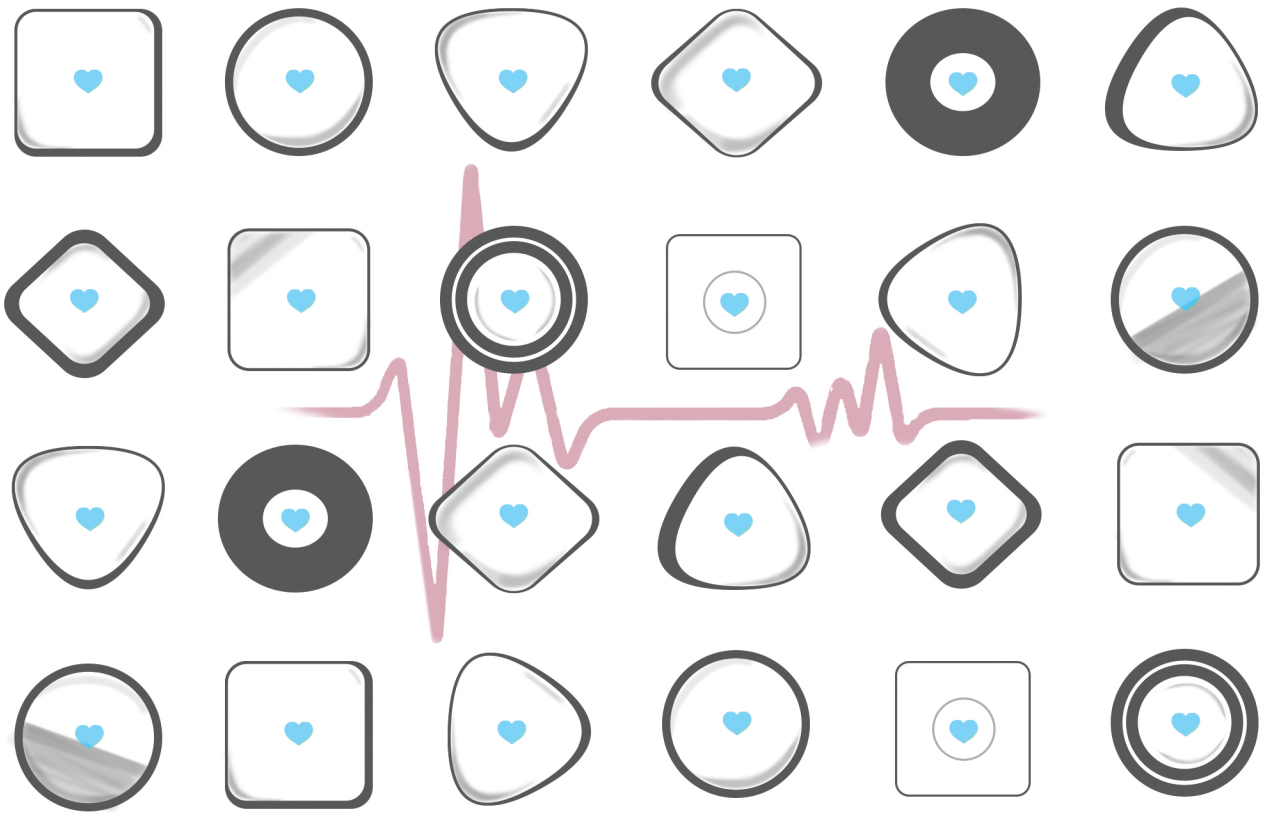
HEAL -GORITHMS:



Understanding the Potential for Bias in mHealth Apps

JULY 2018

CENTER FOR DEMOCRACY & TECHNOLOGY

I. Highlights 2

II. Overview of mHealth 3
A. Data Flows in mHealth 7
B. Demographics of Health App Users 7
C. Regulatory Gaps in mHealth 8
D. Categories of Health Apps and Their Regulatory Status 11

III. Efficacy of mHealth Apps 13

IV. Bias in mHealth Apps 15
A. The Potential for Bias in mHealth 15
B. Development of mHealth Algorithms and Areas Where Bias May Be Introduced 16

V. Recommendations 21
A. Ideating and Designing the App 21
B. Selecting/Designing Health Interventions 22
C. Identifying User Populations and Designing for Inclusivity 22
D. Assembling the Team and Seeking Input 23
E. Designing the User Interface and Features 23
F. Developing the Models that Power Health Interventions 24
G. Testing the Models and Evaluating the App 26

VI. Areas for Further Investigation 27
A. Standardized Metrics 27
B. Evidence-Based Open Datasets 27
C. Policy Framework Addressing Fairness and Equality for Health Data 28

VII. Conclusion 28

Cover designed by Timothy Hoagland

Support provided by Robert Wood Johnson Foundation

CENTER FOR DEMOCRACY & TECHNOLOGY

CENTER FOR DEMOCRACY & TECHNOLOGY

Heal-gorithms: Understanding the Potential for Bias and Discrimination in mHealth Apps

Highlights

For *Policymakers*

- An mHealth app built on biased data or biased assumptions can disproportionately harm – or fail to help – groups that are historically underrepresented or underserved.
- In current U.S. law, protection of information collected through mHealth apps is determined by the provider's terms of service and privacy policy (even where mHealth apps might be covered by HIPAA, that law seeks to ensure the confidentiality of patient data and does not address bias). New rights under the GDPR could impose new requirements on mHealth developers to be able to explain how an app produces health recommendations.
- Policymakers should consider the creation of a data governance framework for fairness and health data equality that can be used by patients, commercial app developers, and lawmakers.

For *App Developers*

- Any bias introduced during data collection, design, development, or testing of an app can produce health interventions that work better for some groups (usually majority groups) than for others (groups that are underrepresented in the training data, historically marginalized and underserved in healthcare and health research, and/or underrepresented among mHealth app developers).
- Biased mHealth outcomes cannot always be traced back to the ML model itself or even the training data. They can also stem from design choices such as the way health interventions are presented to users or the accessibility of a user interface.
- Users of varying ages, ethnicities, incomes, and educational backgrounds cited buried in-app costs, concerns around privacy and security practices, the time burden associated with inputting data, and a lack of interoperability with healthcare providers as causes for their dissatisfaction with and discontinued use of health apps.

For *Business Managers*

- Only 4% of mHealth apps were able to generate a million downloads in 2017, with most apps downloaded less than 5,000 times, and the majority of mHealth apps lack a cohort of active monthly users.
- Though Hispanic and black smartphone users have signalled a strong interest in and comfort with digital health tools, studies show that most mHealth app users skew white, suggesting the existence of a substantial and under-tapped market for inclusive and user-centric mHealth apps.

1401 K Street NW, Suite 200 Washington, DC 20005

- Responsible and ethical app development practices are currently one of the few backstops available to prevent the automated discrimination in mHealth that might stall user growth in diverse communities.

For *Researchers*

- Standard metrics for successful interventions are also needed so that research studies evaluating and comparing different health apps and mHealth interventions can better help developers understand how to design effective tools.
- Research should further explore how to use a "living systematic review" model that can synthesize health research data in real-time so that developers can test their apps against existing knowledge and research.
- To both broaden the appeal of mHealth apps for different populations and better design responsive interventions, research should also consider more deeply how demographics, UI/UX design, and other factors impact the efficacy of interventions delivered by mHealth apps.

I. Introduction

Technology is putting health data directly in the hands of individuals, allowing them to gain new insight into their bodies and minds. The mobile health ("mHealth") industry is at the forefront of this new health frontier, offering a wealth of tools for improving personal health and wellness. These tools include fitness and nutrition trackers, chronic disease management platforms, and cognitive behavioral therapy (CBT) apps for things like smoking cessation and weight loss. The health data generated by consumer-facing apps, devices, and platforms, however, raise important questions about personal privacy and inclusion, particularly when this data is the basis for statistical analysis and prediction that may directly impact an individual's health, and feeds into public health research that informs broader medical practices.[1]

Trends in data-driven personalization have placed an emphasis on applying machine learning techniques to available data to create tailored and adaptable health interventions. Some mHealth apps use automated interventions to mediate access to relevant health information, advocate for a particular course of treatment, or identify patterns indicating the regression or progression of disease.[2]

mHealth apps have the potential to improve access to quality care and information, and there is some evidence that they can help support positive behavioral changes through

[1] Megan Molteni, *Health Care is Hemorrhaging Data. AI is Here to Help*, Wired (Dec. 30, 2017) https://www.wired.com/story/health-care-is-hemorrhaging-data-ai-is-here-to-help/.

