, and educational resources to those who need them. Guided by a conviction that the best ideas, under-resourced, cannot succeed, we conduct research and analysis and convene top experts, innovators, and influencers from different backgrounds and competing viewpoints. We leverage this expertise and insight to construct programs and policy initiatives.

These activities are designed to help people build meaningful lives in which they can experience health and well-being, pursue effective education and gainful employment, and access the resources required to create ever expanding opportunities for themselves and their broader communities.
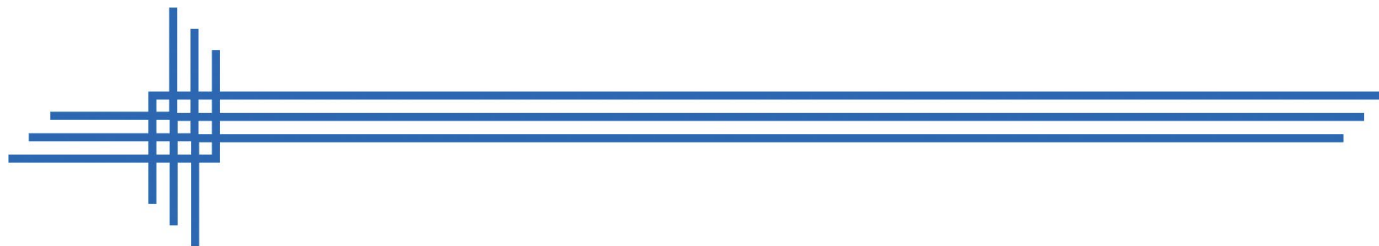
ABOUT FASTERCURES

FasterCures, a center of the Milken Institute, is working to build a system that is effective, efficient, and driven by a clear vision: patient needs above all else. We believe that transformative and life-saving science should be fully realized and deliver better treatments to the people who need them.
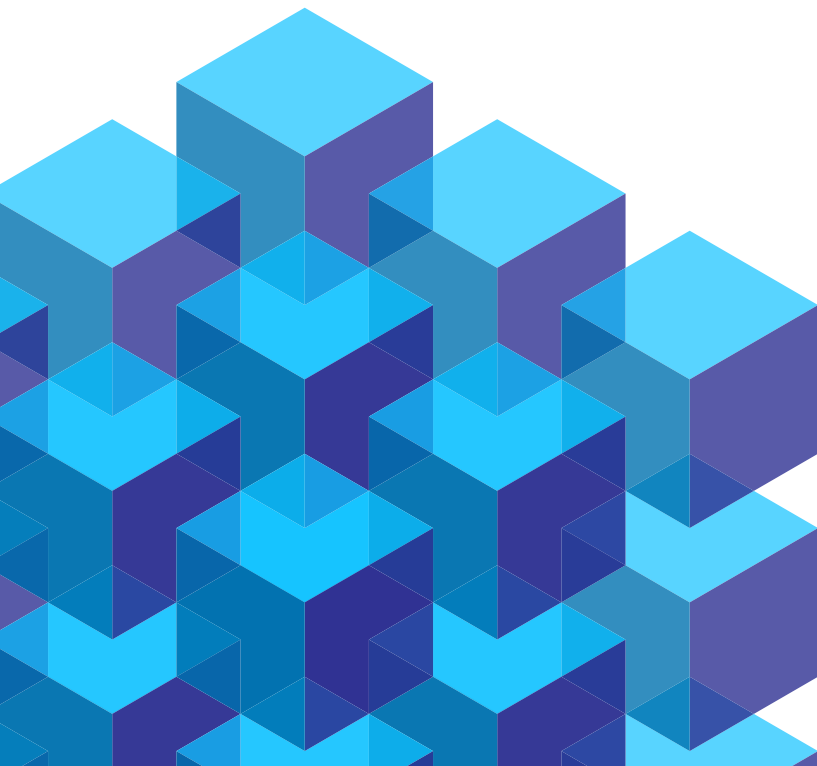
Leveraging Data-Driven Patient Participation to Accelerate Medical Research

