Issues in Information Systems
Volume 24, Issue 3, pp. 306-320, 2023

DOI: https://doi.org/10.48009/3_iis_2024_126

Improving patient care through integration of data privacy and data access

Katherine D'Ordine, *Bryant University, kdordine29@gmail.com*
Suhong Li, *Bryant University, sli@bryant.edu*

Abstract

This research aims to understand the current relationship between data access and data privacy in the health care industry and to find a way that important health care research can still be conducted amidst HIPAA regulations. Using the data collected from a survey and supplementary Tweets data, it was found that a well-regulated process to create a partnership between medical professionals and researchers was a viable solution to work with privacy laws, and to create a cross-functional team of patients and experts across different involved fields that discusses needs and obstacles in order to overcome them. The results show that recommendation for de-identifying patient medical records for use in research was also a potential contributor to the solution. Analysis also led to the discussion of lacking interoperability in health care and the idea that poor data quality and structure pose significant problems for the advancements of medical research, with a recommendation for more interoperability among health care databases to become a priority. This study contributes to the overall knowledge of the relationship between data access and data privacy to improve patient care, and it provides steppingstones for other researchers to expand upon the numerous factors in this relatively unexplored topic.

Keywords: medical research, data privacy, data access, patient care

Introduction

During the age of data everywhere, there are many ways data can be used to significantly benefit people, but it has the potential for being misused. Ethical considerations are imperative to data and are more complex than other ethical discussions since improper use of this data can impact many aspects of the human experience. There is no authority that can give entirely correct information about the proper ways to use data, so it ends up being discerned by the people who use the data to understand the ethics surrounding it (Hand, 2018). Typically, people err on the side of caution when it comes to risk, but it is important to note that risk and benefit must have a balance. The data itself does not come with ethical considerations; it is more about the way that data is used and what types of analysis data will be subject to (Hand, 2018).

Most of these ethical considerations have to do with human subjects data. The E.U. defines this personal data as information that can be connected to a person through direct or indirect means such as through reference of identifying numbers, online presence, location information, socioeconomic details, and other factors. The General Data Protection Regulation (GDPR) states:

> "The protection of natural persons in relation to the processing of personal data is a fundamental… The right to the protection of personal data is not an absolute right… it must be considered in relation to its function in society and be balanced against other

> fundamental rights, in accordance with the principle of proportionality… processing of personal data should be designed to serve mankind." (Hand, 2018)

This part of the regulations shows the difficult decisions about personal privacy and its significance in comparison to promoting a wider good when it may not necessarily be ethical to refrain from giving out readily available data for the benefit of everyone affected (Hand, 2018).

An article written to look at literature surrounding ethical data use for future regulations and ethical discussions looked closely into the biomedical side of data with medical information being sensitive and regulated. The article works with main concerns including data protection and anonymization, who owns the data, and how much people know about their data being used. Additional concerns are listed specifically for biomedical data that are important topics of future research in this area, including:

> "The need to distinguish between 'academic' and 'commercial' Big Data practices in terms of potential harm to data subjects[,] future problems with ownership of intellectual property generated from analysis of aggregated datasets[,] and the difficulty of providing meaningful access rights to individual data subjects that lack necessary resources." (Mittelstadt & Floridi, 2016)

These concerns all relate to the usage of biomedical data and indicate that future research about the topic is imperative for any sort of ethical decisions to be made between privacy and access.When does the benefit to the greater good outweigh an individual's right to privacy of their personal health information? What if there were to be a well-regulated, structured process implemented where research students could partner with medical facilities or organizations to utilize their data for various research projects? Bringing this type of data into research for varying levels of researchers conducting studies may drive improvements in medical and mental health treatments, improving the lives of many in society, but it could also be a cause for concern regarding privacy in used incorrectly; this balance is what needs to be found.

Literature Review

Privacy, its Importance, and How Much We Actually Have

Privacy does not have a universal meaning, but in terms of health care privacy, it refers to the collection, storage, and use of information that can be considered identifiable. This includes rules about what data can be collected of an individual in the first place, as well as what that data can be used for and how the person can use their own data (Institute of Medicine, 2009). It is important to note that security and confidentiality are also used as a synonym of privacy, but they are not the same thing. Confidentiality is more about the rules for people who receive personal information and what they are allowed to disclose to third parties. Security is more about the technical measures implemented to prevent access or modification of data, prevent denial of services attacks, and physically protect the system or computer in which the data is stored (Institute of Medicine, 2009).

Another important aspect of privacy in the health care industry is whether a patient has willingly authorized specific data to be used for specific purposes. Patients may be obligated to sign off on the authorization of the data usage or feel as though they must give consent to releasing information to continue receiving treatment (Gostin & Nass, 2009). This can happen with some types of consent forms because they are designed to protect the medical facility from liability of any sort. Even if they signed off, they may not have fully understood what they signed off for considering the complex details. Privacy should be handled with

care as it also has value to society, giving people comfort in taking part in research activities and studies without their information being released to the public (Gostin & Nass, 2009).
The amount of privacy that people actually have with their medical records is much more complicated than the idea of just keeping it private. Over the course of an average hospital stay, several hundred different people may see some fragment of a patient's records. Direct access to this data is typically needed for information about the care for the patient. Those directly involved in patient care also have access to patient records in order to provide the proper treatment and have a complete understanding of what a patient needs. Other additional services in a medical setting such as labs, therapies, or radiology also need access to this data to perform their jobs safely. On the payment side of medical records, insurance companies and other third-party billings and accounts have access in order to determine how much money they should pay for specific treatments as well as to see how at risk someone is for illness (Scott, 2000). Aside from primary users of health information, secondary users sometimes have access to data for supportive services like risk management, medical schools, the support of legitimate medical facilities and individuals, and medical research. More users that may indirectly have access to health information are those who offer management, marketing, or database implementation services. Governments may also receive data to report infectious diseases or discover types of abuse within the community. They may need to disclose allergies to schools or have information in court situations for people under 18 in custody matters for example (Scott, 2000).