[2] Madhukar H. Trivedi et al., *Assessing Physicians' Use of Treatment Algorithms: Project IMPACTS Study Design and Rationale*, 28(2) Contemporary Clinical Trials 192 (2007), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2793279/. For our purposes, clinical practice guidelines are a related system. Joseph H. Keffer, *Guidelines and Algorithms: Perceptions of Why and When They Are Successful and How to Improve Them*, 47(8) Clinical Chemistry 1563 (2001) http://clinchem.aaccjnls.org/content/47/8/1563.

self-management strategies, such as tracking symptoms, receiving timely medication reminders, or adhering to treatment recommendations. However, they also carry risks of producing interventions based on biased data, historical discrimination, or incorrect assumptions. These risks are disproportionately borne by historically marginalized groups, including people of color, women, and the LGBTQ community, as well as those who are in poverty, unhoused, or disabled. There is some evidence that longstanding systemic health disparities might be ameliorated through accessible health technologies like mHealth apps, which offer mHealth companies the opportunity to improve community health and establish a market advantage. Perhaps adding to the appeal of entering this market, there are few, if any, regulations governing the fairness or inclusivity of automated interventions, lowering compliance costs that can create a high burden to entry for companies. For most mHealth apps in the U.S., there are few, if any, regulations governing the fairness or inclusivity of automated interventions. Responsible and ethical app development practices are currently one of the few backstops available to prevent automated discrimination that might discourage user engagement from different communities.

This report explores the potential for harmful bias in mHealth interventions and considers the impact of such bias on individuals, companies, and public health, ultimately providing recommendations for app developers to ensure that the tools they build are inclusive and nondiscriminatory. This report seeks to advance the conversation about — and implementation of — equity and inclusivity in automated decisions in the health sector in ways that benefit both the public and the companies using data to make decisions by: (a) providing a landscape of the mHealth ecosystem; (b) synthesizing research and investigations to draw out key issues and concerns related to bias in automated decision-making in the commercial health context; and (c) making recommendations that advance identification and mitigation of bias and discrimination in processes that produce commercial health app content.

Part II of this report provides an overview of the mHealth marketplace, covering the types of mHealth apps available, how data flows in and out of these apps, who uses these apps, how these apps are regulated, and how effective these apps are. **Part III** discusses the efficacy of mHealth and suggests that reducing bias is vital to delivering effective health interventions with these tools. **Part IV** examines how and when bias can be introduced into mHealth interventions. **Part V** provides a recommended roadmap of inquiry for developers and others involved in mHealth to identify and mitigate bias. **Part VI** is a review of areas for future research, and **Part VII** is a brief conclusion.

II. Overview of mHealth

mHealth is an umbrella term for health-related mobile devices and services.[3] This report is primarily focused on consumer-facing, commercially available mobile apps that offer health interventions. This can include tools such as fitness, calorie, or sleep trackers; symptom

[3] *See, e.g.*, World Health Organization, mHealth: New Horizons for Health Through Mobile Technologies (2011), http://www.who.int/goe/publications/goe_mHealth_web.pdf.

monitors; chronic disease and mental health management tools; and apps that provide educational information, recommend treatments, create customized fitness and nutrition plans, or even offer local health care providers and specialists. Data and devices covered by the Health Insurance Portability and Accountability Act (HIPAA) are outside the scope of this report. Our discussion and recommendations around bias in mHealth apps focus particularly on apps that use machine learning models to provide health interventions; however, the findings in this report are not limited to machine learning.

The U.S. and global mHealth markets have grown substantially in recent years, and that growth is expected to continue, accounting for billions of dollars of sales annually.[4] A 2017 report found 325,000 health apps available across all major apps stores.[5] One survey found that 63% of Americans would use "digital personal health management" tools to manage health-related issues.[6] Despite these encouraging signs, companies publishing mHealth apps face challenges in both recruiting and keeping active users. One report found that just 4% of mHealth apps were able to generate a million downloads in 2017, with most apps downloaded less than 5,000 times, and the majority of mHealth apps lack a cohort of active monthly users.[7] Users of varying ages, ethnicities, incomes, and educational backgrounds cited buried in-app costs, concerns around privacy and security practices, the time burden associated with inputting data, and a lack of interoperability with healthcare providers as causes for their dissatisfaction with and discontinued use of health apps.[8] It's striking that diverse communities have demonstrated a high interest in and comfort with digital health tools yet they remain underrepresented in mHealth user demographics. It suggests that a substantial and under-tapped market is waiting for inclusive and user-centric developers.

---

[4] *See, e.g.*, Kalorama Information, *mHealth Markets Worldwide: Forecasts to 2020* (2015) (results reported by Sara Heath), *mHealth App Market Sees $400 Million Growth in Five Years*, mHealth Intelligence (Sep. 22, 2015), https://mHealthintelligence.com/news/mHealth-app-market-sees-400-million-growth-in-five-years; Grand View Research, *mHealth Apps Market Size Worth 111.8 Billion by 2025 | CAGR 44.2%*, (Aug. 2017), https://www.grandviewresearch.com/press-release/global-mHealth-app-market; Reuters Brand Feature, *mHealth Market Worth $23 Billion in 2017 and Estimated to Grow at a CAGR of more than 35% over the next three years* (Apr. 18, 2017), https://www.reuters.com/brandfeatures/venture-capital/article?id=4640.

[5] Research2Guidance, *mHealth App Economics*, (Nov. 2017), http://www.uzelf.org/wp-content/uploads/2017/12/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf.

[6] Makovsky, Pulse of Online Health Survey, (Feb. 24, 2015), http://www.makovsky.com/news/fifth-annual-pulse-of-online-health-survey-2/.

[7] Research2Guidance, *mHealth Developer Economics, How mHealth app publishers are monetizing their apps* (March 2018), https://research2guidance.com/wp-content/uploads/2018/03/R2G-mHealth-Developer-Economics-2017-How-To-Monetize-Mobile-Health-Apps.pdf.

[8] Paul Krebs and Dustin Duncan, *Health App Use Among US Mobile Phone Owners: A National Survey,* JMIR mHealth Uhealth 2015;3(4):e101, http://mHealth.jmir.org/2015/4/e101.

CENTER FOR DEMOCRACY & TECHNOLOGY

The 20 most popular health apps on iPhone and Android* are listed in the chart below:

iPhone		Android	
Free	**Paid**	**Free**	**Paid**
Sweatcoin - Sweat for Coin	Full Fitness: Exercise Workout Trainer	Six Pack in 30 Days - Abs workout	The Wonder Weeks
BetterMe: Weight Loss Workouts	The Wonder Weeks	Easy Workout - Abs & Butt Fitness, HIIT Exercises	Couch to 5K®
Calm	AutoSleep Tracker for Watch	Step Tracker - Walking for weight loss, Pedometer	White Noise
MyFitnessPal	7 Minute Workout Challenge	Sweatcoin Pays You To Get Fit	My Macros+ | Diet, Calories & Macro Tracker
Weight Loss Fitness by Verv	My Macros+ | Diet & Calories	Home Workout - No Equipment	Ref Guide for Essential Oils
30 Day Fitness Challenge Log	Couch to 5K® - Run training	Fitbit	Period Tracker Deluxe
Flo Period & Ovulation Tracker	HeartWatch. Heart & Activity	Calm - Meditate, Sleep, Relax	WeCroak
Fitbit	KetoDiet	Calorie Counter - MyFitnessPal	C25K® - 5K Running Trainer Pro
Headspace: Meditation	CRON-O-Meter	Headspace: Meditation & Mindfulness	Ultimate Food Value Diary - Diet & Weight Tracker
Workout for Women: Fitness App	Period Tracker Deluxe	Lose Belly Fat in 30 Days - Flat Stomach	The Blood Type Diet®
Lose It! - Calorie Counter	White Noise	Abs Workout - Weight Loss App, Tabata, HIIT	Pupil Distance Meter Pro | Accurate PD measure

CENTER FOR DEMOCRACY & TECHNOLOGY

iPhone		Android	
Free	**Paid**	**Free**	**Paid**
Nike + Run Club	Ref Guide for Essential Oils	Sleep Music - Relax Soft Sleep Sounds	Young Living Essential Oils Reference Guide
Freeletics Bodyweight	MyFlo Period Tracker	AllTrails: Hiking, Running & Mountain Bike Trails	The Fast Metabolism Diet
Clue - Period & Health Tracker	WeCroak	Carb Manager - Keto & Low Carb Diet Tracker	KetoDiet
AllTrails: Hike, Bike & Run	Sleep Tracker: by Sleepmatic	Six Packs for Man - Body Building with No Equipment	Paleo (io) The Paleo Food List
Weight Loss Running by Verv	Fitness Buddy+: Gym Workouts	Period Tracker Flo, Pregnancy & Ovulation Calendar	buddhify - mindfulness meditation on the go
8fit Workouts & Meal Planner	Streaks	Lose It! - Calorie Counter	Just 6 Weeks
Map My Run by Under Armour	Pocket Yoga	Runtastic Running App & Fitness Tracker	Rainy Mood
Adhere Fitness	Weight Loss Running PRO	BetterMe: Weight Loss Workouts	Runtastic PRO Running, Fitness
Carb Manager: Keto Diet App	Blood Type Diet®	Lose Weight in 30 Days	FitNotes Supporter

**As of June 28, 2018*

Data Flows in mHealth

mHealth apps may be developed using data from a variety of disparate sources, including academic and clinical studies, commercial and government databases, data brokers, and data collected directly from consumers through the use of their mobile devices and apps.[9] These disparate sources of data may be combined and used to train and/or validate a machine learning (ML) model. Developing ML models typically requires large amounts of data.

mHealth apps typically collect information from their users to track health-related metrics to produce personalized interventions or other outputs, and for other commercial purposes such as targeted advertising. This information may be input directly by the user or collected through a mobile phone's sensors.[10] Sensors may be harnessed by mHealth apps to measure an individual's movement (including the number of steps taken), heart rate, and location. An app may request permission to access other device functions, such as the microphone or camera, to enable app features. An app may also collect data unrelated to an app's features, sometimes called "over-privileging," such as information about a user's internet connection or contact lists.[11] Apps may also connect to and collect information from other devices, including health devices for managing chronic conditions.[12] Some mHealth app providers may also sell this data to third parties or share it with their partners or vendors.[13]

Demographics of Health App Users

Studies indicate that mHealth users tend to be young, highly educated, healthy, and have high incomes.[14] A 2013 Pew survey of smartphone owners found that those who are under 50, who are college educated, and/or whose income is $75,000 or more, downloaded a health app more

---

[9] *See, e.g.* Philip Hunter, The Big Health Data Sale, EMBO Rep. (July 11, 2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4967957/; Adam Tanner, *How Data Brokers Make Money Off Your Medical Records*, Scientific America (Feb. 1, 2016), https://www.scientificamerican.com/article/how-data-brokers-make-money-off-your-medical-records/.