Rachel Tunis, FasterCures
Christine Bechtel, X4 Health

INTRODUCTION AND OVERVIEW

The fuel for biomedical innovation is data, and today we are generating more health and lifestyle-related data than ever before. These datasets hold enormous promise for biomedical research because they can be combined and analyzed using technologies such as artificial intelligence to produce new insights essential to improving overall health, building our understanding of disease causation, and potentially discovering new treatments and cures. Data-driven patient participation—the ability for anyone to contribute their health information such as electronic health record (EHR) data, biospecimens, and wearable data for research—has emerged as another tool for research entities seeking to make discoveries that advance medical research and improve care for patients. Further, it directly involves patients in biomedical research, allowing them to help speed its progress in disease areas that they care about, and helps patients engage in and have more control over their health care, which is known to improve health outcomes.

Initiatives facilitating data-driven patient participation in research are rapidly being recognized as tools to better understand COVID-19. Calls for patient contributions of health-care data and other information are louder than ever, from myriad organizations that aim to better understand the virus's threat and course. In response to these calls, some countries have started to collect troves of data from citizens without their permission, whereas other countries are developing new opt-in mechanisms to report health and location data. The effectiveness of these efforts remains to be seen.[1] Now is the time to systematically examine the specific areas where data contribution may meaningfully accelerate research, as well as the

---

[1] Derek Thompson, "The Technology That Could Free America from Quarantine," *The Atlantic*, April 7, 2020, https://www.theatlantic.com/ideas/archive/2020/04/contact-tracing-could-free-america-from-its-quarantine-nightmare/609577/.

potential challenges such as patient privacy and the implications of newly proposed policies and infrastructure. Amassing massive amounts of data with vague goals and weak guardrails will not protect patients and may make them less willing to contribute to research.

THE PROMISE

Widespread data contribution from patients, enabled by appropriate technology and regulations, could position patients at the center of data sharing and collection processes, increasing their influence over the priorities and pace of research. Further, patients could directly benefit from any insights gleaned from research using their data—if effective feedback loops exist between data aggregators and data contributors.

Other potential benefits of widespread data-driven patient participation include the following:

- Public awareness about the uses of health data could improve.

- Researchers could prioritize recruiting and engaging more diverse cohorts, which could lead to not yet known (and perhaps unimaginable) advances in personalized medicine.

- Using trusted platforms, patients could integrate their data into tools or apps to generate information about how to manage their health, both in real time and in the long term.

THE CHALLENGES

Although clearly a worthwhile endeavor to pursue, data-driven participation in research faces several challenges. Data contribution is complex, and the details of its essential processes are often not well known to the public. Further, the goals, implications, and potential benefits and drawbacks of these processes can seem convoluted, especially amid a confusing regulatory landscape and widespread mistrust of health and tech companies that hold and manage data. Major challenges include the following:

1. Gaps in existing regulations that protect consumers' personal data and privacy

No single US law or policy fully covers patient privacy and data rights writ large. For example, health-care companies and the individual health records they maintain are subject to the Health Information Portability and Accountability Act (HIPAA), but commercial entities that hold similar health information are subject

to the Fair Trade Practices Act and other commerce-related laws rather than HIPAA. Therefore, commercial entities can share or sell individually identifiable health information, as long as such practices are stated in their privacy notices. Even those entities governed by HIPAA can freely use de-identified data, despite the fallibility of common de-identification approaches. It is also unclear what laws and policies protect data from which health insights can be inferred, such as online shopping or social media information.

Legal protections matter because the stakes surrounding health data are high. Unlike credit card or social security numbers, health data, if breached, can be used to discriminate against consumers in employment, insurance rates, insurance coverage, and more, as well as expose them to social stigma—particularly in the case of sensitive medical information. A BMJ study found that 79 percent of the 24 top-rated interactive health apps for Android devices shared user data with third parties, who then shared it with "fourth parties."[2] Companies that re-disclose data or are hacked can jeopardize consumer trust and put people at risk of material and social harm from various forms of discrimination.

