THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF RISK MANAGEMENT

RECONCILING GENOMICS ADVANCEMENT WITH RESPONSIBILITY TO HEALTH CARE STAKEHOLDERS THROUGH GOVERNANCE STANDARDS

ALLISON GADDIS
SPRING 2019

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Risk Management
with interdisciplinary honors in Risk Management and Management

Reviewed and approved* by the following:

Forrest Briscoe
Professor of Management & Organization
Thesis Supervisor

David Cather
Clinical Professor of Risk Management
Honors Adviser

Srikanth Paruchuri
Associate Professor of Management & Organization
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The genomics practice is positively and rapidly transforming the health industry through human genome sequencing and individualized medical care, which is denoted as precision medicine. However, the large scale proliferation of human DNA data in genomic databases is raising new risks to involved individuals and their personal data. As precision medicine becomes more widespread, improved regulatory and governance requirements at health care organizations become all the more important in order to responsibly maintain the benefits and innovation derived from genomics. Health care organizations must adjust genomic database governance policies to enhance data responsibility features of data privacy, security, use, and accuracy without impeding the continuation of precision medicine. Through a comparative analysis of societal attitudes and a current genomic database governance structure used in practice, a balance can be found between genomics advancement and responsible organizational data management practices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Forrest Briscoe, for sparking my interest in such an interesting and complex topic. My thesis would not have been possible without his guidance, encouragement, and constant support. My family and friends have offered me invaluable support throughout this process as well. Lastly, I would like to recognize my honors advisors for graciously taking the time to read and review my thesis.

Additionally, the research conducted for this thesis was financially supported by the Smeal College of Business, the Farrell Research Center, and the Rock Ethics Institute.

Chapter 1 Introduction

The genomics field encompasses a complex array of applications, but this thesis focuses primarily on the use of databases containing large amounts of human DNA data and how they enable innovative medical practices and discoveries. The scope of this topic addresses a notable tension in the health industry regarding the governance structure for genomic databases. Doctors and researchers rely on broad data sharing and access in order to innovate the medical field through genomic analysis. However, widescale genomic analysis often results in inappropriate use of data and a variety of risks to data providers. The needs of doctors and genomic researchers and the responsible practices for genomic data privacy, security, use, and accuracy are not currently aligned. It is extremely important to consider public opinion on data management at an organizational level because in order for genomics to advance, people must be willing to contribute their genomic data to health care organizations. The data responsibility framework introduced in this thesis is used to cross reference public attitudes on genomic database governance with the policies used in practice at a health organization's genomics research program. This comparative analysis provides a suggestion for how health care organizations can adjust governance policies and practices in order to achieve both responsible data management and the benefits derived from precision medicine.

This chapter gives an overview of the genomics field, with particular emphasis on data accumulation and genomic databases. It includes a synopsis of how DNA data is sequenced and how genomic databases are utilized in health care organizations. Chapter 2 discusses several risks to genomic data providers, and the related governance structure that currently regulates

these risks in health care. Chapter 3 describes the primary stakeholder groups associated with the genomics practice in health care. It introduces a data responsibility framework for organizations to consider when developing governance standards with responsibility to health care stakeholders. Chapter 4 analyzes the findings from the Genomic Data Governance Survey that pertain to public perception of genomics governance policies in the health industry. Chapter 5 compares survey results to governance policies used at a real health care organization. Chapter 6 details one solution to how health care organizations can use responsible governance to balance proper stakeholder data management with medical innovation made possible by genomics.

Genomics Overview

The human genome contains each individual's unique DNA and genetic information. The genomics practice incorporates a range of functions regarding the human genome. Genomics science uses molecular biology to analyze genome structure and development. Pertinent applications of genomics can include gene editing, gene drives, and the large scale proliferation of human DNA data in genomic databases. As the genomics field continues to become more advanced and common, it is causing a multitude of industry disruption. Genomic data utilization has an impact on hospitals, healthcare organizations, biotechnology and life sciences organizations, academic research teams, commercial activity organizations, employers, for-profit research institutions, and many other organizations [1]. Genomics has both a wide ranging and extremely significant impact because genomic data is being sequenced at a rapid rate. By 2025, it is projected that the amount of genomic data generated in the world will match that of the social

media data generated in the world [2]. The advancement of genomics and its capabilities has a major impact on all industries, but particularly the health services industry.

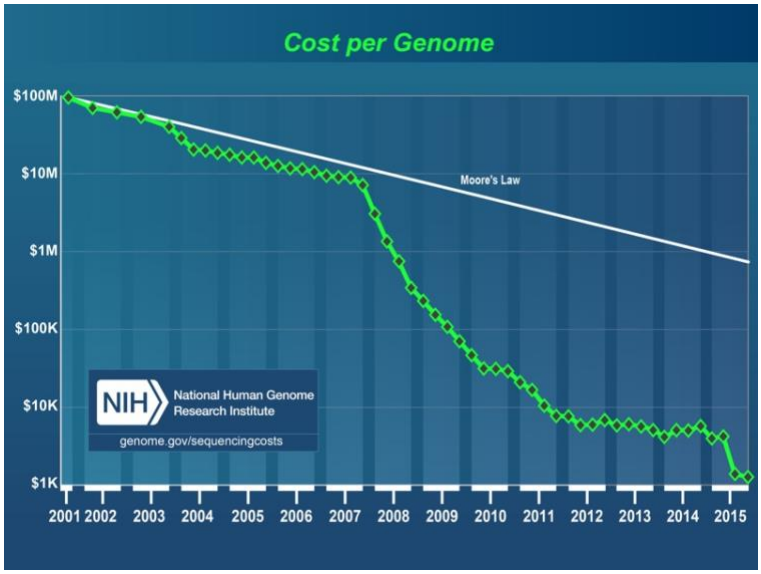
Genome sequencing is a phenomenon that is becoming more prevalent in today's health care services. Medical genomics is the practice of using genomic science to transform current health care practices by analyzing the genome along with a holistic view of the patient's health information. Current health services are beginning to incorporate genomics in order to integrate and interpret an individual's genome sequence when compared to all other DNA data available in a given database. Genomic data can reveal highly comprehensive details about an individual's health, and a comparative analysis against others using big data is what makes clinical conclusions more accurate and precise [3]. Precision medicine is the practice of combining human DNA data analysis with medical recommendations to develop a more specific and personalized diagnosis [3]. Genomic databases facilitate precision medicine because they support genomic data comparison. Precision medicine can be referred to as personal medicine or individualized medicine as well.

This field and genomic databases are enabling a significant amount of positive innovation in health care and health research [4]. These databases are causing more individualized treatment, through precision medicine, in lifesaving ways. Widespread genomic data generated across patients is transforming disease prevention and diagnosis. Quite predominantly, genomic databases are driven by individual genomic data that was sequenced within the health care context [5]. Therefore, the health industry is an important setting to consider.

At present, genomic science is so niche that only a few groups of individuals can interpret such data and the actual sequencing process takes a large amount of time and money [1]. However, current research is rapidly expanding the genomics field. Digital capabilities, machine

learning, and general research are improving the quality and efficiency of genome sequencing and genome analysis. According to the National Human Genome Research Institute (NHGRI), the underlying costs of genome sequencing have dramatically decreased since 2001. Figure 1 below illustrates this decline through cost data at genome sequencing centers funded by the NHGRI [6]. This average cost is expected to continue decreasing as technology and research capabilities enhance the genomics field.

Figure 1. Genome Sequencing Cost Decline



As genome sequencing technology continues to advance in terms of timeliness and affordability [7], sequencing will become more commonly available. The concept of precision medicine broadly encompasses a range of positive applications for health care services and research through human genomic databases. Genomics is already impacting mainstream health care, including a variety of specialties such as dermatology, cardiology, pediatrics, obstetrics, and oncology [8]. DNA data comparison through genomic databases is a critical component to genomic interpretation for precision medicine. Therefore DNA databases are essential in continuing the positive medical innovation derived through genomics advancement.

Genomic database construction and utilization is enabling the possibility of medical innovation, but the storage of individuals' data may have a variety of associated risks for the involved stakeholder groups [4]. Within the health industry, stakeholders include patients, doctors, health researchers, hospital and health care executives, funding communities, the government, society overall, and many more. A patient's biological relatives may be linked to that patient's genomic data without their knowledge [8], which results in a stakeholder group who may be uninformed of their connection. Even for stakeholders who know their personal data is being used, it may not be clear to them exactly how their data will be used, stored, or shared.

There is a lack of strict regulation and governance for genomic databases, despite genomic sequencing becoming more mainstream. Governance policies and practices entail the established organizational structure for genomic data accumulation, usage, sharing, and storage. Health care organizations generally have limited governance standards in place to responsibly manage stakeholder data throughout a variety of possible uses. One governance policy example is that the National Institutes of Health (NIH) requires researchers with NIH funding to acquire express consent from research subjects in order to analyze their genomic data [9]. Health organizations tend to have governance standards of their own for patients, but they can be ambiguous. Generally, broad data sharing within genomic databases currently causes uncertainties such as patient privacy concerns and risks in data security, usage, and accuracy. Genome sequencing is becoming more prevalent in today's society, and enhanced governance practices and regulation will help mitigate stakeholder risk and uncertainty in data usage.

Genomics has a potential impact on all individuals, so it is important to consider societal attitudes and public perception when determining data governance standards that mitigate risk. Genomic capabilities will continue to positively innovate the medical field, which will cause this

technology to become exceedingly more prevalent in the near future. Health care organizations are a very common steward of data accumulated for genomic databases. This arguably creates a sense of urgency for the health industry to adapt organizational governance policies now. Health care organizations must find a balance between allowing medical innovation through the advancement of genomic science and genomic databases, while also maintaining proper responsibility towards health care stakeholder data through governance standards. Research conducted with the Briscoe team suggests that policies should keep patient genomic data usage close to those directly involved. This would be in favor of transparent data control by patients and closely tied researchers rather than broad data sharing between varying organization types and functions. Ideally, proper governance standards will cause mutual beneficence between health care organizations and stakeholders contributing DNA data. By preserving public trust through responsible data management, individuals will be more likely to contribute their personal data to genomic databases.

Genomic Science and Data Generation

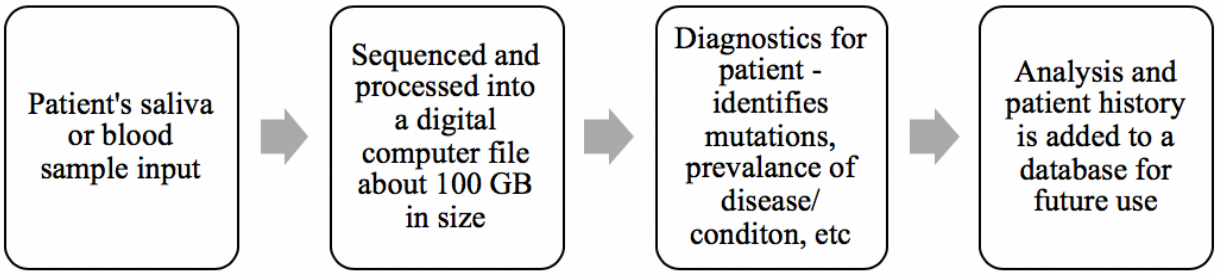
Every human has a unique, biochemical code that is defined by his or her genetic information. An individual's DNA, an acronym for deoxyribonucleic acid, molecules correspond to a biochemical code that forms a sequence called a genome. DNA holds the information necessary for human development and the functioning of cells essential for life, and human cells contain approximately 6 billion base pairs of DNA [3]. All DNA base pairs are contained in the genome, which also includes approximately 20,000 genes that code for proteins intended to complete many biological functions [3]. These proteins determine how we digest our food, store

energy, generate new cells, etc. Overall, information held in the genome controls the expression of physical traits, personal behavior, reactions to the environment, responses to certain foods, and the development of complex diseases. Information held in the genome is synonymously referred to as genomic data or DNA data.

