Cyber Operations and Resilience Program Graduate Projects and Theses

College of Engineering

12-2022

A Cybersecurity Assessment of Health Data Ecosystems

Michelle N. Halsey
Boise State University

A Cybersecurity Assessment of Health Data Ecosystems

by

Michelle N. Halsey

A project completed in fulfillment

of the CORe 591 requirements for the degree of

Master of Science in Cyber Operations and Resilience

Boise State University

December 2022

A Cybersecurity Assessment of Health Data Ecosystems

Michelle N. Halsey
Cyber Operations and Resilience Program
Boise State University
Boise, Idaho

***Abstract*: This paper is an exploratory study that investigates data collected and used by health plans and reviews the laws and regulations governing this data to identify the gaps in protections and provide recommendations for eliminating these gaps. Health insurance companies collect a wide array of data about the people they insure, data that is often only peripherally relevant to the service these companies provide. The data environment currently consists of seven categories of data: personal health information, summary health information, personally identifiable information, financial information, professional information, biometric information, and lifestyle data or social indicators of health. Much of this data is protected under the Health Insurance Portability and Accountability Act (HIPAA) and under an array of other health care laws and regulations; however, there is a category of data not covered by these protections. Lifestyle data or social indicators of health is a category of data that is readily available through digital interactions with third-party platforms, wearable devices, and internet of things devices. This data can be identifiable to the individual but lacks the most basic regulatory and security protections. Weaknesses in HIPAA provide loopholes for data traditionally thought to be protected.**

Keywords: health plan data ecosystem, Health Insurance Portability and Accountability Act (HIPAA), lifestyle data, social indicators of health, shadow health records.

I. INTRODUCTION

Health insurance companies collect a wide array of data about the people they insure, data that is often only peripherally relevant to the service these companies provide. Traditionally, health plans have collected data related to payment and treatment for patients; however, this data has strong federal and state privacy protections which limit how organizations can use this data. Over the last two decades, the dramatic increase in ubiquitous technology has driven a rapid increase in the volume of data that is available that falls outside of these protected areas. Technical advances have enabled the aggregation, compilation, and curation of vast amounts of disparate data sets that were previously not considered together, but when pulled together can be analyzed to identify unexpected associations [1]. Lifestyle data and data generated by Internet of Things devices provide an unprecedented opportunity to understand how patients live and use that information to provide data-driven health care.

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 provided incentives necessary for the healthcare industry to convert to electronic health records, resulting in 96 percent of hospitals and 80 percent of physicians using an HHS certified EHR system [1]. This has accelerated the rise of big data into the healthcare industry, illustrated by a 2018 conference for health insurance companies where companies like LexisNexis, IBM Watson Health, and Optum were key participants. For example, Optum, a company owned by United Health Group has collected data on 150 million Americans which includes medical data like medical diagnosis, tests, prescriptions, and medical costs, and socioeconomic data, which jointly is used to link medical outcomes to socioeconomic data such as gender, race, education level, and net worth. This is expanding the record sets held by health plans to include lifestyle data and data collection from the billions of sensors that gather data about individuals in Internet of Things devices, creating shadow health records that fall outside of regulatory protections. The data landscape is now a complicated web of protected health information, personally identifiable data, summary health data, financial data, lifestyle data, and Internet of Things data.

Supporting the data landscape is a disconnected set of federal and state laws and regulations that govern data privacy in the healthcare industry. The two main healthcare regulations that govern what we traditionally think of as a health record is the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH); however, these laws are supplemented by a series of frameworks, and federal and state regulations to protect this data.

In this paper, we will look at the healthcare data environment from the viewpoint of a health plan, one of the four entities defined by HIPAA and evaluate the legal and regulatory frameworks currently in place with the goal of identifying gaps in data protections for healthcare records. When we look at the data collection that is occurring by health plans, we must first look at the healthcare ecosystem, the risks currently inherent in this ecosystem, the data collected by this system, and the frameworks and regulations used to protect this data. An understanding of this ecosystem will guide the identification of gaps in how this data is protected and what can be done to address these gaps.

The paper is presented with the following sections. Section II outlines background information on the health plan environment within the healthcare industry including the state of cybersecurity breaches within the industry.

Section III provides an overview of the data environment within the industry which includes a definition of the different categories of data collected by health plans. The goal of this section is to define the data ecosystem for health plans. Section IV contains an overview of the legal and regulatory frameworks currently in place to protect data within the industry. In Sections V and VI an analysis of the gaps which currently exist in the data ecosystem legal and regulatory framework are identified and recommendations for closing these gaps are discussed.

II. BACKGROUND

As technology has evolved, various laws designed to protect healthcare related data has limited how companies can store, process, and transmit healthcare related data. At the same time the ubiquity of technology in everyday life has resulted in the generation of immense volumes of data outside of traditional healthcare settings that do not fall into standard regulatory or security frameworks. At any given time, the number of data points for an individual can range from 1,900 to 10,000 points of data [2]. This data is generated by a variety of companies and industries, not just those entities that generally fall under regulatory protections. This has led to the development of shadow health records, data points gathered from a vast array of sources, that can be used to build a health profile. This data may be deidentified at the source, but when combined with data from other sources, can identify patients in less-regulated record sets. For example, IQVIA, a data broker, has combined data from more than 100,000 sources to identify "approximately 400 million comprehensive, longitudinal, anonymous patient records" which enables them to know 85% of the world's prescription records by sales revenue [3]. Because these records are deidentified, they exist outside the protections under current regulations. However, when combined with data from other sources, it would be possible to tie an individual to a prescription captured in this data set.

Over the last decade, the weaponization or theft of data in the Health Care sector rapidly increased. Healthcare companies are a highly attractive and lucrative target for attackers because they store and share a large amount of personal information. Threats against healthcare entities are ongoing threats and not isolated, infrequent events. Common types of security breaches in healthcare include phishing attacks, malware, ransomware, theft of patient data, insider threats, and hacked IoT devices. A 2019 ECRI Institute evaluated the top 10 health technology hazards and found that remote access and malicious changes to data were considered the top technical risk to data [4]. When a breach occurs, the HITECH Act requires covered entities to report to the Department of Health and Human Services to be posted to the Breach Portal. From Mid-2009 to Mid-2022, the number of breaches increased by 588% and impacted over 313 million individuals [5]. Table 1, compiled from the HHS

Breach database reflects the number of people impacted by a breach year over year starting in the fourth quarter of 2009 and continuing through the end of the second quarter of 2022. When we evaluate the percentage increase over each year, we do see drops; however, this is due primarily to large spikes in attacks in the prior year. For example, in 2016, we see an 85% drop in the number of individuals impacted by a beach, but this is due to the 2015 Anthem attack which impacted an unusually large number of people.

Table 1: Health and Human Services Year Over Year Individuals Affected by a Healthcare Breach by Covered Entity Type

Year	Business Associate	Health Plan	Healthcare Clearing House	Healthcare Provider	Grand Total	% Increase
2009	91,400	3,800	-	39,573	134,773	
2010	1,529,729	3,564,344	-	838,203	5,932,276	4,302
2011	8,936,804	89,977	1,250	4,134,127	13,162,158	122
2012	1,146,711	336,265	10,000	1,361,009	2,853,985	(78)
2013	1,058,760	100,655	6,504	5,852,920	7,018,839	146
2014	8,475,565	2,207,239	-	8,390,747	19,073,551	172
2015	3,592,767	102,478,796	-	6,395,157	112,466,720	490
2016	3,612,183	878,905	-	12,213,916	16,705,004	(85)
2017	221,657	391,518	-	4,691,870	5,305,045	(68)
2018	5,997,273	2,833,971	-	5,385,496	14,216,740	168
2019	13,325,671	3,379,671	1,566,938	26,694,803	44,967,083	216
2020	7,780,498	3,679,303	46,232	16,926,581	28,432,614	(37)
2021	5,826,328	2,087,666	17,900	5,229,764	13,161,658	(54)
2022	18,910	15,908	-	301,720	336,538	(97)
Totals	**61,614,256**	**122,048,018**	**1,648,824**	**98,455,886**	**283,766,984**	

If we map this out, we can see that Health Plans (i.e., health insurance companies) accounted for the most disclosures during the thirteen-year period with more than 126-million-member records compromised amounting to more than 43 percent of all breaches. Healthcare providers accounted for the second largest disclosure levels at more than 122 million impacted by a breach. The Anthem hack in 2015 contributed significantly to this number with an estimated 78.8 million patient records breached and an additional 18.8 million non-patient records lost in one breach [6]. This attack is an example of the impact a single breach at a health plan can have on the industry. Ultimately, this breach cost Anthem more than $170.5 million, with 39.5 million paid to state AG's,

$16 million to Health and Human Services Office of Civil Rights, and another $115 million settlement to resolve a class action lawsuit [7]. This is a significant negative impact to a company and the industry.

What is not illustrated in the breach data is the data held in shadow health records as this data is not generally covered by laws and regulations; therefore, reporting requirements do not apply. Shadow health data is data generated outside of covered entities which is then used to make inferences about the healthcare of an individual. However, in some cases, many of the companies collecting data for these shadow health records have themselves been a source of breaches of healthcare data. For example, in 2011, IBM, a business associate of Health Net Inc had a security breach which compromised healthcare data for 1.9 million current and former members when they lost several server drives containing names, addresses, health information, financial information, and social security numbers [5], [8]. In 2015, we see a massive spike in disclosures, with four health plans accounting for the bulk of the data losses during the year all to hacking incidents. Figure 1 illustrates the year over year impact to patients by covered entity type.