In May 2018, the EU implemented a national comprehensive privacy law called the General Data Protection Regulation. Just one month later, California passed the strictest data privacy law in the US, the California Consumer Protection Act, which took effect on January 1, 2020. In the absence of Congressional action, other states are beginning to enact privacy legislation. These laws mandate companies to be fully transparent about their data partners and practices and to allow consumers to opt in or out of data collection, among other things. While disclosure is a step in the right direction, until the US adopts a more comprehensive approach to privacy and data protections that foster trust, consumers may be less inclined to relinquish their data, even for research purposes.

In 2019, several major technology companies were found to be using identifiable patient data to design software intended for use in patient care, without explicit consent from those patients.[3] This discovery furthered the perception that data may not be safe in the hands of third-party companies. Therefore, organizations that hold or manage health data must prove their trustworthiness, at least until more comprehensive privacy laws are enacted in the US.

2   Meg Bryant, "Most Health Apps Share Data with Third, Fourth Parties, BMJ Analysis Warns," *Healthcare Dive*, March 21, 2019, https://www.healthcaredive.com/news/most-health-apps-share-data-with-third-fourth-parties-bmj-analysis-warns/550984/?mc_cid=b507d1938b&mc_eid=a362877c33.

3   Rob Copeland, "Google's Project Nightingale Gathers Personal Health Data on Millions of Americans," *The Wall Street Journal*, November 11, 2019, https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790.

2. Challenges involved in data sharing and interoperability

Recent federal policy has acknowledged the problem that clinical data often do not seamlessly flow among health plans, providers, research programs, and patients.[4] Most patients receive care through multiple providers and systems, and the lack of interoperability, or the ability to electronically exchange data across entities regardless of record system, is a significant challenge for clinical care as well as for research. The 21st Century Cures Act, passed in 2016, made interoperability and patient access to their clinical and claims data national imperatives, requiring that all organizations integrate application programming interfaces (APIs)[5] into their software and services. Further, in March 2020, the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services released final rules that require public and private entities alike to collect specified core data elements in EHRs and to adopt specific data standards to enable standardized data exchange between different platforms and applications through secure APIs. These rules also effectively allow patients to access their own health records for free, using the smartphone or tablet app of their choosing.[6] These new rules have been widely recognized as a step in the right direction toward widespread and meaningful interoperability, but their implementation is in its early stages, and challenges remain. Not all challenges are technical; some of the most formidable of them relate to competition and dismantling long-standing existing norms, which have yet to be fully addressed.

Forging ahead on the widespread implementation of interoperability efforts will be crucial to make data contribution seamless for patients and facilitating organizations.[7] In addition, as data sharing standards and methods are developed, it will be important to ensure that useful data are being collected. For example, social determinants data are not currently mandated as a part of major health data standards but are critical to clinical care and patients' overall health.

4   "Connecting Health and Care for the Nation" (Office of the National Coordinator for Health Information Technology, 2019), https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf.

5   APIs are messengers or translators that work behind the scenes to allow software, apps, websites, and programs to communicate with one another. (Source: HealthIT.gov at https://www.healthit.gov/api-education-module/story_content/external_files/hhs_transcript_module.pdf).

6   "HHS Finalized Historic Rules to Provide Patients More Control of Their Health Data," US Department of Health and Human Services, accessed March 9, 2020, https://www.hhs.gov/about/news/2020/03/09/hhs-finalizes-historic-rules-to-provide-patients-more-control-of-their-health-data.html.

7   Karen DeSalvo and Mark Savage, "The Nationwide Interoperability Roadmap and the ONC's and CMS's Proposed New Regulations: Are We Halfway There Yet?" *Health Affairs*, June 25, 2019, https://www.healthaffairs.org/do/10.1377/hblog20190621.419379/full/utm_source=Newsletter&utm_medium=email&utm_.

