CODE

SUMMARY REPORT

The HHS/CODE Roundtable Series on Health Data Sharing

July 2020

Summary Report: The HHS/CODE Roundtable Series on Health Data Sharing

The Office of the Chief Technology Officer (CTO) at the U.S. Department of Health and Human Services (HHS) and the independent nonprofit Center for Open Data Enterprise (CODE) co-hosted a year-long program to find ways to improve how health data is shared and utilized for the public good. Through three Roundtables involving nearly 200 expert participants, the HHS Office of the CTO and CODE explored data-sharing challenges and opportunities for the emerging areas of Artificial Intelligence (AI) and the Social Determinants of Health (SDOH), and new ways of addressing data privacy concerns across all areas of health research. This summary report reviews the key findings of that year's work and includes recommendations to realize the benefits of health data sharing for a wide range of stakeholders - including patients, doctors, payers, providers, and researchers.

CODE

Table of Contents

Introduction: The Goals of the Project 1

Executive Summary 1

Who Has a Stake in Health Data Sharing? 2

High-Value Types of Health Data 8

Sharing and Utilizing Health Data for AI Applications 9

Applications in Healthcare 9

Challenges in Using AI Data 10

Actionable Next Steps to Improve Health AI 11

Privacy and Health Data Today 13

Risks Associated with Using Different Data Types 13

Supporting and Enforcing Existing Regulations 14

Regulating Privacy in Emerging Data Types 15

Leveraging Data on the Social Determinants of Health 16

What Are the Social Determinants of Health? 16

SDOH Data and Healthcare 17

Overview of General SDOH Data Issues 19

The SDOH Care Continuum: Challenges and Opportunities 19

Recommendations & Solutions 21

Conclusion 22

INTRODUCTION: THE GOALS OF THE PROJECT

Executive Summary

The independent nonprofit Center for Open Data Enterprise (CODE) and the Office of the Chief Technology Officer (CTO) at the U.S. Department of Health and Human Services (HHS) co-hosted a series of three Roundtables to find ways to improve how health data is shared and utilized for the public good. The Roundtables were funded by an Engagement Award to CODE from the Patient-Centered Outcomes Research Institute (PCORI).

Sharing and Utilizing Health Data for AI Applications

The first, a Roundtable on ***Sharing and Utilizing Health Data for AI Applications***, brought together over 70 expert stakeholders from government, industry, clinical research institutions, nonprofit organizations, and academia to discuss opportunities to share and utilize health data for artificial intelligence (AI) in healthcare. The feedback from Roundtable participants helped identify number of core principles and actionable opportunities for organizations that are pursuing an AI strategy. These include: ensuring privacy and consent, investing in technical infrastructure to unlock AI's potential, focusing on data quality and sharing, and leveraging AI to improve efficiency and reduce administrative burden.

Balancing Privacy with Health Data Access

The second Roundtable in the series, ***Balancing Privacy with Health Data Access***, built on these findings and brought together patient advocacy groups along with stakeholders from industry, academia, law, government, and civil society to discuss issues of data privacy. The purpose of this Roundtable was to empower data providers and users to maximize the utility of sensitive health data while providing necessary privacy measures and addressing risk. Much of the discussion focused on the Health Insurance Portability and Accountability Act (HIPAA), which provides the major framework for the protection of health data privacy in the U.S. Based on Roundtable discussions, participants proposed a number of actionable recommendations including to improve individual access to health data, create industry-wide ethical guidelines for consumer generated health data, increase access to social determinants of health, and use technology to improve patient consent for data sharing.

Leveraging Data on the Social Determinants of Health (SDOH)

The third Roundtable, ***Leveraging Data on the Social Determinants of Health (SDOH)***, highlighted current SDOH data use cases, reviewed high-priority data elements and scenarios, and identified actionable opportunities for HHS to support individual-level and population-level approaches to addressing the social determinants of health. The Roundtable brought together stakeholders from industry, academia, federal, state and local government, healthcare, and civil society to help HHS catalyze the use of SDOH data to improve health outcomes. CODE recommended that HHS develop an SDOH data strategy by creating an action plan with these three critical goals: define and standardize SDOH data, create sustainable SDOH data infrastructure, and support local and state-based decision-makers.

Who Has a Stake in Health Data Sharing?

What roles can different groups play? To prepare for each Roundtable, CODE analyzed the different groups with a stake in health data sharing, the ways they can influence the use of health data for good, and the way that health data sharing may impact their work.

Government Stakeholders

Federal, state, and local governmental public health agencies

How does this group influence...	How is this group impacted by...
artificial intelligence? Government officials can define regulation, implement strategy, and incorporate these tools into their workflow. In 2019, HHS awarded a **$49 million** contract to 57 vendors to provide artificial intelligence (AI) and other advanced technology services to the department and other federal agencies.	**artificial intelligence?** Government officials can leverage the benefits of AI to automate administrative tasks. They may be impacted by the ability of AI to reduce costs of diagnosis and treatment. **CMS** has developed an AI Health Outcomes Challenge to accelerate the development of AI solutions that aid clinicians in predicting health outcomes and keeping patients healthy.
data privacy? Government officials have the capability to create their own provisions for data privacy and security. The CDC developed specific **data security and confidentiality guidelines** for HIV, viral hepatitis, sexually transmitted disease, and tuberculosis.	**data privacy?** HHS sets guidelines for data privacy through the HIPAA Privacy Rule and standards around de-identification of sensitive data. In regards to state governments, the **legislative sessions** produced a number of new and updated data protection statutes and regulations.
collection and management of SDOH Data? HHS and its Operating Divisions can act as a convener of other groups that produce datasets that work with the social determinants of health.	**collection and management of SDOH Data?** HHS agencies can leverage the SDOH to better identify eligible members and serve at-risk patient groups. Multiple agencies including the **Department of Housing and Urban Development**, and offices in HHS, including the CDC are **Agency for Healthcare Research and Quality** are developing their own SDOH data collection and use efforts.