Genomes differ by person based on deviations in DNA sequences called "variants" among genes. Decoding a genome sequence includes a process of identifying variants when compared against a reference genome or other genomic data found in a database. Generally, DNA data from two unrelated individuals will only differ 0.1% as an approximate average [3]. This percentage still covers a significant amount of DNA data, especially considering human DNA contains 6 billion base pair sequences. Currently, these methods have a less than 1% error rate [3]. As more genomes are sequenced and entered into large genomic databases, the accuracy and ability of genome interpretation can be improved.

Genome sequencing is a specific and individualized science aimed at collecting human DNA data. The genomic data industry is largely shifting from life science research to both research and clinical diagnoses. Figure 2 below details the general process steps for getting to the clinical diagnostics phase.

Figure 2. Genome Sequencing Process Overview



The first step includes procuring a saliva or blood sample, referred to as a "bio sample" [1,3]. About 100 GB of data is pulled from the bio sample, which is all that is needed to

conceptually reconstruct the individual's genome [2]. This is in raw data form before analysis. The 100GB of data is then analyzed by trained bioinformatics specialists to develop appropriate conclusions and recommendations. Typically, the patient's medical and genetic history are entered into a genomic database to be used with other genomic data for future clinical recommendations [1,3]. The specific patient's genomic data is analyzed to identify genetic variants (mutations) that are based on comparisons across many other people's DNA. This facilitates clinical interpretation of the patient's DNA when compared to the DNA database available [1,3,4].

Genomic databases are repositories of human DNA data, varying in size from small (hundreds of people's genomes) to large (millions of people's genomes). These databases can have a variety of functions and methods for data accumulation [3,4,5]. For example, genome sequencing of individual human DNA data is creating databases for precision medicine to become more widely available. Precision medicine uses widespread DNA data as a database for clinicians to more specifically and accurately make medical recommendations based on an individual's DNA data compared to all DNA data in the database. Data collection and database usage extends beyond the health industry. To accumulate and use massive amounts of genomic information, data sharing can occur between clinical care organizations, research groups, the government, commercial firms, direct-to-consumer (DTC) testing companies, and a variety of other nonmedical organizations who are sequencing human genomes [5]. It is common for limited regulation among these data sharing relationships to result in ease of data transfer beyond original intended use. One way that data sharing is allowed to occur is if a patient's data is simply "de-identified," stripped of identification, and then added to a larger genomic database [10]. Large genomic databases tend to rely on future reuse and analysis of an individual's data,

but data security and individual privacy can often be jeopardized by this further disclosure of information. These databases will continue to have a wide ranging impact as the genomics field continues to progress.

Genomic sequencing devices are highly advanced, but there is room for improvement. Typically, the sequencing lab will be separate from the clinical organization [1,7]. Currently, only bioinformatic specialists can interpret genomic data once it has been sequenced, and this is after very extensive training [7]. Bioinformatic specialists are the scientists working with genome sequences in a lab. The process for analyzing a patient's genome for genetic variants against a reference genome or variant database is automated, but bioinformaticians typically will still manually examine the sequence using a gene browser. Right now, it can take a geneticist up to 15 hours to interpret the genome sequence of one patient [1,7]. However, this may change as genomic sequencing becomes more mainstream and technological capabilities become more advanced. Several companies and research organizations are currently taking steps to innovate the genomics practice. This phenomenon will cause genomic databases to grow as well.

Illumina is an existing company that sequences about 90% of the world's genomic data on their machines [2]. Illumina already uses several digital technology strategies, such as the internet of things and cloud-based platforms. These digital capabilities have enabled Illumina to become a global leader in this practice. For example, Illumina is using Amazon Web Services as a cloud based platform for data storage in order to reduce algorithms and implement hardware acceleration for genomic databases [2]. Amazon Web Services offers a cost advantage due to their massive scalability, and their platform has a global presence already. This mitigates the risk of cross-border genomic data sensitivity because they are equipped to handle such sensitivity already.

Amazon's internet of things strategy consists of a system called BaseSpace, in which these machines will upload the internal diagnostic information from a sequence directly into their cloud based platform [2]. Illumina's sequencing devices are uploading approximately 270 billion data points every year into the BaseSpace database [2]. This system is greatly improving Illumina's inadequate state of data storage centers, which will help enable Illumina to continue to bring positive innovation in this field. Sanjay Chikarmane, the Senior Vice President at Illumina, was interviewed by Digital Health News about the future of genomics. He believes that genome sequencing will become more mainstream in the future, but in order to do so companies must "be able to analyze the data to identify mutations that can be treated, and identify the most effective drugs that can be used to treat patients. And this has to happen in the electronic medical record" [7]. In order for precision medicine to become mainstream and move into standardized health care practices, these databases are extremely important for data integration and analysis.

An interesting case study to note is the Illumina and Genomics England partnership in a 100k genome project that was developed by the UK government [7]. The UK has funded approximately 11 genome centers. Illumina has started an Artificial Intelligence Genomics initiative in response to the 100k genome project, and they are working with Philips and IBM's Watson Health to evolve the AI platform. By sequencing at a massive scale with AI, a genome can be sequenced in less than 24 hours and at a price less than $1,000 [7]. With continued research, their goal is to be able to analyze genomic data with a $100 genome. Frequent genome sequencing and integrated genomic application in many industries will likely become a reality in the near future as technology continues to innovate the field.

Innovation Through Precision Medicine

No one can be quite certain of what the future of precision medicine will be, but advocates think it will offer extremely individualized and more thorough medical care on a per patient basis [4,5]. Using each patient's individual genetic information, behavior, and holistic health information will improve accuracy of clinical care. This will standardize entirely new methods for health services via digitization. The Precision Medicine Initiative, a research initiative involving the National Institutes of Health, defines precision medicine as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" [11]. The increased frequency of human DNA data sequencing, due to cost mitigation and increased technological capability, is causing a rapid proliferation of genomic databases. This allows for precision medicine to become more widely available.

The health industry will benefit from genomic databases with the advancement of digital technology as well. Today's society is moving towards digital technologies, like the internet of things, cloud based platforms, and artificial intelligence, so it is advantageous for genomic technology advancement to be partnering with these capabilities. As sequencing cost and time becomes less of a barrier to genomic science, more accessible and affordable precision medicine will be possible [7]. Genomics application will eventually become a standardized practice in health care. It is estimated that within ten years more than one billion humans will have their DNA sequenced, and genomic technology will be incorporated into a standardized health routine [1]. As more patients have their genomes sequenced, the interpretations of genomic data can become more individualized and accurate through genomic databases. Therefore, a lot of positive

and widespread innovation will enter the health industry as genomics becomes more mainstream within clinical care.

Although precision medicine is a developing approach, it is already transforming modern day medical care. For example, a physician would analyze a patient's genome after the patient presented symptoms of diabetes. If the patient's DNA showed signs of mutations known to render normal diabetes drugs ineffective, based on data in genomic databases, then the physician would know to prescribe a different medication [1,12]. Additionally, a physician examining the genomic data of a patient with chest pains would be able to pinpoint genetic heart-rhythm mutations that would require an alteration in the critical care delivery type [1,12]. Genomic databases are enabling several medically innovative functions through large scale data analysis.

Genomic database utilization will allow a more conceptual and in depth understanding of complex diseases from the analysis of infectious diseases and inherited genetic disorders across large amounts of human DNA data [4]. Scientists and clinicians can develop specific conclusions for disease origins and mechanisms, which in turn can uncover new opportunities for disease prevention and treatment. Genomics is supporting a new molecular methodology of cancer diagnosis, research, screening methods, and treatment as well. For example, patients can be identified more accurately as having a genetic risk for an aggressive form of prostate cancer before traditional screening methods could pinpoint such a risk [1,13]. Another component of precision medicine is using gene editing and gene therapy to prevent and cure diseases based on particular genetic mutations. These methods were used on patients to prevent hemophilia and non-Hodgkins lymphoma in 2016 [1,14]. New discoveries related to complex diseases are continually being made through genomics and precision medicine.

Pharmacogenomics is another generally beneficial factor of precision medicine. On an individual scale, a patient's genome sequence can indicate the most effective drug for that specific individual to use in any necessary health care situation [4]. This permits more efficient treatment decisions, and prevents the prescription of drugs that may have minimal effect. On a larger scale, improved adjustments can be made to overall treatment standards using consistent DNA variants found in broader patient populations. Categories for broader patient populations can include groups of certain demographics, genetic risk factors, phenotypes, and any other relevant health based grouping [9]. These type variations can determine the effectiveness or ineffectiveness of drugs for large patient communities, minimize negative side effects, and possibly develop new drug targets for pharmaceutical companies to develop.

These functions are possible in today's society because of genomics and genomic database usage. As this field advances, the health industry will continue to be positively transformed by the targeted therapies and custom drug prescriptions. With the rise of any new science or technology, several risk based barriers arise as well. It is important to be cognizant of these risks and to handle them responsibly, while maintaining the innovation that comes from genomic data usage for precision medicine. The next chapter discusses current regulation relating to the genomics practice and the relevant risks associated to stakeholders.

Chapter 2 Risk and Regulation of Genomic Databases

Data-Related Externalities

At present, we are living in a new world of data and digital records. This creates challenges for organizations to maintain their stakeholder data. Enhanced genomics capabilities, like sequencing timeliness and cost effectiveness, are causing an increased prevalence of genomics in all industries. This is quickly fueling data generation for genomic databases. Genomic databases are growing as more people have their genomes sequenced for various medical or nonmedical reasons. In 2013, the genomic sector was estimated to have a $25 billion direct contribution to the economy, and a $40 billion indirect contribution to the economy [1,15]. It is expected that this impact will continue growing, and many believe genomics has the potential for global scalability in the near future. This causes risk within the health industry because genomic data contains very personal health information with seemingly unlimited application and use. Even the benefits of precision medicine have uncertainty because the future of genomics is unknown. Genomic database governance is vague, given the wide range of database functions, so many people do not understand how their genomic data is being used.

Many data-related externalities, both positive and negative, exist surrounding the enhancement and escalation of genomics and genomic data storage. Externalities may conceptually include any individual's cost or benefit to the storage of their genomic data due to the current governance structures in place for data protection. Positive data externalities include the range of improvements made through genomic databases and precision medicine. Health information extracted for genomic analysis is extremely personal, and negative data externalities align with the associated risk of such personal information being available. For example within

the health industry, unauthorized parties may gain access to patient health information or data may be used for a purpose beyond knowledge of the data holder. Data use and security in future situations is unknown.

Ultimately, the relative newness of genomic databases creates a need for organizational governance adaptation to combat uncertainties of the genomics field. Several risks to patients and their personal DNA data arise at the expense of innovative precision medicine. Patient privacy and data security risks are the main drawbacks to genomics in health care, and limited regulatory frameworks perpetuate these risks.