3. Challenges with aggregating, integrating, and analyzing data on a large scale

Stakeholders across biomedical research are keenly interested in making connections and gleaning new insights from vast quantities of data. However, they are finding that the process of organizing, cleaning, and linking large and diverse datasets across both health and lifestyle data is a substantial challenge, especially because clinical data are often not interoperable, not collected using uniform standards and definitions, or not structured in a readily accessible way. This challenge is especially true in the context of COVID-19, as hundreds of organizations have scrambled to build out their own solutions and datasets without standards or consensus, resulting in datasets that are disparate and difficult to aggregate.[8] In order for data-driven patient participation to lead to discoveries that accelerate biomedical research and improve patient outcomes, better mechanisms for cleaning and integrating data are needed.

4. Bias and underrepresentation in datasets

Certain historically marginalized populations are underrepresented in datasets, which can negatively affect their health outcomes. For example, algorithms that are trained with data that underrepresent minority groups have been shown to be less accurate in their diagnostics for these groups.[9] Algorithms can also propagate biases from their human programmers. Inaccuracies in artificial intelligence algorithms caused by bias could influence conclusions about disease causality or treatment recommendations, which would be potentially harmful to patients.[10]

Data platforms that successfully increase representation in biomedical research will prioritize outreach to underrepresented groups, maintain transparency, and provide patients with agency over their own data whenever possible to build and maintain trust. Some initiatives, such as the National Institutes of Health's (NIH) *All of Us* Research Program, place a strong emphasis on representation.[11]

---

8   Carolyn Y. Johnson, "Chaotic Search for Coronavirus Treatments Reveals Another Shortcoming in Pandemic Preparedness," *The Washington Post*, April 15, 2020, https://www.washingtonpost.com/health/2020/04/15/coronavirus-treatment-cure-research-problems/.

9   Angela Lashbrook, "AI Driven-Dermatology Could Leave Dark-Skinned Patients Behind," *The Atlantic*, August 16, 2018, https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/; Ruth Hailu, "Fitbits and Other Wearables May Not Accurately Track Heart Rates in People of Color," *Stat News*, July 24, 2019, https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/.

10 Kadija Ferryman and Mikaela Pitcan, "Fairness in Precision Medicine," *Data and Society*, Feburary 2018, https://datasociety.net/wp-content/uploads/2018/02/DataSociety_Fairness_In_Precision_Medicine_Feb2018.pdf.

11 The *All of Us* Research Program Investigators, "The '*All of Us*' Research Program," *The New England Journal of Medicine* 381 (2019): 668-76.

However, attracting a large number of participants from certain underrepresented populations, such as patients who are homeless or who have cognitive impairment, can be challenging. These groups may also be less trusting of the medical system due to historical discrimination.[12] Prioritizing representation in datasets will be crucial to achieving the goal of making discoveries that improve health for everyone through personalized, patient-centered approaches to care.

5. Time and energy required of patients to aggregate and share their data

Effective initiatives will likely require patient engagement over many years. Patient efforts will span contributing the initial data (e.g., donating biospecimens in a lab to completing surveys), maintaining the currency of data sources, linking patient-reported or lab data to the data platform, and responding to ad hoc requests, among other things. Although increased interoperability will smooth the process for individuals who wish to contribute data, data platforms should consider ways to shorten the feedback loops so that patients will see the benefits of contributing their data and will be incentivized to continue to do so.

PROFILES OF PIONEERS

Although organizations that facilitate data-driven patient participation have become more common in recent years, they remain relatively rare, despite large-scale initiatives by the US and other governments to gather health data from individuals with the goal of advancing biomedical research (e.g., *All of Us* in the US and the 100,000 Genomes Project in the UK). Beyond large, centrally funded projects, a handful of startups and smaller companies are entering this field, with varying business models and focus areas. The types of data being collected and analyzed include EHR and clinical data, DNA files from genealogy companies, direct biospecimens such as saliva or blood samples, data from mobile apps or social media, and data from medical devices, fitness trackers, or other wearables.

This section profiles five organizations that are placing patients at the center of the process through which they contribute their experiences and health data to research.

12 Corey J. Williams, "Black Americans Don't Trust Our Healthcare System—Here's Why," *The Hill*, August 24, 2017, https://thehill.com/blogs/pundits-blog/healthcare/347780-black-americans-dont-have-trust-in-our-healthcare-system.