Patients & Patient Advocates

Patients organizations designed to advocate on behalf of patients.

How does this group influence...	How is this group impacted by...
artificial intelligence? Currently, patients are limited in their capacity to provide input on the use of AI in healthcare.	**artificial intelligence?** Patients stand to benefit from the adoption of AI, but also will be the first to pay consequences of misuses of AI.
data privacy? Patient advocates can lobby for increasing ownership over their EHRs and accessing medical notes. They are often the conduits between patients and healthcare providers and exert a greater influence on patient sharing of data. Organizations such as the **Patient Advocacy Foundation** (PAF) are committed to helping patients protect their personal health information and abide by stringent patient privacy policies.	**data privacy?** Most sensitive health data comes directly from patients through EHRs, clinical trial data, surveys, and other sources. Data privacy protects patients from financial discrimination or psychosocial effects of having their data released. Patients should be aware of their data privacy rights, and have access to their own personal health information.
collection and management of SDOH Data? Patient advocates are able to inform decision makers about the value of the SDOH but may not be able to increase the availability of the SDOH.	**collection and management of SDOH Data?** Patients reap the benefits of SDOH by receiving more well-rounded, holistic care from providers. Patients can also access support services outside of traditional providers through referrals and specialized local programs.

Private Sector

Businesses that are working with health data in the healthcare sector.

How does this group influence...	How is this group impacted by...
artificial intelligence? Private tech companies are at the forefront of AI technology, and are responsible for its utility and accuracy. A company known as Sage Bionetworks launched **mPower** as a study using surveys and phone sensors to track symptoms of Parkinson's Disease. The results can help patients, doctors, and caregivers better understand changes over time and the impact of exercise or medication.	**artificial intelligence?** Many private companies are investing heavily in the development of AI applications for health, as demand for these tools is rising rapidly. For example, Researchers at **Google** have developed algorithms to analyze images of retinas and diagnose this disease with over 90 percent accuracy.
data privacy? Companies must comply with HIPAA guidelines, but can shape privacy rules by submitting comments and proposing white papers that suggest amendments to HIPAA. Companies such as the **Consumer Technology Association** have released voluntary health data privacy guidelines for companies that handle health and wellness data.	**data privacy?** Companies must comply with HIPAA standards if their data is in the form of electronic patient health data or claims data, but those that collect real world data through third party apps and wearables fall outside of HIPAA's privacy standards.
collection and management of SDOH Data? Private businesses are aggregating their own databases on SDOH that they sell to prospective healthcare providers and payers. The **SDoH Dashboard Action Tool**, developed by Impaq International includes a comprehensive database containing linked health care measure data with free and publicly available SDoH datasets and is maintained using Tableau.	**collection and management of SDOH Data?** Businesses can deploy SDOH data to better design new products that address health and healthcare disparities. Companies such as **Unite Us** provide comprehensive technology solutions that seek to offer decision support, electronic referral management, care plan management for community based CBOs receiving referrals, and outcomes measurement.

Academic & Clinical Researchers

Researchers who work at either academic institutions or clinical labs.

How does this group influence...	How is this group impacted by...
artificial intelligence? Researchers may be able to greatly influence the use and development of AI with their research, although they do not generally direct the real-world application of AI in medical settings.	**artificial intelligence?** Researchers will be able to use AI to assist in research, and AI may help clinical researchers match with patients for clinical trials, improving their ability to conduct research.
data privacy? The ways that researchers record and publish patient data can influence how privacy is embraced by the wider health community, although they can't always influence the policies that govern privacy and access.	**data privacy?** Clinical researchers must adhere to HIPAA's privacy rules for drug trials, including special use agreements for accessing research data.
collection and management of SDOH Data? Clinical researchers conduct experiments that may draw on data on SDOH, but they are not generally providers of SDOH data.	**collection and management of SDOH Data?** SDOH data helps researchers better understand the environmental causes of disease and identify more holistic possibilities for treatment.

Healthcare Providers

Doctors, nurses, hospitals, and private clinics.

How does this group influence...	How is this group impacted by...
artificial intelligence? Apart from the practitioners that also conduct research, most healthcare providers will be able to decide how to apply AI, but likely not be able to determine its development.	**artificial intelligence?** Healthcare providers will be affected by AI, particularly as AI becomes regularly used in diagnosing illnesses and developing treatment plans. By incorporating AI tools like natural language processing and automation, providers can reduce overall costs and decrease administrative burden.
data privacy? Healthcare organizations advise and provide feedback to government agencies around privacy law. They are often first movers in embracing new technologies in the office.	**data privacy?** Doctors, nurses, hospitals, and private clinics collect large amounts of patient-generated data and enter data into a patient's electronic health record (EHR). These entities need to ensure they follow privacy guidelines on new and emerging patient-data gathering techniques.
collection and management of SDOH Data? Hospitals collect information about patients but do not have extensive information about the communities their patients live in. They can obtain SDOH data by administering surveys to their patients. For example, as part of its **Moving Healthcare Upstream project**, the New York Presbyterian Hospital (NYP) has focused on identifying the psychological and social determinants of health for children ages 0-5 living in Harlem, New York.	**collection and management of SDOH Data?** Similar to doctors, hospitals and clinics recognize that their patients face negative health outcomes that are influenced by factors in their local communities and have an opportunity to offer both clinical and social interventions. Currently, the **Pennsylvania Department of Health** draws on county-level and district-level data that comes from sources such as the Enterprise Data Dissemination Informatics Exchange (EDDIE).

