AN EMERGING RESEARCH LANDSCAPE: NEW ACTORS, DATA SOURCES AND RESPONSIBLE PRACTICES

Horizon Scan

By Camille Nebeker

AUGUST 2021 Prepared for AcademyHealth



✓ ★ AcademyHealth



ABOUT THE PARADIGM PROJECT

The Paradigm Project is a concerted, collaborative effort to increase the relevance, timeliness, quality, and impact of health services research (HSR). Convened by AcademyHealth and funded by the Robert Wood Johnson Foundation, the project is ideating and testing new ways to ensure HSR realizes its full potential to improve health and the delivery of health care. The Paradigm Project is designed to push HSR out of its comfort zone—to ask what works now, what doesn't, and what might work in the future.

Learn more at www.academyhealth.org/ParadigmProject.

1. A PARADIGM SHIFT – WHY NOW?

The Paradigm Project, convened by AcademyHealth and funded by the Robert Wood Johnson Foundation, is generating solutions to improve the ways health services research (HSR) is conceived, conducted, and used. This Horizon Scan highlights relevant issues and opportunities related to novel or emerging methods and data sources, and the innovative ways various actors might use these to generate improved research insights.

Technological advancements afford opportunities to access more data than ever before, which contributes to the changing HSR paradigm. New social media platforms and technological devices (e.g., smartphones, wearables) generate an ever growing and diversifying body of information which, even if collected beyond a health context, can help us understand individual- and population-level health trends.⁸

New forms and uses of data require the creation of new ethical frameworks, governance structures, and a strong, modern, and adaptable data infrastructure.⁸ As technology becomes more accessible and new actors engage in research, the field needs a corresponding trust infrastructure: preserving data privacy, honoring data ownership, protecting intellectual property, and encouraging cross-sector ethics reviews.¹⁵

2. DIVERSE ACTORS, DATA SOURCES AND RESEARCH APPROACHES

Common HSR data sources including surveys, claims data, and electronic health records (EHRs) are increasingly supplemented with patient-generated health data (PGHD) from wearables and in-home or remote monitoring devices.¹⁷ Consumer-level data linked to loyalty cards, social media posts, and internet search histories (collected by health start-ups, tech companies, citizen scientists, etc.) can now be used in HSR. Similarly, the field has begun using new approaches (contrasted with those used for highly rigorous randomized control trials, or RCTs) to supplement existing methods – with the aim of improving research timeliness and generalizability.^{20,21} To maximize impact, researchers must ensure ethical use of new data and methods – and develop an ethical data infrastructure for the HSR ecosystem.²⁶

3. TENSIONS LEAD TO NEW MODELS

Emerging research approaches, such as community-engaged research (CEnR) and citizen science, have changed the characteristics between researchers and community members. In conducting CEnR, researchers work with members of a community (defined by shared condition, location, experience, characteristic, etc.) to develop research questions, design and conduct studies, and/or to communicate results.²⁷⁻²⁸ This method increases the likelihood of producing results with real-world relevance, as community members' input often facilitates development and use of culturally resonant research methodology.²⁹⁻³²

Similarly, citizen science (conducted by individuals without formal research training) has proliferated, with the rise of technological advancements.⁴⁰ Generally, citizen scientists are interested in challenges that have not been prioritized or identified by health systems or researchers.⁴⁷ Without guidance, however, citizen scientists may perpetuate use of methods that are unsafe or unreliable.⁴⁹

4. DIGITAL HEALTH CHECKLIST FOR RESEARCHERS (DHC-R)

This horizon scan presents four case studies in the use of new and unconventional data and methodology in HSR, analyzed via the Digital Health Checklist and Framework for Researchers (DHC-R; Figure 1), a tool that prompts scientists to consider the ethical implications of their research practices.⁵⁰



Figure 1: Digital Health Framework with checklist examples noted within each domain. The DHC-R is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License (2018-2020) and available at https://recode.health/tools/. Published with permission of C. Nebeker.

The DHC-R examines four "domains" of research: access and usability, privacy, data management, and risks and benefits. The DHC-R then describes how the ethical principles of respect for persons, beneficence, justice, and public interest relate to the four research domains.^{50,51}

5. TOPICS, SCENARIOS, ETHICAL CONSIDERATIONS AND ANALYSES

UNCONVENTIONAL DATA SOURCES: THE DARTMOUTH STUDENT LIFE STUDY

Even if not collected for health care or research uses, emerging data sources contain information relevant to individual- and population-level health; they show promise for HSR applications, if accessed and utilized effectively and ethically. This requires researchers to account for bias, obtain informed consent, and establish expectations regarding the use of publicly available commercial data for research.

The Dartmouth Student Life study (n=48) evaluated a smartphone application measuring individual-level behavioral factors such as sleep duration, class attendance, and conversation length. The app was designed to gather data to ultimately predict and prevent health risk behaviors and/or school attrition.

ESTABLISHMENT VS NON-ESTABLISHMENT: PARTICIPANT-LED RESEARCH

This scan defines establishment research as occurring within the traditional academic research sector. Increasingly, facilitated by developments in personal health technologies, patients and general public are conducting their own investigations and studies.^{44,63} This non-establishment research generally occurs in the absence of the formal structures, like ethics reviews, that govern establishment research.^{46,65}

People associated with Quantified Self⁸⁶ designed individualized experiments on how daily activities affected their blood lipid levels. Each individual was trained to use a relevant medical analytical device, and was expected to design his or her own protocol for measuring and analyzing the data. Participants could opt into data sharing with the group, which met regularly to discuss risks, consent, and benefits.

PREDICTIVE ANALYTICS AND THE DIGITAL DIVIDE: THE BIAFFECT STUDY

Disparities in access to new healthcare technologies have created a "digital divide"⁸ that has only been further exacerbated by COVID-19. Even when technology is accessible, some consumers face barriers to its use. If specific populations (e.g., racial minorities, older adults) have limited access to or use of these technologies, they will likely be underrepresented in datasets created via these tools. In one instance, for example, health system leaders used an algorithm to predict missed appointments. While Black patients were found to be more likely to miss their appointments, the tool failed to account for social variables that influenced patients' ability to access healthcare services.^{68,69}

For the BiAffect Study, researchers analyzed patients' keystroke dynamics through machine learning algorithms to identify keystroke patterns representative of neuropsychiatric symptoms.³⁴

UNUSUAL HEALTH DATA REPOSITORIES: THE HUMAN PROJECT

In recent years, new repositories have been created to capture diverse sets of information including participants' purchasing data, local air quality, and movement within their homes.⁷²⁻⁷⁴ Developed under the umbrella of precision medicine, these datasets purposefully capture "unusual" information that allows researchers to study the effects of specific factors and behaviors on health. While these datasets are highly valuable, it can be difficult to obtain informed consent to collect such varied data when the research questions guiding their use have not yet been formed.⁷⁵

The HUMAN Project⁸⁷ planned to collect medical, financial, and social data from 10,000 participants over a ten year period with a goal to examine the interactions between human behavior, environment, and biology. The ethical and social implications were complex and contributed to a delay in beginning the project.

6. DISCUSSION

The rise in digital health investments in response to COVID-19⁷⁹ introduces incredible opportunities for advancing the fields of data science and health research. However, unethical and/or unrestricted use of (particularly of emerging and commercial) data has the potential to perpetuate harm. Mitigating this risk requires strengthening informed consent and understanding the implications of using non-traditional data sources.^{81,82} The trust and governance infrastructure for non-establishment research is evolving, yet similarly nascent. Further development in education and ethics reviews among citizen researchers is needed to ensure the rigor and quality of their activities – and the validity or generalizability of results.⁸⁴

Health services researchers must also consider new ways of evaluating bias and ethics as the data infrastructure continues to evolve. To further build this "trust infrastructure," researchers should reevaluate current mechanisms for obtaining informed consent, managing data, and engaging participants in research. Done successfully, this could create a stronger scaffolding for research work that effectively integrates traditional and emerging data and methods – in order to improve real-world health outcomes.

1. A PARADIGM SHIFT – WHY NOW?

Many, albeit in diverse disciplines, are arguing for a paradigm shift in the academic health research (HSR) ecosystem and challenging the ethics of current practices – particularly in the health services research domain. Practices in question include, for example, a reliance on randomized controlled trials as the only credible path to advance scientific discovery^{1–3} difficulties in accessing quality health data research purposes.⁴ Tensions introduced by productivity metrics for academic promotion⁵ and, the role of regulatory review boards (e.g., Institutional Review Boards in the US and Research Ethics Committees in Europe and Canada) for vetting research involving humans.⁶ For this reason, the Paradigm Project, led by AcademyHealth, is asking the health services research community to reimagine what high quality, timely, ethical and relevant research might look like moving forward. This Horizon Scan speaks to shifts in the research landscape introduced by diverse actors and sectors, digital strategies, evolving methods and the related challenges of creating ethical and respectful processes for advancing credible and trustworthy health research.