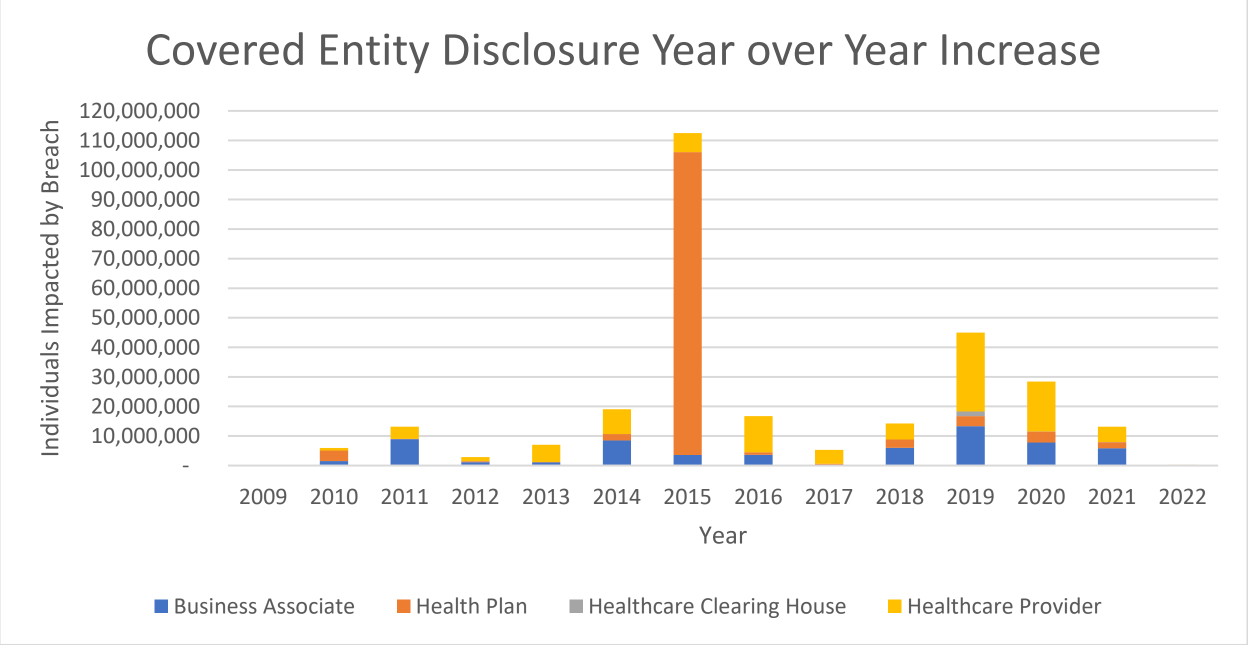

Figure 1: Individuals Impacted by Disclosure by Covered Entity Type Year Over Year

There are six main categories of breach type as defined by HHS in their breach database, with more than 80% of breaches occurring due to hacking or IT incidents. To date, this primarily covers breaches related to protected health information under HIPAA and does not incorporate other types of data collected by health plans from other data sources. Hacking and IT incidents are by far the largest source of data breaches; however, a significant number of individuals are impacted by unauthorized access or disclosure, loss, and thefts, which can occur from insider or outsider threats. Table 2 identifies the number of individuals impacted by each category of breach.

Table 2: Breach Impact to Number of Individuals by Category of Breach

Type of Breach	# Individuals	Percentage
Improper Disposal	1,988,011	0.70%
Loss	9,358,158	3.30%
Theft	26,563,931	9.36%
Unauthorized Access/Disclosure	14,987,157	5.28%
Unknown	3,169,530	1.12%
Hacking/IT Incident	227,700,197	80.24%
Grand Total	**283,766,984**	

The data contained in the breach reports generally encompass data covered under HIPAA protections. The HHS breach reports currently show data that falls within the 18 identifiers defined by HIPAA regulations. Reports from 2009 through 2022 show that highly sensitive data is frequently compromised by covered entities. This data spans personal health information, summary health information, financial data, and personally identifiable information. What is notably absent is lifestyle and Internet of Things data. Data elements that fall within these categories currently operate in a gray area; with a disjointed series of state and federal regulations governing them. The data elements from 4,017 breaches were evaluated to identify the top twenty-five data elements compromised; however, these data points only reflect data in the categories companies are legally required to report and do not accurately reflect the full data environment of the healthcare industry.

Figure 2 captures the top 25 data types compromised by breaches since 2009. The top three elements incorporate primary identifiers that can tie a health record to a specific individual, followed by two elements that encompass personal health information tied to that individual. The data compromised spans across data types, to encompass personally identifiable information, personal health information, and financial information; all exceedingly valuable assets on the data black market.

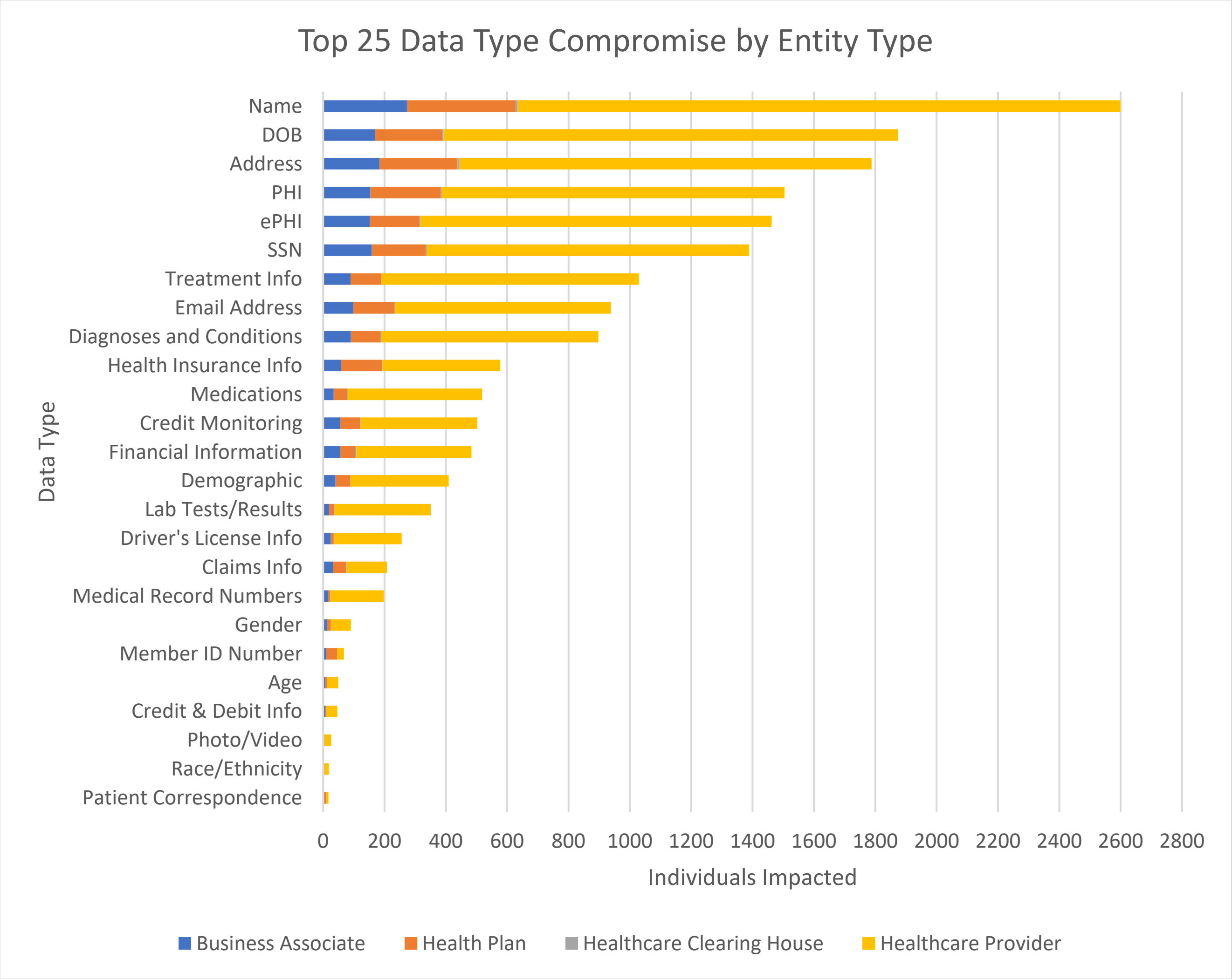

Figure 2: Top 25 Data Types Compromised by Entity Type as Reported in the Health and Human Services Breach Database

Healthcare records are considered highly valuable on the black market because it contains all the personally identifiable information for an individual and when attacks occur, they often gain data for hundreds of thousands of patients. In [13] healthcare records are valued up to $250 per record compared to the average $5.40 for payment card data. Additionally, the criticality of this data to the patient makes this data highly desirable to attackers. A healthcare facility is more likely to quickly pay a ransom in a ransomware attack because access to this data is literally life or death in many cases.

III. DATA ENVIRONMENT

To understand the data environment of health insurance requires a basic understanding of the industry and core functions of a health plan. Health plans allow individuals, employers, and providers to share risks related to health care costs which in the United States consists of both private and public health insurance providers. Private health plans provide health insurance through an employer sponsored plan or they sell individual plans. Public health insurance is provided through federal or state programs such as Medicare, Medicaid, and more recently through the subsidized Health Exchanges. In 2019, 68% of the population were covered under private health care options and 34.1% were covered under some form of public health insurance option, with 55.4% of people covered through an employer [9]. One of the primary functions of a health plan is to ensure claims adjudication between providers, insurers, and the plan member for covered services under the selected plan. As part of this process, health insurance companies collect a broad range of information, including medical information, financial information, physician information, and employment information that must be protected.

Companies are using data from sources as diverse as fitness trackers, web searches, mobile personal health applications, shopping histories, social media posts, genetic databases such as 23andMe, geolocation services, data from pharmaceutical companies, Internet of Things device data, personal information such as names and addresses, health care data, financial data, insurance and social security numbers, contracts, HR data, research and intellectual property of all types, and many other diverse sources [10], [3]. Restrictions on how this data is used has led to health plans branching out to collect data from a variety of other sources. Because of this, insurance companies often have the "most complete and comprehensive digital data about a patient," including medical data, financial data, employer data, personally identifiable information, and a variety of other information [11]. Much of this data is collected by third-party companies who then make this data available to health plans through various analytical engines. For example, IBM has IBM Watson, which holds real-world data assets like patient demographics, clinical data, vitals and biometrics, social history data, laboratory and microbiology data, prescription information, implantable device details, assessments and Pro data, provider demographic data, productivity data, dental data, and many other data sets [12]. Companies like United Healthcare are quietly compiling data that falls outside of HIPAA protections to ostensibly improve patient outcomes while also increasing revenue streams for the organizations. Figure 3 diagrams the data ecosystem as it exists in the industry today by illustrating the data categories in use.