Civil Society

Nongovernmental organizations that work to advance patients' interests.

How does this group influence...	How is this group impacted by...
artificial intelligence? Large NGOs will likely have a large influence on the development of AI, but smaller NGOs will likely have little influence.	**artificial intelligence?** AI may not directly affect the operations of civil society, though it may affect their funding priorities and organizational focus areas, especially around promising health applications.
data privacy? Research and advocacy groups, two key civil society groups, will write white papers about the importance of privacy and generally outline best practices for data privacy. Research Foundations can share de-identified datasets and engage in data-sharing agreements with government agencies.	**data privacy?** There are two key types of civil society organizations in this space: medical foundations and general advocacy organizations. The former must guarantee patient privacy according to rules prescribed by the government while advocacy and research groups often have low impact since they study privacy without having to gather sensitive health data.
collection and management of SDOH Data? Foundations and civil society do not necessarily collect and manage much SDOH data but do assist in its promotion. They publish white papers that outline the value of the SDOH and strategies for their improved access.	**collection and management of SDOH Data?** Civil society groups conduct research and design interventions based on the various factors outlined in the SDOH. Their work improves with increased use of the SDOH.

High-Value Types of Health Data

Throughout the Health Data Sharing Roundtable series, participants identified a number of high-value health data types that can help drive progress in the health care sector, for example by contributing to AI development. Building on the expert feedback gathered at the Roundtables and subsequent research, this section describes six major health data types and the challenges associated with their use.

Administrative and Claims Data generally comes from federal, state, and local government agencies as well as healthcare providers and insurers. This can range from hospital discharge summaries to payment records between insured patients and the **healthcare system**.

Clinical Data is a **broad term** that encompasses different kinds of data generated "in a clinical setting and controlled by a clinician, as opposed to a patient or caregiver."

- **Clinical Trials Data** includes registries and results from publicly and privately funded clinical studies.
- **Electronic Health Record (EHR) Data** is focused on individual patients, and can include information on routine checkups, prescriptions, and medical procedures.

Genomic Data can come in many different forms, ranging from full DNA sequences to individual DNA variants. Recent advances have made it possible to analyze and store data on a person's entire genome sequence. Genomic data is considered highly sensitive and must be shared and used under carefully controlled conditions.

Patient-Generated Data includes "health-related data created and recorded by or from patients outside of the clinical setting to help address a **health concern**." This type of data type is becoming increasingly prevalent through mobile health applications and wearable devices.

- **IoT (Internet of Things) Data** includes data from mobile software applications, voice assistants, and wearable devices such as smart watches.
- **Social Media Data** includes interactions on social media platforms such as Facebook and Twitter.

Social Determinants of Health Data represent "conditions in the environments in which people are born, live, learn, [and] work...that affect a wide range of health, functioning, and quality-of-life **outcomes and risks**."

Surveillance Data is a broad term that encompasses the "ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of **public health practice**."

- **Registry Data** includes data shared voluntarily by individuals that is generally focused around a specific diagnosis or condition such as a particular type of cancer or cystic fibrosis.
- **Survey Data** includes the results of surveys and studies conducted to assess population health.
- **Vitals Data** is generally collected and exchanged between local jurisdictions and the federal government. This **data represents** "vital events," such as births, deaths, marriages, divorces, and fetal deaths.

Applications in Healthcare

Reducing costs and administrative burdens. Utilization of AI tools has the ability not only to reduce costs, but to speed up healthcare administrative operations, like the transcription of medical notes, by eliminating manual data entry, correcting human error, and automatically processing electronic health records (EHR). Natural language processing (NLP), for instance, can be used to extract and interpret handwritten medical notes and text records, and other unstructured data sources. NLP can also support systems like answering free-text queries that require synthesizing numerous data sources. Automation processes, like robot process automation (RPA), and machine learning also have great potential to reduce costs and administrative burdens and are already being utilized in healthcare settings.

Improving population health management. Population health management (PHM) involves using population-level data to identify health risks and treatment opportunities for a group of individuals or community. AI can contribute to PHM by combining, synthesizing, and analyzing datasets from third parties with clinical and/ or patient-generated data. Researchers and health providers can use AI to aggregate longitudinal patient-generated data into larger datasets that tell better stories about the incidence and prevalence of disease.

Connecting patients to resources and care. AI has tremendous potential value in connecting patients with available resources and care. This is especially true for patients living in rural areas with decreased access to quality healthcare. One example is the ability of AI to provide patients with personalized healthcare recommendations. **Sage Bionetworks** has launched a series of mobile research studies, known as **mPower**, to increase understanding of the progression of Parkinson's Disease in individuals by using surveys and phone sensors to track symptoms of the disease.

Development of virtual care plans. With the use of AI, providers have the ability to develop virtual care programs for patients with chronic conditions. Virtual care plans can

especially benefit rural and other resource-constrained areas. There is **increasing evidence** that AI-driven chatbots can address routine patient questions and help doctors communicate with patients about their diagnosis and risk evaluations.

Improving diagnosis and early detection. **Diagnostic errors** are a major problem in the healthcare system, with most patients experiencing at least one diagnostic error in their lifetime. Physicians are leveraging AI to accurately diagnose medical conditions in their patients and treat illnesses at an early stage. AI algorithms draw upon large datasets on medical and social determinants of health to better identify patterns and assist physicians in making diagnoses and developing **treatment plans**.