In an era of emerging technologies combined with troves of individual level data, the opportunity to access novel sources of information not traditionally considered to be health data (e.g., employment, judicial, consumer purchasing behaviors, wellness data) may potentially lead to new individual- and population-level health discoveries. While exciting, this potential can only be realized if we are forward thinking in anticipating and building the appropriate system infrastructures to support a new paradigm for conducting HSR. Unfortunately, this is not a new conversation. Back in 2010, Pittman, a professor in health policy and management, provided recommendations provided recommendations that would set HSR up for operations in 2020, which included steps to address governance, stewardship, data access and methods.⁷ A main point was on the importance of data as the key ingredient to realizing the future possibilities of HSR. For the potential of HSR to impact health policy and practice, data infrastructures would need to be prioritized. Over the decade that followed (2010-2020) we've observed inconsistent, albeit promising movement at the federal level in response to Pittman's data infrastructure recommendations. Namely, creating an accessible federal survey repository, development of a national library and improving the quality and utility of existing survey data.

Since Pittman articulated this HSR vision for 2020, changes in technology, methods and research frameworks have changed and continue to change. These changes include the creation of platforms used for social connection, information access, transportation, housing, coordination and commerce – all providing some form of core services and increasingly used for purposes well beyond the initial intent.⁸ These digital platforms collect an array of individual level social and behavioral data that can be used in generating predictive analytics, or what is increasingly referred to as digital phenotypes.⁹ For example, these data are used by behavioral and social scientists to study sedentary behaviors "in the wild,"^{10,11} gauge stress, and sleep in college age populations,¹² identify suicide ideation among Facebook users,¹³ and driving performance among older adults.¹⁴ These new forms of data collection have led to new frameworks and methods to guide the design of interventions taking place in real-world contexts and, concurrently, shift in how we think about and apply research ethics are gaining attention.⁸

As we enter 2021, the research infrastructure is fragmented and not responsive to the changes in how research is conducted and by whom yet, fortunately, not necessarily beyond repair. What we may now need to consider is, for example, recognition and support of new actors and new approaches for designing research and transparent processes for accessing data sources heretofore not utilized to support health systems or patient care. Moreover, the concept of developing a research "trust architecture" that supports progressive intellectual property and data ownership models, combined with cross sector involvement in ethics review, may foster development of best practices for this emerging frontier.¹⁵ Creation of this architectural scaffolding requires cross-sector collaboration, capacity building across stakeholders and an investment in education as well as the research support needed to inform best practices.¹⁶ These practices may include, yet are not limited to: appropriate risk to benefit assessment, trustworthy data management protocols (i.e., collection, ownership, storage, sharing), respectful privacy practices and models that promote meaningful consent.

2. DIVERSE ACTORS, DATA SOURCES AND RESEARCH APPROACHES

HSR generally speaks to how one's health and access to healthcare maybe influenced and/or related to factors like delivery, financing, quality, costs, safety and other outcomes.^{17,18} Given this relatively broad scope, those involved in studies of health services come from a variety of disciplines including health economics, social and behavioral sciences, clinical disciplines and risk management. For the most part, health services researchers rely on traditional sources of data including, for example, self-report via survey and interview methods, routine metabolic testing, electronic health record entries and claims data.¹⁷ In the traditional research paradigm, the scientific method is the gold standard and one that we trust to produce accurate and, potentially generalizable results. This work requires appropriate research methods and rigor in the design, implementation and reporting of health research — the hallmark of a credible scientific process. One challenge is that the RCT is not appropriate in many settings and, when it is, it takes many years for these studies to demonstrate success, or failure. Not only is it important to recognize and plan prospectively for the challenges of RCTs¹⁹ we recently witnessed how a pandemic leads to adjusting how these studies can be responsive in time-sensitive circumstances.²⁰ Another challenge is that not all people are represented in RCTs, which leads to generation of results that are not useful to all people and stand to benefit a select few.²¹ Increasingly, the process of conducting and reporting research and then translating results to policy and practice is at cross-roads as efforts increase, to improve access to health and healthcare as a human right.^{22,23}

Health data come from myriad sources. Gathering of granular consumer level social, behavioral and environmental data (e.g., loyalty cards, internet search history, social media posts) is commonplace for commercial entities for the purposes of manage business operations. Over the past decade, researchers have increasingly leveraged these and other data sources to answer complex health questions. Facebook is a case in point where natural language processing has been used to flag suicide ideation.²⁴ From that standpoint, it's important for the health services research community to increase awareness of this emerging ecosystem – an ecosystem that includes diverse actors (e.g., health tech start-ups, tech giants, academics, citizen scientists), usual and unusual health data sources and new research methods for approaching data collection.²⁵ New actors, data sources and research methods require key stakeholders to reevaluate current processes and determine how best to identify and evaluate benefits against inherent and potentially unknown risks of harm. Now, more than ever, all involved — especially scientists, must take responsibility for elevating the importance of ethical practices and not defer to dated institutional infrastructure and policy.²⁶

3. TENSIONS LEAD TO NEW MODELS

Recognizing these tensions combined with a curiosity to do better, researchers are exploring and improving innovative new tools and methods to advance scientific discovery including community engaged research, agile science and citizen science – each briefly introduced below.

COMMUNITY ENGAGEMENT AND CAPACITY BUILDING

Community engaged research (CEnR) and community-based participatory research (CBPR) models have emerged and evolved over the past 30+ years to engage communities, generally those at risk of health disparities, in the research process.^{27,28} The practice of CEnR and CBPR has moved forward, given the increase in funding that supports development of authentic community partnerships between researchers and individuals or organizations that combine service delivery and research capacity building. Community engagement means that the people who make up a particular "community" will play an important and active role in problem identification and developing the research questions that they, in collaboration with trained professional researchers, can help to answer. Engagement also means that the study results are communicated and the translation of research to practice benefitting the community is part of the flow. Moreover, involvement of community members (e.g., community health workers and promotores, patient and consumer advocates, civic leaders, cultural liaisons) in these efforts generally leads to culturally relevant and respectful approaches that facilitate successful research partnerships.^{29,30} Further, by solidifying community research partnerships through research capacity building efforts that improve professional skills and knowledge, we are better able to bridge the gap between academic researchers and communities where health disparities are most prevalent.^{31,32}

AGILE SCIENCE

Another innovation over the past decade was recognition of the need for an "agile science" approach suggesting that the research process be iterative with novel approaches tailored to the individual.³³ A model that demonstrates agile science is the Just in Time Adaptive Intervention (JITAI) approach that is increasingly common in digital behavioral science and involves micro-level personalized interventions that are tailored to the individual participant.^{34,35} In addition, there are a number of clinician scientists conducting N=1 single subject, cross over designs with a goal being to tailor and iteratively test solutions per an individual's specific health concern.^{36–38} As interest grows in developing approaches that support precision health, recognizing the need for new methods, frameworks and ethics, including access by unand underserved communities is critical.⁸

CITIZEN SCIENCE

The role of the general public as active data collectors is not novel and, in fact its history is well documented.³⁹ The involvement of citizens in the conduct of health research has become much more visible in the past decades, largely due to technology enabled learning communities.⁴⁰ Learning communities are places where people congregate who share a common interest with a goal of sharing expertise and learning new skills.⁴¹ Those who are members at large of the "citizen scientist" learning community may identify as bio citizens,⁴² lead users,⁴³ personal scientist,⁴⁴ do it yourself or biohacker,⁴⁵ and participant-led researchers⁴⁶ and share a curiosity about solving real-world problems that do not appear to be prioritized by our current health systems.⁴⁷ In some cases, they are interested in improving their health and want to conduct self-tracking and, perhaps, self-experimentation. In other cases, they are addressing serious health issues and are frustrated by their experience with the healthcare system.⁴⁷ Regardless of their motivation, most of these individuals conduct their "research" outside of traditional settings and work within their respective communities to develop norms and standards of operations. While not surprising and yet unfortunate, there are outliers who have publicized unsafe approaches to self-experimentation (see Unnatural Selection⁴⁸), which has led to regulations in California published in 2019 preventing access to CRISPR kits for self-experimentation.⁴⁹

4. UNIQUE CHALLENGES AND OPPORTUNITIES

The preceding examples of community-engaged, agile and citizen science initiatives depict shifts in the traditional research enterprise including how research is designed, who is involved, what tools and strategies are used and potential gaps that may compromise ethical practices and research integrity. Grounded by scenarios derived from actual events and experiences, this Horizon Scan further explores how these shifts are demonstrated across the health research ecosystem. The following case studies are used to portray: 1- unconventional data sources; 2- establishment vs non-establishment research(ers); 3- challenges with non-representative and, subsequently, biased data used in predictive analytics and; 4- large-scale, longitudinal health data repositories that draw from diverse existing and evolving data sources. Each topic begins with an introduction, followed by a scenario and an analysis of ethical challenges. Each topic includes a set of "Reflection Questions" for readers to consider. These questions can be used to prompt discussion regarding various aspects of research including "the actors," the methods used and the extent to which the activities are regulated. While reflection and discussion is important, it is critical that the necessary infrastructures be developed and maintained to realize the potential value to the broader health research community.