To switch gears to privacy regarding patients accessing their own data, medical information is now making an appearance on the internet most notably through patient portals. These allow patients to be empowered to look at their own health records to understand more about their care and challenge details when they feel something isn't right, but they also allow health care providers to access data on a patient from multiple different locations for quicker care. Even with all these policies in using patient medical records, some places still don't protect patient privacy like they state they do, nor have the proper security implemented to follow through on such claims. This is when privacy breaches may occur, with the occasional story surfacing about horrible breaches in patient medical records and may include disclosure of personal information such as STDs, mental illnesses, and genetic disorders, which can cause social, economic, or emotional distress. This can lead to discrimination in the workplace or increased insurance rates due to predisposed illness (Scott, 2000). Some privacy breaches are accidental where records were released on incorrect platforms not behind logins and passwords, but some are malicious and have targeted subjects to harm specific people. There are even some situations regarding people with legitimate access to patient records that use it for the wrong purposes, but this isn't always caught or considered a privacy violation. There is no one right answer to data privacy in the health care industry (Scott, 2000).

Health Research and its Importance

Health research is a clear investigative process that includes the creation, implementation, testing, and evaluation of results in order to create more information for future use and knowledge. Some research is classified as clinical trials where volunteers participate in studies to test different types of applications using newer medical knowledge. Research also includes data-based methods including biological samples and information from patient records. Both methods have yielded impressive information that detect patterns and form new types of medical treatments. With health care data transitioning quickly into a more electronic platform, research is very possible (Institute of Medicine, 2009).

Just like how privacy holds a high level of importance in society, so does research, as it can provide new information about illness, treatments, care practices, and societal needs. It can help determine what the largest health issues are and provide the necessary data to discover patterns and harbor knowledge that can lead to the invention of life-changing medical interventions for overall increased patient health (Gostin & Nass, 2009).

HIPAA Initiation and Data Regulations

Health care data regulations have become stricter with the implementation of the Health Insurance Portability and Accountability Act (HIPAA) in 1996. This act was created to protect sensitive information in health care so patient data could not be given out without explicit authorization from the patient. The exception to this is that data may, in very specific cases, be released for the purpose of research through the Institutional Review Board (IRB) (Department of Health Care Services, 2022). There are 18 types of data that are considered personal identifiers according to HIPAA and, if not fundamentally essential, should be removed from datasets since they are not always required in medical research. They relate to names and contact information, dates, geographic location, addresses, account and identification numbers, internet locations, and biometric identifiers (Department of Health Care Services, 2022). The HIPAA Privacy Rule aims to protect personal health information and require proper storage while still creating a line of usage in health care settings to provide the best care to the patient. However, the Privacy Rule does not apply to medical records that have been completely de-identified, and there is an insignificant risk that the data would be re-identified in some way during the usage of the data (Ness, 2007).

A national study of clinical scientists was conducted on the HIPAA Privacy Rule's impact on medical research, and it was determined that most of the surveyed individuals felt that the Privacy Rule had a significant negative impact on medical research with human subjects due to the added time, costs, and increased unknowns. Only 25% felt that the rule had increased patients' privacy in the process (Ness, 2007). Another study aimed to evaluate HIPAA regulations and the impact they pose on those applying for IRB exemptions by looking into time frames where applications were approved or denied and investigating additional factors. They found that HIPAA seemed to get in the way of research using medical databases and increased work for everyone involved in the process, even with the careful consideration of ethical uses of data. More studies ended because they were not able to meet the requirements, and it was not proven if privacy protection truly increased with this new rule. In the past, using medical records in research has been essential for developing new treatments; it was already time consuming then, and recent rules have not helped to improve that process (O'Herrin et al., 2004).

In a book about the HIPAA Privacy Rule, a committee for health research and privacy worked to develop recommendations in hopes of increasing privacy while still working to create a feasible way to conduct health research with that information. Their main suggestion was that "Congress should authorize HHS [Health and Human Services] and other relevant federal agencies to develop a new approach to ensuring privacy that would apply uniformly to all health research in the United States," (Institute of Medicine, 2009).

In a journal about changing the HIPAA Privacy Rule to improve both research and privacy, they suggest several revisions and discuss the future with these changes. They suggest that the Department of Health and Human Services (DHHS) should improve their information for privacy in research, promote de-identified data usage, be more communicative about research and its purpose, and stay consistent when it comes to research preparation and subjects (Gostin & Nass, 2009).

Another study was conducted on what patients want with the use of their personal medical information when it comes to research by surveying over 600 people with chronic medical conditions or knowledge of familial conditions about their feelings towards privacy of their medical records. This study provided three recommendations for a public policy that can provide balance between privacy and access. The first lies with those who are in research and policy creation; they need to improve their communication to patients and the public about why this kind of research is important for them or their families, and the important role medical records have in the success of this process. Second, they recommend bringing up a blanket consent form with patients for them to understand what their medical records could be used for in research.

Third, it is understood that some patients will never provide consent to the usage of their data regardless of their level of understanding of the study and how low risk a study may be. This should not be seen as a negative, but as a positive, that the policy keeps both the medical care and research programs in a trusting relationship with one another (Kass et al., 2003).