Quite often, a patient's genomic data generated in a medical context is used beyond its original purpose. This may cause infringement on patient privacy. The intent of precision medicine is for clinicians to thoroughly and accurately make medical recommendations and to discover new variants and genetic implications. To do so, an analysis of all DNA data in a genomic database is necessary. Due to broad consent practices, patients are not typically informed of the different ways their data could be used and transferred. The risk culminates in that data could be used in a way the patient is not comfortable with, and the patient may have no knowledge of this use. Genomic databases result in data sharing between clinical care organizations, research groups, and commercial firms. There are many partnerships in place where other organizations can use a patient's genomic data beyond the original intended purpose of such data. A genome sequenced during routine patient care may later be used for academic research or a pharmaceutical company developing a new pharmacogenomics drug [1, 16]. Additionally, if a research study involving genomic data was federally funded, then the National Institutes of Health database is allowed access to the patient's sequenced data [17]. Personal

health information may be disclosed to an external party through the reidentification of genomic data, thus breaching stakeholder privacy.

Not only do health organizations store individual DNA data, but electronic health records containing extended patient health information can be linked in genomic databases as well. This continued extension into personal medical history can cause negative externalities regarding individual privacy due to potential discrimination effects. For example, an individual may be denied health insurance coverage because of a DNA marker that raises the likelihood of developing a particular type of cancer [17]. Any unfair treatment based on genetic differences may result in genetic discrimination.

Patient data security is another risk, and it involves the method by which genomic databases are managed. Large genomic databases revolve around future reuse and analysis of a patient's data, but data security can often be jeopardized by further disclosure of information. Due to HIPAA laws, health care organizations are generally very secure. However, a 2016 Ponemon survey found that 90% of health care organizations surveyed had encountered a data breach within the past two years [1,18]. Of that 90%, approximately 64% of breaches included patient medical records being leaked, which resulted in unauthorized disclosure. Therefore, there is significant risk regarding the vulnerability of patient genetic and medical history due to the current state of data security among health care organizations. Not all genomic data is even classified under HIPAA security. In those cases, there isn't a strict guarantee that all health organizations will prioritize data safety for that type of genomic data.

The next section discusses current regulation of genomic data, and how it is meant to address externalities associated with the exchange of genomic data.

Current Regulatory Guidelines Overview

Genome sequencing and database usage has become more prevalent over the past decade, which has raised the frequency and significance of data related risk to stakeholders. This in turn justifies a need for legal governance reform that will allow for the advancement of genomics, while maintaining responsible practices for stakeholders involved. Limited governance standards for genomic data use and storage only partly address the associated negative externalities. In order to mitigate these negative externalities, an analysis of current regulation is helpful. This section will provide an overview of current regulatory structure in place for genomic data sequencing, storage, and transfer.

The general view within the United States is that data belongs to the holder, and not the individual the data represents [19]. This conceptual ideology varies by industry, where personal data falls into separate categories corresponding to particular regulation. Much of data regulation within the health industry aligns within the Health Insurance Portability and Accountability Act (HIPAA). This act pertains to any organization or group that handles health transactions electronically, and it institutes a protection standard for individuals' personal health information and medical records. Patient health information and electronic health records are secured through HIPAA. Strict guidelines on information storage, usage, and sharing are described in HIPAA. Despite a close association of genomic data and health care, there is no set regulation through HIPAA ensuring genomic data protection. HIPAA only protects genomic data if the patient's medical records are linked to it [1,10]. Therefore, governance structure is largely dependent on the organization maintaining the genomic database with a few guiding factors.

While HIPAA attempts to regulate most of the data used in health care, the Common Rule attempts to regulate data usage in research. Both HIPAA and the Common Rule are within

the scope of the United States, and they may have overlap because health care data is often times used in research projects. The Common Rule is a colloquial name for the Federal Policy for the Protection of Human Subjects. It enforces basic guidelines for Institutional Review Board (IRB) approvals and individual informed consent, and it ensures that human subject research is in compliance with all participating federal departments as well [20]. The Common Rule applies to human research that is government sponsored or federally funded. Its jurisdiction does not include institutions that are not conducting federally funded human research. This could be community hospitals, health care organizations new to clinical research, and commercial genomics companies that do not do government sponsored research. These organizations types may be engaging with genomic databases with a lack of experience with IRB governance.

Privacy externalities involve governance structure for consent as well as individual control over data. A patient must sign a form indicating consent to genome sequencing before genomic data can be collected [1,21]. The consent process indicates a legal approval for the data to be sequenced and further analyzed, but it is not required to detail specific uses. A 2017 update to the Common Rule now allows concept of "broad consent" [21], which is commonly obtained. This is where the patient is informed of the intended current use for genomic data, but under the pretense that it may be generally used for other research purposes in the future. This is permissible because DNA data is typically considered property instead of something to be managed by patient privacy rights. Therefore, organizations are allowed to hold and transfer genomic data for future use if a patient provides broad consent.

The Genetic Information Nondiscrimination Act (GINA) is a United States federal anti-discrimination statute that is relevant to DNA data access and evaluation within genomics field. It aims to allow the public to utilize genetic testing, research, therapies, and technology by

providing protections from discrimination due to genetic information. Genetic information can be defined by genetic test results and family medical history. GINA prevents genetic discrimination within employment and health insurance. However, GINA protections do not apply to life insurance, disability insurance, long-term care insurance, employers with fewer than 15 employees, the U.S. military, and many other organizations [22]. With these exemptions, organizations must address negative data externalities in other ways. Even though HIPAA, the Common Rule, and GINA have limitations within the United States, there is not a regulatory framework internationally that addresses all data concerns.

The European Union (EU) uses the General Data Protection Regulation (GDPR) as a framework for data protection, use, transfer, and privacy for its citizens. GDPR law is considerably more restrictive than U.S. data protection regulation, where individuals in the EU maintain the right to know what data companies have about them [19]. Article 9 in GDPR law specifies processing requirements for many types of personal data categories [23]. However, there is a research exemption within Article 9 regarding genomic data storage and transfer potential. The exemption notes that personal genetic data can be handled without following the overall consent requirements of Article 9 [23]. This justifies relatively easy data transfer to EU based companies for research purposes, mergers and acquisitions, sharing partnerships, etc. There is a lot of variation in regulation among other countries as well. It is best to start with policy adjustment within the United States to combat this variability by safeguarding genomic data wherever it originates.

Another way that data sharing is allowed to occur is if a patient's data is "de-identified," stripped of personal identifiers, and then added to a larger genomic database. This is an alternative method, other than data consent and data security measures, to reduce negative data

externalities. Personally identifiable information (PII) is unique evidence that can link data back to the data holder. According to HIPAA, there are specific elements that are considered PII but there is not a completely exhaustive list [10]. The list includes factors such as, name, date of birth, place of birth, abbreviated social security numbers, etc. HIPAA jurisdiction does not cover genomic data that holds no personally identifiable information after a de-identification process [10]. Additionally, IRB approval can be granted for use and transfer of this genomic data because the Common Rule does not apply to data once it has been stripped of personal identifiers.

As previously stated, it is common for data sharing of genomic information to occur between health care organizations, academic research institutions, and commercial firms with limited regulation forging these partnerships. Regulation allows such data sharing because de-identified data can be reused and transferred for clinical application or research without explicit patient consent. However, it is inherently possible for data to be reidentified as the data holder through analysis of the DNA information alone [24]. Access to even a fraction of a DNA sequence can enable a variety of methods to link that information to the original individual. Re-identification is also possible using eye color, skin color, height, or other brief phenotype information. A study contributed by Craig Venter used DNA data to determine the facial features and voice of the individual who owned the data [25]. Therefore, de-identification from the specific elements noted in HIPAA and the Common Rule that allow data transfer is not enough protection because genomic data is never completely anonymized.

Governance structure for the safeguarding of data affects externalities of both privacy and security. HIPAA mandates cybersecurity requirements within the health industry. The databases comprised of genomic data outside of HIPAA control have loose regulation, and often times are

at the discretion of the organization maintaining the database. Cybersecurity for data excluded from HIPAA jurisdiction is regulated by IRB policies if it is intended for research or federally funded medical programs [1]. The IRB imposes data security requirements, and administers appropriate consequences for inadequate security measures. However, this enforcement does not guarantee that genomic data will be completely secured from unauthorized access.

Himss Analytics conducted a survey in 2016 targeting healthcare IT executives from the largest United States hospitals and health services systems. According to this survey, health care organizations consider adherence to regulatory requirements to be a top component to consider when choosing a provider for cloud services data storage [26]. This does not necessarily mean that all health organizations handling genomic data will thoroughly prioritize data safety. Many other industries are taking steps to update regulatory frameworks given how society is becoming more data centric. For example, the United States Food and Drug Administration is adapting their guidelines on the regulatory requirements for data sharing [27]. Looking towards the future, health industry regulation should define clear standards data storage requirements and how patients are grouped for genomic data sharing to help mitigate data risks.

Current regulation is not equipped to handle how the genomics field is transforming the health industry with precision medicine and individual genomic data accumulation. The next chapter discusses a variety of stakeholder groups that are affected by governance for genomic databases.

Chapter 3 Organizational Stakeholders and Data Responsibilities

Health Industry Stakeholder Groups

All organizations, including health care organizations, have stakeholders. "Stakeholder" is a comprehensive term denoting any individuals, organizations, or general groups that organizational decision making affects in any form. As evidenced in the framework of stakeholder management, stakeholder theory advises that there is a duty for these organizations to manage towards their stakeholders [28]. An organization should operate with the goal of promoting both its mission and the best interest of its stakeholders. This directly relates to theories of corporate social responsibility, which is the idea of acting ethically towards stakeholder groups and the overall common good. This can be initiated through corporate governance at an organizational level [29]. Emphasized corporate social responsibility and stakeholder management is valuable to an organization because it enhances overall reputation as well [30]. For genomics in health care, strong stakeholder management ensures that anyone with access to genomic data is handling it responsibly. This section describes a variety of stakeholder groups related to genomics in health care.

Genomics in health care affects several stakeholder groups even outside of the health care organization setting. Society is a stakeholder in that the advancement of the genomics field is causing advancement to the medical community. Naturally, this will affect our society because essentially all individuals are impacted by health care services in some way. The government, through organizations like the NIH, is involved as a stakeholder group via regulation enforcement and beneficence from medical advances. Investors and funding communities should be considered stakeholder groups as well because their funding can provide necessary means to

research programs regarding genomic databases. There are likely many more general stakeholder groups due to the broad applications of genomics. However, patients and doctors at health care organizations are the primary focus for the remaining stakeholder discussion.

Medical patients, receiving any type of health care treatment, are one of the most notable health industry stakeholders regarding genomics and genomic databases. A large proportion of genomic data is sequenced within the health care context. For patients whose personal DNA data is collected and sequenced, their data is entered into a genomic database for future use. Their data is necessary because genomic databases enable more accurate precision medicine and new medical discoveries based on large amounts of genetic information. It is very important to consider this stakeholder group because the genomics practice essentially could not continue without patient DNA data. General patients whose personal genomic data has not been sequenced are still affected by the genomics practice because of precision medicine capabilities that inevitably affect all individuals through medical care. Also, these patients have potential for DNA sequencing in the future. As noted previously, a patient's genomic data in a database may include a direct link to family health information. Therefore, patient families are a stakeholder group as well. Any individual's genomic data has the potential to be sequenced and used by doctors or researchers at a health care organization. Respondents of the Genomic Data Governance Survey, which is introduced in the next chapter, are members of the general public.