The topics and scenarios feature research led by trained researchers gathering human data using tools like information communication technologies or pervasive sensing devices (e.g., mobile apps, GPS tracing, passive surveillance) as well as studies conducted by people who are not professional scientists who are operating outside of the traditional academic or industry research environments. Research planned and carried out by people without formal academic research training is classified as "non-establishment" research and is generally conducted in un- and under-regulated environments. ^{15,25} The scenario focusing on the use of artificial intelligence models used to improve health service delivery exposes the unanticipated harms associated with predictive analytics and the digital divide. Finally, given the national focus in recent years on advancing precision medicine and the creation of the All of Us Research Program, the last topic/scenario prompts readers to explore how large, multidimensional data repositories are being constructed with a goal of collecting quality data that are largely accessible.

5. DIGITAL HEALTH CHECKLIST FOR RESEARCHERS (DHC-R)

Following each scenario, a brief analysis of the ethical and social implications is presented. For this analysis, a decision-support framework grounded in accepted ethical principles with an accompanying checklist is used to facilitate consideration of factors that influence responsible and respectful practices in this new frontier.⁵⁰ Since these areas of health research are new for many, tools to assist in evaluating risks and benefits, accessibility, possible privacy harms and data management (e.g., security and sharing) protocols are needed. One tool, published in 2019, is the Digital Health Checklist and Framework for Researchers (DHC-R).⁵⁰ This tool was informed through an iterative process involving experts in the digital health research sector that included behavioral scientists, clinician researchers, ethicists, regulatory experts and legal scholars.



Figure 1: Digital Health Framework with checklist examples noted within each domain. The DHC-R is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License (2018-2020) and available at https://recode.health/tools/. Published with permission of C. Nebeker.

The DHC-R anchors four accepted ethical principles (see Belmont and Menlo Reports), with four intersecting domains: 1- Access and Usability, 2- Privacy, 3- Data Management and 4- Risks and Benefits (Figure 1). The DHC-R is a dynamic tool that is constantly evolving and, as of this writing, remains one of the few decision support resources available to those involved with digital health research. The DHC-R checklist is presented as a matrix with the ethical principles of respect for persons, beneficence, justice⁵¹ and respect for law and public interest⁵² on the left column with the four domains listed across the top.² Figure 2 shows the four domains in relation to the principle of Respect for Persons. The principle of Respect for Persons is typically applied to the practice of informed consent and, as such, the prompts pertain to what information may be important to share with a prospective research participant. Given the brevity of each topic scenario presented below, rather than a comprehensive evaluation using the DHC-R, each domain is briefly applied for demonstration purposes.

Digital Health Checklist Domains	Access & Usability	Risks & Benefits	Privacy	Data Management
Ethical Principle	Consent conveys:	Consent conveys:	Consent conveys:	Consent conveys:
Respect for Persons is the ethical principle that guides the practice of informed consent and the presentation of information needed for a person to decide whether to participate in research.	An explanation about the technology used in the study that helps one understand and informed choice (e.g., what, why, how). Plain language is used to describe Terms of Service/Privacy policy. List of relevant definitions using plain language. Access to visual and audio versions of information. A description that indicates there is a possibility of bystander involvement.	A description of the type of potential harm (e.g., physical, psychological, reputational) For potential harms identified, a description of the severity, duration and intensity and, acknowledged if unknown. Strategies for minimizing risks and managing risks that do occur. Statement that indicates possible benefits from knowledge gained during the study.	Nature and scope of personal information collected by the technology. How individual-level data will be shared and with whom. Whether personal data will be de-identified. A description of how a 3rd party may access and use participant information collected during study participation (normally found in a privacy policy when using a commercial device).	Data collection, transfer, storage and security practices. Information about who will have access to data collected by the technology. Whether the research data are controlled by the research team or a third party. Whether the participant will have access to individual-level or group-level data collected during the study.

Figure 2: In this figure, the DHC-R four domains intersect with the principle of Respect for Persons. Each cell includes prompts for the researcher to consider with respect to developing an informed consent process to convey information about the digital health tool that may be of importance to the prospective research participant considering study participation. The remainder of the matrix, not visible in this figure, focuses on considerations for the research protocol development. The DHC-R is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License (2018-2020) and available at https://recode.health/tools/. This segment of the DHC-R is published with permission of C. Nebeker.

2

6. TOPICS, SCENARIOS, ETHICAL CONSIDERATIONS AND ANALYSES

The following scenarios carry elements of the community-engaged, agile and citizen science examples presented above and integrate: 1- unconventional data sources; 2- establishment vs non-establishment researchers; 3- challenges with non-representative and, subsequently, biased data used in artificial intelligence models and; 4- large-scale, longitudinal health data repositories that draw from diverse existing and evolving data sources. Each topic is introduced and followed by a scenario and a brief analysis of ethical challenges. The reflection questions that conclude each segment are rhetorical and designed to prompt self or group discussion of relevant ethical, legal, and social implications including the downstream and unanticipated harms.

TOPIC 1: UNCONVENTIONAL DATA SOURCES

Data sources unconventional to HSR that contain individual and population "health signals" are ripe for exploration. These sources, for example, social media posts and global positioning system (GPS) location data, were not developed for use in health research and, as such, may not be obvious sources for researchers to examine. As these data sources are not necessarily intuitive to health services researchers, it is important to know of and manage the potential biases along with considering the value added if these sources are able to be leveraged and lead to generalizable knowledge. Moreover, if the value becomes evident, it will be critical to think through whether and how informed consent would be obtained or, for public data sets, whether expectations for use in research align with consumer expectations for granting permission for use in health research. Take, for example, tactical operational use of data sources where mobile phone signals and tweets are used to predict earthquakes⁵³ or infer global catastrophes⁵⁴ such that potentially lifesaving actions can be deployed immediately. Similarly, sales of cold and flu products can be used to flag infectious disease outbreaks.^{55,56}

Clearly, it is possible to use somewhat public data sources for operational interventions. When used for health research however, little guidance exists to inform best practices. Take, for example, the "Emotional Contagion" experiment conducted by Facebook researchers.⁵⁷ The researchers experimented with customer newsfeed content to see if less positive news influenced Facebook user posts. This practice experimentation by technology companies is an internal practice typically used to inform changes in product design. However, it turned into a health research study when one of the researchers wanted to share what they had learned about the phenomena known as "emotional contagion" by publishing their findings in a scientific journal. Emotional contagion is an area of study in the field of psychology. The question posed by the Facebook team was whether manipulating the emotional tone of one's newsfeed content could influence their emotions and behaviors. When the research findings are only used for internal product development, the activity is not considered research by federal regulatory standards (see definitions in the federal regulations).⁵⁸ However, when sharing results through a peer reviewed publication, it then qualified as research involving human subjects. As a research study, it violated accepted ethical standards for research with human participants by conducting the research without explicit informed consent of participants. In a study involving the Twitter

platform, researchers reported that some people who tweet expect to provide consent prior to their tweets being mined and used in research studies — even when those tweets are deemed public.⁵⁹ These examples point to a need for guardrails, perhaps in the form of governance and decision-support tools to shape acceptable and respectful expectations for practice - especially when the likelihood of regulation is not realistic anytime soon.⁶⁰

An example of proactive governance is that of The Menlo Report, drafted in 2012 by cybersecurity experts affiliated with the federal Department of Homeland Security. Recognizing the increase in Information and Communications Technology Research (ICTR) would introduce new ethical and social issues that paralleled those experienced in biomedical research, they adapted the Belmont Report principles to address new ICTR issues and challenges.⁵² However, while an excellent example of being forward thinking and proactive, neither the Belmont Report nor Menlo Report principles are required for use by those in the un- and under-regulated and non-establishment research sector.