Research Objectives

This research aims to understand the current relationship between data access and data privacy in the health care industry in order to improve the relationship and enhance patient care through research. The current connection is in opposition and preventing researchers from having access to the necessary information to make advancements in the care of patients' mental and physical health while in the hands of medical professionals. There is also an element of data structure obstacles where this research aims to find ways to improve the structure to comply with privacy through a more established process of de-identifying data for usage. Understanding the data access and data privacy relationship would support the usage of necessary information for researchers, provide better health care treatment success for medical practices, and improve patient outcomes. The following questions will be explored in this study:

1. *What are the challenges regarding data collection/data access in medical fields?*
2. *What are the benefits of partnership between researchers and health care professionals?*
3. *Does partnership between researchers and medical professionals drive better access to health care data?*
4. *Does better access to health care data for analytics research drive improvement in medical and mental health treatments?*
5. *What are public perceptions/sentiment regarding data privacy/HIPAA and medical research?*

Research Methodology

To complete this research, a survey was created and delivered to various medical professionals, other people in health care, analysts, and people in research to get a variety of perspectives on the issue. The survey consists of a few demographic and career questions before entering the main portion. It asks about the subject's familiarity with patient medical records in their job and what parts of the data collection and usage process they have participated in to understand their perspective. Then it asks their agreement with various challenges regarding data collection that they have experienced as well as benefits of a well-regulated partnership process between medical professionals and researchers. Towards the end, it asks about their opinions on two recommendations for steps in this overall process, coming from current existing research.

The first recommendation asks about whether they believe a well-regulated partnership process between medical professionals and researchers would drive better access, and the second asks about whether they believe this legitimate access would actually drive improvements in patient care and treatment options and are referred to as "Research Recommendations" throughout the analysis. All scales that ask for agreement levels are on a 1 to 5 scale, with 1 meaning strongly disagree, 2 meaning disagree, 3 meaning neutral opinion, 4 meaning agree, and 5 meaning strongly agree. Finally, there is an opportunity for the subject to provide other recommendations for overcoming challenges, concerns they have about this process, and anything else they wanted to add to the survey. The goal was to get a wide variety of perspectives on the issue, gauge their knowledge on if given solutions would be successful, and see if any of their recommendations would be feasible to implement and improve the data access and data privacy relationship. The data was collected and transformed for analysis over a period from October-November of 2022.

To enhance the research, 4.1 million Tweets were collected, and basic analysis was done on a subset of these Tweets to understand the point of view and current opinions of the general public on this topic. The key words of "medical privacy" and "HIPAA" were used to filter the final dataset from the collected Tweets into a set with roughly 26,000 Tweets. The inclusion of a filter with the word "research" was used for a subset of around 1,000 Tweets for deeper analysis. A Databricks cluster was used for word clouds and to conduct a sentiment analysis to determine if anyone is talking about this issue or doing anything about it.

Discussion and Results

General Data Demographics and Survey Information

The survey was distributed over a month period from October-November of 2022, and 44 responses were received with 41 of them being usable for analysis. Nearly 50% of the participants were above the age of 50 and the data skewed significantly into the age ranges above 40 years old. The gender split was 33 females and 8 males with 0 participants identifying with any other gender. The range of careers was large with Hospital Administrator, Student in research / analytics / health care, and Nurse being the top three careers. Nearly one-third of participants selected more than one career that applied to them, and there were also several write-in responses including Physical Therapist and Insurance Administrator which added other perspectives to the results.

When asked what stages of the data process they have been part of during their current or former careers, 10 participants had been part of more than one stage, and the most prominent stage with 30 participants was the input stage before data processing as a medical professional, so nearly three quarters of the participants had interacted directly with patients in a medical setting or been part of generating the data from those patients at some point in their careers. The second most prevalent stage with 14 participants was data interpretation and presentation to stakeholders. In terms of the challenges and benefits outlined in the survey, there was a wide range of responses from all career perspectives for each which will be explored in more detail in proceeding sections.

Challenges Regarding Data Collection/Access

The six main challenges identified in this survey were about data quality and completeness, data security and exposing information, data structure and organization, knowledge of a complex data system, regulations / HIPAA, and standard procedures for entering or storing data in an electronic system. They are split into two groups to better investigate trends. In Figure 1, Group 1 is regarding the data challenges and includes data quality, data structure, knowledge, and standard procedures to the left of the black bar. Group 2 includes the remaining challenges focused on privacy and consists of data security and regulations to the right of the black bar. Overall, people seemed to agree that data quality was one of the largest challenges regarding data in their careers with very few people having a neutral option of it and a significantly higher amount of total agreement and strong agreement than with other challenges.

Previous literature indicates that privacy challenges were going to be more prominent, but nearly half of the participants disagreed to some degree that both data security and regulations were hindrances in their careers, with close to 10 people on each privacy challenge remaining neutral about it. There was more agreement with the data challenges, which shows it may not always be data privacy causing the separation between medical professionals and the conducting of health care research; it may be that the data doesn't even exist in a structured way that can be analyzed or used for research in the first place. This was surprising due to the high prevalence of health care professionals and people who had worked directly with patients in medical settings where research stated this was the most prominent obstacle. It would also be more

expected that analysts or researchers would have this type of general opinion, less so the medical professionals.