Doctors, researchers, and health care organizations using genomic databases can be considered direct stakeholder groups as well. Genomic data and databases have a fundamental clinical application. Doctors and researchers make precision medicine possible through their efforts at health care organizations. They rely on genomic databases in order to improve genomic test accuracy and to continue to make new medical discoveries through individualized medical

care. Health care organizations provide the necessary means for doctors and researchers to conduct analysis and to make new medical advancements. Health care organizations also provide a setting for new patient DNA data to be sequenced. Responsible data management is advantageous to health care organizations because it preserves public trust and increases the likelihood of an individual contributing data to genomic databases.

Policies, practices, and routines included in organizational governance can be applied to health care organizations with a general approach. Governance provides a framework that shapes human behavior within an organization. More specifically, governance can guide how assets or resources owned by an organization are handled in terms of use and access. In this context, genomic databases are an important asset to health care organizations with a variety of involved stakeholder groups. The next section discusses responsible principles for organizational data management and how proper governance practices can apply to genomic databases. This has an important direct application to current patients, future patients, and patient family stakeholder groups because that is where data in genomic databases originates. Doctors, researchers, and health care organizations are stakeholder groups that rely on genomic databases to advance the medical community. This illustrates a need to balance data responsibility and medical innovation through a health care organization's governance structure for genomic databases.

Data Responsibility Framework

One recommendation for data responsibility is specified by the U.S. Consumer Privacy Bill of Rights. It contains seven elements of consumer rights intended to standardize protection to data holders. They are Individual control, transparency, respect for context, security, access

and accuracy, focused collection, and accountability [31]. Individual control gives consumers the right to decide what personal data is collected from them and how it is used. Transparency gives consumers the right to access information about privacy and security measures an organization takes. Respect for context gives consumers the right that the original context for providing the data will guide organizational collection, use, and disclosure of that data. Security gives consumers the right to their personal data being protected in a secured fashion. Access and accuracy gives consumers the right to access personal data and correct it if it is inaccurate. Focused Collection gives consumers the right to personal data being collected and retained within reasonable limits. Lastly, accountability ensures that organizations will operate with adherence to the previous elements. It is important to note that like HIPAA, the U.S. Consumer Privacy Bill of Rights does not include "de-identified" personal data. A privacy and big data presidential review group endorsed these elements in 2014, and suggested that they be used as a basis for future legislation [19].

When applying the concepts stated in the U.S. Consumer Privacy Bill of Rights to genomic database policymaking today, there are four key data responsibilities to highlight for responsible governance towards patients and their DNA data. They are privacy, security, use, and accuracy. These categories provide a data responsibility framework for health care organizations to use in future governance evaluation. Privacy encompasses specific policies that affect patient control over data retention, access, collection necessity, or linkage to personal or family medical records. This right is promoted through tighter control of data access at an organizational level and allowing patients to decide what data is collected. Security, much like the U.S. Consumer Privacy Bill of Rights, ensures proper safeguarding of DNA data in genomic databases. Use involves data transfer, retention, use within or beyond original purpose for

sequencing, and general access. Responsible use embraces full transparency on how data may be used in the future. Future use includes the original purpose for sequencing or data transfer for another purpose. Accuracy engages the assurance of correct genomic data per patient and the right to correct data if needed. There are both positive and negative implications to each of the four data responsibilities in this framework. These implications stem from how an organization's policies either enhance or weaken each data responsibility category.

It is suggested that organizational governance based solely on permission from the data holder has potential to suppress genomics innovation [19]. Exercising these four responsibility categories as basic rights to patients and their data still promotes the proliferation and use of genomic databases, so long as use adheres to proper data practices. Governance standards that keep data access and usage close to those directly involved in the original purposed for sequencing data may be a guiding imperative to health care organizations. The four noted data responsibility dimensions provide a framework for responsibly managing stakeholder data in tandem with allowing medical innovation inherent to medical research using genomic databases.

The next chapter details an exploratory study of an emerging phenomenon on public perception towards genomic database governance policies and overall willingness to provide genomic data. This study is not at the hypothesis testing stage yet, but inferences can be made on public expectations for governance standards using the data responsibility framework. DNA data is extremely personal data, so the general public likely will have more restrictive expectations for governance policies in order to contribute personal genomic data. Governance policies that promote controlled data transfer, safety features for genomic databases, and narrow database access will probably be the most favorable to the public. Therefore, public attitudes will likely be consistent with policies that ensure positive representations of the data responsibility features.

Chapter 4 Genomic Data Governance Survey

Background

The Genomic Data Governance Survey, further expressed as GDGS, covered a broad range of content areas and applications of genomics. Forrest Briscoe was the Principle Investigator of this research. The GDGS was financially supported by the Smeal College of Business, the Farrell Research Center, and the Rock Ethics Institute. Its overall purpose was to understand individual willingness to contribute DNA data and to recognize public attitudes towards policies and practices regarding genomic database governance. This research was a part of an ongoing study to determine best practices for organizational decision makers to consider when forming genomic data governance structure in the future. The GDGS was nationally representative, so the reported conclusions on public preferences are within the societal context of the United States.

Overall, results reflected in the GDGS are unique in comparison with other surveys on willingness to provide DNA data to research driven databases. Many other credible surveys note a respondent pool with well above 50% to be willing to contribute DNA data [32,33]. The GDGS concluded much smaller percentages for participant willingness. For example, 12% of survey participants stated altruistic willingness for free, 51% of survey participants stated willingness in combination with additional compensation, and 38% of survey participants stated an unwillingness to contribute even if additional compensation was available.

An integral element of the GDGS was to explore how several relevant governance policies across many different organization types affect individual willingness to provide DNA data. These policies were intended to be widely applicable, at a high level, due to the diverse

uses and characteristics of genomic databases. Respondents provided reaction data to two broad categories of governance practices, technical and organizational, that are contained in this survey. Technical governance policies describe data storage, and this includes database design, infrastructure, and software. Broad organizational governance policies are also included in the GDGS. These policies depict human resource practices, such as training for employees and appropriate access to data.

Another survey section intended to gauge insight on public opinion towards genomic application within the health industry. Specifically, who should be responsible for database policy making and who should be allowed genomic data access at a hospital or other health care organization. Health care organizations are a main contributor of genomic data to the genomic databases guiding precision medicine, and patient communities are a prevailing stakeholder group within the health space. At the rate that genomics is affecting the health industry, all survey respondents can be considered stakeholders either currently or in the near future.

This thesis only analyzes GDGS survey data collected on public attitudes towards governance policies specifically within the health care or hospital setting. The GDGS results are essential in evaluating how health care organizations can responsibly manage DNA data based on a commitment to stakeholders and public values. As previously noted, genomic databases are critical to advancing the genomics field, which in turn innovates modern medicine. A balance between proper genomic database governance and genomics advancement can be found by prioritizing certain types of governance standards. The GDGS conclusions on governance with the highest importance to the public will determine which responsible practices to prioritize. More responsible genomic database governance will preserve public trust in the genomics

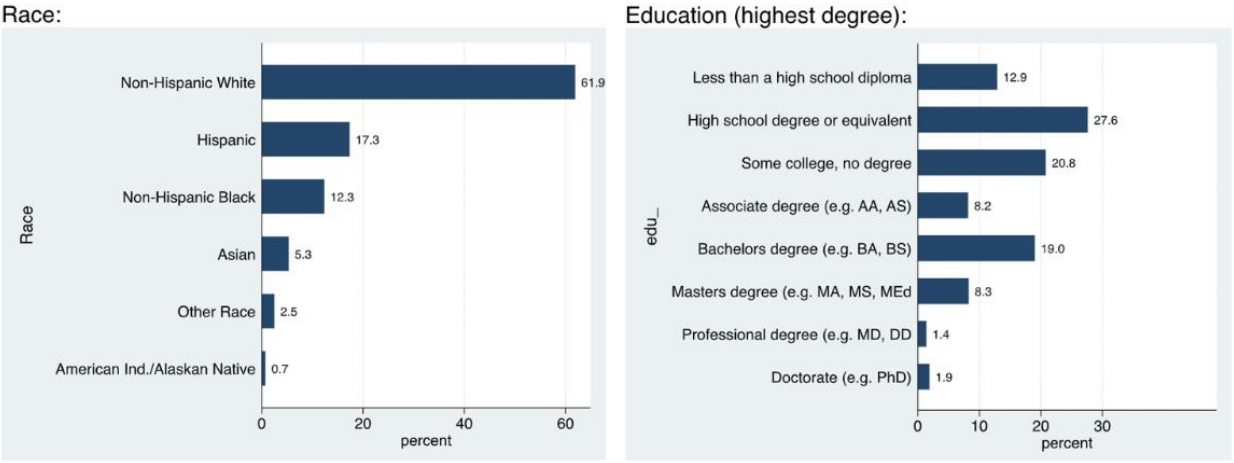
practice as well. Therefore, responsible governance should also allow the advancement of genomics because more individuals will feel willing to contribute personal DNA data.

Design and Methodology

The Genomic Data Governance Survey was a nationally representative survey, which was administered by Qualtrics through the Qualtrics online platform. It was completed by members of the United States public through the Qualtrics Panels survey recruitment service. Screening questions were used to ensure national representativeness encompassed dimensions such as respondent gender, age, employment status, race/ethnicity, and education level.

The survey questions were pretested in the Penn State Smeal Behavioral Research Lab with 174 Penn State undergraduate students in October 2018. The pretest pilot survey was administered through the Qualtrics online platform as well. The official survey completion through Qualtrics occurred from November 27th, 2018 to December 20th, 2018. Survey results are comprised of n=2020 online survey respondents. Within the respondent pool, 49.65% reported female, 48.71% male, and 0.94% prefer not to say, and 0.69% prefer to self-describe. The median age of respondents was 46 and the mean was 46.01, with a standard deviation of 16.96 and a range of 18-90. Approximately 59.65% of respondents reported being employed. Figure 3 below represents the race/ethnicity proportions across respondents, and the education levels across respondents. Each respondent watched a 3 minute video that covered unbiased information on the science, brief history, and applications of the genomics field. Including this video, the average survey completion time was 18.2 minutes in total.

Figure 3. GDGS Race and Education Level Distribution



Two segments of the Genomic Data Governance Survey are relevant to this thesis. The first evaluates reactions to general genomic data governance policies after respondents were primed with a health care organization description, and the second is a module on genomic data access and policy-making in hospitals.

In the general governance policy section of the survey, respondents were randomly assigned to one of five organization types to serve as a reference frame for questions. The possible organization types were intended to encompass a holistic representation of how individuals could be prompted to provide personal genomic data. The respondents could receive one of the following organizational descriptions:

- *Genetic Data Inc. is a U.S. for-profit technology corporation.*
- *GreatCare Hospital is a U.S. non-profit hospital system.*
- *Genomics & Health Research Lab is located at Middle State University.*
- *BioPharmaCo is a global for-profit drug company.*
- *The National Institutes of Health (NIH) is a U.S. federal research agency.*

Ultimately, this section was intended to gauge public attitudes towards a variety of governance policies that could be in place regarding genomic data at any given organization. Respondents were asked about their willingness to contribute their DNA data to their assigned organization based on each of 12 general governance policy statements. To measure this willingness, response options for each policy included a 5 point Likert scale ranging from "reduce willingness greatly" to "increase willingness greatly." Policy statements appeared in a randomized order to minimize response bias from priming effects, ordering effects, or subject fatigue. For the remainder of this thesis, the governance policy reaction data will only include respondent data extracted from the *GreatCare Hospital* organization type (N=394).