1. ALL OF US RESEARCH PROGRAM

Key Facts:

- The *All of Us* Research Program was developed as a key element of the Precision Medicine Initiative, which was launched by the Obama Administration in 2016 and allocated $130 million to the US NIH to build a national, large-scale research cohort.

- As of January 2020, more than 311,000 participants have enrolled, while 242,000 participants have completed the initial steps of the program. Greater than 80 percent of these participants are from groups historically underrepresented in biomedical research.[13]

- As of March 16, 2020, *All of Us* appointments and events have been paused in response to the COVID-19 pandemic.[14]

Program Goal:

- To "engage one million or more volunteers living in the US to contribute their health data over many years to improve health outcomes, fuel the development of new treatments for disease, and catalyze a new era of evidence-based and more precise preventive care and medical treatment."[15]

Data Collected:

- Data collected through *All of Us* include responses to health questionnaires, EHRs, physical measurements, information from Fitbits or other wearables, and biospecimens.

- Questionnaires explore factors such as lifestyle, socioeconomic, and environmental characteristics.

Additional Characteristics:

- *All of Us* seeks to ensure that its cohort sufficiently represents the diversity of the United States and therefore includes characteristics that may not have previously been emphasized in research, such as race, ethnic group, age, sex, gender identity, sexual orientation, disability status, access to care, socioeconomic status, educational attainment, and geographic location.

---

14 Josh Denny, "Coronavirus Update from '*All of Us*'," National Institutes of Health, March 16, 2020, https://allofus.nih.gov/news-events-and-media/announcements/coronavirus-update-all-us.

15 "Program FAQ," National Institutes of Health, *All of Us* Research Program, accessed May 1, 2020, https://allofus.nih.gov/about/faq.

2. FIGHT TO END COVID (GENETIC ALLIANCE, XCURES, LUNADNA)

Program Goal:

- The Fight to End COVID initiative was developed in March 2020 by Genetic Alliance, xCures, and LunaDNA to gather information about personal experiences with COVID-19 so that scientists can study long-term health effects, treatment effectiveness, and best methods to prepare for future infectious diseases.[16]

Data Collected:

- Patients complete a five-minute questionnaire about their experiences with COVID-19. Data are privacy protected, remain in the platform, and can be revoked or deleted at any time.

- Patients can connect their EHR data and patient portals to their account so that their data flow into the system automatically.

Additional Characteristics:

- Patients can share their experiences through the website's blog and will be able to make community-specific queries in a forthcoming platform.

- Researchers have already demonstrated interest in accessing the initiative's aggregated data. Updates on progress and discoveries will be relayed directly to participants.

---

16 "Your Voice is Urgently Needed in the Fight," FightoendCOVID.com, accessed May 1, 2020, http://www.fighttoendcovid.com/?fbclid=IwAR1hLamQRNl1Pu53e3mdzT3cR2NryqybcatQ6kk2w_gxmsgI6oRdXMCaoDc.
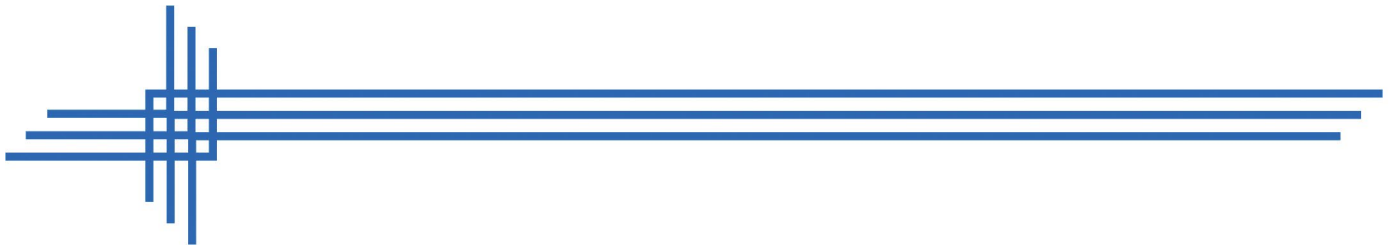
3. PATIENTSLIKEME

Key Facts

- PatientsLikeMe (PLM), a personalized health network for sharing and learning from patient data, was created by Jamie and Ben Heywood in 2004 after their brother Stephen was diagnosed with ALS. Today, more than 650,000 patients representing nearly 3,000 different conditions are members of PLM.[17]

Program Mission:

- "To improve the lives of patients through new knowledge derived from shared real-world experience and outcomes."[18]

Data Collected:

- Patients provide their health information on their profile, including health conditions, drug histories, and side effects, and they can share personal experiences and tips with other users on the site.