[10] Mobile phone sensors include an accelerometer, a gyroscope (which measures which way the phone is oriented), a magnetometer, a global positioning system (GPS), a barometer, and an ambient light sensor. David Nield, *All the Sensors in Your Smartphone, and How They Work*, Gizmodo (July 23, 2017), https://fieldguide.gizmodo.com/all-the-sensors-in-your-smartphone-and-how-they-work-1797121002.

[11] Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney, *Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps*, Technology Science, (October 10, 2015), https://techscience.org/a/2015103001/index.php#Demonstration.

[12] *See, e.g.*, Quell, https://www.quellrelief.com.

[13] *See* Jinyan Zang et al., *supra* note 11 (According to an analysis tool provided by the authors, on July 17, 2018, Nike+ Running app on iOS sent the following user information to third parties: location to apple.com, wunderground.com, and exacttargetapis.com; and gender to urbanairship.com.

[14] Jennifer K Carroll et al., *Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach*, J Med Internet Res. (Apr. 2017), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5415654/; Susannah Fox and Maeve Duggan, Tracking for Health, Pew Research Center (2013) http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_TrackingforHealth%2520with%2520appendix.pdf.

often than other groups.[15] Another study based on data from 2015 also concluded that "the main users of health apps were individuals who were younger, had more education, reported excellent health, and had a higher income."[16] However, a 2015 Pew survey found that both black and Hispanic smartphone owners were more likely than white smartphone owners to use their phones to research a health condition. 73% of Hispanic smartphone owners and 67% of black smartphone owners used their phone in the preceding year to investigate a health condition, compared to 58% of white smartphone owners.[17] Finally, studies from 2017 observed that age and education were the primary significant variables for predicting whether a person had adopted a mobile device at all (a prerequisite for using mHealth apps).[18] Again, these investigations suggest a mismatch between user interest and needs in the current mHealth market, one that may shed light on low download rates and sluggish growth among active monthly mHealth app users.

Regulatory Gaps in mHealth

In most cases, mHealth apps are regulated lightly, if at all. The primary U.S. health privacy law, HIPAA, applies only to data transferred to or from "covered entities" (e.g., health plans, health care clearinghouses, and health care providers) and their business associates.[19] Health apps that are downloaded and used by consumers, without the direction or association of a covered entity, do not fall under HIPAA.[20] All of the top 20 health apps on both iPhone and Android fall outside the scope of both HIPAA and the Food and Drug Administration's (FDA) regulatory authority, the principal regulator of health care devices. This leaves the status and protection of information collected through mHealth apps up to the provider's terms of service and privacy policy. Even where mHealth apps might be covered by HIPAA, that law seeks to ensure the confidentiality of patient data and does not address bias.

Recent legislation has sought to promote mHealth development by lowering regulatory barriers related to data and device practices. The 21st Century Cures Act exempted certain kinds of software from the definition of a medical device under the Federal Food, Drug, and Cosmetics

---

[15] Fox & Duggan, *supra* note 14.

[16] Carroll et al., *supra* note 14.. However, the study indicated that "individual sociodemographic factors" like gender, race, and educational attainment were only "mild to moderately" predictive of engagement with mobile health apps.

[17] Monica Anderson, Racial and Ethnic Differences in How People of Color Use Mobile Technology, Pew Research Center (Apr. 2015), http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/.

[18] Carroll et al., *supra* note 14.

[19] 45 C.F.R. § 160.103 (2013).

[20] Department of Health and Human Services, *Health App Use Scenarios & HIPAA* (Feb. 2016), https://hipaaqsportal.hhs.gov/community-library/accounts/92/925889/Public/OCR-health-app-developer-scenarios-2-2016.pdf. The Apple & Google App Stores allow users to change countries, and therefore broaden or restrict access to certain apps that may be regulated by the country identified by the App Store selection.

Act.[21] This exemption includes software that is intended "for maintaining or encouraging a healthy lifestyle and is unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition."[22] The FDA has also signalled that it will not regulate some apps that do qualify as medical devices, because the agency has determined that they pose a lower risk to the public.[23] Apps in this category include those that facilitate behavioral change. According to FDA guidance from 2015, "the majority of mobile apps on the market at this time fit into these two categories" (i.e., non-medical devices or low-risk medical devices).[24]

The Federal Trade Commission (FTC) may also exercise authority over software companies through its Section 5 authority over unfair or deceptive practices.[25] For example, if an mHealth provider were to make false promises or omit material information about its app, including its data privacy and security practices, the FTC could bring an enforcement action for deceptive trade practices. The FTC could also take action against unfair trade practices of mHealth providers.[26] While it is possible that the FTC could one day take enforcement action against biased, unreliable, or discriminatory algorithms under a theory of deception or unfairness, the Commission has not yet used its Section 5 authority in this way. The FTC did issue a rule that would require health apps to notify users in the event of a data breach.[27]

---

[21] 21st Century Cures Act, Pub. L. No. 114-255, § 3059(a), 130 Stat. 1033, 1130 (2016) (Codified as amended at 21 USC 360), https://www.congress.gov/bill/114th-congress/house-bill/34/text#toc-H73E766CC72AC4EB8AAF36670EB540164.

[22] *Id.*

[23] Food and Drug Administration, *Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff*, (2015) (hereinafter FDA), https://www.fda.gov/downloads/MedicalDevices/.../UCM263366.pdf.

[24] *Id.* (parenthetical added). The FDA's existing approval process for health devices and products is problematic with regard to developers' use of analytics systems. The approval framework assumes that devices and products, for example, can be tested in one defined and unchanging way. Today's sensors and devices are built to provide continuous data streams, and machine learning models can be designed to change continuously in response to new data or feedback. Applying a static test to determine compliance would not offer much protection against bias. In part to address challenges like these, the FDA launched a program in 2017 called "FDA Pre-Cert" aimed at easing regulatory burdens around approval for health IT software developers. U.S. Food & Drug Admin., Digital Health Software Precertification (Pre-Cert) Program (https://www.fda.gov/medicaldevices/digitalhealth/digitalhealthprecertprogram/default.htm).

[25] 15 U.S.C. § 45.

[26] For a practice to be unfair, it must cause an injury that is 1) objectively substantial, 2) without offsetting benefits, and is 3) not reasonably avoidable by consumers. This standard can be difficult to meet, and some commenters are wary of the FTC expanding the scope of its unfairness enforcement. *See* Maureen K. Ohlhausen, *Section 5 of the FTC Act: principles of navigation*, J. of Antitrust Enforcement 1 (2013), https://www.ftc.gov/system/files/documents/public_statements/section-5-ftc-act-principles-navigation/131018section5.pdf; J. Howard Beales, *The FTC's Use of Unfairness Authority: Its Rise, Fall, and Resurrection*, Marketing & Pub. Pol'y Conference (2003), https://www.ftc.gov/public-statements/2003/05/ftcs-use-unfairness-authority-its-rise-fall-and-resurrection

[27] For a description of the rule, *see* Federal Trade Commission, Complying with the FTC's Health Breach Notification Rule, 16 CFR § 318 2010, https://www.ftc.gov/tips-advice/business-center/guidance/complying-ftcs-health-breach-notification-rule).

U.S. states also have laws that may be relevant to health data but most address information privacy and are not designed to address bias. Some states have laws governing medical records that may be stricter or broader than HIPAA. One prominent example is the California Medical Information Act (CMIA), which extends confidentiality requirements to "[a]ny business that offers software or hardware to consumers, including a mobile application or other related device that is designed to maintain medical information" for a variety of purposes including "the diagnosis, treatment, or management of a medical condition of the individual."[28] However, like HIPAA, the CMIA is concerned with the confidentiality of medical data, not the fairness or inclusivity of analyses or recommendations that rely on that data. States also have consumer protection laws that empower state authorities, such as the attorney general, to bring lawsuits against companies that harm consumer welfare, including in the healthcare field. These laws tend to be similar to the FTC mandate to combat "unfair and deceptive practices," the scope of which varies by state.[29] New York City has an ordinance creating a task force to study government use of automated decision systems, including how they might be biased,[30] but as of now, no state has passed a law to directly address algorithmic bias in a purely private-sector context. At least in the U.S., mHealth apps and the algorithms that power their recommendations fall into a regulatory gray zone and have received comparatively little scrutiny compared to traditional healthcare providers.[31] This puts companies and app developers on the front lines of protecting sensitive health information and ensuring that mHealth interventions are reliable, inclusive, and equitable.