All respondents were randomly assigned to one of three different module types as well. Each module focused on genomic data functions in the workplace, new businesses and entrepreneurial ventures, and the health industry.

The health care module (N=672) gauges perceptions of governance policies for DNA data specifically within the context of hospitals and health care organizations. Respondents gave information on their perception of genomic data governance within the health care organization context. This module was framed so that respondents knew their answers on ideal governance structure would indicate an increased likelihood for willingness to contribute their genomic data if that structure were in place. They answered a variety of questions aimed at gauging who should have access to DNA data within a hospital, and who should be involved in making DNA data governance policies at a hospital. Respondents operated under the assumption that their local hospital would be collecting their DNA data and also connecting it with personal medical records. Each question included several categories of people who could either have access to

DNA data or be responsible for developing data governance parties at a hospital. Options were presented in list format, and respondents were able to select any number of categories.

The GDGS data were imported from Qualtrics through Stata 15. Analyses and conclusions were generated using Excel.

Governance Policies and Sharing

Organizational governance structure must reconcile an allowance of medical innovation through genomics advancement and precision medicine with proper responsibility towards stakeholders and their data. In order to establish this balance, it is helpful to consider how governance policies enable or do not enable organizations to keep genomic data close and control it responsibly. According to the data responsibility framework, organizational governance with close control on genomic data access and use will result in the most responsible practices towards those involved with the data. A tension arises in that precision medicine, the culmination of medical innovation, is directly supported by wide data access and sharing through genomic databases. An analysis of how governance aligns with data responsibility and either widens or narrows data access will help to determine which governance policies to prioritize at health care organizations.

The Table 1 below maps each of the 12 GDGS governance policies to the data responsibilities of privacy, security, use, and accuracy. Plus signs indicate the policy enhances the particular data responsibility, and minus signs indicate the policy weakens the data responsibility. Additionally, each policy is marked to either facilitate wider access to data or more narrow access to data.

Table 1. Governance Policy Feature Map

Policy	Privacy	Security	Use	Accuracy	Widens Access	Narrows Access
1. *Individuals have the right to request that their DNA data be deleted from the database at any time*	+		+			X
2. *DNA data are not sold, rented, or shared with any other organizations*	+	+	+			X
3. *Individuals will be asked permission for each specific use of their DNA data in the future*	+		+			X
4. *Government requests for access to DNA data are refused without a warrant*	+	+	+			X
5. *State-of-the-art IT security are used for all DNA data and other customer data*		+				X
6. *All employees sign an ethical "code of conduct" which includes safeguarding of DNA data*		+	+	+		X
7. *An independent company audits and certifies DNA data security*		+		+		X
8. *Access to family medical records is required, and these will be linked to DNA data*	-		-	-	X	
9. *Members of the general public serve on a committee to decide how genomic data will be used*			+			X
10. *Individuals' DNA data are stored indefinitely*	-		-		X	
11. *Copies of all DNA data (without individuals' names) are deposited into a government database*	-	-	-		X	
12. *Access to DNA data is sold to pharmaceutical firms (without requesting further permission)*	-	-	-		X	

Policy 1 includes a system for retracting DNA data so it can be found and deleted if the individual requests it. Policy 2 prohibits any transaction involving the transfer of DNA data

outside of its original sequencing purpose. Policy 3 requires permission for future use either within original organization's clinical care or research or if data is shared with another organization. Policy 4 includes actively updating security systems for databases as needed. Policy 5 indicates that employees must sign this code of conduct to access and use the genomic database. The warrant noted in Policy 6 suggests a reasonable and legal need to access the database. Policy 7 would likely involve a data protection contracting company. Policy 8 suggests that the family medical records would be stored in the same databases with the same protections as DNA data. Policies 9 and 10 are fairly straightforward. Policy 11 involves de-identification of data but the government gains access automatically to other health history information as well. Policy 12 occurs without notifying individuals that their data use extended beyond its original purpose, and this is made possible through broad consent.

Generally, there is a consistent pattern between how each governance policy reflects on the four data responsibilities and the associated data access type. The policies that enhance at least one data responsibility are connected to a more narrow access to data. Whereas, the policies that weaken at least one data responsibility are connected to a more broad access to data.

Table 2 below represents GDGS respondent reaction data to genomic data governance policies. The column that reflects the respondents for whom the policy increases willingness contains the proportional amount of respondents indicating "increase willingness greatly" or "increase willingness somewhat" per policy. Whereas, the column that reflects a decreased willingness contains the amount of respondents who answered "decrease willingness greatly" or "decrease willingness somewhat." Those who answered "neutral" on the 5 point Likert scale are not included in these percentages. Policies are listed in a descending order based on how much each policy increased individual willingness to contribute genomic data.

Table 2. Public Reactions to Governance Policies Used by Hospitals

Policy	Respondents for whom policy increases willingness	Respondents for whom policy decreases willingness
1. *Individuals have the right to request that their DNA data be deleted from the database at any time*	72%	7%
2. *DNA data are not sold, rented, or shared with any other organizations*	70%	7%
3. *Individuals will be asked permission for each specific use of their DNA data in the future*	67%	10%
4. *Government requests for access to DNA data are refused without a warrant*	56%	14%
5. *State-of-the-art IT security are used for all DNA data and other customer data*	54%	10%
6. *All employees sign an ethical "code of conduct" which includes safeguarding of DNA data*	52%	10%
7. *An independent company audits and certifies DNA data security*	48%	13%
8. *Access to family medical records is required, and these will be linked to DNA data*	31%	34%
9. *Members of the general public serve on a committee to decide how genomic data will be used*	27%	29%
10. *Individuals' DNA data are stored indefinitely*	22%	29%
11. *Copies of all DNA data (without individuals' names) are deposited into a government database*	22%	40%
12. *Access to DNA data is sold to pharmaceutical firms (without requesting further permission)*	10%	66%

Policies 1 through 7 are highlighted in green because a significantly higher amount of respondents reported an increased willingness to provide personal genomic data than a decreased willingness. Increased willingness ranges from 48% to 72% of respondents, whereas decrease willingness ranges from 7% to 13% of respondents. Also, these seven policies are positive representations of privacy, security, use, and accuracy. Overall, they each strengthen at least one aspect of the four noted data responsibilities. Also, they each cause greater restraint on genomic data access and an increased control at an organizational level.

Policies 8 and 10 through 12 are highlighted in red because a significantly higher amount of respondents reported a decreased willingness than an increased willingness. Decreased

willingness ranges from 34% to 66% of respondents, whereas increased willingness ranges only from 10% to 31%. These four polices are negative representations privacy, security, use, and accuracy as well. Each policy weakens at least one of these four data responsibilities because they promote an ease in data sharing, privacy infringement, external data transfer, and indefinite use. They allow for a much wider range of access to data, which may permit data use beyond the original intended purpose.

Policy 9 is not highlighted at all because it is considered a positive representation of organizational data responsibility, but only 27% of respondents report and increased willingness to contribute genomic data. It enhances data responsibility by placing value on public opinion regarding genomic data governance, and yet it falls towards the end of the list. It is not highlighted because it does not appear to be relevant to the public or a preferred governance item for health care organizations to have.

Public view on this tension between data responsibility and medical innovation is clear. Governance policies that strengthen organizational data responsibility (1-7) through tighter control and closer access are favorable to the public. These policies tend to increase data contribution willingness, but may stifle medical innovation. Governance policies that weaken organizational data responsibility (8 and 10-12) through less organizational control over data and more distant data access are not favorable to the public. These policies tend to decrease data contribution willingness, but increase the potential of medical innovation. Regardless, these results suggest a general guideline for health care organizations to consider in evaluating responsible governance structure. Organizational governance structure that keeps genomic data generally within the original purpose for which genomic data was sequenced will allow for

responsible management of stakeholder data. This will also increase contribution willingness, which will likely improve genomic databases and precision medicine through additional data.

The GDGS health care module had one aspect that evaluated public attitudes towards groups that could have access to genomic data specifically within health care organizations. This concept has direct implications on all four data responsibility features. Also, it helps to determine who is considered "close" to the data holder with more narrow data access. This module gauged public opinion based on three different health care scenarios. The first presented scenario was:

> *Imagine that your local hospital is collecting your DNA data and linking it to your medical record. They are doing this in order to advance biomedical research and improve health care. Your doctors can have access to your DNA data, just as they have access to your medical record.*
>
> *You are now being asked for permission to include your DNA data and medical records in a larger database that will be shared with <u>other</u> doctors and researchers. The shared database would not include individual names, although it could be possible to re-identify your records within the database using certain techniques.*
>
> *For each group listed below, how likely are you to give permission to access your DNA data and medical records?*
> *__ Doctors and researchers from <u>your</u> health care organization?*
> *__ Doctors and researchers from <u>other</u> US-based health care organization?*
> *__ Employees of a US government research agency?*
> *__ Doctors and researchers from organizations based in other countries?*
> *__ Employees of a global pharmaceutical firm?*

The willingness to share DNA data and medical records with a hospital or health care organization varied considerably by the particular group that would gain permission access to the data. Respondents selected an option on a 5 point Likert scale ranging from very unlikely to very likely. Table 3 below depicts the five different groups who could possibly be granted genomic data access. The "likely to share" column includes the proportion of respondents who selected very or somewhat likely, and the "unlikely to share" column includes the proportion of

respondents who selected very or somewhat unlikely. Again, categories are listed in a descending order based on reported likelihood to share genomic data.

Table 3. Likelihood of Providing Genomic Data Access Permission

Group Category	Likely to share	Unlikely to share
Doctors and researchers from your health care organization?	50%	28%
Doctors and researchers from other US-based health care organization?	33%	41%
Employees of a US government research agency?	23%	56%
Doctors and researchers from organizations based in other countries?	19%	60%
Employees of a global pharmaceutical firm?	17%	60%

Respondents were most likely to share their data with doctors in the respondents' own local health care organization (50% likely). Doctors and researchers from another health care organization have a relatively high likelihood (33% likely) as well. Whereas, respondents were least likely to share their data with organizations based outside of the US (60% unlikely) or with global pharmaceutical firms (60% unlikely). This would suggest that it is important to stakeholders for their data access to be kept close and within a health care context. There appears to be an aversion to more distant data access that extends farther beyond the intent for providing the health care organization with genomic data. Not surprisingly, this data suggests similar findings that support narrow access to genomic data in order to increase organizational responsibility and individual willingness to contribute DNA data. This in turn enhances data privacy, data security, appropriate data use, and data accuracy. However, data use limitations through stricter access controls will lower the amount of data that can be used to advance

genomic databases. This may negatively affect the ability of researchers and clinicians to innovate the medical field through precision medicine. Thus illustrating the tension again.