Drug and therapy development. Drug development is a costly and time-consuming process. AI can help improve drug development through the entire development lifecycle, from identifying gaps in current therapeutics to bringing new products to market.

Challenges in Using AI Data

Roundtable participants identified numerous legal, cultural, and technical challenges associated with sharing and utilizing health data for AI applications. While some of these challenges are specific to AI development, others are general issues that impact all applications of health data.

Legal Challenges

- **Inconsistent restrictions on data use.** Among the legal challenges, roundtable participants noted that health data types have different legal and regulatory constraints on their use. For example, administrative and claims data, clinical data, and certain types of public health data, such as survey data, can include sensitive, individual-level **information**. The use of these data types is often restricted under existing privacy frameworks such as HIPAA.
- **Concerns about intellectual property.** Roundtable participants also discussed the challenges and potential benefits of using and sharing proprietary data and algorithms. Data collected in drug development trials, through private-sector health surveys, or in other ways could benefit researchers and organizations in the health sector developing AI applications. In a similar way, proprietary AI models could be developed for greater accuracy if the algorithms they use were shared.

Cultural Challenges

- **Underlying bias in health data.** There are widespread concerns about bias and lack of diversity in health data, which can have serious consequences when utilized for AI development.
- **Algorithmic Bias.** Negative uses of data can be combated by setting standards for data security and supporting efforts to mitigate algorithm bias and digital content forgeries, such as “deep fakes.”
- **Data silos and administrative hurdles.** While HHS is developing more efficient ways for its operating agencies to share data, it is still difficult for HHS agencies to share data with each other, and it can be

difficult for organizations outside of government to obtain data from HHS. Culture change is needed to reduce the administrative hurdles that prevent timely data sharing.

- **Overly restrictive interpretations of HIPAA.** While HIPAA is intended to protect patient privacy, it does allow data sharing and use under specific conditions. Roundtable participants suggested that HHS could provide more guidance on what is and is not permissible under HIPAA in different contexts.

Technical Challenges

- **Limited technical capacity for data management and analysis.** Roundtable participants inside and outside of government noted the need for more staff with data science training. In particular, both government and the private sector need more experts in AI and its application to health data and issues.
- **Inadequate IT infrastructure for hosting and analyzing large datasets.** AI applications require large quantities of data, and large computational capacity, to train and test algorithms. Both HHS and the stakeholders that work with government may need to upgrade their infrastructure to meet these challenges.
- **Poor data interoperability.** Across the healthcare system, large amounts of data are structured in different ways, preventing stakeholders from easily exchanging and integrating this information. Roundtable participants attributed these challenges to a lack of common data standards and issues with enforcement where standards do exist.

Actionable Next Steps to Improve Health AI

The Health AI Roundtable produced a number of actionable next steps that government and the private sector can take to improve health AI. Most of these actions could be undertaken by individual government agencies or companies, or through public-private partnership.

1 Invest in IT Infrastructure and Expertise to Support AI

AI demands a robust information technology (IT) infrastructure, including data infrastructure, and staff with the skills to apply it. Both infrastructure and expertise must be able to manage the large amounts of data needed to support AI as well as the development of advanced AI applications. Government agencies and private-sector companies alike will need to make comprehensive technology investments and build their AI expertise to support their AI strategies.

2 Ensure Access to Data for AI While Protecting Privacy

Concerns about privacy are paramount in the application of individual health data. While the use of health data in EHRs and other medical records is governed by federal and state legislation, other data types like IoT data are only regulated through “terms of service”

agreements developed by the private sector. Both government and private-sector health data providers will need to ensure that sensitive information is not disclosed or misused when these data sources are applied. The section below on Privacy and Health Data describes a range of strategies to use.

3 Use Standards to Improve Data Quality and Interoperability

Data for AI applications should be clean, timely, accurate, and standardized. There are numerous challenges in integrating data and metadata from multiple sources. The current lack of interoperability is a major obstacle to applying data for AI development. Common standards for data collection and management, whether developed through government leadership, by industry groups, or through a public-private effort, can ensure that data and metadata are accurate and consistent across organizations and across healthcare applications, using a shared library of variables that are applied across datasets.

4 Develop Easy-to-Use Agreements for Data Sharing

AI applications are most effective when they can integrate large amounts of data about diverse facets of health. However, researchers inside and outside of HHS often have difficulty accessing the data they need. To share data from other sources, researchers must have data use agreements (DUA) that abide by HIPAA regulations, whether HHS is sharing data with outside researchers or whether different operating divisions within HHS are sharing data with each other. Model DUAs can go a long way to solving this challenge.

5 Clarify Appropriate Use of Patient-Generated Data

Patient-generated data includes data collected through sensors and wearables, and through social media and mobile applications. Large amounts of this data are collected under "terms of service" agreements and are being used by entities that are not covered by **HIPAA**. As interest in patient-generated data increases, there is a need for clearer rules around its appropriate use, particularly in the context of AI development. This clarification could come through legislation or other government action, through industry self-regulation, or through public-private collaboration.

Risks Associated with Using Different Data Types

Many of the benefits of accessing datasets that include sensitive personal health information (PHI)are well documented, including improved preventive care, clinical outcomes, and coordination of care. Similarly, many of the major health data types pose similar challenges and issues that can negatively impact patients and communities.