The General Data Protection Regulation (GDPR) that recently took effect in the European Union (EU) could significantly impact mHealth apps that collect or process data on EU residents. The GDPR's stringent requirements on data processing, which includes how data is collected, used, shared, or retained,[32] will impose new requirements on mHealth companies that serve EU residents and rely on personal data. Early qualitative surveys have suggested that mHealth apps are deficient in their management of private information under the GDPR.[33] Implementation of more technical aspects of the new regulation could prove more challenging.

Specifically, Article 22 of the GDPR gives data subjects, which include the consumers of mHealth apps, the right not to be subject to automated decision-making, including profiling, without

[28] Cal. Civ. Code § 56.06(a) (2017), http://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=56.06.&lawCode=CIV.
[29] Public Health Law Center, *Consumer Protection: An Overview of State Laws and Enforcement* (2010), http://www.publichealthlawcenter.org/sites/default/files/resources/phlc-fs-agconsumer-2010.pdf.
[30] Automated decision systems used by agencies, Int. No. 1696-A (2018), http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0; see also Galen Stump, New York City's for Accountable Algorithms, Fordham Intell. Prop. Media & Ent. L.J. (Apr. 4 2018), http://www.fordhamiplj.org/2018/04/04/new-york-citys-push-for-accountable-algorithms/.
[31] *See* Nicolas P. Terry, *Big Data and Regulatory Arbitrage in Healthcare, in* Big Data, Health Law, and Bioethics 56 (I. Glenn Cohen, Holley F. Lynch, Effy Vayena, & Urs Gasser eds. 2018) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2821964.
[32] Commission Regulation 95/46, art. 4, 2016 O.J. (L119) 1 (EC).
[33] A. Papageorgiou, M. Strigkos, E. Politou, E. Alepis, A. Solanas and C. Patsakis, *Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice*, in IEEE Access, vol. 6 9390-9403 (2018).

appropriate safeguards.[34] Profiling-based decisions using special categories of information like health, sex life, or sexual orientation require either a data subject's explicit consent or the processing must be in the substantial public interest.[35] The GDPR also gives data subjects a right to "meaningful information" about automated decisions (often referred to as the "right to explanation"), though there is still uncertainty around when information is required and what "meaningful information" includes.[36] The right to meaningful information under the GDPR could impose new requirements on mHealth developers to be able to explain how an app produces health recommendations. Recital 71 of the Regulation, which seeks to provide context but is not itself law, mentions the use of "technical and operational measures to . . . prevent, *inter alia*, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status, or sexual orientation."[37] Despite the law's commitment to fairness and transparency, scholars have expressed doubt about whether the full suite of GDPR provisions can fully combat algorithmic bias.[38] Thus, the impact of the GDPR on mHealth apps remains unclear.

Categories of Health Apps and Their Regulatory Status

Most consumer-facing health apps fall into the following general categories:

Health Reference Apps: These apps allow users to look up health information dynamically downloaded from the internet or pre-loaded at the time of app installation. Individuals can use these apps to learn about medical conditions or research providers in their area.[39] Some apps, like ZocDoc, recommend providers based on user reviews and other factors, and allow individuals to schedule appointments through the app.[40] These apps are unregulated by the FDA or HIPAA.

Fitness Tracking Apps: These apps track individuals' health and fitness data over time. Apps like Fooducate,[41] ShopWell,[42] and Zipongo[43] use user inputs about food intake, weight, and lifestyle to recommend healthy foods and recipes. Apps such as Pact allow users to track their fitness

---

[34] Commission Regulation 95/46, art. 22, 2016 O.J. (L119) 1 (EC), https://gdpr-info.eu/art-22-gdpr/.
[35] *Id.* (4).
[36] Compare Sandra Wachter, Brent Mittelstadt, & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, Harv. J.L. & Tech. (forthcoming 2018) https://arxiv.org/ftp/arxiv/papers/1711/1711.00399.pdf with Bryce Goodman & Seth Flaxman, *European Union regulations on algorithmic decision-making and a "right to explanation"*, 38 AI Magazine 1 (2017), https://arxiv.org/pdf/1606.08813.pdf. *See also* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to Explanation' is Probably Not the Remedy You are Looking For*, 16 Duke L. Tech. Rev. 18 (2017).
[37] Commission Regulation 95/46, rec. 71, 2016 O.J. (L119) 1 (EC), https://gdpr-info.eu/recitals/no-71/.
[38] *See* Goodman & Flaxman, *supra* note 36, https://arxiv.org/pdf/1606.08813.pdf.
[39] *See* Anne Marie Helm & Daniel Georgatos, *Privacy and mHealth: How Mobile Health Apps Fit into a Privacy Framework Not Limited to HIPAA*, 64 Syracuse L. Rev. 131, 170 (2014).
[40] ZocDoc, https://www.zocdoc.com/.
[41] Fooducate, https://www.fooducate.com/.
[42] ShopWell, http://www.shopwell.com/mobileapp.
[43] Zipongo, https://meetzipongo.com/.

activity and recommend individual fitness plans and exercises.[44] BabyBump relies on user-entered information to calculate a due date and provide weekly tips for expecting mothers.[45] Some apps may "integrate alarms, timers, reminders, and other interactive features, while others try to 'gamify' health care by using rewards systems, point-tracking, and challenges to encourage healthier behavior."[46] Tracking apps also include those linked to wearable devices, including Fitbit[47] and Garmin,[48] that passively collect information on individuals and recommend new fitness goals. Some smartphones also come with built-in tracking features, such as the Apple iPhone Health App.[49] This category of apps is also not covered by HIPAA or FDA regulations.

Diagnostic Apps: Diagnostic apps use information provided by individuals to calculate results based on "commonly used reference information" or "simple calculations routinely used in clinical practice."[50] For example, symptom checkers, like iTriage[51] and WebMD's Symptom Checker,[52] allow users to input their symptoms, which are run against a database to determine potential medical issues. Depression Test is an app that administers a standard questionnaire to diagnose depression that follows criteria found in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[53] Some apps even allow users to submit pictures that are then algorithmically assessed to generate potential diagnoses. For example, Miiskin is an app that allows users to track changes on their skin over time to detect signs of melanoma, rather than diagnose the illness.[54] Importantly, the FDA would have enforcement authority over apps that use images to make diagnoses. HIPAA would not apply unless the app were administered by a covered entity or business associate.

Disease Management Apps: These apps help patients manage diseases, typically based on user-entered information and established practices and guidelines. This type of app would, for example, help "heart disease patients create a diet based on published nutritional guidelines."[55] Such apps could also download information from a medical device and use it for basic disease management without falling under FDA enforcement. Though many of these apps could fit the FDA's definition of a medical device, either on their own or as an accessory to a regulated

---

[44] See Catherine Shu, *GymPact, the App that Pays You for Working Out, is Relaunching as Pact with New Diet Features*, TechCrunch (Jan. 1, 2014), https://techcrunch.com/2014/01/01/pact/.
[45] See Twitter, @babybumpapp, https://twitter.com/babybumpapp?lang=en.
[46] Nathan Cortez, *The Mobile Health Revolution?*, 47 U.C.D. L. Rev. 1173 (2014).
[47] Fitbit, https://www.fitbit.com/home.
[48] Garmin, https://buy.garmin.com/en-US/US/wearabletech/wearables/c10001-c10002-p1.html.
[49] Apple iOS Health, https://www.apple.com/ios/health/.
[50] FDA, *supra* note 23, at 17.
[51] iTriage, https://www.itriagehealth.com/avatar (note that this app may no longer be on the market).
[52] WebMD Symptom Checker app, https://www.m.webmd.com/symptomcheckerapp.
[53] Depression Test Questionnaire, Patient.info https://patient.info/doctor/patient-health-questionnaire-phq-9 (last visited July 3, 2018).
[54] Miiskin, http://miiskin.com/.
[55] *See* Vincent J. Roth, *The mHealth Conundrum: Smartphones & Mobile Medical Apps - How Much FDA Medical Device Regulation Is Required*, 15 N.C. J.L. & Tech. 359, 424 (2014) (citing Scott D. Danzis & Christopher Pruitt, Note, *Rethinking the FDA's Regulation of Mobile Medical Apps*, 9 No. 4 ABA Scitech Law 26, 27 (2013)).

device, as noted above, the FDA has decided not to regulate these mobile health apps unless their "functionality could pose a risk to a patient's safety if the mobile app were to not function as intended."[56] An app that calculated changing medication dosages (which would pose a safety risk if it malfunctioned) would warrant FDA regulation. As with diagnostic apps, in the absence of a covered entity or business associate HIPAA would not apply to data collected through these apps.

Apps that address mental health issues in novel ways fall into this category.[57] For example, Sleepio (beta), administers cognitive-behavioral therapy (CBT) to treat insomnia.[58] Users are given assignments over time to improve their sleep and update information about their sleep and lifestyle habits in the app. Woebot, available on Facebook Messenger, assesses a user's emotional profile through a series of questions and applies a CBT algorithm to recommend activities that improve mood.[59] Some apps, like Ginger (beta) and BiAffect (beta) use patterns of smartphone use, including factors like a drop in the frequency of outgoing text messages, typing speed, use of spell check, how hard keys are pressed, and backspace usage, to diagnose mental health disorders or identify episodes.