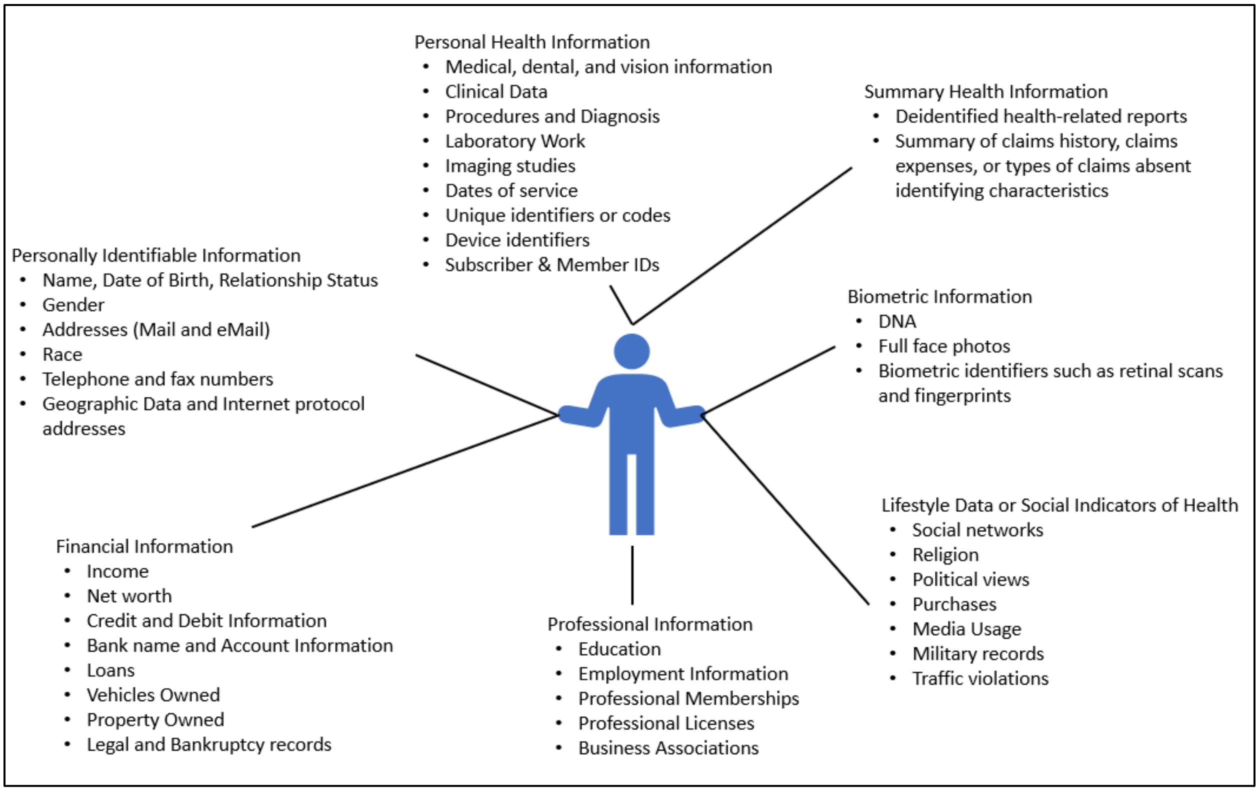
Personal Health Information
• Medical, dental, and vision information
• Clinical Data
• Procedures and Diagnosis
• Laboratory Work
• Imaging studies
• Dates of service
• Unique identifiers or codes
• Device identifiers
• Subscriber & Member IDs
Summary Health Information
• Deidentified health-related reports
• Summary of claims history, claims expenses, or types of claims absent identifying characteristics
Personally Identifiable Information
• Name, Date of Birth, Relationship Status
• Gender
• Addresses (Mail and eMail)
• Race
• Telephone and fax numbers
• Geographic Data and Internet protocol addresses
Biometric Information
• DNA
• Full face photos
• Biometric identifiers such as retinal scans and fingerprints
Lifestyle Data or Social Indicators of Health
• Social networks
• Religion
• Political views
• Purchases
• Media Usage
• Military records
• Traffic violations
Financial Information
• Income
• Net worth
• Credit and Debit Information
• Bank name and Account Information
• Loans
• Vehicles Owned
• Property Owned
• Legal and Bankruptcy records
Professional Information
• Education
• Employment Information
• Professional Memberships
• Professional Licenses
• Business Associations

Figure 3: Health Data Ecosystem Data Type Categories Collected by Health Plans

A. Protected Health Information (PHI) Data

The privacy rule in HIPAA protects all individually identifiable health information stored, processed, and transmitted by a covered entity. The Health Insurance Portability and Accountability Act (HIPAA) was specifically designed to protect personal health information as attacks on HIPAA-covered data can have a profound impact on victims. HIPAA defines eighteen identifiers that can turn health information into protected health information:

- Names
- Dates, except when only the year is provided
- Telephone numbers
- Fax numbers
- Geographic data identifying the location of a patient
- Social Security numbers
- Email addresses
- Medical record numbers
- Account numbers
- Health plan beneficiary numbers such as subscriber or member identifiers
- Certificate/license numbers
- Vehicle identifiers and serial numbers including license plates
- Web URLs
- Device identifiers and serial numbers
- Internet protocol addresses
- Full face photos and comparable images
- Biometric identifiers, such as retinal scan, fingerprints, or DNA
- Any unique identifying number or code

In personal health information there is some overlap with data elements that are categorized as personally identifiable information (PII) data and biometric information.

B. Personally Identifiable Information (PII) Data

According to the Department of Labor, personally identifiable information is any data or information that directly or indirectly infers the identity of an individual to whom that data or information applies [14]. In addition to the eighteen elements defined by HIPAA, this can include data elements such as gender, race, date of birth, employment data, academic data, and other descriptors for an individual. Personally identifiable information is generally considered biographical data about the individual.

C. Biometric Information

Biometric information is information that defines body measurements and calculations that are common to healthcare. This data also encompasses unique anatomical, physiological, or behavioral characteristics that can be used to authenticate the identity of an individual. Biometrics may include data elements such as fingerprints, iris patterns, facial features, voice recognition, signature, gait, keystroke dynamics, DNA, and other characteristics. Much of this data is captured in secondary systems which are then compiled into databases that are made available across a broad range of industries. For example, companies like Ancestry and 23 and Me are compiling genetic data about individuals, personal assistants like Alexa and Siri utilize voice recognition, and surveillance systems are starting to utilize facial recognition technology. All this data is collected into databases which are compiled and sold on secondary markets.

D. Summary Health Information (SHI) Data

As defined by Title 45 §164.504, summary health information is individually identifiable health information that "summarizes claims history, claims expenses, and type of claims" but has been sanitized of protected health information as defined in Title 45 §164.514 [12]. Summary health data is collected by health plans and is generally utilized for high-level reporting. Companies like IBM are developing programs that de-identify healthcare data for the purposes of performing healthcare analytics; however, this data is derived from personal health information and personally identifiable information and can be traced back to the individual with minimal effort.

E. Financial Data

The average health plan provides services to a variety of stakeholders by providing medical, dental, and vision insurance products to individuals, employer groups, and government agencies. Health plans, like many companies, collect a variety of financial data, including but not limited to credit and debit card data, bank account data including bank name and bank account number, salary and tax information, information contained on credit reports, property data, and other financial data points. This data enables companies to build social profiles of individuals.

Additionally, health plans hold financial data for employers who carry health insurance through the health plan, for healthcare providers who provide medical care, and insurance brokers who sell health plan products. The compromise of the systems containing financial information could result in hundreds of millions of dollars in losses for stakeholders associated with the health plan.

F. Internet of Things Data

Internet of things devices generally collect three categories of data: consumption data, physical activity data, and physiological data which creates patient generated health data which can be shared [15]. Networked devices, collectively known as the Internet of Things, generate an unprecedented amount of information about the habits, personalities, preferences, and everyday actions of individuals; however, there are a subset of these devices, called wearable devices, that capture personal health information that can directly be tied to an individual [16]. Wearable technology gathers real-time health related data such as heart rate, brain activity, respiration, body temperature, stress levels, blood pressure, oxygen saturation levels, insulin levels, calorie consumption, sleep patterns, and many other health related data points. This data is captured in conjunction with geolocational information and fed into software applications that are designed to provide tailored feedback to the individual and often transmit this data from their mobile devices to social networks and other applications. Some devices additionally collect data from accelerometers, gyroscopes, vibration monitors, and GPS chips which give the manufacturers data about physical activity of the individual.

While these devices give consumers unparalleled capabilities to monitor their personal health and fitness information, they pose some risks. Hardware devices, like fitness trackers, do not exist in isolation, they are combined with software that collects, manages, shares, and views data on mobile phones and this data is then transmitted into cloud applications [15]. These devices generate complete personal health records that enable continuous passive tracking of data 24/7 by companies that are not bound to health care privacy regulation. The continuous network connections provided by these devices give an opaque, unregulated view into the private life of the user. The users of these devices have little awareness of what data is collected, and how that data is stored and used, and permanent deletion is near impossible. Additionally, these transmissions are rarely encrypted, which puts them at risk.

As noted in [16], this data is largely unregulated and “provides priceless insight to marketers, advertisers, retailers, insurers, employers, financial service providers, and social contacts” and is stored within networks with weak security. Health plans are partnering with wearable manufacturers to make these devices available to their members. For example, in August 2019, Blue Cross Blue Shield Association announced they were partnering with Fitbit to provide the devices at a reduced cost to their members. To date, the data generated by these devices are not regulated; but there is a high black-market value for acquiring data from wearable devices, making it an increasingly valuable target for hackers. In general, wearable devices are not covered by HIPAA under current regulations which limit protections to "covered entities" and does not include wearable technology manufacturers. While these manufacturers are trying to create connections between the information gathered by their products and doctors, they typically use targeted markets to consumers to keep track of their health and manufacturers avoid officially marketing the devices in a manner that would convert them to medical devices that can share information with physicians. This keeps these manufacturers outside of the definition of a "business associate" and thus outside of the purview of HIPAA.