Incomplete Data. Healthcare providers frequently make clinical decisions based on data that may not be complete or that lacks additional context from complementary data. While many physicians make decisions based purely on available patient electronic health record (EHR) data or claims data, they may miss potential insights and context from genomic data or data on the social determinants of health. Moreover, technical restrictions can lead to incomplete or inaccurate data.

Possibility of Re-identification Through Other Datasets. Technical approaches have been a key strategy to preserve the anonymity of individuals and safeguard sensitive PHI. Data scientists and researchers have made advances in de-identifying data, which involves removing key identifiers from data, or anonymizing data, which entails changing the way variables are coded in a dataset. However, these technical fixes may not be enough to completely safeguard data privacy. Recent research has demonstrated how companies and researchers can take anonymized datasets and re-identify individuals with a high degree of accuracy when the anonymized data is combined with other third party data, a problem called the "mosaic effect."

Inappropriate Data Sharing and Use. Personal health data may be misused by third party providers that violate informed consent or by data brokers that illegally obtain PHI without a patient's permission or knowledge. While data flows between entities covered by HIPAA are common, patients are not often informed about when their data moves from one entity to another. HIPAA defines six approved categories of use for sharing an

individual's PHI. However, many patients are unaware of the limits of approved data use, and may not immediately realize when their right to privacy is being breached.

Confusion About Data Safeguards and Protection. Many participants in the Privacy Roundtable emphasized that both patients and providers are frequently confused or misunderstand how PHI is protected under HIPAA. Many patients are unaware that data-gathering mobile applications, medical devices, and automatic transcription voice assistants are sold and managed by entities that are not covered under HIPAA. Moreover, many patients are unaware of the specific situations in which a healthcare entity covered by HIPAA's rules is allowed to access PHI. Lastly, patients and providers both may be misled by the privacy policies of third-party companies that do not have to abide by data safeguards.

Supporting and Enforcing Existing Regulations

While HIPAA and other existing regulations have some limitations, they provide a framework that is familiar to healthcare providers and that could be applied more strongly. Roundtable participants proposed several measures for supporting and enforcing the provisions that **HIPAA** already includes.

Improve individual access to health data. The Right of Access is an important tool to empower patients and give them access to different types of PHI data safeguarded under HIPAA. The **Office of Civil Rights** in HHS is responsible for enforcing this rule and ensuring that providers and payers comply when patients request this data. Despite this, many patients are denied access to their data or unaware that they are allowed to request this data from providers. Patients should be aware of their health data rights in regards to access, sharing, and utilization.

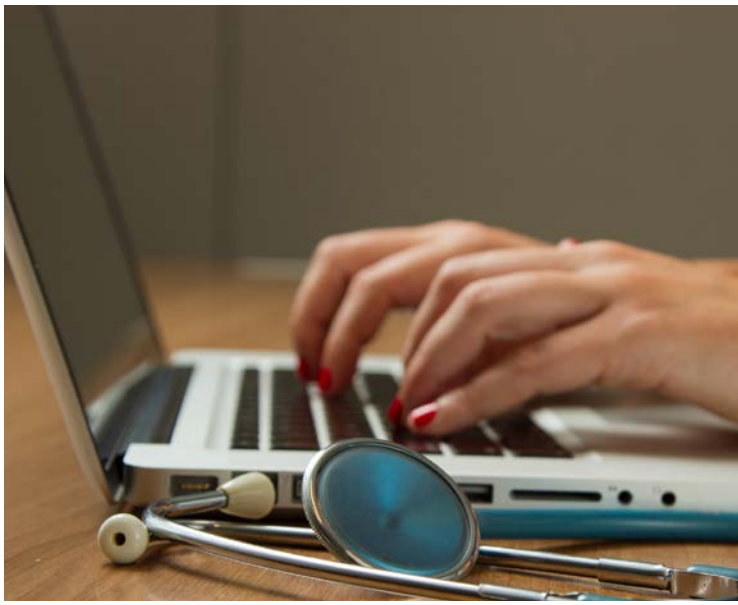
Hold health related "business associates" accountable. The passage of the **HITECH Act** required that business associates comply with the security and privacy rules of HIPAA just like covered clearinghouses, providers, and payers. Despite this advancement, these business associates are not directly regulated: the covered entities they work with are responsible for ensuring that their business associates follow the rules. Companies should have to ensure that they're business clientele HIPAA and privacy regulations. HHS should improve its direct monitoring, and enforcement of business associates, to ensure that these business associates adhere to the standards of data de-identification, limited data collection, and the range of accepted and incidental data uses.

Help start-ups comply with HIPAA's requirements. HIPAA sets a high standard for covered entities that gather data with PHI, and costs for HIPAA compliance can be prohibitive. Obtaining the HITRUST certification to confirm compliance can cost tens of thousands of dollars. As one solution to help companies avoid these high startup costs, the CMS has established a **Virtual Research Data Center** (VRDC) to provide a secure portal to efficiently use de-identified CMS data that is approved for wider use. This "containerized" approach creates a HIPAA-compliant virtual sandbox where small companies can submit and run their tech applications on the portal without ever having to download the data in ways that would require them to be HIPAA-compliant. Small businesses should use proper techniques to access sensitive health data, and HHS could also create some sort of research data center.

Regulating Privacy in Emerging Data Types

Despite the various improvements that can be made to HIPAA, patients are now accessing data that falls outside its purview. Technology companies that manufacture and sell fitness wearables, mobile applications, and home assistants are not considered "covered entities" under HIPAA. Moreover, despite its increasing use, data on the Social Determinants of Health is also not regulated by HIPAA. These recommendations outline strategies for HHS and other partners to extend the kinds of protections provided by HIPAA to other data types.