- In 2016, PLM began to collect and connect biospecimens to patients' reported data to create more advanced and comprehensive datasets. This research program is called DigitalMe.

Additional Characteristics:

- PLM was acquired by UnitedHealthGroup and became a part of its R&D unit in June 2019.

- Although a for-profit organization, PLM has always been transparent about how it uses patient data, who its partners are, and how it makes money.

17 "About us," PatientsLikeMe, accessed May 1, 2020, https://www.patientslikeme.com/about.

18 Ibid.
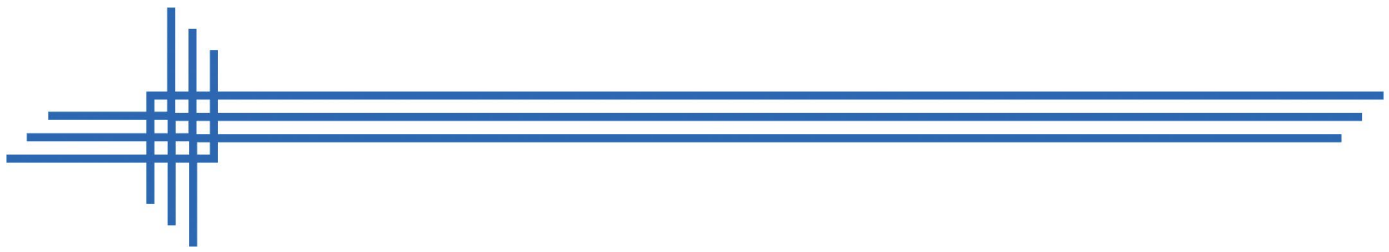
4. TIDEPOOL

Key Facts

- Tidepool is a nonprofit organization that provides software and programs to patients in the diabetes community so that they can easily make use of data from their diabetes devices.

- Tidepool's Big Data Donation Project shares Tidepool data with diabetes researchers. Tidepool then shares the research findings with the patient community and donates 10 percent of proceeds from its industry partners to an organization of the data donor's choice. Tidepool users opt into the program, anonymously contributing their de-identified data.

Program Vision:

- Tidepool is "dedicated to making diabetes data more accessible, actionable, and meaningful for people with diabetes, their care teams, and researchers."[19]

Data Collected:

- Patients can upload blood glucose, insulin, and carbohydrate data from their insulin pump, continuous glucose monitor, and blood glucose meter, using Tidepool Uploader on their Mac or Window PCs and Tidepool Mobile on their iPhones. The data are made available to researchers, universities, and industry innovators.

Additional Characteristics:

- The Tidepool Web portal syncs data from all available devices and sources and displays them in an intuitive interface so that patients and providers can view their trends by the hour, day, or week and manage their diabetes more effectively.

- Patients control their Tidepool data and can choose to share with anyone on their care provider team, as well as friends or family.

---

19 "About Tidepool," Tidepool, accessed May 1, 2020, https://www.tidepool.org/about.
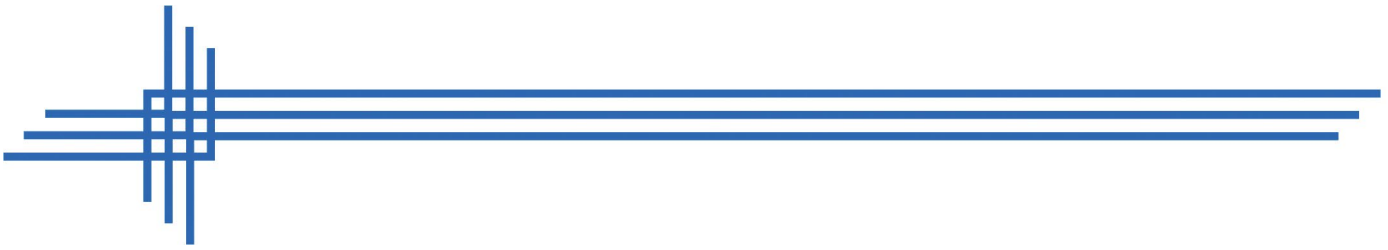
5. WAR ON CANCER