G. Lifestyle Data

Many health insurers and data brokers are collecting and utilizing lifestyle data as social determinants of health (SDOH) to influence outcomes and efficiency in health care and drive what patients pay for health insurance. An Axiom Pulse Survey notes that two-thirds of hospitals actively use or want third-party consumer and lifestyle data, with more than half wanting to integrate SDOH data into electronic health records [17]. The U.S. Department of Health and Human Services defines SDOH data as information that may impact a person's "health, well-being, and quality of life" and includes information pertaining to economic stability, education access and quality, health care access and quality, neighborhood and environment, and social and community factors [18]. The World Health Organization (WHO) defines SDOH as data that describe the conditions in which people are "born, work, live, and age" and include political and social systems, economic policies, work-life conditions, food or housing insecurity, and developmental agendas resident in the area in which live [19]. These definitions are critical for defining the type of data incorporated into the lifestyle data bucket. Much of this information can be gathered from the public domain and unlike community level data available from the Area Health Resource Files, County Health Rankings, or the US Census Bureau, lifestyle data can be attributed at the individual level [20]. Optum Health, LexisNexis, and IBM Watson are leading the market in gathering SDOH data and making this data available to health plans.

Optum, a company tied to UnitedHealth Group, combines clinical and claims information with social media interactions with the stated intent of improving clinical outcomes. Marketing material for Optum identifies lifestyle data as one of the data components along with demographic data, health behaviors, and health needs to calculate the COVID Area Vulnerability Index to determine patient volumes over time and is designed to aid health officials at all levels prepare for and manage public health emergencies [21]. This calculation uses risk scores in three areas: mobility/density/SDOH, morbidity rates, and the adequacy of health resources to provide guidance on the distribution of resources into an area during a health emergency [22]. A significant volume of individual-level SDOH data must be gathered by Optum to perform these calculations and provide these tools. A second company, LexisNexis, is developing data repositories that use "442 non-medical personal attributes to predict medical costs" from a cache of 78 billion records gleans from 10,000 public and proprietary data sources and LexisNexis Risk Solutions provides data from a broad array of sources to a variety of industries, including health plans [23]. This data may include information auto and home insurance policies, real estate, collections decisions, marriage and divorce records, email addresses, professional licenses, business associations, legal and bankruptcy reports, education, phone records, social security records, military records, traffic violations and a broad array of other data points. Much of this data is protected to a limited basis under the Fair Credit Reporting Act (FCRA). The third company, IBM Watson, surveys 80,000 Americans a year to gather information on their lifestyle, attributes, and behaviors with the goal of identifying social and economic factors in an area. This survey pools responses to assess the health of people within an economic area. This data is protected by a series of disjointed state laws.

IV. LEGAL AND REGULATORY FRAMEWORKS

The healthcare industry is one of the more heavily regulated industries because of the high volume and sensitivity of the data that is stored and transmitted. A variety of frameworks, regulations, and standards exist to protect data held by organizations across industries, including the healthcare industry. In the healthcare industry, the Health Insurance Portability and Accountability Act (HIPAA) is the prevailing framework; however, there are other frameworks and regulations that impact how data is secured across the industry. Rules such as HIPAA Security Rule define the basic requirements healthcare organizations must comply with, but they do not provide guidance on how to comply with these security needs [10].

A. HealthCare Interoperability Standards

The industry has several interoperability standards that govern how data is processed and transmitted. This includes the HL7 Fast Health Interoperability Resources (FHIR), Logical Observation Identifiers Names and Codes (LOINC®), SNOMED CT, and CDS hooks standards. LOINC® is a widely used standard for codes and terminology used for laboratory test orders and results, SNOMED CT are a collection of medical terms and codes assigned to records, and CDS hooks are a specification for clinical decision support. FHIR is an interoperability standard that defines how healthcare information is exchanged between systems and forms a lynch pin in the communications of healthcare related data between providers and health plans. The FHIR standard provides a module on how to protect security and privacy of a FHIR server; however, this standard does not provide the technical approach to security, it provides a building block to create secure systems based on the following use cases [24]:

- Security and privacy ensure that data is secured using encryption and privacy is ensured using privacy principles such as privacy by design that considers individual preferences when developing the system.
- Authorization and access control should be varied according to the access scenario where access is granted for authorized users and denied for unauthorized users. FHIR uses REST API and oAUTH access tokens to provide authorization and access controls. Policies are defined to ensure query parameters return results only when the requestor is authorized to see the data.
- User identity and access context defines the use of role-based, context-based, and attribute-based controls to ensure the requestor only receives the data to which they have access.
- Security and privacy audit logging is provided through tamper-proof logs that capture privacy and security related events as well as interactions with the REST APIs used by the system.
- Account of disclosures and access reports ensure the patient understands how their data is collected, processed, used, stored, disclosed, and transmitted. HIPAA provides a restrictive disclosure reporting that is leveraged by the FHIR system but provides the report in a readable format. The report defines who accessed what data, where and when the data was accessed, and why it was accessed.
- Privacy consent is captured by the FHIR system and governs the rules for “collection, use, and disclosure of health data” by the patient. This consent flag may be used by a variety of devices to enable interoperability with healthcare systems as well as Internet of Things and other devices.
- Provenance records the who, what, when, where, and why a data record was created, and records are created, updated, or deleted from the system. Since the FHIR system is used to compile large volumes of health care data from multiple sources, understanding where the data is sourced is critical for ensuring data integrity.

- Digital and electronic signatures are provided to ensure authenticity, integrity, and non-repudiation of the record set. These signatures tie into the provenance component of the FHIR system.
- De-identification, anonymization, and pseudonymization is provided to limit what requestors can see of a data record. This reduces privacy risks by only providing the data elements necessary to perform the service needed.
- Test data security considerations are incorporated into the standard to ensure when development work and testing are completed, the data used is representative of the data types needed by the systems but does not provide “production” data that could compromise patient privacy.

B. National Institute of Standards and Technology (NIST) Frameworks for Improving Critical Infrastructure

NIST provides multiple frameworks that can be used by health plans to guide how they protect the data in their systems. The three primary frameworks are the NIST Frameworks for Improving Critical Infrastructure, NIST Special Publication 800-53 Security and Privacy Controls for Federal Information Systems and Organizations, and the NIST Cybersecurity Framework. The NIST Framework for Improving Critical Infrastructure that provides guidelines for securing information technology (IT), industrial control systems (ICS), cyber-physical systems (CPS), and Internet of Things (IoT) devices within critical infrastructure. This can guide how data gathered from medical devices can be protected. The NIST 800-53 is a comprehensive framework that provides a broad focus for integration of privacy controls into security controls for entities, including those in the healthcare sector. This standard is required by government agencies and organizations that do business with government agencies; however, compliance with this standard is typically voluntary for private sector businesses that do not contract with government entities.

This framework maps into the HIPAA and ISO 27001 frameworks, making it a key framework used in the industry. The NIST Cybersecurity Framework is a key method for identifying weaknesses in a security system and is composed of three components: the core, the implementation tiers, and the profiles. The core organizes cybersecurity goals into five phases: identify, protect, detect, respond, and recovery. The implementation tiers categorize cybersecurity effectiveness into four tiers ranging from partial (tier 1) to adaptive (tier 4). The third component identifies opportunities for improvement by comparing organizational objectives to the framework core. The HIPAA Privacy Rule and the HIPAA Security Rule collectively establish national standards to protect health related information that is electronically stored or transmitted.

C. Center for Internet Security (CIS) Critical Security Controls

The CIS controls define 18 actions to protect against cybersecurity attacks presented based on priority, with the most critical controls occurring first in the standard. These controls map to many frameworks, regulations, and standards used within the healthcare industry, including the NIST Cybersecurity Framework, NIST 800-53, ISO 27000 series, HIPAA, PCI DSS, and FISMA. As with the NIST frameworks, these controls are not mandatory, but they provide a mechanism for streamlining cybersecurity within an organization and are most effective when used in conjunction with other controls.

D. Health Insurance Portability and Accountability Act (HIPAA)

The Health Security Act was later added to Title II, Subtitle F of HIPAA to add administrative simplification to HIPAA by focusing on two areas: improving the efficiency and effectiveness of healthcare by standardizing electronic exchange of administrative and financial transactions; and protecting security and privacy of health information during transmission [25]. Under this law, Health and Human Services has the authority to develop standards for administrative and financial transactions that must be used by health plans and healthcare providers. These standards cover medical code sets, security, electronic signatures, privacy, and the development of unique identifiers and their use. The Health Security Act was the precursor to and would eventually evolve into HIPAA. The Health Security Act proposed the development of scalable, comprehensive security requirements that did not specify technology, but did require implementation of the most appropriate technology based on the entity.

HIPAA defines how personal health information (PHI) can be used by specific entities within the healthcare industry and is primarily a regulation governing the risk associated with highly sensitive and valuable healthcare data [26]. Compliance to HIPAA is mandatory for health plans, heath care providers, health care clearinghouses, and business associates. This Act is a compilation of three main rules: the HIPAA Privacy Rule, the HIPAA Security Rule, and the HIPAA Breach Notification Rule. The HIPAA Privacy Rule under 45 CFR Part 160 and subparts A and E of Part 164, requires safeguards to be in place to protect all protected health information and gives individuals rights to examine and correct this information; the HIPAA Security Rule under 45 CFR Part 160 and subparts A and C of Part 164 requires administrative, physical, and technical safeguards to be in place to protect data that is created, received, used, or maintained by the entity holding the data for electronic PHI data; and the Breach Notification Rule, under 45 CFR §§ 164.400-414 defines and requires notification protocols when a breach occurs in an organization [27]. The goal of HIPAA is to address commonly exploited vulnerabilities that cause breaches of protected health information.