What should consumers do? The disclosure of consumers' PHI not only affects the individual, but can impact those related to them as well. HIPAA and other Federal rules and regulations fail to address the risks posed to the family members, who's genetic makeup is also analyzed. Genomic data can be used to trace blood relatives, as is the case with company, 23andMe, regardless of whether the relative has consented or not. Consumer should be aware of the risk E.g. 23 and Me.



What should IOT manufacturers do? IOT manufacturers, such as Mozilla, should be as transparent as possible in their development processes and have special regulations developed for them to abide by. Consumers can be the ones to evaluate privacy.

In addition, there are more steps that could be taken to improve the way this data is managed.

Policymakers should adopt legislation to broaden data privacy rights. There is currently no federal oversight of consumer-generated health data. Many participants stated that this problem should not be left to industry self-regulation, but that the House and Senate should pass comprehensive legislation to properly regulate the appropriate use of patient-generated data. One possible route is the 2019 Protecting Personal Health Data Act introduced by Senators Murkowski and Klobuchar, which would create a comprehensive set of policies to regulate the use and sharing of consumer- generated health data.

The private sector should create Industry-wide Ethical Guidelines for Consumer-Generated Health Data. Even with legislation in place, the private sector will need to coordinate its efforts to adopt best practices for preventing individual discrimination or group harm from misuse of health data. Companies should collaborate to produce a set of ethical guidelines that govern the use of patient-generated data. This framework could build on existing models such as the Future of Privacy Forum, Consumer Technology Association, or the CARIN Alliance. For example, the MITRE Framework for the Use of Consumer-Generated Data in Healthcare outlines a set of Principles, Values, and Guidelines for companies using consumer generated data. Moreover, companies should inform their consumers about these guidelines and publicly commit to following them.

What Are the Social Determinants of Health?

The social determinants of health (SDOH) have been defined as the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems that shape the conditions of daily life. These conditions impact everyone, and may have either positive or negative effects. Data on the SDOH, described in the table below, are increasingly important to healthcare providers, researchers, government agencies, and the people they serve.

Figure 1: Social Determinants Categories and Sample Public Data Sources

SDOH Category	Sample Determinants	Sample Public Data Source
Economic Stability A person's ability to pay bills on time, create longer term budgets, and stay out of debt are functions of their economic stability.	■ Access to employment ■ Hourly wages ■ Level of personal debt	■ Department of Labor Unemployment Statistics ■ Internal Revenue Service Income Tax Rates
Neighborhood and Physical Environment Neighborhood may impact a person's health outcomes due to subpar housing, the threat of crime, environmental hazards, and poor walkability.	■ Housing ■ Access to transportation ■ Public parks ■ Safety ■ General geography	■ Zillow's walkability and home ownership rates ■ DOT's data on proximity to public transportation ■ National Parks Service data on parks and recreational areas
Education Limited educational opportunities and inadequate schooling can limit the comprehension of important clinical recommendations for individuals, and make it difficult to understand diagnoses.	■ Literacy rates ■ Vocational training opportunities ■ Early childhood education ■ Rates of higher education	■ Access to employment ■ Hourly wages ■ Level of personal debt

SDOH Category	Sample Determinants	Sample Public Data Source
Food Access to healthy foods and subsidy programs like SNAP and WIC are indicators that can provide important insights into nutrition and health.	■ Access to fresh fruits and vegetables ■ Number of grocery stores in a neighborhood ■ Number of residents who access food pantries	■ WIC and SNAP rates by neighborhood[1] ■ Proximity to a grocery store ■ Access to a food pantry or other emergency food service
Climate and Environmental Data Additional SDOH factors that are not confined to neighborhood borders including climate, weather, and air also have a major impact on an individual's health.	■ Air and water quality ■ Access to clean water ■ Exposure to power plant emissions	■ U.S. Environmental Protection Agency (EPA) air quality and toxic waste data ■ U.S. Department of Energy (DOE) data on power plants.

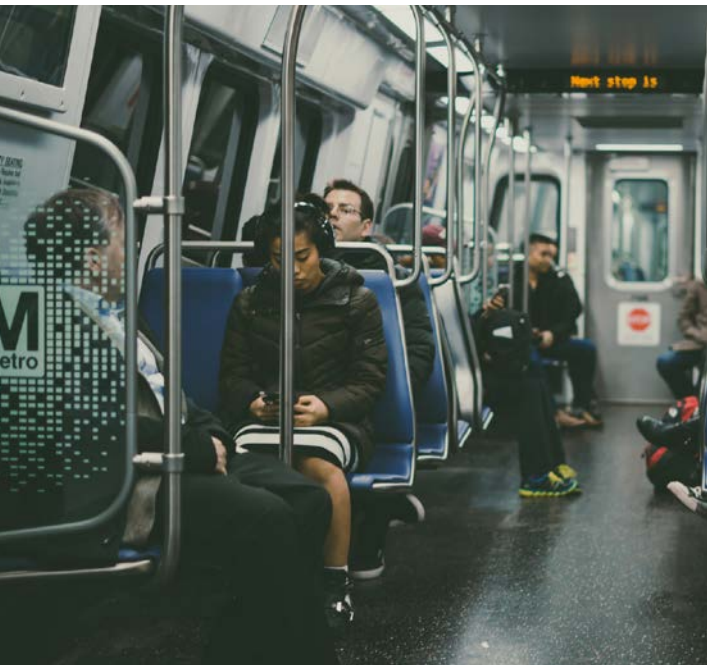
SDOH Data and Healthcare

Health-focused organizations across the private sector and civil society are now realizing the potential of SDOH data. Healthcare providers and payers are using the data to better understand and address their patients' needs. Health management companies are applying the social determinants of health to predict health risks like opioid overdose and target preventive interventions. Other companies are serving as SDOH data providers, creating platforms of SDOH data for research and healthcare applications. Civil society groups are serving as conveners for a variety of stakeholders and seeking to increase coordination to improve patient care. They are demonstrating how SDOH data can provide a holistic understanding of factors that affect an individual's risk of disease and response to treatment, including measures as diverse as the air quality of a person's neighborhood or their proximity to a grocery store.