III. Efficacy of mHealth Apps

Though mHealth apps may have the potential to improve equitable access to health information and empower people to manage their own health and well-being, it's not clear whether they are effective at producing good health outcomes. Since mHealth is a relatively new phenomenon, so are systematic studies analyzing apps' efficacy. A 2013 survey of mHealth apps found that apps and interventions were not adequately tested for efficacy, and thus it was "still unknown . . . whether mHealth leads to better overall health outcomes."[60] A 2017 meta-analysis of mHealth apps for diabetes found that "the quality of reviews varied considerably and most of them had important methodological limitations."[61] Standard metrics for successful interventions are needed so that research studies evaluating and comparing different health apps and mHealth interventions can better help developers understand how to design effective tools. Much more research is necessary in this area.

---

[56] FDA, *supra* note 23, at 7.
[57] *See* Emily Anthes, *Mental Health: There's an App for That*, Nature Magazine (Apr. 6, 2016) available at https://www.scientificamerican.com/article/mental-health-there-s-an-app-for-that/.
[58] Sleepio, https://www.sleepio.com/.
[59] Woebot, https://woebot.io/ (last visited July 3, 2018).
[60] Santosh Kumar et al., Mobile Health Technology Evaluation: The mHealth Evidence Workshop, Am. J. Prev. Med. 45(2) 228–36 (Aug. 2013), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3803146/.
[61] Spyros Kitsiou et al., *Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews*, PLOS One (2017) (Nearly half of the SRs (n = 6; 40%) are characterized by important methodological limitations and risks of bias that impair their internal validity and limit their usefulness for clinical and policy decision-making purposes).

Some studies have found evidence of improved outcomes from mHealth apps for conditions ranging from type II diabetes to mental health.[62] At best, however, these studies show mixed results, which vary depending on the type of intervention (e.g., text message reminders vs. activity logging), amount and frequency of engagement, interaction with healthcare professionals, and other factors. For example, one meta-analysis found that mHealth apps could modestly improve diabetes patients' glycemic control, but the results varied depending on whether the app relied on text reminders alone or combined texts with self-monitoring tools and health education, and on the frequency of intervention (e.g., daily vs weekly).[63] The authors also found that, contrary to prior research, there is "insufficient evidence to determine whether mHealth interventions that are adaptive in nature or based on behavioral change theories are more effective than others."[64] Another study reported that text message interventions can help with smoking cessation, but found "mobile technology-based interventions for diabetes control that have statistically significant effects are small and of borderline clinical importance."[65]

Even where mHealth apps can improve health outcomes, the benefits may not be equitably distributed. Lack of access to the internet or smartphones, disability, and lack of English-language proficiency or literacy are all barriers to taking advantage of mHealth apps. What we know about the demographics of mHealth app users suggests that mHealth benefits are concentrated among higher income, younger, and healthier people. Studies of mHealth efficacy typically do not show efficacy rates broken down by demographic groups.

As the next section will discuss, efficacy can be confounded by bias in mHealth apps. Any bias introduced during data collection, design, development, or testing of an app can produce health interventions that work better for some groups (usually majority groups) than for others (groups that are underrepresented in the training data, historically marginalized and underserved in healthcare and health research, and/or underrepresented among mHealth app developers). Decades of race discrimination in America have resulted in wide disparities in access to quality health care and information. According to a report from the Joint Center for Political and Economic Studies, communities of color lack critical information about obtaining

---

[62] *Id.*; Daniel R. Bateman et al., *Categorizing Health Outcomes and Efficacy of mHealth Apps for Persons With Cognitive Impairment, a Systematic Review*, 19(8) J Med Internet Res 301 (2017) https://www.ncbi.nlm.nih.gov/pubmed/28855146; Amy Leigh Rathbone, *Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review*, 19(11) J Med Internet Res 399 (2017) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5727354/. *But see* John M. Jakicic, Kelliann K. Davis, Renee J. Rogers, *Effect of Wearable Technology Combined with a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial*, 316 (11) JAMA 1161 (2016), https://jamanetwork.com/journals/jama/fullarticle/2553448 ("Among young adults with a BMI between 25 and less than 40, the addition of a wearable technology device to a standard behavioral intervention resulted in less weight loss over 24 months. Devices that monitor and provide feedback on physical activity may not offer an advantage over standard behavioral weight loss approaches.").

[63] Kitsiou et al., *supra* note 61.

[64] *Id.*

[65] Caroline Free et al., *The Effectiveness of Mobile-Health Technology-Based Health Behaviour Change or Disease Management Interventions for Health Care Consumers: A Systematic Review. PLoS Med. 2013;10(1):e1001362. doi: 10.1371/journal.pmed.1001362.* http://dx.plos.org/10.1371/journal.pmed.1001362.

and maintaining consistent preventative care than white communities, and to be stymied in receiving quality care by limited resources and access.[66] High-quality, inclusive, and accessible mHealth apps could potentially improve health disparities, so it's important to understand the challenges developers must navigate to create inclusive and equitable mHealth tools.[67]

IV. Bias in mHealth Apps

The concept of health interventions predates mobile technology, but mHealth apps create the possibility for more frequent, granular, and widely accessible health interventions. From medication reminders to diet and exercise recommendations to activity logs, the value of health apps is their ability to prompt or nudge users to track and self-manage their health with the intent of changing their behavior. The decisions that go into designing mHealth interventions can therefore have an outsized impact on people's health and well-being. An mHealth app built on biased data or biased assumptions can disproportionately harm – or fail to help – groups that are historically underrepresented or underserved. The trend toward using machine learning models to create more personalized or adaptable health interventions can exacerbate the risk of bias or make bias more difficult to discover and correct.

The Potential for Bias in mHealth

Designing an mHealth app requires a series of decisions – from defining the app's objective to selecting training data and validating models – all of which impact the app's outputs, how it will be used, and its effects (or non-effects) on users' health. These design choices can amplify and perpetuate existing societal bias and inequity. There is a growing body of literature around bias in decision systems, and particularly in machine learning.[68] The term "bias" has several

[66] Nicol Turner-Lee, Brian Smedley & Joseph Miller, Minorities, Mobile Broadband and the Management of Chronic Diseases, Joint Center for Political and Economic Studies (April 2012), *available at* http://www.jointcenter.org/research/minorities-mobile-broadband-and-the-management-of-chronic-diseases.

[67] Ernsting C. et al., *Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey*, 19(4) J Med Internet Res 101 (2017) https://www.ncbi.nlm.nih.gov/pubmed/28182803 (Finding that among German mHealth users, "[a]lthough there were many smartphone and health app users, a substantial proportion of the population was not engaged."); Keiki Hinami et al., *Patient-level exclusions from mHealth in a safety-net system*, 12(2) J Hospital Med. 90 (2017) https://www.ncbi.nlm.nih.gov/pubmed/28182803; Laura Siga Stephen et al., *Processes and Recommendations for Creating mHealth Apps for Low-Income Populations*, 5(4) JMIR mHealth and Uhealth 41 (2017) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5394264/; Andrew Broderick & Farshid Haque, *Mobile Health and Patient Engagement in the Safety Net: A Survey of Community Health Centers and Clinics*, Commonwealth Fund (May 13, 2015), https://www.commonwealthfund.org/publications/issue-briefs/2015/may/mobile-health-and-patient-engagement-safety-net-survey-community.

[68] *See, e.g.*, Joy Buolamwini & Tamnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 Proceedings of Machine Learning Research 1 (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Cal. L. Rev. 671 (2016), http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf; Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (2016); Julia Angwin et al., *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*,

overlapping meanings, all of which hold some relevance for the mHealth context. In statistics, bias refers to "systematic error introduced [into a model] by selecting or encouraging one outcome or answer over others."[69] For example, as described in AI Now's 2017 report, "the idea of 'selection bias' refers to errors in estimation [lower accuracy rates] that result when some members of a population are more likely to be sampled than others."[70] As Kadija Ferryman and Mikaela Pitcan explained in their Fairness in Precision Medicine report:

> *In medical and social science research, bias has been defined as any tendency that prevents unprejudiced consideration of a question or advances prejudice in favor of or against one group compared with another. The definitions of bias in both computer science and medical/social science research share an acknowledgement that bias implies error resulting in one group being favored over another.*[71]

Bias also has "normative meanings," referring to "judgement based on preconceived notions or prejudices, as opposed to the impartial evaluation of facts."[72] For the purposes of this report, bias can refer to systematic error introduced into a model or to outcomes that prejudice certain groups (and the design decisions that lead to those outcomes). This explanation from the AI Now report is applicable here:

> *[I]n practice there is rarely a clear demarcation between the statistical and the normative definitions: biased models or learning algorithms, as defined statistically, can lead to unequal and unfair treatments and outcomes for different social or racial groups.*[73]

Biased mHealth outcomes cannot always be traced back to the ML model itself or even the training data. They can also stem from design choices such as the way health interventions are presented to users or the accessibility of a user interface. The next section discusses opportunities for the introduction of harmful bias in the mHealth context.