HIPAA is codified in the Code of Federal Regulations, Chapter 45, Parts 160, 162, and 164. Part 164 of the Act specifies the security requirements to which health care entities, including health plans, must comply to ensure the confidentiality, integrity, and availability of electronic protected health information. HIPAA requires healthcare entities to perform an accurate and thorough risk analysis, implement a risk management strategy, regularly monitor information access, perform regular workforce awareness and training activities, have incident plans, disaster recovery plans, and data backup plans in place, and implement physical and technical safeguards [28]. Physical safeguards include requirements for workstation use and security and access controls and technical safeguards include using unique assigned identifying numbers, automatic logoffs, encryption, and decryption [29]. A second key function of HIPAA is to fully define responses to and the penalties for a data breach. The framework does an excellent job of identifying what must be protected but falls short of how to protect these assets [11]. To address this gap, NIST developed a crosswalk with the Cybersecurity Framework and HIPAA to enable entities to identify the necessary activities to protect personal health information.

HIPAA avoids specifying technologies that must be used to protect the data held by covered entities, but rather specifies process and outcome requirements that focus on what is reasonable for the entity. When determining what security measures an entity should implement, they consider four factors: “the size, complexity, and capabilities of the entity; the probability and criticality of potential risks to electronic PHI held by the entity; the cost of reasonable security measures; and the technical infrastructure, hardware, and software security capabilities held by the entity” [30]. Under the security rule, there are two types of specifications to be considered: those rules that are required and must be implemented, and those rules that are addressable under “reasonable and appropriate” means based on the environment. The security rule consists of administrative safeguards, physical safeguards, technical safeguards, organizational requirements, and documentation requirements.

Table 3: HIPAA Summary of Requirements Related to Administrative Safeguards, Physical Safeguards, Technical Controls, Organizational Requirements, and Policies, Procedures, Documentation

Administrative Safeguards	
Security management process	This safeguard requires risk reviews, including a risk analysis, risk management, and a formal review of information system activity. Entities should develop and implement a sanction policy for violations.

Assigned security responsibility	The person responsible for developing and implementing security policies in accordance with the Security rule in HIPAA should be clearly identified.
Workforce security	Policies and procedures clearly defining who may access electronic PHI and for what purpose should be defined. The entity should have clearly defined policies and procedures for assigning and terminating access to sensitive data.
Information access management	Policies and procedures should be implemented for who can access information and for what purpose. These policies and procedures should be consistent with the security rule.
Security awareness training	A security awareness training program should be implemented for all employees based on their roles and responsibilities. This training should encompass common security practices such as password management procedures, login monitoring to sensitive systems, and on-going security training relevant to industry risks.
Security procedures	Procedures should be developed that clearly identify the process for identifying and responding to breaches or suspected breaches.
Contingency plan	A data backup plan, disaster recovery plan, and incident response plan should be developed.
Evaluation	Technical and non-technical evaluations are performed periodically.
Business associate contracts	Utilize business associate contracts that document that a business associate has the proper organizational requirements in place relevant to the security rule.
Physical Safeguards	
Facility access controls	Entities should have physical access controls in place that limit the access to systems and facilities that house sensitive data.
Workstation use	Policies and procedures are in place that identify how workstations which access sensitive data are used, the way those functions are performed, and the physical configuration of the workstation.
Workstation security	Entities initiate physical safeguards for workstations that access electronic PHI and restrict access to authorized users only.
Device and media controls	Policies and procedures for disposition and reuse of data recording media, including media used for data backups are in place.
Technical Controls	
Access controls	Entities implement access controls, including implementing unique user IDs, automatic log-off, encryption and decryption measures, and emergency access procedures.
Audit controls	Entities implement audit controls for hardware, software, and procedural mechanisms which identify who accessed sensitive system, when the system was accessed, and what was done while using the system.
Integrity	Safeguards are in place that prevent modification of sensitive data held by the entity.

Person or entity authentication	Procedures are in place to verify the identity of persons accessing sensitive data.
Transmission security	Safeguards are in place to protect data while in transit.
Organizational Requirements	
Business associate contracts	Reasonable measurements are in place to ensure business associates utilize reasonable protection mechanisms when receiving electronic PHI.
Group health plans	Requirements specific to group health plans are identified and documented.
Policies, Procedures, Documentation	
Policies and Procedures	Reasonable and appropriate policies and procedures are implemented to comply with the security standard.
Documentation	Policies and procedures, reports of actions and activities, and assessments required by the rule are documented and maintained for six years.

HIPAA has been amended by several laws over the years, such as the Genome Information Nondiscrimination Act, the Coronavirus Aid, Relief, and Economy Security (CARES) Act, and the 21st Century Cures Act. Title 1 of Genome Information Nondiscrimination Act (GINA) amended HIPAA to protects against discrimination based on genetic information when underwriting health insurance policies. This amendment incorporated genetic data as part of the personal health record. The CARES Act was developed in response to the COVID-19 pandemic and contains privacy and security components for data related to health care. This Act provided funding for public health surveillance and analytics infrastructure with the intent of securing the public health data. There was a rapid increase in cyber-attacks on health care providers and plans during the pandemic due to the rapid increase in use of technology to continue business operations. Many health plans utilized remote access to their systems to facilitate health care. CARES provided financial assistance to small businesses, including health plans, for training on cybersecurity risks and how to mitigate these risks. The Cybersecurity and Infrastructure Agency was allocated funding for improving critical healthcare infrastructure. The CARES Act amended the Public Service Act to conform with HIPAA by requiring breach notification and consent requirements. While this act did not provide security details, it was critical to providing funding to enable increased security and privacy protocols. Finally, the 21st Century Cures Act, included components to ease regulatory burdens related to electronic health records and health information technology systems, and prohibited blocking information if information sharing was within the bounds of HIPAA and state privacy laws.

The HIPAA privacy rule is not comprehensive, as it only applies to four distinctly defined covered entities: health care providers, health plans, health clearinghouses, and their business associates. Any entity outside of this definition is not bound by HIPAA regulations. A second weak point are numerous, broadly worded exceptions that exist in the law. [1] notes there are twelve exceptions where identifiable health care information may be shared without consent or notification to the patient: "(1) where required by law; (2) for public health activities; (3) about victims of abuse, neglect, or domestic violence; (4) for health oversight activities; (5) for judicial and administrative proceedings; (6) for law enforcement; (7) about decedents; (8) for cadaveric organ, eye, or tissue donations; (9) for some types of research; (10) when there is a serious threat to health or safety; (11) for special government functions, including national security; and (12) for worker's compensation." This leaves a significant number of exceptions to the law that can be exploited.

E. Health Information Technology for Economic and Clinical Health Act (HITECH)

The HITECH Act updated HIPAA in 2009 and later in the 2013 Omnibus Final Rule. This regulation identifies "required" implementation specifications and "addressable" implementation specifications. A required specification must be implemented by the organization while the addressable specifications provide the discretion for organizations to implement the specification or based on an assessment, implement an alternate specification based on the circumstances [26]. This specification also further expands the requirements for administrative, physical, and technical safeguards implemented by the organization. The administrative safeguards consist of non-technical measures, such as defined processes and procedures and technology usage, an organization uses to protect the data; the physical safeguards physical components and includes physical access control safeguards and facility security planning, and technical safeguards are the processes used to store, process, and transmit electronic PHI [26].

The goal of the HITECH Act is to promote the use of electronic health records to improve the efficiency of medical treatment. This Act contains four subtitles, two of which directly influence cybersecurity standards. Subtitle B covers the testing of health information technology and Subtitle D covers the requirements around improving privacy and security of healthcare IT functions and the relationship between HITECH and other laws [31] HITECH forced business associates to become HIPAA compliant and introduced tougher penalties for HIPAA violations. HITECH requires more stringent notification requirements and requires the HHS Office for Civil Rights to publish healthcare data breach information, thereby strengthening HIPAA. HITECH provided incentives for companies to shift to the use of electronic records to increase interoperability, but it also increased the cybersecurity risk for these entities.

HITRUST created the Common Security Framework (CSF) which combined risk management and security controls from HIPAA, the National Institute of Standards and Technology (NIST), the International Standards Organization (ISO), and Process Safety Information (PSI)" into one framework [32]. HITECH requires any company that provides federal health benefits to comply with NIST standards and mandates audits of healthcare providers and sets the standards for information security companies must follow. HITECH continued the shift to electronic medical records, however, most of the focus was on the meaningful use function, with little focus on patient privacy, security, usability, or interoperability on the part of health care companies [33]. The HITRUST CSF framework is designed to enhance HIPAA regulations by utilizing risk analysis and risk management and combining elements of other frameworks. HITRUST leverages components of ISO, NIST, PCI Security Council, and HIPAA standards as part of the framework to provide baseline security controls.

F. American Recovery and Reinvestment Act (ARRA) of 2009

The American Recovery and Reinvestment Act of 2009 was initiated to address the economic downturn of 2008; however, it did contain provisions which modified HIPAA. Prior to this Act, an individual could request a covered entity not share their PHI; however, the covered entity was not required to comply. The ARRA required covered entities to comply with this request except in cases where a provider was communicating with a health plan for payment or other operational functions. Second, the ARRA required the Secretary of Health and Human Services to issue guidance to define the "minimum necessary standard" more fully for requests for and disclosures of PHI [34]. Prior to ARRA, this standard was open for interpretation by the covered entity. The third element was an expansion of the accounting rule that required a covered entity to provide an accounting of disclosures made even if those disclosures are made for payment, treatment, or healthcare purposes. Fourth, the Act gave the individual the right to access their electronic healthcare record on request. Finally, the Act included several requirements around the expansion of the electronic health record by driving toward an interoperability standard for data exchange between covered entities.

G. Affordable Care Act (ACA)

The focus of the Affordable Care Act was concerned with making healthcare understandable and affordable for patients. There are provisions in the Act which define operating rules for information formats and transmission formats to enable more uniform communications between healthcare providers and health plans. The ACA attempts to standardize the communication standards used to exchange information between entities in the healthcare industry, which ultimately impacts the design and implementation of information systems used to transmit data.

H. Consolidated Appropriations Act (CAA)

The Consolidated Appropriations Act of 2016 and the Cybersecurity Act of 2015 required the creation of an automated system which enables voluntary sharing of cybersecurity threats between health care providers and the federal government. As part of the CAA, a task force comprised of health industry stakeholders, cybersecurity experts, and HHS approved federal agencies was developed for addressing how to improve cybersecurity in healthcare. This task force is responsible for providing recommendations for regulations and guidelines impacting the healthcare industry. This includes guiding covered entities to implement NIST CSF standards.