[1] WIC stands for "The Special Supplemental Nutrition Program for Women, Infants, and Children" and SNAP stands for Supplemental Nutrition Assistance Program.

Stakeholder Group	SDOH Data Use
Federal, State, and Local Government	Government officials draw on data sources from surveys, programs, and reports to develop a better understanding of their health systems' performance and any resource gaps in the communities they serve. Government agencies can also be data providers, organizing and curating granular geographic data at the city or county levels, or providing it through the U.S. Census Bureau or other federal departments.
Civil Society Alliances and Groups	Organizations like **The National Alliance to impact the Social Determinants of Health** (NASDOH) bring together cross-sectoral efforts by engaging organizations across the health care system and the business sector to articulate the value proposition of addressing SDOH.
Patients and Advocates	An individual's health status is heavily influenced by social needs and social risk factors. Patients are the primary recipients of the benefits of utilizing SDOH data and are increasingly interested in learning about risks and resources in the communities where they reside. They may seek support in accessing healthcare, from individual-level transportation assistance to connecting with community-led remote clinics.
Technology Companies	Technology companies often use software solutions to provide granular SDOH data, analytics, and care coordination for healthcare companies that are seeking to understand their patients. These companies draw on a combination of commercial and publicly available data sources that can include clinical, financial, and social data.
Healthcare Payers	To gain better understanding of the conditions that impact patient utilization of care, healthcare payers are partnering with technology companies to identify particularly vulnerable communities and developing risk profiles of certain communities using population-level data.
Healthcare Providers	Healthcare providers often interact with patients in clinical care settings and can generate individual-level data from these clinical encounters. Healthcare providers often aim to collect data on SDOH through electronic health records (EHRs) and employ screening tools to gain a better understanding of the contextual and social conditions affecting their patients.

Overview of General SDOH Data Issues



Several generalized concerns about data cut across all dimensions of care. These include ethical and privacy considerations, lack of standards, and gaps in data and data sharing.

Ethical Data Gathering and Privacy Challenges. Developing social risk profiles could result in community redlining and individual stereotyping. Privacy concerns become a significant issue as a growing number of stakeholders are using individual-level commercial data that may be vulnerable to re-identification.

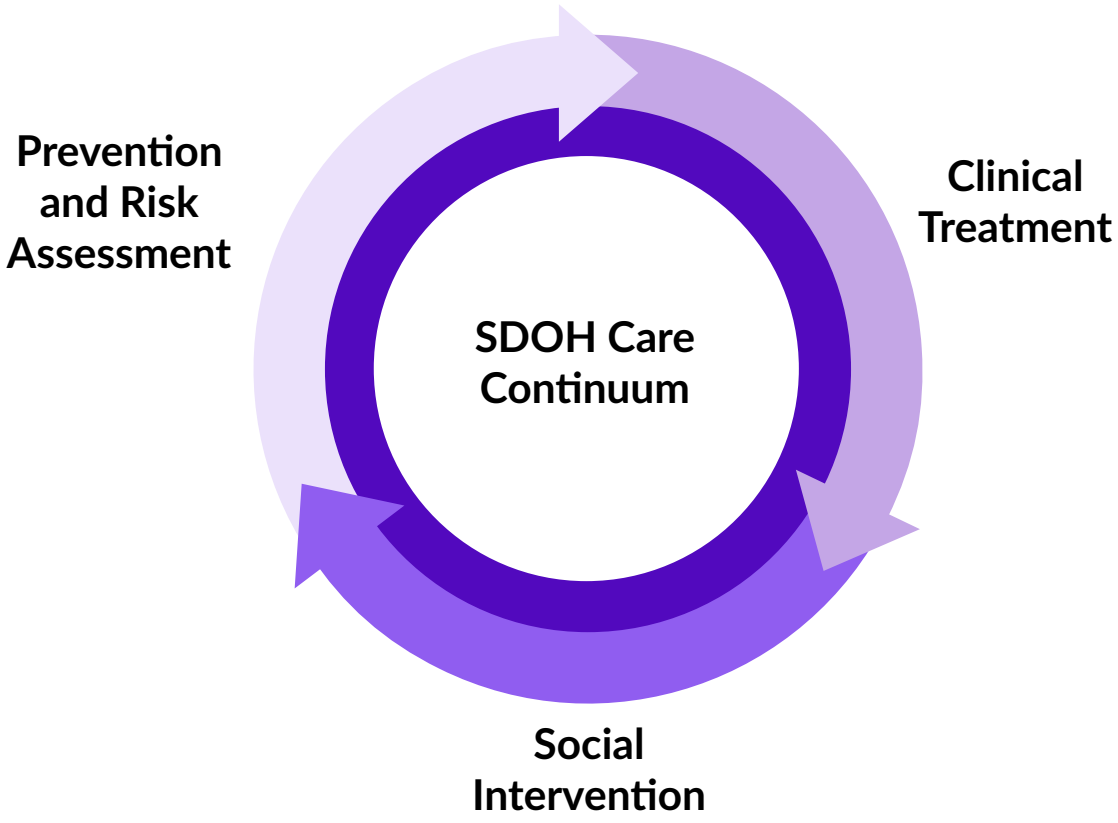
Lack of Data Standards and Interoperability. The growing interest in SDOH data has produced more and more measures of different health risk factors and indicators. Their lack of data standardization impedes interoperability and may create confusion among the wide variety of tools and resources available to providers.