Involvement in Governance Policy Making

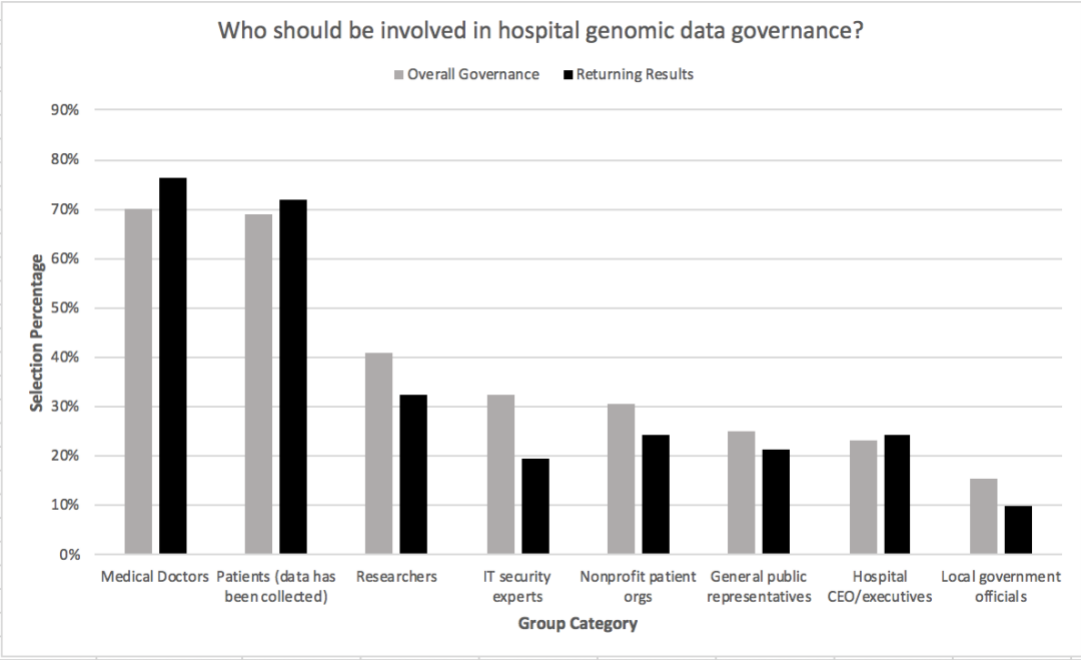
Another aspect of organizational governance is the structure for who is responsible for governance policy making. The different groups of people who could be involved in governance policy making have different background knowledge and priorities when it comes to handling genomic data. Therefore, each group has implications for governance outcomes regarding the four data responsibility features and data access type. The last two scenarios presented to respondents in the GDGS health care module involve group categories who could possibly be involved in the development of governance policies for genomic data.

The first question below explores overall governance decision making and the second question explores decision making on the return of relevant test results based on DNA data.

> *In the future, DNA data may be collected as part of your routine medical care. Hospitals will have to make policies about how to handle this data and the associated genetic information on their patients. Thinking about your local hospital, who should be involved in making these policies? Choose all that apply:*
>
> - *Medical doctors*
> - *Researchers*
> - *Patients whose data have been collected*
> - *Representatives of the general public*
> - *Local government officials*
> - *Hospital CEO or other hospital executives*
> - *Nonprofit patient orgs. (e.g. American Cancer Society, American Heart Association, Autism Speaks)*
> - *IT security experts*
>
> *Hospitals will also have to make policies about what to do when DNA data reveals that a patient may be at risk for a disease. For example, policies about when to give this new information to patients or their families. Again thinking about your local hospital, who should be involved in making these policies? (same categories)*

Respondents were prompted to select any, or none, of the group categories that they believed should be involved in each scenario. Selection frequency results per scenario are displayed in Figure 4 below. The "Overall Governance" data represents the first question above, and the "Returning Results" data represents the second question above. The group categories are ordered in descending order based on reported percentages for the results in the overall governance question.

Figure 4. Involvement in Governance Policy Decision Making



When asked who should be making governing policies regarding DNA data management and the return of clinically relevant results, respondents generated relatively consistent results for both situations. The most frequently cited groups were medical doctors, patients whose data has been collected, and researchers. While the least frequently cited groups were local government officials, hospital CEO or other hospital executives, and representatives of the general public. It is important to note that not all respondents selected even the most frequently cited groups. This

indicates a general wariness towards involvement in genomic data governance policy making. However, the data varies enough to still suggest particular themes for who the public feels should and should not be involved in policy making.

Overall, these outcomes suggest that respondents are most comfortable when policies are made among those who are closely involved with the data. This is either by providing the data or being directly involved with the original purpose for sequencing the data in the health context. Additionally, this may suggest that participants would be unwilling to contribute their DNA data if the more-distant people and organizations were involved in policy making for genomic data governance. This represents a similar phenomenon to previous conclusions. Based on public opinion, responsible governance practices at health care organizations should primarily involve those closely tied to the data holder and advancing the medical community. Policy decisions made by medical doctors and researchers, as opposed to the other groups, can be considered a more narrow involvement. This implies that these groups have an appropriate background to handle data responsibly, according to public perception.

The next chapter discusses how genomic data governance policies are reflected in practice at a major academic medical center affiliated with a very large United States university.

Chapter 5 Governance Policies in Practice

Comparison with The Hershey Medical Center

The Briscoe team sought to compare the GDGS governance policy results with governance policies that real health care organizations are either considering to use or are using in practice already. The health care organization used for comparison is the Penn State Health Milton S. Hershey Medical Center, which has a variety of services. It encompasses a teaching hospital, medical school, and research center. Their genome sciences facility incorporates genomic consultation, instrumentation, and research services [34]. Specifically, the Penn State Personalized Research for Innovation, Discovery, and Education (PRIDE) program houses genomics research. Information to conduct this comparison with GDGS results was gathered through discussions with a bioethicist employed by this organization an interview with the PRIDE program director.

Table 4 below directly compares survey findings on public preference with the information gathered from this interview. Policies 1 through 12 are the same general genomic governance policies in the GDGS. Policies 13 through 15 reflect policies from the GDGS health care module, and they are framed as group categories that were beyond who was favorable to respondents. The Hershey Medical Center's governance structure is mapped to whether the GDGS policies exist, partially exist, or do not exist. Policies 1 through 7 were satisfactory to GDGS respondents. They enhance the four data responsibility features and narrow data access. The "exists" column is highlighted green for these policies, and the "does not exist" column is highlighted red for these policies. Policies 8 and 10 through 15 were not satisfactory to

respondents. They weaken the four data reasonability features and widen data access. The "exists" column is highlighted red for these policies, and the "does not exist" column is highlighted green for these policies. Again, policy 9 is not highlighted green or red because it did not appear to be relevant to the public. Responsible governance that aligns with public preference in the GDGS results would include markers in the green boxes only.

Table 4. Governance Policies in Practice vs Public Preference

Policy	Exists	Partially exists	Does not exist
1. Individuals have the right to request that their DNA data be deleted from the database at any time	X		
2. DNA data are not sold, rented, or shared with any other organizations			X
3. Individuals will be asked permission for each specific use of their DNA data in the future			X
4. Government requests for access to DNA data are refused without a warrant		X	
5. State-of-the-art IT security are used for all DNA data and other customer data	X		
6. All employees sign an ethical "code of conduct" which includes safeguarding of DNA data		X	
7. An independent company audits and certifies DNA data security			X
8. Access to family medical records is required, and these will be linked to DNA data		X	
9. Members of the general public serve on a committee to decide how genomic data will be used			X
10. Individuals' DNA data are stored indefinitely			X
11. Copies of all DNA data (without individuals' names) are deposited into a government database		X	
12. Access to DNA data is sold to pharmaceutical firms (without requesting further permission)	X		
13. Data access permission is granted beyond doctors and researchers at the focal health care organization & other US health care organizations	X		
14. General genomics policy decision making involves groups beyond medical doctors, sequenced patients, and researchers		X	
15. Return of results policy decision making involves groups beyond medical doctors, sequenced patients, and researchers		X	

The Hershey Medical Center seems to have a moderately conservative approach to genomic database governance. Policies 1, 5, and 10 directly align with public preferences. The inclusion of 1 and 5, and the exclusion of 10 positively supports the data responsibility features of data privacy, security, and use. These governance policies enable data to be kept close and with narrow access. Policies 2, 3, 7, 12, and 13 do not align with public preferences. The exclusion of 2, 3, and 7, and the inclusion of 12 and 13 has negative implications for the data responsibility features of data privacy, data security, and data use. These governance policies have a higher likelihood of wider access to genomic data. These two groupings suggest this organization's governance structure is fairly balanced. Additionally, this organization's governance structure has partial components of several policies, which are marked in the "partially exists" column. Many of the governance policies in the GDGS are very specific. Information was collected on how the organization's governance could only partially align with some GDGS policies through the interview with the PRIDE program director. There is slightly more weight on policies in partial existence that are not in public preference. Overall, this organization supports positive and negative data implications through partial components of the GDGS policies in a balanced fashion as well.

The remainder of this section includes a holistic description of information gathered from discussion with the PRIDE program director. This information provides continued explanations for why each policy was marked as being in existence, partial existence, or no existence.

The Hershey Medical Center's institute for personalized medicine is intended to provide tools to clinicians to determine if there are genetic contributions to the way individuals respond to treatment. The generated data is contained in a secure data center, and the program is "not collecting genomic sequences in a random fashion to use later… they are all collected and used

in terms of the specific questions they are asking." Their studies are designed around 25 to 30 investigations of different diseases or treatment. They obtain consent from patients they think would be relevant to those particular questions. They then apply the relevant genomic tools, whether that is genotyping or a panel of genome or exome sequencing.

Their broad goals for genomic databases include an "intent for the genomic data to be linked to medical record data," according to the PRIDE program director. A patient's genomic data is put into a red cap database, which is a type of SQL database, and the data is linked through an identifier code that can then be linked to the patient's medical records. This information is only accessible to their honest brokers, who are three senior managers at the organization. These brokers ensure that valid reasoning supports any instance of retracing the identifier code back to the patient that the genomic and health data is linked to. Research investigators do not have automatic access to identifier codes.

The PRIDE program has two committees responsible for making general policies related to genomic database governance, an access committee and a return of results committee. The access committee acts on requests to obtain samples, and this is relevant when investigators are looking for particular samples that would be useful for addressing the questions they are interested in. According to PRIDE program policy, samples can be acquired by honest brokers with Institutional Review Board (IRB) permission. They may need to create their own IRB protocol, in which case they would not have to go through the access committee in order to gain information because the IRB becomes the access committee. The return of results committee is tasked with determining how genomic data results should be returned to the individual patient. Researchers do not often engage with this committee because variant calling is done in the context of the study, so return of results would already be built into the study. To conduct

research, they "generate a sequence that may be relevant to variants of a particular disease, but they never set up a query to determine if those variants are present in an individual patient." Individuals involved on both committees only include researchers, clinicians, ethicists, and administrators.

The Chief Information Officer organizes a security subgroup who is responsible for IT security at the Hershey Medical Center. Any data that the PRIDE program generates is subject to the same security controls that are placed on medical records, and this is all controlled through the IT group. The level of security matches the level that is required for HIPAA compliance. As indicated in previous chapters, at Hershey there is also an ongoing discussion on whether genomic data is private health information or identifiable in itself. Their current point of view is that this data is treated as secure as patient medical records, but when providing to external organizations then that organization's governance structure takes over. For example, data could be shared with the NIH where genetic information is not considered private health information.

The Hershey Medical Center's consent process begins with direct communication with the potential consenter. This can occur either face to face with an individual authorized with IRB protocol to obtain consent or through a video link. Their consent form for genomic data collection states that data will be shared with other organizations, and will be available for future research use under different circumstances. This is parallel to the overall concept of broad consent that was described in previous chapters. The circumstances for future data transfer are ambiguous. The consent form states that data is "subject to discovery." Hershey maintains the right for sharing genomic data with the NIH, the government, or anywhere with a genomic database. Additional patient permission is not required before data transfer can occur.