Additional Provided Concerns and Challenges

There were spaces provided to participants where they could list additional challenges and concerns regarding the current situations they face and/or regarding potential suggestions or recommendations provided in the survey. A data analyst in health care who also has experience as a hospital administrator brought up an issue with resources for producing reports containing this type of data, especially since many parts of databases are not accessible to most people and only certified personnel are allowed to write information to those official databases. This perspective shows how even if there was access to the data, it may not be structured or plentiful enough in an organized way for it to be used in research to produce meaningful results. If resources are scarce, then it is less likely to be done in a structured way. A physician and OT were also worried about a similar issue but more regarding errors in data and not having knowledge of how reliable the data input sources are. An analyst/administrator brought up an additional point about research and resources with it being very difficult to provide data to a group that may be working on a similar research effort as another group while keeping the access equal to both groups without causing any issues or competition for resources. All these concerns give way to a lot of additional research that could be done to provide suitable solutions.

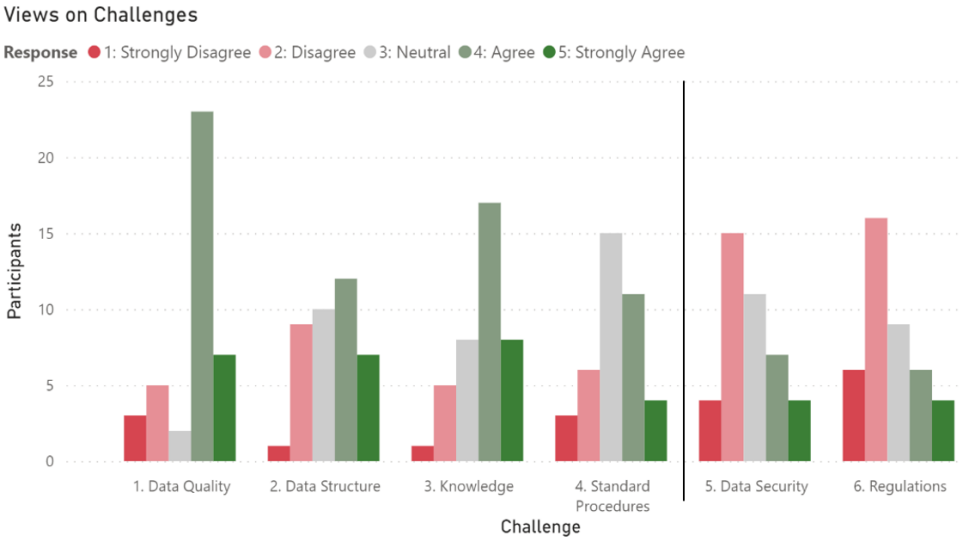

Figure 1: Challenges

On the topic of interoperability and usage of systems throughout different medical facilities, several participants in health care fields brought up concerns with its lack of presence in health care. One stated that different programs contain completely different types of data, such as for emergency departments, inpatient stays, operating rooms, outpatient visits, laboratory notes, or radiology and imaging centers. This requires data to be collected from numerous locations and joined, which begins to bring in the issue of data quality and cleanliness. Another stated that it is very difficult to find all the details needed at the right level since EMR (electronic medical record) systems have so many different styles and fields, that some do not communicate well with one another. A third person mentioned misunderstood data with the complexity that the analysis would bring and was worried it wouldn't even be possible to directly correlate that this research would truly improve patient outcomes. All these concerns speak more to the idea that privacy access is not the only hurdle in this situation; it is also the fact that the data may not even be usable for accurate results even if it were accessible.

A few concerns or other challenges that were brought up were regarding specifics in privacy. A few participants brought up the higher risk of data breaches, ransomware attacks, and leaking of individual patient information as their main concerns, and others mentioned overall confidentiality and HIPAA especially in mental health data. Two high-level professionals in health care companies brought up situations where increased access could cause poor representation in data or some demographics to be targeted due to the fact that their health information is now open to different types of sources than it was before, which are factors that would need to be managed properly if access were to be increased without a well-regulated process for specific access purposes. A participant who works in drug development mentioned that some data is proprietary to begin with, and there is no way for someone to have access to that data for any purpose yet. This may cause a few issues if access is given to someone's data, but they are part of a clinical study with data that is not accessible. Having separate systems or places where data goes may be a partial solution, but it would take a lot of time to implement effectively into busy hospital systems.

In terms of patient-facing concerns and patient interactions, a medical practice administrator felt as though having a more structured data process and more procedures for inputting data takes away from connecting with patients. If the health care professional is occupied by taking notes and inputting things properly, it may take away from the patient experience and cause them to not feel as though they are being heard and supported through their treatments. Some patients are bringing up concerns about an increase in health care data access or structure because they don't want to feel like a statistic nor feel pressured to be part of research if they feel their options are only to accept it or be "backing away from physician care." This would need to be rectified and methods would be needed to ensure that patient connection and individuality is still maintained in the process.

Perceived Benefits of Partnership Between Researchers and Medical Professionals

The participants were asked about perceived benefits of partnership between researchers and medical professionals and the results are shown in Figure 2. There are four total benefits that were included in the survey, and they are split up into two groups. Group 1 is the patient benefits regarding how a patient may be benefited by a well-regulated process between health care professionals and medical researchers for data access and includes the patience experience and the treatment options as the two benefits to the left of the black bar. Group 2 is regarding the provider benefits for this recommendation and includes provider / researcher collaboration and provider treatment success as the two benefits to the right of the black bar.