Incomplete Data. There are data gaps at each phase of the data collection lifecycle which hinder new technological applications such as machine learning and the development of impactful interventions.

Lack of Financial Alignment. Throughout the entire SDOH care continuum, stakeholders have confronted hurdles in leveraging the power of the social determinants of health without the proper financial incentives.

The SDOH Care Continuum: Challenges and Opportunities

A continuum of care model focused on SDOH could help stakeholders think strategically about how to employ SDOH and other data to deliver better healthcare. CODE has proposed an SDOH Care Continuum for clinical settings that draws from existing models. It includes three stages: prevention and risk assessment, clinical treatment, and social intervention and referrals.



Prevention and Risk Assessment	Clinical Treatment	Social Intervention
Risk assessment and prevention leverages predictive analytics and risk mapping to identify patient populations that are at greater risk than other populations for developing certain diseases or chronic conditions based on their social needs.	Patients in clinical settings can be screened for social needs. These screenings determine key risk factors and recommend treatment based on conditions.	Social interventions include referrals or treatments that aim to address a key social need such as food insecurity or housing instability. These interventions usually are with social services organizations.

Recommendations & Solutions

Based on the input of many Roundtable participants and CODE's additional research, this project recommends that **HHS develop an SDOH Data Strategy** that comprehensively maps the space and helps coordinate action among different stakeholders. Such a comprehensive strategy could improve interoperability and data access, address privacy concerns, and broadly set the agenda for improved use of SDOH data. Each of the following sections includes recommendations that could be a part of that broader SDOH Data Strategy.

Define and Standardize SDOH Data. Data standards and commonly agreed upon definitions serve as the foundation for data interoperability, privacy and protection, and ethical data use, including the principle of data minimization.[2] As more and more data is gathered about financial, housing, environmental, and contextual indicators, Roundtable participants proposed a series of recommendations to ensure that SDOH data is standardized and streamlined for use. The recommendations include the improvement and alignment of open source assessment tools, creation and adoption of data standards and definitions, and the development of a data governance body. Multiple efforts are being taken in the academic sector, such as the **Gravity Project** to address these priorities. HHS should also work to create data standards for the SDOH.

Work With and Support Community Based Organizations Collecting SDOH Data. As value-based care and population health are implemented across the healthcare system, more and more stakeholders will engage in the collection, aggregation, and sharing of SDOH data. This will require increasing access to data infrastructure, such as online data storage, the integration of SDOH data into EHRs, and the ability for small and larger stakeholders alike to streamline data collection. The actionable recommendations suggested were to empower and strengthen community based organizations (CBOs), and improve financial alignment for health care payers, providers, and CBOs.

Develop Better State and Local Data Collection. A large portion of healthcare delivery takes place at the local level through health clinics, small providers, and community health centers. HHS can play a unique role in expanding the resources and representative data available to state policymakers and practitioners to enhance policy decisions and funding allocations. This is best achieved by developing granular, representative snapshots of communities and creating toolkits for state officials seeking to improve their SDOH strategies. To achieve this priority, hyper-local snapshots of communities should be developed, and state-level SDOH strategy toolkits should be created. Civil society organizations can help support and highlight important efforts around the country.

[2] The data minimization principle emphasizes that data collection and the amount of data used for any particular project is only what is necessary to accomplish the needed tasks. This reduces the possibility of unnecessarily gathering potentially sensitive information about an individual.

CONCLUSION

The Health Data Sharing Roundtable series brought together a diverse group of stakeholders to identify the challenges, opportunities, and needs that come with using health data. Participants of the ***Roundtable on Sharing and Utilizing Health Data for AI Applications*** were able to develop core principles and actionable opportunities. These principles include ensuring privacy and consent, investing in technical infrastructure to unlock AI's potential, focusing on data quality and sharing, and leveraging AI to improve efficiency and reduce administrative burden.

Participants of the second ***Roundtable on Balancing Privacy with Health Data Access*** further built upon the first roundtables findings, and developed ways to maximize the utility of sensitive health data while providing necessary privacy measures and addressing risk. The actionable opportunities that came out of this roundtable include, improve individual access to health data, create industry-wide ethical guidelines for consumer generated health data, increase access to social determinants of health, and use technology to improve patient consent for data sharing.

The last Roundtable in the series, ***Leveraging Data on the Social Determinant of Health***, focused on the current and potential uses of SDOH data in different stakeholder groups. Based on Roundtable findings, CODE suggested three critical goals for the development of an SDOH federal data strategy to HHS. These goals include defining and standardizing SDOH data, creation of sustainable SDOH infrastructure, and support of local and state-based decision-makers.

Sharing high-value health data is an ongoing challenge. We need new policies, protocols, and strategies to make this essential data available with maximum utility and minimal risk. The Center for Open Data Enterprise will continue to bring together government, academia, the private sector, civil society, and patients themselves to find new ways to reach that goal. We are launching a website, HealthDataSharing.org, to provide in-depth results from the Roundtable Series and to serve as an ongoing resource and hub for this work. And we welcome comments, ideas, and opportunities for collaboration through the website or by writing to us at **contact@odenterprise.org**.

Notes

CODE