I. Payment Card Industry Data Security Standards (PCI DSS)

This framework was developed specifically to protect financial data related to the credit and debit cards. PCI Compliance is not an obvious factor to consider when developing a cybersecurity plan for a health insurance company; however, health insurance companies possess a great deal of payment information for individuals and business entities. Health plans manage transactions between vendors, health care providers, brokers, employer groups, government agencies, healthcare clearinghouses, and members. Transactions take place through back-end processes, web portals, and through the customer care centers; all of which generates large amounts of financial data. Where payment operations are performed, PCI Compliance will impact cybersecurity protocols for the organization. Companies that manage card transactions must build and maintain secure networks and systems, protect cardholder data, maintain a vulnerability management program, implement strong access control measures, regularly monitor, and test networks, and maintain an information security policy in accordance with PCI standards [35].

J. ISO 27001 and 27002 Series Certifications

The ISO 27001 and ISO 27002 certifications are international standards that are indicators of a mature cybersecurity program in an organization. These frameworks are used by organizations that handle sensitive data. 27001 used to identify and mitigate risks related to handling sensitive information and 27002 used to build security practices and policies, asset management, physical and environmental security, operations management, access control, business continuity, and information security.

K. Federal Information Security Management Act (FISMA)

FISMA requires organizations to document their network integrations and digital assets, monitor their IT infrastructure, and regularly evaluate risks. FISMA strengthened the use of continuous monitoring by covered entities, increased focus on reporting and compliance on issues caused by security events rather than planning for security events, reduced reporting burdens, and emphasized cost-effective security using risk-based policies by incorporating the National Institute of Standards and Technology (NIST) suite of risk management standards and guidelines to implement the risk-based policies [36]. FISMA combined with the NIST Risk Management Framework and associated risk standards are designed to provide a minimum foundation for cybersecurity for data and information held by government agencies as well as state agencies and private organizations that manage federal programs who collect or maintain information by or on behalf of an agency" [36]. The Act sets minimum standards and standardizes risk processes in accordance with recommendations by NIST.

Under FISMA, organizations must meet the following the provisions: compile an inventory of their systems and integrations, information must be categorized according to the FIPS 199 standard, a system security plan is maintained and updated, security controls relevant to the organization are implemented and continuously monitored, risk assessments are performed, data is encrypted at rest and in transport, and annual security reviews must be performed [37]. Organizations are required to inventory and document the systems used within the agency, the information they collect and transmit, and any integrations between the systems and the networks. The systems holding the data must be categorized based on the level of risk in accordance with the FIPS 199, the Standard for Security Categorization of Federal Information and Information Systems. Data at rest and in transit should be properly encrypted according to its categorization. A security plan is created and maintained which defines the security controls the organization has in place, the documented security policies, and the timetables for additional controls to be added. The organization is required to implement the relevant controls to the organization and systems as defined in NIST Special Publication 800-53. The implemented controls should be continuously monitored to capture status reporting, configuration management, security controls, and changes made to systems [38]. Regular risk assessments, in accordance with NIST Special Publication 800-30, are conducted to assess the effectiveness of security controls in place and to determine if any additional controls are needed. These risk assessments cover organizational risks, business process risks, and information system risks. The sixth core feature of FISMA requires an annual audit to determine if the organization is meeting minimum security requirements.

L. Federal Trade Commission (FTC)

In healthcare, the primary function of the FTC is to prevent anti-competitive behaviors; however, they can also regulate data security standards. For example, the FTC developed a framework that makes healthcare organizations liable for poor data security practices even absent harm to the consumer. Since 2008, the FTC has filed more than 50 suits related to data security practices related to organizations failure to implement reasonable cybersecurity hygiene practices, which resulted or could potentially result in lost data. More recently the FTC provides guidance on minimum requirements for protecting against ransomware. These requirements include implementing education and awareness programs for preventing phishing schemes which result in ransomware attacks, utilizing good cyber hygiene principles, back up data often, and develop and test business continuity and incident response plans [26].

M. State Regulations

In the United States, each state may have different levels of regulation governing health care data and data from other sources. One of the strongest is the California Consumer Privacy Act (CCPA) which went into effect in 2020 and gives consumers greater control over their personal information. This Act applies only to California residents but has the potential to impact residents in other states. It does not apply to protected health information already covered by HIPAA, but it will apply to data contained in shadow health records [3]. The CCPA also provides detailed safe-guard requirements and provides a deeper definition of "research" that limits data used for research to data that is deidentified and not used for commercial purposes. One of the main weaknesses of data privacy in the United States is the disjointed state-level approach utilized for data protection. No comprehensive federal standard exists for protecting non-healthcare related data; therefore, it is up to each individual state to provide data regulations.

V. GAPS AND RECOMMENDATIONS

Health plans are expanding on the data they use to make health care cost decisions in response to legislation such as the Affordable Care Act that prohibits the use of pre-existing conditions in determining eligibility or calculating costs of health insurance. A combined NPR and ProPublica article note oceans of data in the public domain are available from data brokers who are feeding this data into complicated computer algorithms to predict what health care would cost for an individual [23]. During analysis of the data environment and the accompanying frameworks and regulations used to protect this data, three gaps were identified.

A. Gap 1: HIPAA Weaknesses

One of the weaknesses of HIPAA is data sharing occurs with big data companies under the auspices of the "business associate" agreement. One example of this is Project Nightingale, a project between Ascension chain of medical facilities covering 50 million patients, and Google. As part of this project, Ascension made all the intact electronic health records in their systems available to Google for the stated purpose of using artificial intelligence to analyze the data and improve medical outcomes, all without notice or consent from patients [1]. Google and Ascension executed a "business associate" agreement, making this data exchange legal under HIPAA. This is not a unique situation as many other large health care providers are partnering with companies like Amazon, Apple, and Microsoft to perform healthcare analytics with the goal of improving medical outcomes. The overly broad list of exceptions leaves too many loopholes through which identified healthcare data can be shared. Second, Under HIPAA, an individual can request their data be shared with a third party. This is often done through a legal compulsion or by using economic leverage. For example, an employer can request health information in exchange for an incentive saving on health insurance, life or long-term care insurance companies can require access to medical data to provide a lower premium, or worker's compensation or veteran benefit programs may compel sharing of health data to gain access to programs. In [1] 25 million compelled authorizations occur each year in the United States. HIPAA leaves those who are subject to economic leverage or legal compulsion open to compelled approvals and once that data is voluntarily shared with a non-covered entity, it is no longer covered under HIPAA.

B. Gap 2: Wearable Devices and Internet of Things Data Propagation

In [16], the collection of personal data is not regulated and there is a patchwork of federal regulations and antiquated agency guidelines that overlap, dovetail, or contradict state laws and regulations. No clear judicial framework exists to address the privacy concerns related to data generated by wearable devices. The Food and Drug Administration (FDA) is the only agency to provide guidance on the security of the data through their Mobile Medical Applications Guidance, however, this guidance is limited, and they have indicated they do not intend to regulate wearable devices, leaving this a gray area. Companies lack guidance on where personal fitness trackers end and personal medical devices begin. Regulations like the Health Insurance Portability and Accountability Act (HIPAA) protect health information, not lifestyle or financial information, or information generated from wearable devices. Metadata and other non-health data to indicate a patient's health status generated by companies that produce products like smartwatches, fitness trackers, health wearables, mobile apps, and other products are not covered by HIPAA or other regulating bodies [39].

The large volume of data propagated by Internet of Things devices creates an interconnection between "trusted" and "untrusted" devices that can enable cyber-physical attacks. Most organizations, including health plans do not understand the data life cycle and how this data can be used against consumers. HIPAA requires technical safeguards be in place to protect against attacks, however, because Internet of Things devices are often held by third parties that are not covered under HIPAA, this data is often not adequately protected. There are billions of sensors that gather data on individuals each day. For example, an individual may have Alexa, a digital assistant device owned by Amazon. They may ask questions of this assistant about how to treat a common ailment. This data is collected and then sold by Amazon to various companies. Because Amazon does not fall within the covered entity definition protected by HIPAA, these inquiries are not protected under current legislation. Given the rapid advances in technology, regulation and security of these devices have not kept pace with the kinds of data collected by these devices. The data from these devices is often collected, stored, and transmitted using unsecure mechanisms, to mobile and cloud platforms.

C. Gap 3: Lifestyle Data

A third key gap revolves around lifestyle data collected by virtually every company that has a digital presence. Much of this data lives in the public domain and is being gathered by large data brokers who then sell the data and intelligence generated from that data to health plans who use it to make decisions about what services an individual may access or how much those services will cost. There are risks to incorporating lifestyle data into the pool of data health plans use to provide services to their members. Collection of lifestyle data presents security issues, privacy issues, regulatory concerns, lack of guidance on how to use this data, and a lack of infrastructure to address complaints related to the use of this data in decision making [40].

The third-party platforms used to gather this information have significant security issues. Data in this category is collected by companies like Google, Microsoft, Social media websites, and other corporate websites; data that is frequently compromised. For example, in March 2020 a breach at Microsoft impacted 30,000 organizations, LinkedIn had a breach that exposed personal data of 93% of all their members, a compromise at Yahoo in 2013 impacted three billion accounts, and the WannaCry virus impacted more than 400,000 servers worldwide at a cost of $4 billion and counting [41]. The only protection consumers in the United States have been the terms and conditions which most do not read and a disjointed collection of federal and state regulations and frameworks.

D. Recommendations

45 CFR § 160.103 states that HIPAA applies to healthcare information “created or received by a health care provider, health plan, employer, or healthcare clearinghouses” which puts these third-party devices outside the scope of HIPAA regulation. The current HIPAA security rule fails to

protect large swaths of patient information and should therefore be modified to provide stronger protections while maintaining the scalability and flexibility provided in the current legislation [26]. Data collected by wearable technology and Internet of Things devices pose significant privacy and security concerns. These devices are often developed by companies that do not fall within the purview of HIPAA, HITECH, or any other regulations. They generate large volumes of personally identifiable information that is captured within network servers of the companies that create the devices. Additionally, this data is priceless to marketers and is often repackaged and sold for use in behavioral advertising campaigns.