To understand how research occurs in under- and unregulated sectors, several years ago, a paper was published in the New England Journal of Medicine that described weight gain and loss associated with holidays in Japan, Germany and the United States.⁶¹ To conduct this analysis, consumer data were obtained from the Withings corporate database that was populated by people who used its wireless weight scale (WS-50).⁶¹ Curious about consumer preferences to provide consent when their weight or other health data would be used for research or clinical purposes, academic researchers partnered with Withings to deploy a brief survey.⁶² The survey asked customers if they would be willing to share data collected by the Withings product to advance scientific research or to support their clinical care. Responses were mostly favorable, especially if the data were deidentified. A few participants (247 respondents representing 15% of the total responses) provided comments generally focused on a desire to provide consent, receive compensation for data use and assurance that privacy and consent regulations were respected. An interesting finding was that many respondents did not realize that data collected by Withings were already being used for health research without their explicit consent. The tradition of prospective informed consent to participate in research occurs in regulated human subjects research. In corporate research, the permissions (should they exist) are located in the terms of service agreements, which are not typically read by those accessing the product. In this particular study, the majority disagreed that they routinely read service agreements carefully.62

The takeaway message is that when regulations do not require prospective and accessible informed consent, consumers may feel as though their rights have been violated and, subsequently, may be less likely to trust researchers. This example with Withings is one that could easily be generalized to Twitter, Facebook, LinkedIn, Google and other corporate entities. While organizations are involved in conducting research with people or data obtained from people, it is important to recognize that organizations conducting what would qualify as research with human participants in regulated "establishment" research operate outside of these requirements intended to protect human research participants.²⁵ This is largely due to the fact that regulations in the US are tied to whether the organization receives federal funding to conduct research – if not, the Common Rule⁵⁸ does not dictate practice.

TOPIC 1: REFLECTION QUESTIONS

How might we leverage passive data sources (search engines, shopping, fitness trackers) not typically considered for health research applications, for individual and public health benefit?

How might these and other data sources un- or under-utilized in traditional settings be accessed respectfully and used fairly and responsibly to democratize access to care by those currently marginalized due to geographic location or other social determinants?

When consumer data are used in health research, what is our obligation to obtain prospective and accessible informed consent as well as share research results with those who contribute?

Topic 1 Scenario: Unconventional Data Sources Title: College student mental health and the Dartmouth Student Life study. https://studentlife.cs.dartmouth.edu/

Classifiers: federally regulated, unconventional data sources (GPS, microphone)

The Student Life study tested a smartphone application and sensing system to infer student behavior with a goal of understanding signals of student stress. Could faculty and administrators potentially use these data to prevent dropouts or health risk behaviors? Students (n=48) were enrolled spring quarter and downloaded the app. Machine learning algorithms were used to assess sensor data to passively make inferences about student behavior including sleep, social and physical activity. The app was programmed to record sleep, conversations, physical activity, location and movement. Student participants responded to ecological momentary assessments (EMAs) of their mood and stress level and completed validated mental health surveys. Rather that reporting class attendance, the GPS could identify if the student was physically present in the course in which they were enrolled, at the library or at the gym etc. The microphone could measure the length of time conversing. The study captured behavioral trends across the 10-week term and observed changes in sleep, socializing, conversation length, workout frequency and course attendance. The research team gleaned these data without needing to interact with the participant, with the exception of the brief EMAs conducted periodically to assess mood or stress.

DHC-R Brief Analysis:

Access/Usability: Students were required to have an android smartphone to participate. This requirement may have prevented some students from participating.

Privacy: Studies like this are able to gather volumes of granular and potentially sensitive data through the mobile application, which can be used to answer important health related questions. That being said, passive sensor technologies and communication platforms are either producing or collecting granular and sensitive data that are not always under the protective privacy policies (e.g., FERPA and HIPAA) and could reveal stigmatizing health issues.[REF]

Data Management: The investigators of the Student Life study have made the resulting raw data set open to the public, which can be accessed at: <u>https://studentlife.cs.dartmouth.edu/dataset.html</u>. There is a disclaimer that for privacy reasons, data that can reveal a participant's identity have been removed. That would appear to be a respectful data management practice; however, GPS data remain available. The GPS coordinates can be used to identify where a person was sleeping – the exact longitude and latitude, which provides an address that can be linked to the person who lived at that address and, subsequently identify an individual who participated. If this is an oversight, that may mean that IRBs and researchers are not aware of the sensitive nature of GPS data nor how location data can be used to identify an individual. If that's the case, educating all stakeholders involved in the digital health research decision-making (e.g., risk assessment, consent, etc.) must be prioritized. As we see pervasive health studies increasing, ethics boards, legal scholars, privacy and data management experts may be needed to prospectively shape protocols that include appropriate research participant protections.

Risks/Benefits: A research benefit can only be realized if the data collected are valid and reliable. As the app developed for this study was created by the study team, it is important to assess the quality of data collected to be certain that the data are trustworthy. If the app is confirmed to produce valid and reliable data, a next step is to evaluate possible risks of harms including the type, duration, severity and intensity. As the study is collecting sensitive individual level data 24/7, the study team should have a plan in place if they observe a possible anomaly that may be indicative of a crisis. For future actions, faculty and administrators who consider using an app of this sort to gauge student health and, subsequently plan interventions should be certain about the tool's accuracy in measuring what it claims to measure.

With respect to informed consent, it is appropriate to question whether prospective participants actually understand the nature and granularity of data gathered using this particular app. If the students' technology and data literacy are low, it would be important to provide examples that depict the individual level of data collected during the consent process to increase understanding; otherwise, their awareness of potential harms may not register.

TOPIC 2: ESTABLISHMENT VS NON-ESTABLISHMENT

In addition to new data sources, it is important for HSR to recognize that the heath care ecosystem is changing and expanding beyond the establishment – that is, the traditional academic research sector. In addition to corporate research carried out using consumer data (e.g., financial institutions, transportation, social network platforms), patients and everyday citizens are increasingly taking their health and healthcare into their own hands.^{44,63} In some cases, this liberation for citizens/patients stems from persistent frustrations with the healthcare system in addressing their needs.⁴⁷ For others, it is driven by a curiosity of wanting to learn how to influence their health and wellbeing by doing selfstudies.⁶⁴ Access to the internet combined with direct-to-consumer digital tools, including mobile apps and pervasive sensor technologies, have made it possible for people to understand and, subsequently, improve their personal health. Those who may be called "citizen scientists" might begin by solving a personal health problem but then find ways of sharing practices with others and forming communities (see Open APS², Quantified Self³, Precision Health Ecosystem⁴). For the most part, these individuals and communities operate outside of the traditional, regulated research environment and without the infrastructure checks (e.g., ethics review) that can foster safe and rigorous research. This is not to say that non-establishment research lacks rigor or integrity, but that there are system gaps that need to be recognized as these efforts are growing and lack the governance needed to guide safe practices.^{46,65}

"The need for innovative and responsive solutions has never been greater and more urgent than we have seen with the pandemic. Just One Giant Lab or JOGL is one initiative that gained traction rapidly as its community of citizen scientists collectively began to design research and share results.⁶⁶ JOGL is a research and innovation platform that made it possible for experts and laypeople to answer research questions (i.e., develop COVID-19 diagnostic tests, validating mask efficacy) and work on meeting the needs of frontline healthcare workers (i.e., making masks and other forms of personal protective equipment). Working outside of the establishment created problems from regulatory and legitimacy perspectives. The issues of legitimacy are not new, and regulatory requirements are rigid for good reason. However, the need exists to identify pathways for citizen scientists to have innovations vetted for safety and efficacy, and shared more easily so that others can benefit. However, pathways for sharing safe and effective innovations with others who can benefit are needed.

Recently, a trust architecture was proposed to facilitate development of community norms that support ethical, accessible and trustworthy DIY science and bridge the gap between establishment and nonestablishment scientific pursuits.¹⁵ The trust architecture is developed by combining formal and informal structures designed to shore up legitimacy and trust for an endeavor (e.g., financial institutions, drug manufacturing) while mitigating possible harms.¹⁵ With a trust architecture, the idea is to 1- build from what is working and not force establishment norms (e.g., IRBs); 2- recognize that the conventional path is not the only path, and; 3- consider how to facilitate motivators and reduce barriers, like intellectual property rights, to foster innovation during emergencies.¹⁵

TOPIC 2: REFLECTION QUESTIONS

How might we support the growing under-regulated, non-establishment, uncoordinated research enterprise?

What can we learn from the non-establishment research (e.g., commercial entities, citizen science, DIY) to move from system-driven healthcare to patient- and community-driven health research, policy and practice?