There is a direct link between the bio samples provided and the genomic data generated from them. Samples are tied to a code which is tied to the medical record, and there is a straight forward link between samples and data through an honest broker. Additionally, PRIDE collects life style data through a questionnaire after obtaining consent. Patients are asked a few questions on basic background information outside of medical records. This background information may include tobacco use, general family history information, whether or not the patient has a twin, history of a particular ailment, etc. These questions may be specifically tailored to the type of study being conducted. The gathered information is tied to the individual's genomic data.

Genomic data in Hershey's databases are linked to a unique number, and patients are not identified by name. Study staff link health information and records to the provided sample for tracking purposes. Patients reserve the right to leave the study at a later date, after which a retracing process will occur using the unique identifier number to find the original bio sample and data that will be deleted. Patients do have the right for information to be deleted; however, it will not be expunged from use in previous studies. Patients can follow either of two mechanisms to request data deletion, a written request or a phone call that is followed up with a written request. Deletion from their database occurs almost immediately after the request is made.

There is no specified cut off time for genomic data usage. There is an ongoing discussion regarding how long data will be archived, but that is more of an IT related discussion than a compliance or legal issue. Hershey ensures that data is backed up for at least seven years, but there is not much need to expunge data beyond that. The PRIDE program director notes that "the only reason to expunge data would be financial reasoning." However, over time it is becoming cheaper to store large amounts of data. They do not have a statement of how long they keep data beyond expectations for retaining research data for the mandated timeframe.

Hershey's approach to staff compliance for genomic data is not different than their approach to other data forms. There is a lot of compliance and mandated trainings relevant to general health care regulation, but there is no specific protocol for compliance genomic data. They do have general compliance training and refresher courses every few months regarding general compliance, HIPAA, Title IX, etc. There is no specific compliance structure related to genomic data because it has yet to be classified under HIPAA jurisdiction. If it does become relevant to HIPAA, then a variety of required compliance trainings would result for employees with genomic data access. The PRIDE director mentioned that "we try to resist as much as possible, so if [additional compliance] remains voluntary then I don't think it will happen unless it becomes an issue that a specific institution from the outside is mandating more training."

Hershey has a general research code of conduct, but employees do not operate under a code of conduct specific to genomic data. Their first line of defense is that data has restricted access, and the study's overseer or principal investigator of personalized medicine must have specific permission to access data. Thereafter, it is up to the principal investigator to ensure that the researchers with data access understand that this is not data to be shared on a public forum.

Hershey has an additional policy where a data sharing agreement is required before transferring data to an external organization. The data sharing agreement must be written and signed by both parties. Specifications to this agreement are made on a case by case basis, but this may include an explanation of what the organization is and what the data is being used for. Data sharing outside of the U.S. is regulated by a data sharing agreement as well. Research administration at Hershey oversees this process.

The PRIDE program's genomic databases are maintained within the Hershey Medical Center, and they are protected with the same security that protects patient medical health records.

They do not utilize cloud based storage, like Amazon Web Services, even though it may be more secure than their own database. Additionally, they do hot hire an outside auditor for genomic data security, and they are not specifically covered under any institutional insurance policy. They do not have governance regarding the use of external vendors for genomic database storage or security. The PRIDE program director specifically stated that "unless there is a substantial change in the attitude here, it won't change in the near future, but it's possible in the long term." External auditing and database usage at Hershey is largely driven by cost considerations and by history of how secure these third party database and storage sites tend to be. For now, their database security is managed by their IT department.

The next chapter summarizes relevant findings from Chapters 4 and 5, and discusses potential future directions of this research.

Chapter 6 Discussion

Conclusions

Medical innovation brought by precision medicine depends on genomic data sharing and wide ranging access to genomic databases. The risks to patients and their data due to broad genomic data transfer and access must be addressed through a framework for responsible governance. The data responsibility framework presented in this thesis illustrates four key data responsibilities and an overall general theme to guide genomic database governance at health care organizations. This framework suggests that responsible organizational governance includes closer control on genomic data access and an enhancement of patient data privacy, security, use, and accuracy.

When the data responsibility features were mapped to governance policies that could be used in practice, the policies that enhanced data responsibility also narrowed access to genomic databases. Whereas, the policies that weakened the data responsibility features also had a tendency to widen access to genomic databases. The GDGS findings were compatible with the data responsibility framework. Results suggest that the strongest public interest is in governance policies that enhance the four data responsibility features and narrow access to data. The policies that weaken the four responsible data responsibility features and allow broad data access were the least favorable to respondents. Additionally, survey results suggest a governance structure that conflicts with the data responsibility framework may prevent individuals from contributing their genomic data to these databases.

One of the primary stakeholder groups is the genomic data providers and their biological relatives. In the health care context genomic data providers are patients who have been

sequenced or may be sequenced in the future. Anyone from the general public has genomic data that could be acquired by a health care organization. The GDGS findings mainly represent this stakeholder group because respondents were members of the general public. Another clear stakeholder groups are researchers and doctors who are using genomic databases to improve the medical field through precision medicine. Funding communities that support genomics research, the government, and IT groups that maintain genomic database security are important stakeholder groups as well. Society as a whole can be considered a stakeholder group because of how the genomics practice is revolutionizing modern medicine, which is intended for all.

Governance structure has a varying impact on different stakeholder groups. The GDGS results help to evaluate governance practices that maintain responsibility towards health care stakeholder groups, particularly genomic data contributors. However, many stakeholder groups would not benefit if only general public preference were enforced through governance health care organizations. For example, this includes doctors and researchers who rely on widespread data access in order to improve accuracy of genomic analysis and to enable precision medicine. Conversely, if the ideal governance structure indicated by GDGS results was completely ignored then all stakeholder groups would not benefit. The primary negative impact would be on genomic data contributors, in this case the patients, because their personal data would not be handled responsibly. This in turn would decrease data contribution willingness. All stakeholder groups would likely be negatively impacted from a slowed growth of genomic databases because large databases are essential to medical innovation and precision medicine.

Governance structure should promote an enhancement of all four data responsibility features without restricting the potential for innovation and new discoveries through database usage. Based on current health industry governance structure in practice, even stronger

responsibility to patients and their data is possible through slight policy shifting. Organizations should adjust governance to prioritize data responsibility enhancing policies and to avoid policies that weaken data responsibility. This can be done in a balanced manner by emphasizing the "partially exists" column in Table 4.

A form of partial existence should be found for the policies that enhance data responsibility and do not exist in practice and for the policies that weaken data responsibility and do exist in practice. Policies 2, 3, and 7 should in theory enhance data responsibility when in existence. By shifting them from entirely not in existence to a looser form of being possible on a case by case basis then genomics advancement will still be possible while also upholding data responsibility. This shift, along with the continuation of the responsible governance policies in existence, would improve organizational control over data and enable data to be kept close with narrow data access. Policies 12 and 13 exist in practice, but weaken data responsibility. These policies impose substantial constraint on data privacy, security, and responsible use. However, these situations can be extremely useful to the medical community due to more information being available for analysis. Organizations should attempt to implement structure that generally avoids the negative implications of the situations noted in policies 12 and 13 through shifting to more involved regulation and observation, at an organizational level, of data access. This shift, along with continued avoidance of the other policies that weaken data responsibility, will hold organizations accountable to proper data handling techniques without restricting genomics advancement.

The upside of investing in responsible and balanced governance at a health care organization is that it benefits stakeholders and the organization overall. The downside of weak governance at a health care organization is the privacy and security risks to data providers, and

slowed advancement of genomic databases given a decreased willingness to contribute genomic data. Altogether, this recommendation advocates for health care organizational governance structure to emphasize patient rights consistent with the four data responsibilities. It allows for responsible management of stakeholder data, while not completely eliminating the continuation and use of genomic databases. Proper implementation of the data responsibility enhancing policies and transparency towards data holders should allow for genomics advancement. Ideally, this approach will ultimately further the genomics field because more individuals will feel willing to contribute personal data to an organization with a control focused governance system. All health care organizational stakeholders will benefit from this because of the continued innovation genomics is bringing to all through precision medicine.

Next Steps

The findings from the GDGS and their compatibility with organizational data responsibilities through governance offers an interesting framework for best practices regarding genomic database governance. This thesis offers one suggestion for how responsible governance practices can be reconciled with genomic database advancement and precision medicine. However, the future of precision medicine and the genomics field overall are uncertain. There are more questions to be answered, both within and outside of the health care context.

Within the health care context, not many of the policies in the GDGS addressed the data accuracy responsibility. Currently, members of the general public are unable to identify if their reported genomic data is actually their own or not. This raises concerns for data accuracy because the individual cannot personally track this information. Typically, only researchers and

trained individuals are able to confirm DNA data accuracy per individual genome sequence. Future analysis should include how governance structure can guarantee correct DNA data.

Many comments from the genomics research program director at the Hershey Medical Center indicate that several policies in practice are driven by cost considerations. The health care industry overall is largely propelled by cost considerations. Each governance policy may have a particular monetary impact on an organization. Future analysis could map each governance policy from the GDGS to an overall cost factor of implementing that policy. This may evaluate the feasibility of executing publicly preferred policies at an organization.

Other industries have different guiding regulation and may have different public attitudes regarding genomic database governance. The conclusions drawn from this thesis about genomic database governance in health care and its impact on precision medicine may not be directly applicable to other industries. It could be beneficial to conduct a similar evaluation of GDGS results on another organization type compared to that industry's governance in practice.

Outside of the United States, there may be differences in public preference for how genomic data should be handled at an organizational level. Regulatory frameworks that manage personal data tend to vary widely by country. This may cause issue with multi-national sharing partnerships, which are becoming very common through global partnerships, mergers, and acquisitions. Future investigations should more specifically analyze how governance can responsibly address the inevitable possibility of multi-national genomic data sharing.

Appendix A Hershey Medical Center Request for Information

Topic: Policy making

0--What are the institution's broad goals with respect to the genomic database?

1--What committee or process is used for making general policies related to genomic database governance? Which types of individuals are involved (e.g. administrators, researchers, clinicians, representatives of patients, advocacy groups, IT security, government, general public, other)?

2--Through what committee or process are IT security investments made for genomic databases? Who is involved? Are these decision processes the same for genomics and other types of data?

Topic: Consent

3--Can you please share the consent form(s) and any other documents that you use for participants who provide their genomic data (and other biobank materials)?

4--When individuals provide their genomic data, are they told whether it will be shared with other organizations (commercial, scientific, etc.)? Will they be asked permission for specific future uses of their data?

5--Are they told how long their data will be retained? Do they have the right to later request that their data be deleted? If so, how long does it take for such a request to be fulfilled, and how does this work (in general terms)?

Topic: Compliance

6--For employees given access to genomic data, what policies and/or practices are used to screen, train, monitor, and/or sanction in relation to compliance/risk management? Do employees sign a 'code of conduct' relevant to the context of genomic data?

7--Is the institution's approach to staff compliance for genomic data any different than its approach to other forms of data? How?

8--Have you hired an external auditor or outside vendor to assess your policies and procedures related to data security, privacy, or other aspects of your genomic database? Is the genomic database covered explicitly under an institutional insurance policy?

Topic: External data access

9--Do you have written policies covering which different groups outside your institution can have access to genomic data? Are there separate policies related to access for different purposes? For academic versus commercial institutions? From organizations outside the U.S.?

10--What is your policy (if any) regarding government requests for access to the genomic database? From funding agencies? Law enforcement agencies? Courts? Others?

11--How do you decide about the use of external vendors for genomic database storage, security, compute, etc.? What contractual arrangements do you require (in general terms)? For what purposes (if any) do external vendors have permission to use the data?