Key Facts:

- War on Cancer is a social media app for cancer patients, survivors, and their loved ones. The app serves as a safe space for patients to share their experiences, ask questions, and connect with one another.[20]

- In addition to a Social Profile, app users have a Health Profile, which they can use to manage their treatments and symptoms, match with clinical trials, and contribute to research.

Data Collected:

- If they choose to contribute their data to research, patients can complete surveys through the War on Cancer app that tracks patient-reported outcome and experience measures.

Additional Characteristics:

- Once users have consented to sharing their data with industry partners, War on Cancer strives to return any results to patients and keep them informed about how their data are being used.

---

20 "About us," War on Cancer, accessed May 1, 2020, https://waroncancer.com/about/.
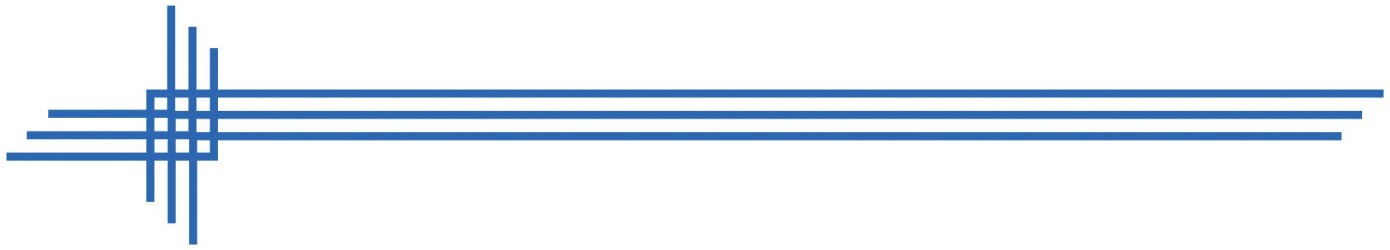
CONCLUSION

There is no question that patient participation in research is of paramount importance so that patients can help drive research agendas and reap the benefits of new discoveries. However, much work remains to achieve the goal of providing patients with personalized insights about their health derived from their contributed data. Rethinking partnerships between data platforms and consumers to ensure a higher level of transparency may help, because a lack of trust among the public is a major concern of companies that hope to use data to accelerate research.

Ultimately, development of new policies and data platforms will require the utmost thought and care, especially in the context of COVID-19. In this new context, these policies and platforms may be developed more quickly and with less oversight than normal; they will also affect hundreds of millions of people and have already disproportionately disadvantaged the most marginalized groups in society. Furthermore, the policies and platforms we develop are not likely to simply disappear after the immediate threat of the pandemic passes; rather, they will lay the groundwork for the policies and infrastructures of the future.[21] We must get this right.

---

21 Yuval Noah Harari, "Yuval Noah Harari: The World After Coronavirus," *Financial Times*, March 20, 2020, https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75.

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Rachel Tunis is an associate at FasterCures, a Center of the Milken Institute, primarily providing research support to core programmatic areas related to health data and technology. Tunis enjoys using social science research approaches to better understand complex problems in the biomedical research system. She holds a BA from the College of William & Mary in sociology and French.

Christine Bechtel is the co-founder of X4 Health, a purpose-driven organization that is carving new pathways to a better system, fueled by human-centered innovation. Bechtel is a well-known consumer advocate with expertise in areas ranging from social impact design and movement building, to health data, patient and family engagement, and quality measurement and improvement. An avid golfer, she lives in Maryland.