Pro Publica (May 23 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), https://www.fatml.org/.

[69] Kadija Ferryman & Mikaela Pitcan, Data & Society, Fairness in Precision Medicine at 10 (Feb. 2018), https://datasociety.net/pubs/pm/DataSociety_Fairness_In_Precision_Medicine_Feb2018.pdf.

[70] Alex Campolo et al., AI Now 2017 Report at 14 (2017), https://ainowinstitute.org/AI_Now_2017_Report.pdf.

[71] Ferryman & Pitcan, *supra* note 69, at 10–11 (citing Christopher J. Pannucci and Edwin G. Wilkins, *Identifying and Avoiding Bias in Research*, Plastic and Reconstructive Surgery 126, no. 2 619–25 (August 2010)).

[72] Campolo et al., *supra* note 70, at 14.

[73] *Id.*

Defining the Problem/Defining the Objective

Before any data is crunched or algorithm written, someone (or a group) has to decide what kind of app to build and *why*. What is the app's objective? What problem is it trying to solve or what value will it provide? This step is perhaps the most directly influenced by the developer's (and potentially the company's or society's) subjective goals or assumptions. For example, Cathy O'Neil, author of *Weapons of Math Destruction*, discusses how different definitions of "success" can change outcomes:

> *The dinners I make for my family on a daily basis require the data of the ingredients in my kitchen and the amount of time I have to cook. My definition of success is, a meal is successful if my kid eats vegetables. It's very different from if my youngest son were in charge. He'd say success is if he gets to eat lots of Nutella. But I get to choose success. I am in charge. My opinion matters. That's the first rule of algorithms.*[74]

In mHealth, these decisions are largely being made by developers who do not represent the diversity of their users.[75] Several reports have called attention to the design of period tracking apps, which have tended to focus on the goals of getting pregnant and pleasing one's (presumptively heteronormative) partner.[76] Some apps prompt the user to share her menstruation data with her partner, so that he can monitor her ovulation as well.[77] In 2014, Apple was criticized for leaving menstruation out of Apple Health's metrics altogether.[78]

Data Bias

Datasets can be biased in many different ways. They can over- or under-represent certain groups, they can be incomplete, and they can be collected under discriminatory or unreliable circumstances (as is the case with law enforcement arrest data). Ferryman and Pitcan have identified ways in which historical biases have impacted health research datasets:

---

[74] Cathy O'Neil, Ted Talk, https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end

[75] AI Now p. 16-18.

[76] Daniel A. Epstein et al., *Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools*, Proc. SIGCHI Conf. Hum. Factor Comput. Syst. (May 2017), http://www.depstein.net/pubs/depstein_chi17.pdf; Jennifer Langston, *Period Tracking Apps Failing Users in Basic Ways, Study Finds*, U. Wash. News (May 2, 2017), http://www.washington.edu/news/2017/05/02/period-tracking-apps-failing-users-in-basic-ways-study-finds; Miranda Hall, *The Strange Sexism of Period Tracking Apps*, Motherboard (July 25, 2017), https://motherboard.vice.com/en_us/article/qvp5yd/the-strange-sexism-of-period-apps.

[77] Hall, *supra* note 76.

[78] Rose Eveleth, *How Self-Tracking Apps Exclude Women*, Atlantic (Dec. 15, 2014), https://www.theatlantic.com/technology/archive/2014/12/how-self-tracking-apps-exclude-women/383673/.

> *In the United States, biomedical research has long failed to include representative samples of women and minority populations. When the data analyzed comes from narrow populations, findings may not be generalizable to all patient subgroups.*[79]

The problem of exclusion from medical research has been self-perpetuating. Guidelines for tests, such as lung cancer screenings, were created based on unrepresentative population samples, such that subsequent data collection (from those screenings) has continued to exclude black people.[80] This is also exacerbated by disparities in access to healthcare, which disproportionately impact people of color and people in poverty.

As previously touched on, the mHealth ecosystem risks perpetuating this cycle of underrepresentation. mHealth users tend to be young, affluent, and educated, and they are the individuals whose data is used to generate new insights that feed back into ML models. Biased data can impact validation (testing a model for accuracy), as well as model training. If accuracy is tested on a test set that underrepresents minority groups, the resulting overall accuracy rates will not be true for those groups. Joy Buolamwini and Timnit Gebru found that face recognition systems had much lower classification accuracy rates for darker skinned black women than for white men because the systems were tested on a dataset that underrepresented black women, especially those with darker skin tones.[81]

Feature Selection

Features are the data variables that factor into an algorithm's outputs. For example, the features in a model for recommending healthcare providers might be location, age, symptoms, and type of insurance. Features can be identified by the developer or through machine learning techniques. Feature selection can lead to biased outcomes if features are unevenly distributed across different groups. For example, heart disease tends to progress differently in women than in men,[82] so a diagnostic model that incorporated features for predicting heart disease that are more common in men than in women could be less accurate for women.

Feature selection is limited by the datasets that developers use. For example, you can't use age as a feature if you have no age data. Because of proprietary or regulatory barriers, a lot of health data is not accessible to companies, and there is a risk that developers might be tempted to use what is easily available rather than what is evidence-based when selecting features for a model. The sensors on smartphones and other mobile devices, for example, only collect certain metrics, such as acceleration and location. Because these metrics are easy to collect,

---

[79] Ferryman & Pitcan, *supra* note 69, at 23. Studies have shown that women face disparities in care; they are more likely to die of septic shock, for instance; Rebecca Robbins et al., Health App Use Among US Mobile Phone Users: Analysis of Trends by Chronic Disease Status, 5(12) JMIR mHealth and uHealth e197 (2017) http://doi.org/10.2196/mHealth.7832; "bias through invisibility"

[80] *See* Ferryman & Pitcan, *supra* note 69, at 24.

[81] Buolamwini & Gebru, *supra* note 68.

[82] Harvard Heart Letter, Heart Attack & Stroke: Men vs. Women (April 2014), https://www.health.harvard.edu/heart-health/heart-attack-and-stroke-men-vs-women.

developers might be more inclined to use them regardless of whether they are actually helpful for making health decisions. Perhaps this is why almost all fitness trackers focus on the number of steps a person takes, even though that is only one small aspect of fitness and the data may not be that accurate.[83]

Training the Model

ML models "learn" not only from the training data itself but from labels given to the data, often by humans. In supervised learning, models learn features and patterns from examples that are labeled by humans, whose biases and assumptions can affect their labeling. For example, natural language processing (NLP) researchers often use supervised learning on labelled examples of text, such as hate speech or terrorist propaganda. But people's perception of what hate speech or terrorist propaganda is can be influenced by their culture, prejudices, and other factors, so NLP researchers often have problems getting reliable labels. In healthcare, NLP is being developed for clinical decision support, which can involve tasks such as processing doctors' notes and predicting whether a patient is depressed, but doctors' notes are notoriously difficult for machine learning since they often lack standard grammatical structure and contain a lot of shorthand.[84]

Model Selection and Intelligibility

Machine learning involves identifying patterns in training data and using those "rules" to sort new data and create outputs. ML models can range from simpler decision trees to more complex neural networks.

The intelligibility (or interpretability) of an ML model refers to the ability (usually of a model's developers) to understand how the model is making decisions. Intelligible models, compared to "black box" models, make it easier to detect and mitigate bias. Sometimes tradeoffs must be made between intelligibility and accuracy, but Microsoft researchers Rich Caruana et al. argue that healthcare models should prioritize intelligibility.[85] They presented a case study of a model for predicting outcomes for patients with pneumonia. The model was designed to help hospitals decide whether to send patients with pneumonia home (low risk of death) or admit them (high risk of death). They were surprised to find in testing that the model was predicting a lower than average risk of death for asthma sufferers who presented with pneumonia, indicating that they could be sent home. What the model did not know is that asthma patients with pneumonia have better outcomes because they are almost always admitted and treated,

[83] Sean Downey, *Why Fitness Tracker Calorie Counts Are All Over the Map*, Wired (Aug. 8, 2017). *See also* Megan Molteni, *The Problem With Fitness Studies Based on Activity Apps*, Wired (July, 11, 2017) ("the handful of algorithms that Apple and other phone manufacturers and app developers employ to package that raw data into easy-to-use step counts can't accurately capture the huge variety in people's walking mechanics.")
[84] *See* Jennifer Bresnick, *What is the Role of Natural Language Processing in Healthcare?*, HealthIT Analytics, https://healthitanalytics.com/features/what-is-the-role-of-natural-language-processing-in-healthcare.
[85] Rich Caruana et al., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission* (2015), http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf.

since their risk of death is inherently higher. The researchers were only able to discover this bias because they created an intelligible model.[86]

User Experience (UX)/User Interface (UI) Design

The decisions that mHealth app developers make when designing an app's UI and UX can also incorporate biased assumptions that may disadvantage some groups of users. Apps without certain affordances, such as dictation, may be more difficult for people with disabilities to use. Core technical issues that might also impact different users such as sporadic wireless disconnections, differences in device features, and differences in battery life.