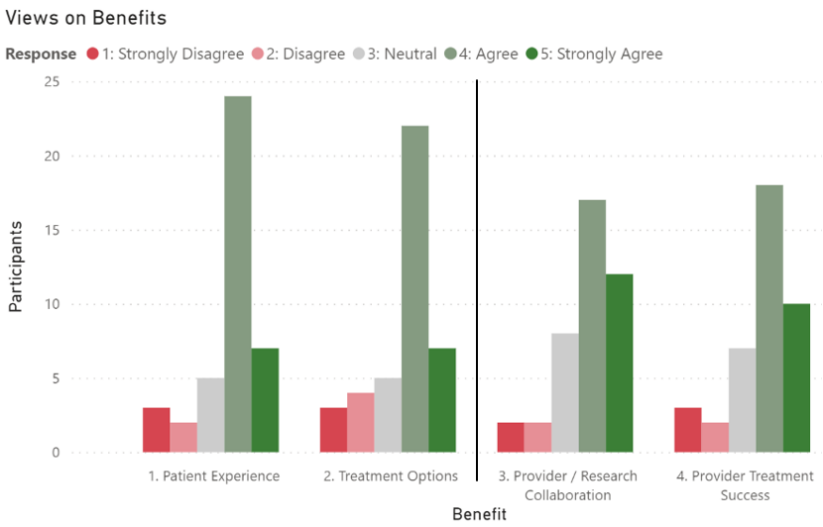

Figure 2 Benefits

Overall, most participants agreed or strongly agreed with all the benefits. Patient benefits seemed to have a high portion of people who agreed, and there were some people who strongly agreed that there were provider benefits. There were only a few people that disagreed or strongly disagreed with any of the benefits, and it spanned across several types of careers; it was not traced down to a specific category of careers. There were more neutral opinions about provider benefits than neutral opinions about patient benefits, which brings back the idea that there were a lot of patient-facing participants who took the survey and were looking out for patients and their needs.

Impact of Partnership Between Researchers and Medical Professionals on Data Access

In the survey, participants ranked how much they agreed that an increased partnership and well-regulated process between health care professionals and people in medical research would increase access to necessary health care data for research, shown in Figure 3. From research for this topic, it seemed that this might be a good suggestion, so it was important to ask for the feasibility of this. 80% either agreed or strongly agreed with a 4 or 5 respectively, with the remaining portion being neutral; no one disagreed.

Impact of Data Access on Medical and Mental Health Treatment

They were also asked about whether they thought that this increased access would actually lead to improvements in patient treatment and experience, shown in Figure 4. From research it seemed like this proper access could potentially have this effect of bringing improvements, 88% of the participants either agreed or strongly agreed with a 4 or 5 respectively, and the remaining were neutral with only 2 people disagreeing.

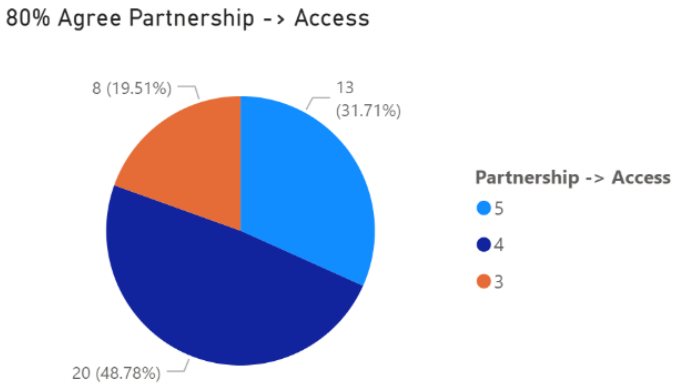

Figure 3: Partnership and Data Access

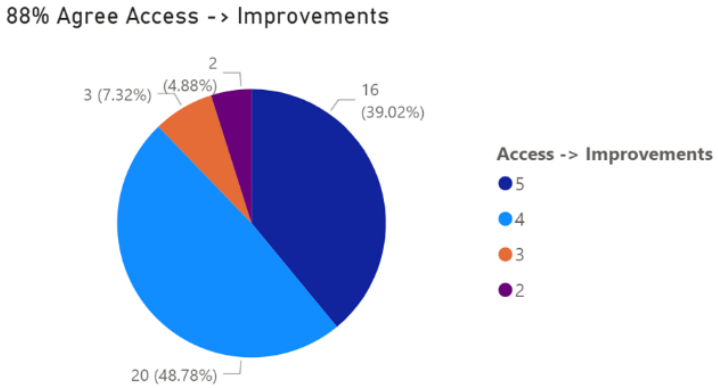

Figure 4: Data Access and Improvement in Medical/Mental Health Treatment

Additional Provided Comments and Recommendations

Before jumping into the conclusions, in the additional comments and recommendations section of the survey, a few participants chose to provide some recommendations or extra information. For example, a pharmacist who strongly agreed with most of the challenges mentioned that even they don't have access to a lot of information, it is only what insurance companies give them since they are not connected to medical offices. However, through partnerships with insurance companies, pharmacies can get more information about patient treatments and medications to "suggest additional therapies, thus making more profit for the pharmacy." Though not directly related to the questions at hand, it brings up an interesting suspicion about whether medical data is as private as it is discussed to be since it may be shared for the profit of another segment of the health care industry. That being said, it is also data that when in the right hands, could potentially improve the regimen or treatment for a patient. Since it can go both ways, this topic is something that would be very interesting to explore in future research.

A medical practice administrator stated that they agree that a collaborative effort to increase access for legitimate research would be extremely beneficial to both patients and providers / medical practices. Regarding a well-regulated and structured partnership process, one participant with knowledge from several applicable careers felt that even though it is currently tough for people in research to obtain access to the protected data they need to provide meaningful improvements to the health care industry, this type of program could be a great option and is feasible. Another person summed up an aspect of this research quite well in stating, "Risks aside, I think democratization of data accessibility can help harmonize practices and costs across the US health care system." There may even be unintended positive effects to this type of change to health care and medical research industries that may be indirect, but still significant. Another professional with several levels of health care career experience gave a general recommendation for how to go about successfully getting data for legitimate research and having it positively impact patient care. They stated:

> "The most important one I can think of is educating and informing healthcare providers on the utility and application of the data. Also, in the mental health field - you may want to work with large state-run credentialing boards to reach a broader audience of mental health workers. Many are in small to private practice and their clinical data would vary significantly from say college counseling centers, inpatient units, IOPs or community mental health centers."