HIPAA and HITECH jointly provide administrative, physical, and technical controls; however, they do not provide information on how to differentiate between data in motion or data at rest, nor does it provide information on how and when the data should be encrypted, it merely notes the data should be encrypted [26]. Modifications should be made to HIPAA to mandate stronger technical safeguards by defining more stringent storage and backup requirements, provide stronger encryption requirements and limit who has access to encryption keys, provide differentiated requirements for data in transit and at rest, prohibit use of generic usernames, mandate access triggered breach notification. Further the definition of an entity in HIPAA should expand on the definition of business associate to manufacturers of devices and software that generate data captured by the healthcare industry.

One of the goals of HIPAA is to provide flexible and situational discretion for how covered entities implement security requirements. Sweeping modifications to HIPAA are not necessary, however, the rules should be modified to define how covered entities should comply with requirements to avoid situations where organizations use the vagueness of the rules to justify implementing minimal security measures to protect data held by healthcare entities [26]. This can be accomplished by modifying HIPAA regulations to adopt the NIST data security standards to ensure the organization is utilizing the most up-to-date security practices in the organization. Risk assessments should be performed in accordance with the NIST Risk Management Framework. The risk management framework overlaps HIPAA implementation standards, therefore, it would not be onerous to implement. One area where HIPAA could be expanded is to expand the definition of “covered entities” to include manufacturers of devices that collect, store, and transmit health related information. HIPAA should be modified to expand on the definition of “covered entity” to include third party companies that provide data for the purposes of providing precision medicine to patients. This expansion should include the collection and dissemination of shadow health records and the privacy standard should follow the record beyond the covered entity.

As noted earlier, the advancement of wearable technology is occurring much more rapidly than the legislative process. Establishing regulations or laws that govern this technology would be challenging as the laws would be dated at the time of implementation. In [16] the implementation of a cabinet level position to regulate Internet privacy and data security, which would include the data transmitted by Internet of Things devices and enable regulations to keep up with technology advancements, is recommended. Rather than creating a new agency, expanding on the capabilities of NIST to provide guidance on how this data can be securely stored and transmitted would be less onerous on the industry.

The FTC developed Fair Information Practices (FIPs) with the rise of computerized data systems. These FIPs were originally intended to govern data collected while performing commerce activities, however, these FIPs could be expanded to incorporate health care. In the FIPs, the Federal Trade Commission identifies five core principles when it comes to protecting privacy: notice or transparency, consent, access, redress, integrity, and security [42]. First, no data collection system should exist that is secret therefore consumers should be given notice when their data is collected to maintain transparency. Second, to ensure data collected for one purpose should not be used for another purpose the entity should require consent before collecting data on the consumer. Third, the consumer should have access to their data, meaning they should know what information is being collected and how that data will be used. Fourth, the consumer should have the right to correct inaccuracies in their data. Fifth, the consumer should be assured the organization collecting data maintains the integrity of the data and stores, transforms, and transmits data in a secure manner. A sixth item should be added to the FIPs, the right to refuse. The consumer should have the right to refuse the collection and/or sale of their personal data.

Health and Human Services should expand on the HHS Breach Database to incorporate categorization of risk groups into known, unknown, and semi-known; and identify common solutions to the threat based on the risk level for data generated by Internet of things and wearable devices that collect healthcare related data [26]. Additionally, manufacturers or business associates that hold data generated by Internet of Things devices, wearable devices, or lifestyle data that contributes to the data held in shadow health records should be required to report breaches to the HHS Breach Database.

NIST should be empowered to develop a framework for healthcare big data that defines the data environment in the healthcare industry. A framework for healthcare big data should include four distinct layers of protection: the health data sources layer, the big data technology layer, the big data analytics layer, and the application layer [43]. The health data sources layer should define the sources of data that are incorporated into an electronic health record and should include sources that make up the shadow health record. The second layer, the big data technology layer provides security recommendations for protecting the technology that is used to collect, transform, and transmit data through the healthcare ecosystems. Included in this area should include healthcare interoperability standards, internet of things devices, medical devices, and wearable technology specific to healthcare. The third area should provide guidelines for the big data analytics layer of the model which ties into the NIST Big Data Working Group efforts to define interoperability guides with the goal of developing a big data analytics guideline for healthcare specific data. The fourth layer considers the applications that will incorporate this data into the implementation aspects of healthcare. This is where you use the data to provide value-based care and medical research. The framework supporting healthcare big data should take the highly sensitive nature of the data into account while also considering the lifestyle and IoT data that may be incorporated to build shadow health records. The goal is to provide an industry specific set of guidelines for working with the data.

VI. CONCLUSION AND FUTURE WORK

The pervasiveness of lifestyle data and wearable devices are largely unregulated and as such are creating the perfect storm of privacy weaknesses where consumers are clueless about the volume and range of information collected, data is stored across a labyrinth of interconnected networks with limited security, and the market for the collection of this type of data is growing 40 to 60 percent per year [16]. Cybersecurity attacks are on the rise in all industries. A February 2021 survey of healthcare found 34% of healthcare organizations were hit by ransomware in the last year with 65% noting the attacks were successful [44]. From 2009 to 2021, cybersecurity attacks increased by 588% and impacted over 313 million individuals, with more than 80% of breaches occurring due to hacking or IT incidents at an average breach cost of $7.13 million, the highest cost across industries [5], [41]. Health and human services categories these attack vectors into six categories: unauthorized access and disclosure, theft, loss, improper disposal, hacking and IT incidents, and unknown. Most of these incidents can be traced back to human error.

Health care data is a high value asset as they generally store a broad range of data as part of an electronic health record. An electronic health record contains data elements such as protected health information, personally identifiable information, and summary health data, highly sensitive information that is highly regulated.

Health plans store medical, financial, and personal identification information for patients, financial and identifying information for health care providers and insurance agents including licensing information, and financial and identifying information for many businesses. These organizations also collected financial data as a general course of business, which is also covered by various regulations. Over the last decade, companies have expanded the collection of data into non-traditional areas such as the pooling of Internet of Things data, data from medical devices, and lifestyle data. These secondary resources used to create shadow health records are not traditionally protected by the laws and regulations that protect healthcare data.

Legislation and regulations covered under HIPAA, HITECH, ACA, CAA, ARRA, PCI DSS, healthcare interoperability standards, the NIST framework and CIS Critical Security Controls, FISMA, FTC guidance, and state regulations influence how health insurance companies secure their data. HIPAA identifies patient privacy rights and has significant penalties for data disclosure. This rule governs the use and standards of individual health information and controls how information is properly protected [45]. HITECH drove the shift from paper to electronic medical files and contains two sub-titles which directly influence cybersecurity standards. The ACA defined standardized communication standards for electronic records and the CAA contributed to a system for sharing threats with the federal government. ARRA developed a minimum necessary standard for disclosure under HIPAA, which replaced the prior open interpretation of the rule. PCI compliance governs how payment information is managed in financial transactions. Healthcare interoperability standards govern how data is collected and transmitted and are intended to provide a consistent format to enable electronic communications of data. NIST provides a variety of cybersecurity frameworks, while standard across the industry is not legally mandated, giving many companies loopholes around implementing data protection policies. Additionally, CIS Critical Security Controls provide a high-level framework, which again are not mandated for data protection. Finally, each state has differing levels of controls and protections for data that can be used to create a shadow-health record from alternate data sources. All these laws and regulations contain information use and security provisions that can impact the cybersecurity strategy for an organization.

There are three critical gaps in the protection of data in the health plan ecosystem. First there are several HIPAA weaknesses that need to be considered. HIPAA and the regulations that followed apply primarily to four entities: health care providers, health plans, business associates, and healthcare clearinghouses. The regulation is open to interpretation by each of these entities and weakened by broadly worded exceptions, and typically does not apply to third parties' collection of peripheral data. The second gap relates to wearable devices and data propagation from Internet of Things devices. These devices contain sensors that generate billions of data points per year, yet they have very little regulation around how this data can be collected, stored, transmitted, or sold. Additionally, the devices connect with software and mobile and cloud platforms using unsecure methods. The third main gap is around the collection of lifestyle data. Lifestyle data is collected from virtually every digital interaction in which an individual participates and is then packaged and sold to a variety of companies, including health plans. The data from wearable and Internet of things devices, and lifestyle data is typically not regulated and the people to whom the data relates are often unaware of what data is being collected and how that data is used by large companies.

Addressing these gaps will involve expanding the definition of a covered entity under HIPAA and providing strong language around the protection of data from primary and secondary sources when exchanged with health plans. HIPAA should mandate stronger technical, physical, and administrative safeguards and should expand on what data triggers a breach notification. Currently, the regulation only identifies eighteen key data points that trigger a breach notification, this should be expanded to include peripheral data that is collected by a health plan that can be traced back to an individual user. A goal of HIPAA and other regulations in this area was to provide flexibility and adaptability based on the environment in which a covered entity operates. This goal is a critical element of the regulation and can be achieved by incorporating mandates to use security frameworks to build out security protocols for protecting data. The law could do this by expanding on the capabilities of NIST to provide guidance for this industry by empowering NIST to develop a framework for healthcare big data which includes primary and secondary sources of data. When it comes to wearable and Internet of Things devices, the FTC Fair Information Practices could be expanded to include informed consent when data is collected using third party devices. Finally, expanding the requirement around breach reporting and penalties would introduce accountability into big data in healthcare.

Future work in this area should investigate three key areas: an in-depth investigation into lifestyle data and how it is used by health plans; an in-depth investigation into wearable devices and internet of things devices that generate shadow healthcare data and is collected by health plans; and how the data collected by health plans can be used to compromise organizations and individuals. The investigation into lifestyle data, and data generated by wearable devices and other internet of thing's devices should evaluate how the data is collected, disseminated, and utilized to circumvent laws and regulations governing healthcare data.