How are these efforts supported and what are the models for expanding support to foster sustainable community-driven health?

https://openaps.org/what-is-openaps/

https://quantifiedself.com/

https://precisionhealthcareecosystem.org/

Topic #2 Scenario: Establishment vs Non-establishment

Title: Participant-Led Research

https://bmjopen.bmj.com/content/9/4/e025633.info

Classifiers: non-establishment, under-regulated, conventional data sources (blood testing)

Several years ago, a few dozen people associated with Quantified Self decided to participate in a participant-led research project to learn about blood lipids. Specifically, they were interested in how their daily blood lipids fluctuated in response to various activities (e.g., sleep, diet, exercise, etc.). Each person designed their own self-experiment and collected finger prick blood samples using a CardioChekTM device. Prior to beginning their study, a training took place so that each person was familiar with the protocol for taking and disposing of samples and using the device. There was also a group meeting held to discuss potential risks and benefits of participating in the self-study. This meeting served as the "self-consent" process. Each person developed a protocol for recording and analyzing their data There was an option to share data on a group spreadsheet. The study lasted about 6 weeks. Participants were interviewed at the end of the study to learn about what went well, what improvements could be made and whether there were any surprises.⁶⁷

DHC-R Brief Analysis:

Access/Usability: The plan involved webinar style trainings and postings on a communications platform where participants received instruction on how to use the device along with protocols for collecting blood samples and disposing of materials. The device and materials were shipped to people who participated alogn with written instructions.

Privacy: PLR participants were identifiable on the communication platform but, could control what information they shared with others in the group. They were able to meet as part of the group or individually with the PLR organizers.

Data Management: Data collected by each individual were then recorded on their personal spreadsheet. They could choose to contribute their data to a group spreadsheet that would be visible to other PLR participants or, limit sharing. What this means to HSR is that individual level data may be collected and shared using digital tools or old-fashioned pen/paper journals and, these data may remain at the individual level or be combined with other individual level data. The issue for researchers would be similar to accessing any other existing data set where data quality is essential. How to improve data quality is a persistent challenge.

Risks/Benefits: Risks were identified through two group discussions – the first occurred prior to the start of the PLR and the other, a few weeks after the PLR projects began. The goal was to not only identify possible risks but, to also explore solutions to manage those risks. Possible benefits were also discussed. At the end of the PLR, one-on-one interviews were held with each participant to identify whether the risks and benefits discussed prior to the study aligned with their actual experience.

With respect to informed consent, the group attended a presentation on the purpose of informed consent and what is typically disclosed during a study that is covered by federal regulations. While a formal consent was not developed for this PLR, the concept of self-consent was discussed by all during convened group meetings. That being said, what if a health services researcher came across these data through an open access repository? Organizations like Open Humans were created for this purpose – to allow the opportunity for people to create and share individual level data. In addition to hosting these studies, the organizers publish expectations and responsibilities for conducting research on the platform, including ethics review, and data access, use and sharing protocols.

TOPIC 3: PREDICTIVE ANALYTICS AND THE DIGITAL DIVIDE

It is also important to recognize the <u>reality of a digital divide</u>. The challenges of technology access and connectivity are non-trivial as they are the source of downstream harms, including unknown harms that we cannot easily anticipate.⁸ We make assumptions that technology is accessible and that everyone has a smartphone, but this is not true. The challenges of digital equity have been magnified with COVID-19. Not all have access to technologies, and those who do may have difficulty connecting due to cost or low technology literacy. This is important because it is through consumer use of products that big data sets are created that, subsequently, can be used as training data for the artificial intelligence learning process.

When algorithms are trained on data that are not representative of the population, then the resulting predictive analytics will be flawed. Missing data, by virtue of groups not being included (i.e., minorities, older adults) result in non-representative data sets, which then lead to biased and flawed algorithms that cause harm. The digital divide can emerge in the regulated, establishment and under-regulated, non-establishment settings and have equally detrimental outcomes. Take for example the use of health information technologies like the electronic health record (EHR). The EHR was developed to improve efficiencies in healthcare systems, including billing of services rendered. When the EHR is used to determine care needs rather than care costs, problems arise in the form of perpetuating disparities. In research published in *Science*, Obermeyer et al. (2019) describe how unequal access to care influences how much is spent caring for Black patients when compared to White patients.⁶⁰ From an operational perspective, when members of the health system operations team noticed the impact of no-shows on the bottom line, they wondered if an AI model could address the problem by predicting no-shows. If they could anticipate a no-show, they could then double book that timeslot. Imagine what happens when the model shows that People of Color tend to have a higher incidence of not making their appointments and, as a result, are now double booked?⁶⁹

The problem depicted here is one of models not factoring in the important social determinants that influence why a person may not make their appointment, or for that matter, even seek out healthcare. Rather than address the social circumstances, the model exacerbates existing disparities without actually recognizing the person behind the no-show. That said, this use of EHR data is considered legitimate administrative use by authorized providers. Yet, the provider who decides to solve the no-show problem via AI may not foresee the potential harms. To prevent the data controller from creating AI models that cause harm in the name of improved efficiencies, a prospective vetting process of planned data uses might be a solution. Given that those conducting HSR may not have legitimate access to these data, perhaps perhaps decision-makers (i.e., scientists, administrators, data controllers) are consulted to identify the precursors to no-shows and help to design interventions that take into account the social determinants that may contribute. Moreover, if there are ethical approaches to permit the sharing of these data with the HSR community that take into account the regulations and privacy expectations for data access, problems with access can be addressed as a research question to then inform a healthcare management problem.

TOPIC 3: REFLECTION QUESTIONS

How might we train, support and incentivize our clinicians and healthcare professionals to collect and factor in the social determinants of health that are critical to blending the best of technology with social justice and equity?

How do we get ahead of a future where the "health" technologies are potentially transformative, but health professionals resist adoption due to a lack of confidence in both the technologies and their lack of relevant training?

How do we ensure that "big data" sources and the promise of AI are not contributing to health disparities?

Topic 3 Scenario: Predictive Analytics

Title: Digital Phenotyping Mental Health

Relevant resource: <u>https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1377-7</u> and Keywise: <u>https://keywise.tech/</u> and <u>https://www.biaffect.com/</u>

Classifiers: semi-establishment (academic research/start-up company), regulated, unconventional methods

We interact with our phones, computer mouse and keyboards and how we each do that is unique to the point that we all have a digital signature. This digital signature was something that Eric Horvitz and colleagues speculated could potentially reveal a diagnosis of Parkinson's disease.⁷⁰ Led by a research team at the University of Illinois at Chicago, a technology has been developed that uses the patient's digital signature to study mood and cognition, including bipolar disorder. Patients who enroll in the BiAffect Study, which is approved by an IRB, agree to download an app to their smartphone, which then replaces their native smartphone keyboard with a virtual keyboard that passively collects keystroke dynamics data (i.e., not what you type but how you type it). Through the capture of keystroke metadata and application of machine learning algorithms to the data, the BiAffect team is able to capture individualized patterns that serve as a digital phenotype of neuropsychiatric symptoms.⁷¹

DHC-R Brief Analysis:

Access/Usability: The BiAffect app is available on iOS with an Android version set for release soon. The app operates passively on the one's smartphone and, as an open science project, US based participants can join the study by completing the informed consent process on their smartphone from the comforts of their home.

Privacy: Participant identity is generally not linked to the metadata collected. For example, for analytic purposes, the backspace key and space bar data are collected but not the individual character data to make anonymization of data possible. Community level data are accessible to participants and are returned to participants via the BiAffect team's Twitter account.

Data Management: The data are uploaded via a secure encrypted protocol to the study server. Participants receive results of the studies in which they participate via the community dashboard with some participant-level data being returned including keyboard activity, self-monitoring information and results of some in-app survey questions and tests. It is unclear if data from these studies are accessible to other researchers, including those within the HSR community.

Risks/Benefits: Several studies have been published that report feasibility and validity of neuropsychiatric symptoms with more studies in progress to validate the keystroke dynamics. Testing feasibility is an important step to creating a reliable and valid measurement tool. Should this technology prove effective, clinician researchers will be able to use predictive analytics, personalized to a particular patient, to improve symptom management and overall treatment.

TOPIC 4: UNUSUAL HEALTH DATA REPOSITORIES

The promise of precision medicine has led to a number of national and global initiatives to create somewhat <u>unusual health data repositories</u>.^{72,73} In the US, the National Institutes of Health's All of Us Research Program (AoURP) began recruiting its one million person cohort in 2018 with hopes of collecting biological, behavioral and environmental longitudinal data, including sensor derived data, and creating the nation's largest data/biorepository.⁷⁴ Concurrent with the launch of the AoURP, scientists in New York City were designing the HUMAN project – a platform for hosting big data human studies including a longitudinal study that would enroll a representative sample of 10,000 NYC residents. Both data repository programs planned to involve participants for 10 or more years and, with the resulting data collected and housed in the repository, answer some of our most pressing questions about human health.

Creating large biorepositories is not necessarily new; however, what is unique about these two precision medicine initiatives is the use of atypical data sources to make inferences about health. For example, accessing purchasing data via loyalty cards, remote home sensors to capture air quality, wearable sensors to passively obtain individual's movement within the home, accessing judicial and employment records to infer quality of life. These data sources are normally siloed yet, when imported into a networked system, they can provide a rich set of data points that can be leveraged to better understand factors influencing individual and community health. By creating a multidimensional data repository, we may be able to answer important individual and societal health issues. The vision is that the data repository would be accessible to traditional "establishment" researchers as well as the "non-establishment" community to answer a variety of research questions.