This brings up an important point about who to work for especially in terms of improving mental health outcomes of patients and alludes more to the point mentioned earlier where communicating with professionals from numerous fields cross-functionally would allow experts to share knowledge and create a solution.

Secondary Twitter Analysis about Medical Privacy and HIPAA with Medical Research

A secondary analysis was conducted with Tweets to investigate the public opinion of various concepts contained in this research. The 26,000 Tweets mentioned earlier were utilized for the purposes of this analysis, all of them were relating to either "HIPAA" or "medical privacy." Figure 5 shows a time series of the total Tweets collected by month.

The number of Tweets collected across the years showed a significant upward trend when 2020 started and continued climbing even more in 2021 and 2022, showing that the COVID-19 pandemic caused a lot more discussion of medical privacy and HIPAA among the public. This was a bit surprising given that the total

number of Tweets collected in the 4.1 million decreased after 2020, so it seems as though there is a higher ratio of discussion about medical privacy in general than before; it is not just an inflation of the number of overall Tweets from the pandemic or other worldly events.

Text Analysis with Word Clouds

After removing stop words and keywords used to filter tweets (HIPAA and Medical Privacy), a word cloud was created as shown in Figure 6. It can be seen that the words that appeared the most regarding HIPAA and medical privacy were "law," the use of the "#hipaa" hashtag, "right," "health," "record," "violate," and "protect." When looking at some of the full Tweet text that contained these words, there are differing opinions of whether medical privacy and HIPAA actually protect patients, whether they are being violated in regard to certain events such as COVID-19, and what the laws actually say about patient medical record protection. Overall, it provided knowledge that people were in fact discussing these topics, but not a lot could be said yet about how people felt about medical privacy in terms of research.

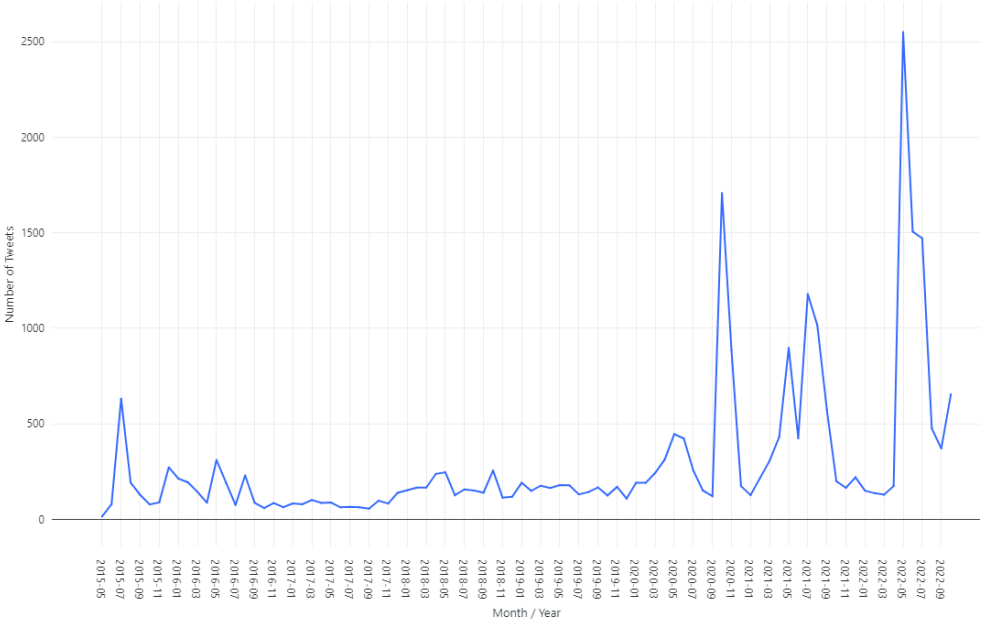

Figure 5: Time Series for Number of Tweets Discussing HIPAA or Medical Privacy

Figure 6: Overall Word Cloud

Figure 7: Earliest Most Prevalent Terms from 2015

To dive a little deeper into the word usage across different times frames, the word clouds were split by year to determine trends based on real world events. "#Hipaa" was the number one term found in Tweets from 2015 all the way through 2019, but then became much less significant from 2020 to 2022 and seemed to drop off during the start of COVID-19. Figure 7 highlights the earliest terms that were prevalent in 2015 to show later how much they changed over time.

"Law" was prevalent in the first couple of years but became the most common word in 2021 when many laws were being discussed about the pandemic and other controversial health care topics. Something interesting to note was that in 2015, the discussion seemed to be more heavily surrounding violation and invasion of privacy and the rights people had to their privacy, and 2022 focused heavily on laws and protection of health records; most other years focused on general patient record discussion. When the pandemic began in 2020, there was more discussion of respect, expectations, and wanting to talk about what was going on. In 2020, a user became the most common word in the text content of these Tweets, "@kurteichenwald" run by Kurt Eichenwald, a writer that has been extremely active in health care matters especially during COVID-19, and many people were communicating with him through Twitter about the events.

In 2022, the most prevalent words were a bit different than the collection of words from previous years, and as briefly stated before, were concerned more about rights, violations, laws, and data leaks than most other years as shown in Figure 8.