UX/UI design can also reflect biased assumptions about users. A study of period tracking apps found that many of them had traditionally feminine designs ("predominantly pink or using flower and heart images") and used iconography that assumed the user was female-identifying and heterosexual.[87] Another study found that only a third of users aged 65 and older were comfortable using the digital health services to manage health conditions and the majority of this age group only occasionally accessed health information via email, landline, or cell phone.[88] One diabetes app usability study found that "[t]he most important points to keep in mind when designing an app for the elderly are [] the size, visibility, and comprehensibility of buttons and symbols.[89]

Understanding how a user might feel as they face decisions related to their health is another helpful frame for developers to incorporate into their UI/UX design process. The Decisional Conflict Scale, for example, used by the North American Nursing Diagnosis Association, may help develop user-centric decision models that reflect how people approach risk-related choices. According to the user manual, the tool measures ". . . a) uncertainty in choosing options; b) modifiable factors contributing to uncertainty such as feeling uninformed, unclear about personal values and unsupported in decision making; and c) effective decision making such as feeling the choice is informed, values-based, likely to be implemented and expressing satisfaction with the choice."[90]

---

[86] *Id. See also* Rich Caruana, Microsoft Research, Friends Don't Let Friends Deploy Black-Box Models: The Importance of Intelligibility in Machine Learning for Bias Detection and Prevention (2017), https://www.ischool.berkeley.edu/events/2017/friends-dont-let-friends-deploy-black-box-models-importance-intelligibility-machine.

[87] Epstein et al., *supra* note 76.

[88] Helen Levy, Alexander T. Janke, and Kenneth Langa, *Health Literacy and the Digital Divide Among Older Americans*, Journal of General Internal Medicine, (March 2015), https://doi.org/10.1007/s11606-014-3069-5

[89] Masa Isakovic et al., *Usability Pitfalls of Diabetes mHealth Apps for the Elderly*, J. Diabetes Research (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4807066/.

[90] https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf

V. Recommendations

Based on the potential for bias in mHealth apps outlined above, we have provided a roadmap for developers[91] to identify and mitigate bias. The roadmap is based in part on CDT's Digital Decisions tool, which provides a generalized set of questions that should be addressed during each phase of model development.[92]

The mHealth roadmap below is a more specific set of inquiries tailored to the objectives, considerations, and risks particular to developing an mHealth app. The questions are organized loosely into development phases, but these phases (and the questions within them) may be overlapping and may not occur in any particular order. The roadmap focuses on bias and inclusivity and does not comprehensively address all issues of concern for developers, including privacy, data security, or safety. Because mHealth app bias is context specific, we have found that offering directed but open-ended inquiries are often more useful than prescriptive advice because they allow developers to consider the specific risks raised by their particular app. Future work should apply this roadmap to interrogate specific mHealth apps – a process that would require openness and cooperation from the mHealth industry.

Ideating and Designing the App

- Is the health issue you are trying to address documented in health research or other credible sources?

- What is your motivation to address this health issue? Are you seeking to fill a gap in mHealth resources? Do you have any incentives as a company (e.g., gaining a large number of users, advertising, selling premium services) that could conflict with providing the highest quality health interventions to users? What checks are in place to mitigate these risks?

- What does success look like? Can you define and measure success based on observed health outcomes (e.g., users of a weight loss app lose weight and maintain weight loss), or will you only be able to measure users' activity on the app (e.g., users log calories and exercises that indicate a lifestyle conducive to weight loss)? Do the metrics you will have access to correlate with actual positive health outcomes, supported by health research? Does measuring success require the assumption that users' inputs are accurate?

[91] Model or app developers need not be the only ones to ask these questions. Product and business managers, investors, attorneys, policy teams, external auditors, and regulators may also want to use this roadmap. The roadmap can also be a useful guide for companies or governmental bodies seeking to procure mHealth technology from an outside vendor.

[92] Center for Democracy & Technology, Digital Decisions Tool, https://cdt.info/ddtool/.

Selecting/Designing Health Interventions

- Are the health interventions (such as Behavioral Change Techniques) in your app evidence-based (are they supported by credible health research)?

- How can you ensure that your company's financial incentives do not undermine the efficacy of your app's health interventions? For example, do you rely on evidence found in health research to set the frequency and content of notifications and user interactions, or do financial incentives (such as advertising, licensing, or subscription dollars) guide those decisions? For example, some experts have supported "gamification" features in apps, such as the use of scores or streaks, because of the established link between reward systems and positive behavior change[93] but the line between encouraging user engagement and seeding digital addiction can be blurry.[94]

- What assumptions are you making about users in designing interventions? Do your interventions assume that users have access to certain devices, equipment, or healthcare and other services? Have you made assumptions about your users' age, sex, gender, education level, ability, sexual orientation, relationship status, income, or location? For example, does your pregnancy app make recommendations that involve the baby's "father" or the mother's "partner"?

Identifying User Populations and Designing for Inclusivity

- What type of user are you designing for and how will they access and use the app? Is your app designed to reach as many users as possible, or are you targeting a specific subset of users (e.g., people with a particular medical condition)? Consider whom you might be unintentionally excluding, such as low-income users (if your app costs money or requires infrastructure), users who do not have smartphones (if your app only has a mobile interface), or LGBTQ users (if your app makes assumptions about users' gender, sex, or relationships). Consider how the design and features of your app will affect different groups of users, especially those who may have particular vulnerabilities.

- How do you evaluate and take your users' needs into account in the design and improvement of your app? Do you conduct focus groups, solicit feedback through the app, or collect input in other ways? The Health Information Management Systems Society (HIMSS) recommends designing an app with users as an integral part of the process from the start.[95]

---

[93] Eric Wicklund, *mHealth Games Offer a 'Fun' Way to Boost Patient Engagement*, mHealth Intelligence, https://mHealthintelligence.com/features/mHealth-games-offer-a-fun-way-to-boost-patient-engagement (last visited July 5, 2018).

[94] David Sloan, *Customer engagement vs digital addiction*, Medium blog, (March 15, 2017), https://medium.com/data-in-motion/customer-engagement-vs-digital-addiction-d542712dc1d8.

[95] Mark Groshek, et al., *mHealth App Essentials: Patient Engagement, Considerations, and Implementation*, HIMSS website (Feb. 20, 2015) ("An app should be developed with actual users as part of the design process," stated

Assembling the Team and Seeking Input

- Is your team diverse (especially in terms of race, gender, LGBTQ representation, ability, language, and age), particularly among those who get to make design decisions? If not, how can you seek and incorporate input from groups that are not represented on your team?

- Does your team represent your users and/or the population affected by the health problem you seek to address? If not, how will you seek and incorporate input from affected groups?

- Does your design or product team include the medical expertise necessary to evaluate whether your health interventions are medically sound, culturally aware, and evidence-based? For example, an app aimed at improving an individual's nutritional intake might flag whether a person's background implicates widespread lactose intolerance (such as in Native American communities, where lactose intolerance is estimated to impact 75% of the population). How might you involve outside medical experts in the design and development of your app to help address these issues?

Designing the User Interface and Features

- How usable is your app? Is it accessible to users who have low vision, blindness, hearing impairments, cognitive impairments, or motor impairments? mHealth apps should be designed and tested for usability and should be compatible with accessibility features such as screen readers.[96] Material Design provides a useful guide to accessible design.[97] Accessibility factors include color contrast, clearly labeled elements or actions, and readable fonts.[98] Are your apps available in in multiple languages (or able to be translated)? Does the app work for users who have access to lower bandwidth internet connections or limited device functionality (that may not support streaming video, for example)?

- Does your usability testing include users who are disabled? Including users early in the UX/UI design process may help developers identify unobvious differences in user experience, such as avoiding unnecessary scrolling for individuals with limited motor

---

HIMSS. "A new app should be piloted with enough time to correct defects, test the support plan, and ensure staff comfort with the application and its capabilities before it is launched broadly.")

[96] For information on how to design user-friendly apps that are compatible with the iOS and Android accessibility tools, *see* iOS Human Interface Guidelines, https://developer.apple.com/design/human-interface-guidelines/ios/overview/themes/; Android Developer Guides, Making Apps More Accessible, https://developer.android.com/guide/topics/ui/accessibility/apps.

[97] Material Design, Accessibility, https://material.io/design/usability/accessibility.html#understanding-accessibility.

[98] *Id.*

ability or not bombarding users with limited visual acuity with too many text or graphic-based choices or controls.[99]

Developing the Models That Power Health Interventions

Selecting the Models

- What types of models will you use to process user data and trigger health interventions such as reminders, alerts, and recommendations? Many factors – which are beyond the scope of this report – go into choosing the best model for a task.[100] One important consideration for any health model is whether it is intelligible (or interpretable) – meaning that someone with access to the model and the relevant expertise can interpret how the model makes decisions. Sometimes tradeoffs must be made between accuracy and intelligibility, but model intelligibility in health is particularly important for catching flaws in a model's decision-making that can lead to harmful outcomes.

Selecting/Creating Training Datasets

- Where did your training data come from and whom does it represent? Was it collected in ways or under circumstances to lead to bias? For example, data from clinical trials may underrepresent minority groups, and data obtained from other mHealth apps may exclude people who do not have smartphones (primarily people in poverty, children, and elderly people). Can you mitigate sampling bias by broadening your sample to better represent the population? Including more dark-skinned and female faces in training data, for example, has led to more balanced accuracy (across phenotypes) in IBM's Watson Visual Recognition service for facial analysis.[101] This may increase costs, particularly if you have to collect new data, but inclusion is critical given that mHealth outputs may be used in future medical decision-making.