The creation of a research resource of this scope and magnitude challenges our existing infrastructure in new ways. For example, can informed consent be truly informed when the specific research questions are not yet known? Researchers asked this question and learned that those involved in the recruitment and enrollment of biorepository participants viewed the process of consent as a form of engagement.⁷⁵ While staff could not provide specific study information, they saw it as an opportunity to begin a relationship with the prospective contributors to the repository.⁷⁵ Similarly, AoURP developed guiding principles that include a commitment of sharing research results back to repository contributors and consider participants to be partners in the research endeavor. The idea of returning research results is new and guiding principles and development of corresponding best practices are only now being shaped.^{76–78}

TOPIC 4: REFLECTION QUESTIONS

When involving people in long-term data collection when the actual research questions are not yet known, what methods of informed consent are meaningful and lead to informed participants?

What are the short and long-term engagement strategies and how are those determined?

Increasingly, we are looking for how to engage participants in learning about their contributions via accessible individual and group level results. How can we appropriately and respectfully return research results in ways that present value to research participants?

What can be learned from failed attempts (National Children Study)⁸⁸ and successful endeavors (Million Brazilian Cohort)?⁸⁹

Topic 4 Scenario: Health Data Repositories

Title: The HUMAN Project

https://www.thehumanproject.org/

Classifiers: establishment, regulated, unconventional and conventional data sources

Until now, large-scale longitudinal studies of people have focused on specific domains of inquiry and subsets of populations. The HUMAN Project was designed to look at dynamic patterns and feedback mechanisms between human behavior, biology and environment across the lifespan of 10,000 people. Learning how these domains interact will enable the development of research, treatments and policies. Study participants will provide or agree to release data over a 10+ year period that includes:

- Traditional Medical: genome, proteosome, metabolome, microbiome, blood chemistry, EMR, environmental toxins, household air quality
- · Financial: wealth, labor, taxes, swipe-level purchases, cash flows
- Social Network: SMS, MMS, phone, email, search terms, geotracking, MAC/Bluetooth locations
- · Surveys: behavior, education, personality, mental health, education
- Family Interactions: tracking parent/child interactions and elder family interactions
- Criminal Justice: legal, incarceration

Some of these data currently exist. Other data will be captured using sensor technologies and ecological momentary assessment strategies. The study team will store all data in a state-of-the-art secure vault for future access by scientists. Families will be informed about the study in person and, once enrolled, will be reconsented annually. When children reach legal age, they will have the option of continuing or not and, if not, of choosing whether to allow the retention of data previously collected for future analysis.

DHC-R Brief Analysis:

Access/Usability: The study will recruit a representative sample of people living in NYC leading to generalizable results. The study materials have been translated to languages spoken by the majority of people. Body worn or remote sensor technologies will be explained to the participant. If necessary, technologies placed in the home will be installed by the study team. The consent process includes illustrations depicting how personal data are gathered, stored and shared using language accessible to people reading at a 6th grade level and illustrations developed to clearly convey complex topics.

Privacy: The checklist prompts reflection on privacy expectations and the need to protect personal information from 3rd party access and exploitation. Moreover, the need to guard against bias so that profiling and social harms like discrimination are minimized. The consent would need to contain information about the personal information collected and how those data may be shared, including any additional risks associated with privacy policies if a 3rd party vendor is used for data collection (e.g., ecological momentary assessment).

Data Management: While the family may consent to participating, any visitors to the home would potentially be included in the data collection by remote sensor technologies. Bystander rights would need to be considered and managed prospectively. In addition to managing potential bystander data, the protocol should address how data are managed and whether the participant will have access to individual level data. The data management system needs to be checked for any vulnerabilities and participants will need to be informed of how a potential data breach will be handled. Protocols are needed to inform participants if any data will be transferred to the electronic health record and, of course, make sure the data storage is HIPAA compliant and meets minimum encryption standards. Children, when reaching the age of majority, will have the ability to consent and, if desired, have any previously collected data removed from the project.

Risks/Benefits: Many of the data sources were not intended to be used as health data and, as such, any inferences made must be verified as valid, to the extent possible. Creating a knowledge network, such as what would be created via the HUMAN database, is potentially a game changer with respect to understanding human health and the factors that influence individual and community level health. The promise of precision health is only possible if the risks of harm are examined carefully and via an ongoing process. For example, if a social scientist is interested in passively observing human dynamics and obtains data from sensors placed in the home or worn by the family members, it may be possible to infer child abuse based on movement patterns over time. Before these data are useful, the tools used to capture data must be valid and reliable or, they could be more harmful than useful.

7. WHAT'S NEXT?

COVID-19 has led to a seismic digital takeover with 1 billion telemedicine events since March 2020 and an ever-increasing reliance on technologies to access food, education, entertainment, supplies, services and information. Whereas in January 2020, we saw increasing criticism of the investments in digital health, 2020 closed out as a banner year for digital health investments.⁷⁹ The impact of sociotechnical health systems and new ways of leveraging data will continue to impact health research and care in the foreseeable future.

The examples provided in this Horizon Scan have highlighted new potential benefits along with harms associated with new data sources, new actors and organizations that vary with respect to formal research training, regulatory structures and ethics acculturation. Moreover, potential risks associated with using unrepresentative training datasets that exacerbate harms combined with the ethical and

responsible access and use of data sources that were never intended for use in health research should cause pause. With this shifting research ecosystem, the need to evaluate our infrastructures and update or replace obsolete processes is critical – that is if we are to evolve and advance important health research with the plethora of tools and sources in existence and emerging. As noted, current methods for qualifying and assessing risks of harm are typically not evidence-based and we rely on an IRB process that produces variable and subjective outcomes – often delaying important health research.^{8,80} Innovating and testing new methods for ethics review are clearly needed. Moreover, current models for obtaining informed consent are not effective, data management protocols are inconsistent, and little is known about how to authentically engage participants as partners in the research experience.^{81,82}

Accessing data from non-traditional sources may be leveraged and adapted for HSR, but what are the access barriers? And, once the data are accessible, what is the process for evaluating potential risks? For example, if we know that the use of keyboard kinematics is potentially useful in digital phenotyping from regulated research studies,^{71,83} who should determine if it is appropriate to gather these passive data without the full consent of consumers in our public health settings? Is it appropriate to use keyboard kinematics to assess mental health, or Facebook posts to flag suicidal ideation? While these data sources exist, they were not developed for the purpose of making health predictions or deploying interventions, and obtaining effectiveness data from those guarding proprietary secrets makes a transparent assessment process difficult. That's not to say these sources are off limits but, we – researchers, ethicists, legal scholars, regulatory experts and the public – do need to be mindful of potential harms that could in fact undermine the possible benefits.

Other critical issues to consider are governance structures. In some cases, governance is non-extant or inconsistent across sectors. The emergence of non-establishment activity under the broad citizen science umbrella is an area that would benefit from support, including education and a form of ethics review.⁸⁴ What supports are needed to increase access, to democratize governance, to equalize power — and to support agency, training, governance, and ethics? How should stakeholders collectively foster safe, ethical and respectful use of consumer generated data for individual, population and public health purposes? To answer these questions, the Citizen Science Association formed an Ethics Working Group, and the National Academies workshop on emerging issues in bioethics included presentation on the challenges and opportunities for personal science and dedicated a special issue of its journal to this topic.⁸⁵ In the past year, there have been discussions at establishment and non-establishment conference venues calling attention to the emerging DIY community including challenges faced and efforts to develop a community accessible ethics review program. These challenges are among those to be addressed as we consider the new bioethics and research ethics in the 21st century.¹⁶

8. CONCLUSION

Novel and emerging technologies present opportunities to access data for use in health services research. These data may not have been collected for the purposes of health research and, as such, additional ethical and social considerations come into play to assess and minimize potential harms and unintended consequences. Data from diverse sources are only valuable in the health sector if representative, and bias is mitigated to the extent possible. Ethical principles abound to guide the development and use of AI, including recognizing and managing bias, accountability, transparency, rigor, agency and trust and trustworthiness. These principles are meaningless without associated practices that are accepted and adopted by the broader community of stakeholders. Furthermore, we can no longer defer ethical reflection and decision making to IRBs or their equivalents. The research community is collectively responsible for creating an infrastructure in which decisions are guided by ethical principles. By owning this responsibility, we can develop standards of practice that promote trust and trustworthiness – the foundation from which we can advance better health for all.