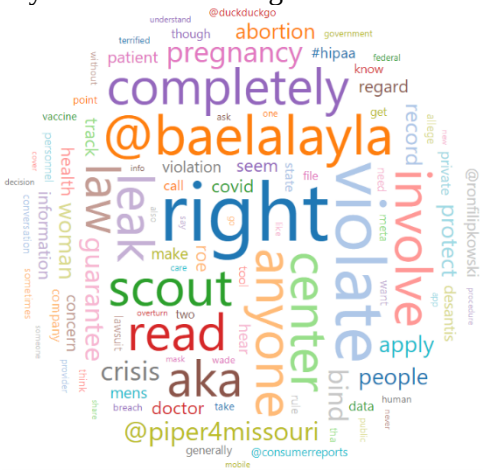

Figure 8: Latest Most Prevalent Terms from 2022

Figure 9: Most Prevalent Terms in Research Related Tweets

This change makes sense considering the number of controversial health care topics in laws and privacy being discussed. That being said, there were still no particularly telling details about research, studies, or clinical trials and their relationship to data access; nothing had appeared in word clouds or listed in any other aggregate analysis. This could mean that even though they have a lot to say about medical privacy and HIPAA, as expected by the small amount of peer reviewed research on this topic, not a lot of people have been talking about its connection to medical research and improvements in patient care, at least on Twitter.

When a specific subset of data was used that contained anything regarding research, there were less than 1,000 Tweets out of the over 26,000 from the main dataset and the result is shown in Figure 9. In addition to the hashtag "#hipaa," some of the most common words included "patient," "record," "data," "healthcare," "law," "health," "right," "use," "need," "information," "share," and "access." A portion of

these words have been most prevalent in all Tweets about medical privacy and HIPAA, but some specific to the research-geared Tweets included "need," "use," "right," "share," and "access." This is the kind of information that was being investigated, and it seems as though the people who are discussing research might be mentioning it in reference to access and the need to use or share health care data. This also means it could be in reference to people who do not believe access should be given for medical research, but a more in-depth analysis would need to be done with a significantly greater quantity of Tweets to be certain. In a cursory look at the text fields themselves, it was clear that though there were few people discussing research and privacy / access, they had very strong and differing opinions on it. Some brought up that HIPAA laws allow for research already and that they don't believe these records should be accessed for this research, while most others brought up that HIPAA has created unnecessary roadblocks to improving medical treatments and improving patient care / quality of life. This is where the sentiment analysis comes in to get a general idea of how people feel about HIPAA and medical privacy.

Sentiment Analysis

A sentiment analysis was conducted on all Tweets across 8 years, and it was found that there is slightly more emphasis on negative Tweets than positive Tweets, making up about 40% and 34% respectively, with neutral Tweets taking up just over a quarter of Tweets (26%). This was also broken down by year in Figure 10 similarly to how the word clouds were visualized, to see if specific time frames had more frequent words or strong opinions.

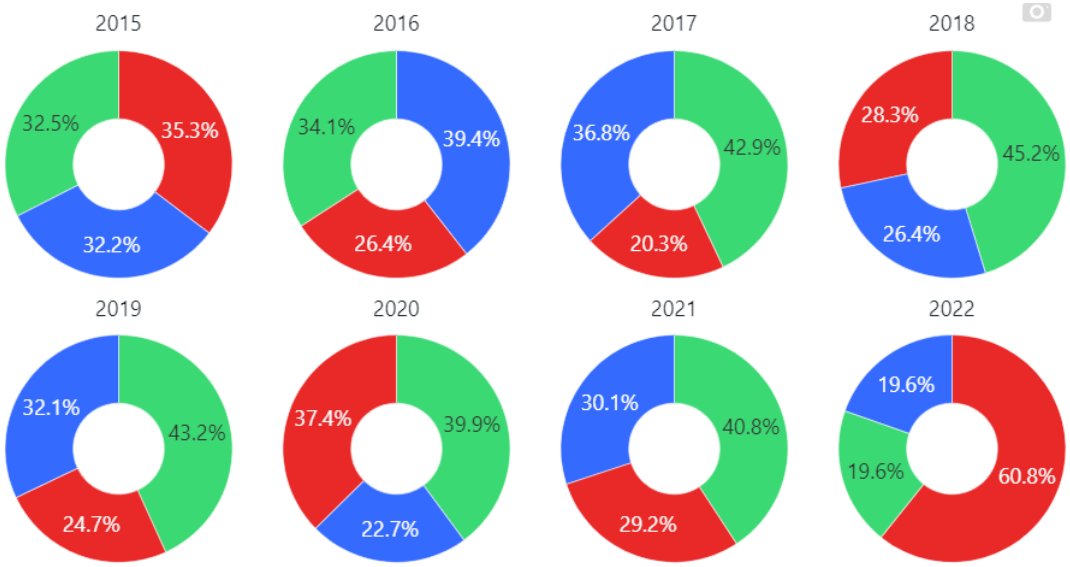

Figure 10: Sentiment by Year (Red indicates Negative, Green = Positive, Blue = Neutral_

Overall, from 2015 to 2019, the ratios were relatively split in thirds give or take a few percentage points, with positive Tweets becoming more prominent from 2017-2019. 2020 showed significantly less neutral Tweets as people became polarized by the COVID-19 pandemic, and there were more negative Tweets than were taken away from the neutral Tweet percentage, showing a significant decrease in positive opinions towards medical privacy and HIPAA. The ratios returned to a similar level in 2021 as they were in 2019, but then in 2022, the largest change occurred with over 60% of the Tweets in 2022 being negative regarding medical privacy and HIPAA laws, with an even split between positive and neutral Tweets for the remaining portion. Several controversial health care topics were prevalent in the news and media at this time which could explain the dramatic change.