Further investigation into how this data is utilized by health plans and how it can be compromised by attackers should define how this data could be weaponized against companies and individuals. The goal of this research would be to provide further targeted recommendations for legal and regulatory improvements.

VII. REFERENCES

[1] Rothstein, M. A. (2021). Big Data, Surveillance Capitalism, and Precision Medicine: Challenges for Privacy. Journal of Law, Medicine & Ethics, 49(4), 666–676. https://doi-org.libproxy.boisestate.edu/10.1017/jme.2021.91.

[2] Research Brief. "Disinformation That Kills: The Expanding Battlefield of Digital Warfare." CB Insights. October 21, 2020. https://www.cbinsights.com/research/future-of-information-warfare/. [Accessed September 16, 2022].

[3] Price II, W. N., Kaminski, M. E., Minssen, T., & Spector-Bagdady, K. (2019). Shadow health records meet new data privacy laws. Science, 363(6426), 448–450. https://doi-org.libproxy.boisestate.edu/10.1126/science.aav5133

[4] F. T. Jaigirdar, C. Rudolph and C. Bain, "Risk and Compliance in IoT- Health Data Propagation: A Security-Aware Provenance based Approach," 2021 IEEE International Conference on Digital Health (ICDH), 2021, pp. 27-37, doi: 10.1109/ICDH52753.2021.00015.

[5] "Breach Portal: Notice to the Secretary of HSS Breach of Unsecured Protected Health Information," U.S. Department of Health and Human Services Offices of Civil Rights. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf. [Accessed September 12, 2022].

[6] Jessica Davis. "7 Largest Data Breaches of 2015", Health Care IT News. December 11, 2015. https://www.healthcareitnews.com/news/7-largest-data-breaches-2015. [Accessed September 18, 2022].

[7] Heather Landi. "Anthem to Pay $39M to State AGs to settle landmark 2015 data breach", Fierce Healthcare. September 30, 2020. https://www.fiercehealthcare.com/tech/anthem-to-pay-39m-to-state-ags-to-settle-landmark-2015-data-breach. [Accessed September 18, 2022].

[8] K. Robertson. "Health Net reports security breach of up to 1.9 million members." Sacramento Business Journal. March 14, 2011. https://www.bizjournals.com/sacramento/news/2011/03/14/health-net-reports-security-breach.html. [Accessed October 12, 2022].

[9] Katherine Keisler-Starkey & Lisa N. Bunch. "Health Insurance Coverage in the United States: 2019", United States Census Bureau. September 15, 2020. https://www.census.gov/library/publications/2020/demo/p60-271.html. [Accessed October 18, 2022].

[10] "Adopting the NIST Cybersecurity Framework in Healthcare." White Paper. Symantec. Mountain View, CA. 2018. https://docs.broadcom.com/doc/adoping-the-nist-cybersecurity-framework-in-healthcare-en. [Accessed September 16, 2022].

[11] Derek Mohammed. "US healthcare industry: Cybersecurity regulatory and compliance issues." Journal of Research in Business, Economics and Management 9, no. 5 (2017): 1771-1776.

[12] Title 45 Public Welfare. Code of Federal Regulations. (2022). https://www.ecfr.gov/current/title-45. [Accessed October 12, 2022].

[13] T. Taylor. "Hackers, Breaches, and the Value of Healthcare Data." SecureLink. June 30, 2021. https://www.securelink.com/blog/healthcare-data-new-prize-hackers/. [Accessed October 12, 2022].

[14] "guidance on the Protection of Personal Identifiable Information." U.S. Department of Labor. https://www.dol.gov/general/ppii. [Accessed September 15, 2022].

[15] J. Mnjama, G. Foster and B. Irwin, "A privacy and security threat assessment framework for consumer health wearables," 2017 Information Security for South Africa (ISSA), 2017, pp. 66-73, doi: 10.1109/ISSA.2017.8251776.

[16] Arnow, G. (2017). Apple Watch-Ing You: Why Wearable Technology Should Be Federally Regulated. Loyola of Los Angeles Law Review, 49(3), 607–633.

[17] Infographic. “The Power of Consumer and Lifestyle Data in Healthcare.” Pulse Survey. Acxiom, LLC. 2018. https://www.acxiom.com/resources/infographic-the-power-of-consumer-and-lifestyle-data-in-healthcare/. [Accessed September 15, 2022].

[18] Healthy People 2030. “Social Determinants of Health.” U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Retrieved from https://health.gov/healthypeople/priority-areas/social-determinants-health.

[19] “Social Determinants of Health.” World Health Organization. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1. [Accessed September 15, 2022].

[20] M. Rezaeiahari. “Moving Beyond Simple Risk Prediction: Segmenting Patient Populations Using Consumer Data.” Frontiers in Public Health. July 15, 2021. https://doi.org/10.3389/fpubh.2021.716754.

[21]. “Expanding your reach through consumer marketing.” Optum Case Study. 2021. https://cdn-aem.optum.com/content/dam/optum3/optum/en/resources/case-studies/cas-expanding-your-reach-through-consumer-marketing-case-study.pdf. [Accessed September 15, 2022].

[22] Abstract. “COVID-19 Area Vulnerability Index.” OptumServe. WF2616134. 2020. https://cdn-aem.optum.com/content/dam/optum4/resources/pdf/sub-folder/wf2616134-30.7-ohb-vulnerability-index.pdf. [Accessed September 15, 2022].

[23] M. Allen. “Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates.” ProPublica. July 17, 2018. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates. [Accessed September 14, 2022].

[24] “Security and Privacy Module.” HL7 FHIR. Release 4B. http://www.hl7.org/fhir/secpriv-module.html. [Accessed October 12, 2022].

[25] W. R. Braithwaite, "The federal role in setting standards for the exchange of health information," Proceedings Pacific Medical Technology Symposium-PACMEDTek. Transcending Time, Distance and Structural Barriers (Cat. No.98EX211), 1998, pp. 340-343, doi: 10.1109/PACMED.1998.769952.

[26] Krisby, R. M. (2018). Health Care Held Ransom: Modifications to Data Breach Security & the Future of Health Care Privacy Protection. Health Matrix: Journal of Law-Medicine, 28, 365–401.

[27] “HIPAA for Professionals.” Health and Human Services. https://www.hhs.gov/hipaa/for-professionals/index.html. [Accessed September 16, 2022].

[28] “HIPAA Administrative Simplification Regulation Text”, U.S. Department of Health and Human Services Office for Civil Rights, 45 CFR Parts 160, 162, and 164. Unofficial Version, as amended through March 26, 2013. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf. [Accessed October 8, 2022].

[29] Clemens Scott Kruse, Benjamin Frederick, Taylor Jacobson, D. Kyle Monticone. Cybersecurity in healthcare: A systematic review of modern threats and trends. Technol Health Care. 2017;25(1):1-10. doi: 10.3233/THC-161263. PMID: 27689562.

[30] Oatway, D. (2004). HIPAA Security Is Next. Nursing Homes: Long Term Care Management, 53(1), 37–40.

[31] “What is the HITECH Act?”, HIPAA Journal. https://www.hipaajournal.com/what-is-the-hitech-act/. [Accessed October 18, 2022].

[32] “How do HIPAA, NIST, and HITRUST CSF Work Together?”, 360 Advanced. July 13, 2021. https://360advanced.com/how-do-hipaa-nist-and-hitrust-csf-work-together/. [Accessed October 18, 2022].

[33] Jay G. Ronquillo, J. Erik Winterholler, Kamil Cwikla, Raphael Szymanski, and Christopher Levy. "Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information." JAMIA open 1, no. 1 (2018): 15-19.

[34] Text - H.R.1 - 111th Congress (2009-2010): American Recovery and Reinvestment Act of 2009. (2009, February 17). http://www.congress.gov/bill/111th-congress/house-bill/1/text.

[35] “PCI DSS Quick Reference Guide”, Security Standards Council. Version 3.2.1. 2018. https://www.pcisecuritystandards.org/documents/PCI_DSS-QRG-v3_2_1.pdf?agreement=true&time=1637810302162. [Accessed October 5, 2022].

[36] “Federal Information Security Modernization Act (FISMA) Background.” National Institute of Standards and Technology. https://csrc.nist.gov/projects/risk-management/fisma-background. [Accessed September 4, 2022].

[37] Nate Lord. “What is FISMA Compliance? 2019 FISMA Definition, Requirements, Penalties, and More.” Data Insider. December 1, 2020. https://digitalguardian.com/blog/what-fisma-compliance-fisma-definition-requirements-penalties-and-more. [Accessed September 4, 2022].

[38] Alexander Gillis. “Federal Information Security Management Act (FISMA).” Tech Target. September 2020. https://www.techtarget.com/searchsecurity/definition/Federal-Information-Security-Management-Act. [Accessed September 4, 2022].

[39] D. -Y. Kim, L. Elluri and K. P. Joshi, "Trusted Compliance Enforcement Framework for Sharing Health Big Data," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 4715-4724, doi: 10.1109/BigData52589.2021.9671834.

[40] White Paper. “Digital marketing optimization: Strategies for engaging health care consumers online.” Optum Health. WF130301, 2016. https://www.optum.com/content/dam/optum3/optum/en/resources/white-papers/WF130301_Digital_Marketing_Strategy_White_Paper.pdf. [Accessed September 12, 2022].

[41] R. Sobers. “166 Cybersecurity Statistics and Trends [updated 2022].” Varonis. July 8, 2022. https://www.varonis.com/blog/cybersecurity-statistics. [Accessed October 20, 2022].

[42] Landau, S. (2015). Control use of data to protect privacy. Science, 347(6221), 504–506. https://doi-org.libproxy.boisestate.edu/10.1126/science.aaa4961.

[43] M. Sheeran and R. Steele, "A framework for big data technology in health and healthcare," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017, pp. 401-407, doi: 10.1109/UEMCON.2017.8249095.

[44] “Ransomware Trends 2021”, HHS Cybersecurity Program, Office of Information Security, ID 202106031300. https://www.hhs.gov/sites/default/files/ransomware-trends-2021.pdf. [Accessed October 8, 2022].

[45] “Summary of the HIPAA Privacy Rule”, Health and Human Services. July 26, 2013. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html. [Accessed October 18, 2022].