References

- 1. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomized controlled trial. Am J Public Health. 2008;98(8):1359-1366. doi:10.2105/AJPH.2007.124446
- Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: A review of alternatives in mhealth clinical trial methods. JMIR mHealth uHealth. 2016;4(3):e107. doi:10.2196/mhealth.5720
- Henry D, Tolan P, Gorman-Smith D, Schoeny M. Alternatives to Randomized Control Trial Designs for Community-Based Prevention Evaluation. Prev Sci. 2017;18(6):671-680. doi:10.1007/s11121-016-0706-8
- 4. Whicher D, Rosengren K, Siddiqi S, Simpson L. A N AT I O N A L AC A D E MY O F M E D I C I N E S P EC I A L P U B L I C AT I O N THE FUTURE OF HEALTH SERVICES RE-SEARCH Advancing Health Systems Research and Practice in the United States.; 2018. https://lccn.loc.gov/2018048491. Accessed February 6, 2021.
- Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. Lancet (London, England). 2014;383(9912):156-165. doi:10.1016/S0140-6736(13)62229-1
- Bloss C, Nebeker C, Bietz M, et al. Reimagining human research protections for 21st century science. J Med Internet Res. 2016;18(12). doi:10.2196/jmir.6634
- Pittman P. Health services research in 2020: Data and methods needs for the future. Health Serv Res. 2010;45(5 PART 2):1431-1441. doi:10.1111/j.1475-6773.2010.01149.x
- Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision Health: The Role of the Social and Behavioral Sciences in Advancing the Vision. Ann Behav Med. 2020;54(11):805-826. doi:10.1093/abm/kaaa018
- 9. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol. 2015. doi:10.1038/nbt.3223
- 10. Doherty AR, Kelly P, Kerr J, et al. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. Int J Behav Nutr Phys Act. 2013;10(1):22. doi:10.1186/1479-5868-10-22
- Jankowska MM, Schipperijn J, Kerr J. A Framework for Using GPS Data in Physical Activity and Sedentary Behavior Studies. Exerc Sport Sci Rev. 2015;43(1):48-56.
- Wang R, Chen F, Chen Z, et al. StudentLife : Assessing Mental Health , Academic Performance and Behavioral Trends of College Students using Smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ; 2014:3–14.
- Barnett I, Torous J. Ethics, Transparency, and Public Health at the Intersection of Innovation and Facebook's Suicide Prevention Efforts. Ann Intern Med. 2019;170(8):565. doi:10.7326/M19-0366
- Seelye A, Mattek N, Sharma N, et al. Passive assessment of routine driving with unobtrusive sensors: A new approach for identifying and monitoring functional level in normal aging and mild cognitive impairment. J Alzheimer's Dis. 2017;59(4):1427-1437. doi:10.3233/JAD-170116
- Rasmussen, L.M., Guerrini, C.J., Kuuiken, T., Nebeker, C., Pearlman, A., Ware. S.B., Wexler, A., Zettler P.J. Realizing Present and Future Promise of DIY Biology and Medicine through a Trust Architecture. Hastings Cent Rep. 2020;November-D.
- Addie S, Hackman M, Choi K, Wizeeman D, Beachy S. An Examination of Emerging Bioethical Issues in Biomedical Research. National Academies Press; 2020. doi:10.17226/25778
- 17. Lohr KN, Steinwachs DM. Health Services Research: An Evolving Definition of the Field. Health Serv Res. 2002;37(1):15-17. doi:10.1111/1475-6773.01020
- Steinwachs DM, Hughes RG. Health Services Research: Scope and Significance. Agency for Healthcare Research and Quality (US); 2008. http://www.ncbi.nlm.nih. gov/pubmed/21328761. Accessed December 29, 2020.
- 19. Nichol AD, Bailey M, Cooper DJ. Challenging issues in randomised controlled trials. Injury. 2010;41(SUPPL. 1). doi:10.1016/j.injury.2010.03.033
- 20. van Dorn A. COVID-19 and readjusting clinical trials. Lancet (London, England). 2020;396(10250):523-524. doi:10.1016/S0140-6736(20)31787-6
- Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16(1). doi:10.1186/ s13063-015-1023-4
- Castillo CHM, Garrafa V, Cunha T, Hellmann F. Access to health care as a human right in international policy: Critical reflections and contemporary challenges. Cienc e Saude Coletiva. 2017;22(7):2151-2160. doi:10.1590/1413-81232017227.04472017

- Jones GH, Kantarjian H. Health care in the United States-basic human right or entitlement? Ann Oncol. 2015;26(10):2193-2195. doi:10.1093/annonc/mdv321
- 24. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed Inform Insights. 2018;10:117822261379286. doi:10.1177/1178222618792860
- 25. Nebeker C. mHealth Research Applied to Regulated and Unregulated Behavioral Health Sciences. J Law, Med Ethics. 2020;48(Supp. 1).
- Pagoto S, Nebeker C. How scientists can take the lead in establishing ethical practices for social media research. J Am Med Informatics Assoc. 2019;26(4). doi:10.1093/jamia/ocy174
- Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: Assessing partnership approaches to improve public health. Annu Rev Public Health. 1998;19:173-202. doi:10.1146/annurev.publhealth.19.1.173
- Minkler M. Ethical challenges for the "outside" researcher in community-based participatory research. Health Educ Behav. 2004;31(6):684-697. doi:10.1177/1090198104269566
- Hohl SD, Thompson B, Krok-Schoen JL, et al. Characterizing Community Health Workers on Research Teams: Results From the Centers for Population Health and Health Disparities. Am J Public Health. 2016:e1-e7. doi:10.2105/ AJPH.2015.302980
- O'Brien MJ, Squires AP, Bixby A, Larson SC. Role development of community health workers: an examination of selection and training processes in the intervention literature. Am J Prev Med. 2009;37(6):262-296.
- Nebeker C, Kalichman M, Talavera A, Elder J. Training in Research Ethics and Standards for Community Health Workers and Promotores Engaged in Latino Health Research. Hastings Cent Rep. 2015;45(4):20-27. doi:10.1002/hast.471
- Nebeker C, Giacinto RE, Pacheco BA, López-Arenas A, Kalichman M. Prioritizing Competencies for "Research" Promotores and Community Health Workers. Health Promot Pract. March 2020:152483992091354. doi:10.1177/1524839920913548
- Hekler EB, Klasnja P, Riley WT, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med. 2016;6(2):317-328. doi:10.1007/s13142-016-0395-7
- Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med. 2018;52(6):446-462. doi:10.1007/ s12160-016-9830-8
- Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Heal Psychol. 2015;34(Suppl):1209-1219. doi:10.1037/hea0000306
- Duan N, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: A pragmatic clinical decision methodology for patient-centered comparative effectiveness research. J Clin Epidemiol. 2013;66(8 SUPPL.8):S21-S28. doi:10.1016/j.jclinepi.2013.04.006
- Schork NJ, Goetz LH. Single-Subject Studies in Translational Nutrition Research. Annu Rev Nutr. 2017;37:395-422. doi:10.1146/annurev-nutr-071816-064717
- Whitney RL, Ward DH, Marois MT, Schmid CH, Sim I, Kravitz RL. Patient Perceptions of Their Own Data in mHealth Technology–Enabled N-of-1 Trials for Chronic Pain: Qualitative Study. JMIR mHealth uHealth. 2018;6(10):e10291. doi:10.2196/10291
- Bonney R, Shirk JL, Phillips TB, et al. Citizen science: Next steps for citizen science. Science (80-). 2014;343(6178):1436-1437. doi:10.1126/science.1251554
- Wiggins A, Wilbanks J. The Rise of Citizen Science in Health and Biomedical Research. Am J Bioeth. 2019;19(8):3-14. doi:10.1080/15265161.2019.1619859
- Index H, Theories L, Definitions M, Books P. Communities of Practice (Lave and Wenger). Search. 2010:8-9. about:reader?url=https%3A%2F%2Fwvw.learning-theories.com%2Fcommunities-of-practice-lave-and-wenger.html.
- Pauwels E, Denton S. The Rise of the Bio-Citizen.; 2018. https://www.wilsoncenter.org/sites/default/files/rise_of_biocitizenfinal.pdf. Accessed September 9, 2019.
- von Hippel E. Lead Users: A Source of Novel Product Concepts. Manage Sci. 1986;32(7):791-805. doi:10.1287/mnsc.32.7.791
- Wolf GI, De Groot M. A Conceptual Framework for Personal Science. Front Comput Sci. 2020;2:21. doi:10.3389/fcomp.2020.00021
- Rasmussen LM, Guerrini CJ, Kuiken T, et al. Realizing Present and Future Promise of DIY Biology and Medicine through a Trust Architecture. Hastings Cent Rep. 2020;50(6):10-14. doi:10.1002/HAST.1194
- 46. Grant AD, Wolf GI, Nebeker C. Approaches to governance of participant-led