Topic: Linked data

12--Is the genomic database (directly or indirectly) linked to medical records? Records of family members? Other types of data? If so, how are these connected (in general terms)?

REFERENCES

[1] Briscoe, Forrest, and Barbara Gray. "INNOVATIONS IN MEDICAL GENOMICS: HOW TO ENABLE ADVANCES WHILE MANAGING PRIVACY AND SECURITY RISKS?." (2017).

[2] Dickinson, Alex, Director. *Illumina Massively Scales Its DNA Sequencing Technologies Using AWS. Amazon Web Services*, Illumina, 2018, aws.amazon.com/solutions/case-studies/illumina/.

[3] Snyder, Michael. *Genomics and personalized medicine: what everyone needs to know*. Oxford University Press, 2016.

[4] Collins, Francis S., and Harold Varmus. "A new initiative on precision medicine." *New England Journal of Medicine* 372.9 (2015): 793-795.

[5] Topol, Eric. *The patient will see you now: the future of medicine is in your hands*. Basic Books, 2015.

[6] "The Cost of Sequencing a Human Genome." *National Human Genome Research Institute (NHGRI)*, National Institutes of Health, 6 July 2016, www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/.

[7] Hoeksma, Jon. "Genome Sequencing Will Go Mainstream Using Artifical Intelligence." *Digital Health*, Digital Health 2017, 22 Feb. 2017, www.digitalhealth.net/2017/02/100-dollar-genome-possible-with-ai/.

[8] Middleton, Anna, et al. "'Your DNA, Your Say': global survey gathering attitudes toward genomics: design, delivery and methods." *Personalized medicine* 15.04 (2018): 311-318.

[9] Clayton, Ellen W., et al. "A systematic literature review of individuals' perspectives on privacy and genetic information in the United States." *PloS one* 13.10 (2018): e0204417.

[10] "Methods for De-Identification of PHI." *HHS.gov*, US Department of Health and Human Services, 6 Nov. 2015, www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#coveredentities.

[11] "What Is the Precision Medicine Initiative? - Genetics Home Reference - NIH." *U.S. National Library of Medicine*, National Institutes of Health, 19 Mar. 2019, ghr.nlm.nih.gov/primer/precisionmedicine/initiative.

[12] Schwartz, Peter J., Lia Crotti, and Roberto Insolia. "Long-QT syndrome: from genetics to management." Circulation: Arrhythmia and Electrophysiology 5.4 (2012): 868-877.

[13] Cooperberg, Matthew R., et al. "Combined value of validated clinical and genomic risk stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy cohort." *European urology* 67.2 (2015): 326-333.

[14] Grupp, Stephan A., et al. "T cells engineered with a chimeric antigen receptor (CAR) targeting CD19 (CTL019) have long term persistence and induce durable remissions in children with relapsed, refractory ALL." (2014): 380-380.

[15] Battelle Technology Partnership Practice. *The Impact of Genomics on the U.S. Economy*. United Medical Research, June 2013, web.ornl.gov/sci/techresources/Human_Genome/publicat/2013BattelleReportImpact-of-Genomics-on-the-US-Economy.pdf.

[16] Devos, Lauren, et al. “The Genomics Inflection Point: Implications for Healthcare.” *Rock Health*, Rock Health, rockhealth.com/reports/the-genomics-inflection-point-implications-for-healthcare/.

[17] “Genetic Discrimination Fact Sheet.” *National Human Genome Research Institute (NHGRI)*, National Institutes of Health, 6 Apr. 2015, www.genome.gov/10002328/genetic-discrimination-fact-sheet/.

[18] “Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data.” *Ponemon Institute*, Ponemon Institute, LLC, 12 May 2016, https://www.ponemon.org/blog/sixth-annual-benchmark-study-on-privacy-security-of-healthcare-data-1

[19] Schneier, Bruce. *Data and Goliath: The hidden battles to collect your data and control your world*. WW Norton & Company, 2015.

[20] Office for Human Research Protections. “Federal Policy for the Protection of Human Subjects 'Common Rule.” *HHS.gov*, US Department of Health and Human Services, 18 Mar. 2016, www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

[21] Jaschik, Scott. “New 'Common Rule' for Research.” *Inside Higher Ed*, Inside Higher Ed, 19 Jan. 2017, www.insidehighered.com/news/2017/01/19/us-issues-final-version-common-rule-research-involving-humans.

[22] Green, Robert C., Denise Lautenbach, and Amy L. McGuire. "GINA, genetic discrimination, and genomic medicine." *New England Journal of Medicine* 372.5 (2015): 397-399.

[23] Pormeister, Kärt. “Genetic Data and the Research Exemption: Is the GDPR Going Too Far?” *International Data Privacy Law*, vol. 7, no. 2, 4 May 2017, pp. 137–146., doi:10.1093/idpl/ipx006.

[24] von Thenen, Nora, Erman Ayday, and A. Ercument Cicek. "Re-identification of individuals in genomic data-sharing beacons via allele inference." *Bioinformatics* 35.3 (2018): 365-371.

[25] Lippert, Christoph, et al. “Identification of Individuals by Trait Prediction Using Whole-Genome Sequencing Data.” *PNAS*, National Academy of Sciences, 19 Sept. 2017, www.pnas.org/content/114/38/10166.

[26] Himss Analytics. “2016 Cloud Survey.” *Level(3) Communications*, 2016, www.level3.com/-/media/files/brochures/en_cloud_br_cloudsurvey.pdf.

[27] “FDA Releases Draft Guidelines to Clarify Regulatory Framework Supporting Precision Medicine.” *GenomeWeb*, GomeWeb LLC, 15 Dec. 2017, www.genomeweb.com/molecular-diagnostics/fda-releases-draft-guidelines-clarify-regulatory-framework-supporting.

[28] Freeman, R. Edward, and David L. Reed. "Stockholders and stakeholders: A new perspective on corporate governance." *California management review* 25.3 (1983): 88-106.

[29] Freeman, R. Edward, and John McVea. "A stakeholder approach to strategic management." (2001).

[30] Siniora, Dina. "CORPORATE SOCIAL RESPONSIBILITY IN THE HEALTHCARE SECTOR." *The 4th Annual Graduate Student Research Symposium.[online] Available at: https://ddc.duq.edu/gsrs/2017/proceedings/2 [Accessed 4 December 2018]*. 2017.

[31] The White House Office of the Press Secretary. “We Can't Wait: Obama Administration Unveils Blueprint for a ‘Privacy Bill of Rights’ to Protect Consumers Online.” *National Archives and Records Administration*, National Archives and Records Administration, 23 Feb. 2012, obamawhitehouse.archives.gov/the-press-office/2012/02/23/we-can-t-wait-obama-administration-unveils-blueprint-privacy-bill-rights.

[32] Sanderson, Saskia C., et al. "Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US." *The American Journal of Human Genetics* 100.3 (2017): 414-427.

[33] Ewing, Altovise T., et al. "Demographic differences in willingness to provide broad and narrow consent for biobank research." *Biopreservation and biobanking* 13.2 (2015): 98-106.

[34] “Genomic Analysis.” *Genomic Analysis - Penn State College of Medicine*, Penn State College of Medicine, 2019, med.psu.edu/core/genomics.

ACADEMIC VITA

ALLISON L. GADDIS

alligaddis@gmail.com

EDUCATION

The Pennsylvania State University, Schreyer Honors College — **University Park, PA**
Smeal College of Business | B.S., Risk Management – Enterprise Risk Management — *Class of May 2019*
Minor in Psychology

India Travel Program – Schreyer Honors College Signature Travel Experience — **New Delhi, India**
Summer 2016
- Presented quantitative data, using my statistical background, on waste management practices in the US and India
- Actively immersed myself in cross cultural student interactions and activities to build a broader global perspective

PROFESSIONAL EXPERIENCE

PricewaterhouseCoopers — **Chicago, IL**
Health Industries Advisory | Consulting Intern — *June 2018 – Aug 2018*
- Facilitated discussion during client's Global Health Foundation development workshop, and compiled consumer data and benchmarking work on foundations of similar size, monetary value, etc. while determining next steps
- Reported client's transaction data in 25 affiliate countries according to each country's timeline and governance structure
- Compiled report metrics to enable contract negotiations for a $10 million, 5 year global spend transparency program

Transamerica Financial Life Insurance Company — **Baltimore, MD**
Asset Liability Management and Hedging | Actuarial Intern — *May 2017 – Aug 2017*
- Developed a dynamic hedging program for fixed annuities in Excel using Greek measurements and other risk exposure techniques in order to pinpoint areas of cost mitigation and growth potential by making hedge adjustments
- Assessed business strategies to create a Derivative Use Plan Control Overview, which optimized management's quality and time efficiency in strategic planning of quantitative limits, derivative reporting, governance structure, and compliance
- Initiated the transfer of several billion dollars in confidential assets from a surplus portfolio to a product-based portfolio

Penn State Undergraduate Speaking Center — **University Park, PA**
Public Speaking Mentors | Tutor — *Jan 2017 – May 2019*
- Collaborate with students on content development and research across a multitude of academic and interest areas
- Provide guidance for students on delivery, rhetoric appeal, visual aids, and structure for speeches and presentations

United States Congress — **Washington, D.C.**
House of Representatives | Congressional Aide Intern — *June 2014 – July 2014*
- Directed day to day activities in the congressional office of Bennie G. Thompson – Department of Homeland Security
- Managed the front office, wrote constituent letters, and attended legislative hearings to gain government exposure

LEADERSHIP EXPERIENCE

Penn State Homecoming — **University Park, PA**
Merchandise Committee Director | Executive Council — *Nov 2017 – Dec 2018*
- Coordinated a team of 8 captains and 20 committee members to generate $13,450 in revenue at a 40% profit margin
- Established an improved data analytics standard, and realized a 35% increase in sales revenue from the prior year
- Enhanced Homecoming's brand pervasiveness by redesigning 4 existing items and introducing 2 new apparel items

University Relations Committee Captain | Involvement Liaison — *Mar 2017 – Nov 2017*
- Collaborated with 10 others to organize 25 community locations for 300 student volunteers during our Day of Service
- Lead a committee of 25 first year students, whom I interviewed and selected, to promote Homecoming events

Royalty Committee Captain | Administrative Assistant — *Mar 2016 – Nov 2016*
- Organized the nomination process and interviewed 50 applicants for 10 Royalty Court positions with 8 co-captains
- Communicated budget and event information with the other 15 committees to facilitate effective collaboration

Schreyer Honors College — **University Park, PA**
Schreyer Scholar Ambassador Program | Philanthropy Committee — *Aug 2017 – May 2019*
- Represent the Honors College as a Schreyer ambassador at our organized special events for alumni, donors, faculty, etc.
- Conspire with SAT members, Penn State associations, and Schreyer administration to promote philanthropic initiatives
- Mentored 10 first year Schreyer Scholars to ease their college transition and to promote integrity and professionalism
- Presented best practices for developing a global perspective and securing an internship as a panelist to 300 students

Penn State Dance Marathon — **University Park, PA**
Communications Committee | Intra Committee Liaison — *Oct 2015 – Feb 2019*
- Managed information booths, lost and found centers, a committee food drive, and sleep shift stations during THON
- Integrated morale enhancing activities to facilitate unity in our 30-person committee and build a more cohesive team

AWARDS AND SKILLS

- *Awards:* Girl Scout Bronze Award, Silver Award, Gold Torch Award; India Study Abroad Contrast/Cohesions Competition (2nd place)
- *Skills:* Working experience with Excel, Power-Point, and Word