- Was your training data drawn from a standardized methodology considered reliable and support by clinical evidence, such as from a Randomized Clinical Trial (RCT), implementation study, or newer methods such as the Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT)?[102] Health data is frequently acquired through clinical databases such as physioNet bank, MIMIC data sets and MIT, data sets that all use RCTs.

---

[99] *See* Emily Anne Fulton, et al., *Lost in Translation: Transforming Behaviour Change Techniques into Engaging Digital Content and Design for the StopApp*, Healthcare (July 2018), doi: 10.3390/healthcare6030075.
[100] *See, e.g.*, Hadi Banaee, Mobyen Uddin Ahmed & Amy Loutfi, *Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges*, Sensors (Basel) (2013), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3892855/.
[101] Ruchir Puri, IBM Research, Mitigating Bias in AI Models (Feb. 6, 2018), https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/.
[102] David Mohr, et al., Continuous evaluation of evolving behavioral intervention technologies. American Journal of Preventive Medicine (Oct. 2013) 45(4):517-23, doi: 0.1016/j.amepre.2013.06.006.

- Is your training data suitable for your objectives? For example, is the sample size large enough to support reliable predictions? Does the context of collection match the context of use? Using contextual information to assist in data mining allows for a more complete understanding of the suitability of a data set for your purposes.[103] Contextual information might include a user's weight, height, age, sex, history of vital signs, as well as information derived from a patient's electronic health record (EHR) (with permission and in accordance with HIPAA guidelines for business associates).

Selecting Features and Deciding Which Health Metrics to Collect

- Are the features used to train the model supported by evidence found in medical research? Machine learning techniques may discover novel correlations between certain features and outcomes for which there is no documented causal relationship. In this case, app providers should consider conducting or commissioning research trials to test these correlations before using them as a basis for health interventions.

- Are your features evenly distributed across demographic groups? If not, resulting interventions could have disparate accuracy rates.

- How did you decide which health metrics to collect from users? The most easily available metrics (e.g., those that are readily available through mobile phone sensors such as accelerometer or gyroscope data) are not necessarily the best metrics for producing accurate health interventions.

Labeling Data and Training the Models

- If you're using supervised learning, who is labeling the training examples? Do the labelers have domain expertise? Crowdsourcing is often used for labeling in machine learning, but experts tend to perform better in labeling technical health data.

- How might labelers introduce bias into the model? For example, when labeling examples for sentiment analysis, different people might have different ideas about what is a "positive" or "negative" sentiment based on cultural background or personal experiences. Can you attempt to mitigate labelers' individual biases, for example, by providing detailed labeling instructions? Do your labelers represent demographic diversity?

[103] John Ovretveit, Paul Shekelle, et al., *How does context affect interventions to improve patient safety? An assessment of evidence from studies of five patient safety practices and proposals for research.* BMJ Quality & Safety (2011), 20(7):604–610. https://qualitysafety.bmj.com/content/20/7/604.long.

Selecting/Developing Test Datasets and Testing the Models

- Do you have reliable and representative datasets for testing your models? Were the test sets created or labeled and reviewed by domain experts to ensure reliability? Do the test sets represent the demographic diversity of your users and of the population at-large?

- How much error is acceptable in your app, and how should you balance different types of error (for example, are false positives preferable to false negatives given the risks of each)? Can you analyze errors to figure out why they happened? Depending on the types and amount of error, it may warrant providing warning labels or other notice to users about conditions under which the app's interventions may be inaccurate.

- Have you user-tested your app's health interventions? Do your testers represent diversity in terms of demographics, abilities, and other relevant factors?

- If your model is designed to continuously learn from feedback, can you introduce human review to ensure that the model is maintaining accuracy and is not learning incorrect rules?

- How do you incorporate user feedback? Do you have a mechanism for users to report outputs that are incorrect or unhelpful for them?

Analyzing the App's Outcomes

- Can you design a systematic study to test the efficacy of your mHealth interventions? The design of efficacy studies should be tailored to the app's objectives and metrics. In their mHealth Evidence Workshop report, Santosh Kumar et al. proposed several potential research designs to evaluate the efficacy of mHealth interventions.[104]

- Can you design a disparate impact test for your health interventions? There is a growing body of literature presenting disparate impact testing methods that can be useful for detecting when facially neutral algorithms systematically treat some groups differently.[105] Much of the literature on disparate impact focuses on legally protected classes, since the term "disparate impact" comes from antidiscrimination law. However, mHealth developers should not constrain themselves to particular legal definitions of protected classes when testing for disparate impact. For example, much of U.S. federal law does

[104] Kumar et al., *supra* note 60, at Table 3.

[105] *See* Michael Feldman et al., *Certifying and Removing Disparate Impact*, Proceedings of the 21st ACM International Conference on Knowledge Discovery and Data Mining 259-268 (2015), http://hdc.cs.arizona.edu/papers/kdd_2015_bias.pdf; Cynthia Dwork et al., *Fairness Through Awareness*, Proceedings of Innovations in Theoretical Computer Science (2012), https://arxiv.org/abs/1104.3913.

not treat people in poverty as a protected class, but mHealth developers should investigate the potential disparate impacts of their health interventions on people in poverty.

VI. Areas for Further Investigation

There is much fertile ground for further research related to improving the efficacy, usability, and inclusiveness of mHealth apps.

Standardized Metrics

The efficacy of mHealth app interventions remains in question in part due to a lack of methodologically-sound studies on the subject and due to a lack of standardized metrics for successful interventions and outcomes across the mHealth app industry. Research should investigate the utility of open-source databases designed to help create standards and common metrics for mHealth. Open mHealth, for example, is an open-source data repository for mHealth researchers and practitioners.[106] Facilitating a more robust dataset would require more openness and willingness to share data on the part of companies developing mHealth apps, which would also improve and expedite more research in the space in general. Scholars have also pointed to Interrupted Time-Series Design (ITSD) as a possible model for designing non-randomized mHealth efficacy and testing studies. In the ITSD model, large amounts of data are collected before and after the "treatment" to measure an "interruption" that would indicate efficacy.[107] To both broaden the appeal of mHealth apps for different populations and better design responsive interventions, research should also consider how demographics, UI/UX design and other factors impact the efficacy of various interventions delivered by mHealth apps.

Evidence-Based Open Datasets

Developers facing pressures to get their apps to market, and to sustain user engagement within their apps, may be dissuaded from seeking out training data supported by evidence-based methods, such as Randomized Control Trials, as these datasets are typically less available and less responsive to a continuous, real-time data generation environment. Launching an RCT can be costly and time-consuming, pushing developers toward using the same datasets (identified above in the roadmap), which may reinforce embedded bias. Research should further explore how to use a "living systematic review" model that can synthesize health research data in

[106] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3803146/#R36

[107] "The premise is that administration of the treatment should produce an interruption to the pre-treatment time series. The interruption can be found along any of three dimensions: form of the effect (the level, slope, variance, cyclicity); permanence of the effect (continuous or discontinuous); and immediacy of the effect." Kumar et al., *supra* note 56 at 236.

real-time. This model might offer a valuable approach for developers seeking to test their app against existing knowledge and research.[108]

Relatedly, continuous, real-time data generation requires data mining methods that are affordable, straightforward, and allow for more interrogation of correlations between variables that would affect outputs. More research should be done on how methods such as rule-based, decision tree, and statistical techniques might be harnessed to quickly handle the data processing requirements. Developers should also consider using statistical methods that allow for more understanding of variables within and between subject effects, such as sequential multiple assignment randomized trial (SMART). SMART allows developers to choose core treatment aspects to look at closely, then randomize individual data at decision points based on factors like privacy or feasibility.

Policy Framework Addressing Fairness and Equality for Health Data

As the collection of health data becomes more ubiquitous and granular in everyday life, traditional measures used to protect user privacy may not be enough to reduce broad health disparities and build a culture of health. This is especially true as developers turn to demographic data to build new personal health technologies. This data may not be protected under the law, yet poses unique risks of bias in targeted content for individuals. Without access to HIPAA-covered data, developers of these technologies are turning to demographic data to build the machine learning analytics models that deliver personalized health-related recommendations and interventions to users. Many measures of what we typically consider demographic data, such as ethnicity, gender, and economic status, are also the attributes, if used in a model, that have a high probability of introducing bias into targeted content.

It may be valuable to create a data governance framework for fairness and health data equality that can be used by patients, commercial app developers, and lawmakers to set responsible, respectful standards for data within the health framework and policy to reflect the ethics of the use of this data on a broad scale.

VII. Conclusion

Health disparities in the United States aren't a forgone conclusion. Advances in accessible and inclusive technology like mHealth may offer a path to improved health for marginalized communities. In the absence of policy, adding safeguards against discrimination will convey a direct impact to users and allow companies to enter ripe but untapped markets for mHealth apps that may well improve all of our health and opportunities.

[108] Julian H. Elliot et al., *Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap*, 11(2) PLOS Medicine e1001603 (2014).