22

research: a qualitative case study. BMJ Open. 2019;9(4):e025633. doi:10.1136/ bmjopen-2018-025633

- Nebeker C, Weisberg B, Hekler E, Kurisu M. Using Self-Study and Peer-to-Peer Support to Change "Sick" Care to "Health" Care: The Patient Perspective. Front Digit Heal. 2020;2(2):2. doi:10.3389/fdgth.2020.00002
- Unnatural Selection: the eye-opening Netflix docuseries on gene editing I Documentary I The Guardian. https://www.theguardian.com/tv-and-radio/2019/ oct/17/unnatural-selection-netflix-docuseries-gene-editing. Accessed February 14, 2021.
- Today's Law As Amended. https://leginfo.legislature.ca.gov/faces/billCompareClient.xhtml?bill_id=201920200SB180&showamends=false. Accessed February 14, 2021.
- Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med. May 2019. doi:10.1093/tbm/ibz074
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (NCPHS). Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Vol 44.; 1979.
- Dittrich D, Kenneally E. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. U.S. Department of Homeland Security; 2012.
- St Louis C, Zorlu G. Can Twitter predict disease outbreaks? BMJ. 2012;344(7861). doi:10.1136/bmj.e2353
- Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. PLoS Med. 2011;8(8):e1001083-e1001083. doi:10.1371/journal.pmed.1001083
- Pivette M, Mueller JE, Crépey P, Bar-Hen A. Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. BMC Infect Dis. 2014;14(1):604. doi:10.1186/s12879-014-0604-2
- Liu TY, Sanders JL, Tsui FC, Espino JU, Dato VM, Suyama J. Association of Over-The-Counter Pharmaceutical Sales with Influenza-Like-Illnesses to Patient Volume in an Urgent Care Setting. PLoS One. 2013;8(3):59273. doi:10.1371/ journal.pone.0059273
- Kramer A, Guillory JE, Hancock JT. Experimental Evidence of Massive-scale emotional contagion through social networks. Proc Natl Acad Sci. 2014;111(24):8788-8790. doi:10.1073/pnas.1412469111
- 58. Services USD of H and H. Code of Federal Regulations Title 45, Public Welfare; Part 46, Protection of Human Subjects.
- Fiesler C, Proferes N. "Participant" Perceptions of Twitter Research Ethics. Soc Media + Soc. 2018;4(1):2056305118763366. doi:10.1177/2056305118763366
- Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med. 2019;17(1):137. doi:10.1186/s12916-019-1377-7
- Helander EE, Wansink B, Chieh A. Weight Gain over the Holidays in Three Countries. N Engl J Med. 2016;375(12):1200-1202. doi:10.1056/NEJMc1602012
- 62. Nebeker C, Leavy T, Roitman E, Steinhubl S. Ethical Considerations When Creating Evidence from Real World Digital Health Data. medRxiv. 2020.
- 63. Lewis D, Leibrand S. Real-World Use of Open Source Artificial Pancreas Systems. J Diabetes Sci Technol. 2016;10(6):1411. doi:10.1177/1932296816665635
- 64. Lupton D. Self-tracking, health and medicine. Heal Sociol Rev. 2017;26(1):1-5. doi :10.1080/14461242.2016.1228149
- Nebeker C. mHealth Research Applied to Regulated and Unregulated Behavioral Health Sciences. J Law Med Ethics. 2020;48(1):49-59. doi:10.1177/1073110520917029
- Just One Giant Lab Co-Founder Leo Blondel on the Power of Community and Open Source During COVID-19 - Creative Commons. https://creativecommons. org/2020/04/28/just-one-giant-lab-blondel-on-the-power-of-community-opensource-during-covid-19/. Accessed November 28, 2020.
- Grant AD, Wolf GI, Nebeker C. Approaches to governance of participant-led research: A qualitative case study. BMJ Open. 2019;9(4). doi:10.1136/bmjopen-2018-025633
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science (80-). 2019;366(6464):447-453. doi:10.1126/science.aax2342
- Discrimination By Artificial Intelligence In A Commercial Electronic Health Record—A Case Study I Health Affairs. https://www.healthaffairs.org/do/10.1377/

hbiog20200128.626576/full/. Accessed February 15, 2021.

- Jonathan Rabinovitz. Your Computer May Know You Have Parkinson's. Shall It Tell You? Stanford Magazine. https://medium.com/stanford-magazine/your-computermay-know-you-have-parkinsons-shall-it-tell-you-e8f8907f4595. Published 2018. Accessed February 19, 2019.
- 71. Zulueta J, Piscitello A, Rasic M, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study. J Med Internet Res. 2018;20(7). doi:10.2196/jmir.9775
- 72. The "All of Us" Research Program. N Engl J Med. 2019;381(7):668-676. doi:10.1056/nejmsr1809937
- 10,000 New Yorkers. 2 Decades. A Data Trove About 'Everything.' The New York Times. https://www.nytimes.com/2017/06/04/nyregion/human-project-newyork-city-study.html. Accessed November 28, 2020.
- 74. Sabatello M, Appelbaum PS. The Precision Medicine Nation. Hastings Cent Rep. 2017;47(4):19-29. doi:10.1002/hast.736
- Bromley E, Mendoza-Graf A, Berry S, Nebeker C, Khodyakov D. From "Informed" to "Engaged" Consent: Risks and Obligations in Consent for Participation in a Health Data Repository. J Law, Med Ethics. 2020;48(1):172-182. doi:10.1177/1073110520917007
- Principles for the Return of Individual Research Results: Ethical and Societal Considerations - Returning Individual Research Results to Participants - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK525079/. Accessed November 28, 2020.
- Nebeker C, Leow AD, Moore RC. From Return of Information to Return of Value: Ethical Considerations when Sharing Individual-Level Research Data. J Alzheimer's Dis. September 2019:1-8. doi:10.3233/jad-190589
- Khodyakov D, Mendoza-Graf A, Berry S, Nebeker C, Bromley E. Return of Value in the New Era of Biomedical Research—One Size Will Not Fit All. AJOB Empir Bioeth. 2019;10(4):265-275. doi:10.1080/23294515.2019.1666175
- COVID-19 supercharged digital health funding in 2020 to reach record levels: report I FierceHealthcare. https://www.fiercehealthcare.com/tech/covid-19-supercharged-digital-health-funding-2020-record-21-6b-invested-report. Accessed February 15, 2021.
- Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth. 2017;8(4):266-276. doi:10.1080/23294515.2017.1403980
- Bloss C, Nebeker C, Bietz M, et al. Reimagining Human Research Protections for 21st Century Science. J Med Internet Res. 2016;18(12):e329. doi:10.2196/ jmir.6634
- Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth. 2017;8(4). doi:10.1080/23294515.2017.1403980
- Leow A, Stange J, Zulueta J, et al. 247. BiAffect: Passive Monitoring of Psychomotor Activity in Mood Disorders Using Mobile Keystroke Kinematics. Biol Psychiatry. 2019;85(10):S102-S103. doi:10.1016/j.biopsych.2019.03.261
- Resnik DB, Elliott KC, Miller AK. A framework for addressing ethical issues in citizen science. Environ Sci Policy. 2015;54:475-481. doi:10.1016/j.envsci.2015.05.008
- 85. Rasmussen LM, Cooper C. Citizen Science Ethics. Citiz Sci Theory Pract. 2019;4(1):1-3. doi:10.5334/cstp.235
- Grant, A. D., Wolf, G. I., & Amp; Nebeker, C. (2019, April 1). Approaches to governance of participant-led research: A qualitative case study. BMJ Open. Retrieved October 6, 2021. https://bmjopen.bmj.com/content/9/4/e025633.
- The human project. The Human Project. (2017, November 20). Retrieved October 6, 2021. https://www.thehumanproject.org/.
- U.S. Department of Health and Human Services. (2015, April 16). Statement on the National Children's Study. National Institutes of Health. Retrieved October 6, 2021. https://www.nih.gov/about-nih/who-we-are/nih-director/statements/ statement-national-childrens-study.
- Pita, Robespierre & Sena, Samila & Cabral, Liliana & Teles, Carlos & Almeida, Bethania & Barreto, Marcos & Barreto, Mauricio & Ichihara, Yury. (2020). Enabling the use of huge administrative data to health research: The CIDACS' 100 Million Brazilian Cohort platform study case. https://www.researchgate. net/publication/340716592_Enabling_the_use_of_huge_administrative_data_ to_health_research_The_CIDACS'_100_Million_Brazilian_Cohort_platform_ study_case

ABOUT THE AUTHOR

Camille Nebeker, MS, EdD, is an Associate Professor of Medicine at the University of California, San Diego.

SUGGESTED CITATION

Nebeker, C. "An Emerging Research Landscape: New Actors, Data Sources and Responsible Practices." AcademyHealth, August 2021.