Conclusions and Recommendations

The proper balance for the relationship between HIPAA/medical privacy and data access for medical research is extremely important to find, but also a difficult one with people having very strong opinions about how an improvement in its relationship should be executed if at all. It was concluded that a viable way to go about this type of massive industry change would be to have a well-regulated structured process to create a symbiotic relationship between medical professionals and researchers that primarily improves patient care and outcomes. That being said, many participants displayed concern with certain data-specific challenges, and they contributed equally if not more than data access hurdles due to data privacy in terms of hindering advancements in medical research. Without structured data and interoperability, there is no data to be shared. Without data in an analyzable state, nothing can be used in the first place; this is what needs to be fixed first. Overall, the hypotheses were proven correct, with an additional finding holding true regarding the data quality and structure challenges being more prevalent than the privacy challenges, though the privacy challenges remain prominent.

A potential solution to both being contributing obstacles would be the implementation of a process at the initial point of data collection that is more structured, interoperable with other related health care facilities, and also set up with the 18 HIPAA personal identifiers flagged in databases / data repositories where they can be removed for purposes of research and assigned a unique, unrelated identifier. This implementation process would come with the need to increase communication with anyone affected by the new systems or processes to inform them and increase their trust in the new system. Many survey participants at varying levels of knowledge of this topic and participation in these kinds of issues in their career agreed that this process would bring a variety of benefits to both patients and providers. Most also agreed that the partnership concept would lead to access, and that access would lead to improvements in medical research, which shows the viability of these recommendations as well as its potential to be effective in improving patient outcomes both physically and mentally / emotionally. These effects could span as far as mitigating health issues or helping patients manage life after medical care if they develop any form of medical PTSD. A hospital such as Massachusetts General Hospital in Boston, Massachusetts could have better structured records and partner with a research institution like the Warren Alpert Medical School of Brown University in Providence, Rhode Island to provide legitimate researchers with necessary information to increase medical knowledge and make advancements in patient treatments.

Limitations and Future Research

Due to the nature of this research, there were some limitations encountered throughout the process that may have affected the results and conclusions elaborated on above. First, there were not a significant number of informational sources or studies found that were already discussing this specific type of relationship. Second, in terms of the primary source of information, only 41 usable surveys were used to conduct analysis, and having at least 100 would have led to stronger results and clearer delineations between opinions. None of these participants were patients or patient advocates partially due to the presence of these data access issues, so that side of the argument may not have been represented as well. Third, there were not a lot of Tweets collected that discussed medical privacy and research, which in and of itself is a data point that it is not discussed by as wide a range of people as previously predicted with these topics most likely being discussed more individually unrelated to one another. Twitter is also public, and it was difficult to find out the background of the users in the dataset to find out where their opinions were stemming from, unlike in the survey data where there was a lot more information about who the person was and why they felt the way they did. As a side note, the spelling of HIPAA is sometimes incorrect and written as “HIPPA,” so any Tweets about HIPAA that were spelt incorrectly were not included in the dataset, and it could be

something to note for the future. That being said, there still may not have been enough data points in the Twitter analysis to come to substantial population wide conclusions.

Future research can take this idea and produce a more in-depth study with more questions about the data structure obstacle instead of keeping it mainly about data privacy and HIPAA, and it could even include new obstacles or general views about new solution proposals as they are created and tested. Opening this data collection to a much wider population and gaining significantly more participants responding to these issues over time would help gain more perspective and a more comprehensive idea of what is being discussed or proposed as a solution to these complex issues. Doing a more extensive Twitter analysis with more Tweets could also lead to strong results and more patterns that could also be investigated in future research.

References

Gostin, L. O., & Nass, S. (2009). Reforming the HIPAA Privacy Rule: Safeguarding Privacy and Promoting Research. *JAMA*, *301*(13), 1373–1375. https://doi.org/10.1001/jama.2009.424

Hand, D. J. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data*, *6*(3), 176–190. https://doi.org/10.1089/big.2018.0083

Kass, N. E., Natowicz, M. R., Hull, S. C., Faden, R. R., Plantinga, L., Gostin, L. O., & Slutsman, J. (2003). The Use of Medical Records in Research: What Do Patients Want? *The Journal of Law, Medicine & Ethics: A Journal of the American Society of Law, Medicine & Ethics*, *31*(3), 429–433.

Medicine, I. of, Services, B. on H. C., Policy, B. on H. S., & Rule, C. on H. R. and the P. of H. I. T. H. P. (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press.

Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In B. D. Mittelstadt & L. Floridi (Eds.), *The Ethics of Biomedical Big Data* (pp. 445–480). Springer International Publishing. https://doi.org/10.1007/978-3-319-33525-4_19

Ness, R. B., & Joint Policy Committee, S. of E., for the. (2007). Influence of the HIPAA Privacy Rule on Health Research. *JAMA*, *298*(18), 2164–2170. https://doi.org/10.1001/jama.298.18.2164

O'Herrin, J. K., Fost, N., & Kudsk, K. A. (2004). Health Insurance Portability Accountability Act (HIPAA) Regulations. *Annals of Surgery*, *239*(6), 772–778. https://doi.org/10.1097/01.sla.0000128307.98274.dc

Scott, C. (2000). Is Too Much Privacy Bad for Your Health—An Introduction to the Law, Ethics, and HIPAA Rule on Medical Privacy. *Georgia State University Law Review*, *17*(2), 481–530.

Services, D. of H. C. (2022). *List of HIPAA Identifiers*. Retrieved November 1, 2022, from https://www.dhcs.ca.gov/dataandstats/data/Pages/ListofHIPAAIdentifiers.aspx