Your Mental Health Data is For Sale!

The Impact of Data Brokers on the Exchange and Use of Mental Health Data and the Related Implications on the Personal Privacy of Depressed and Anxious Individuals

Joanne Kim
A thesis submitted to the Department of Public Policy for honors
Duke University
Durham, North Carolina
2022
Acknowledgments

A sincere thank you to my thesis advisors, Professor David Hoffman and Professor Ken Rogerson, for continuously supporting me through each milestone, providing feedback and insights, and for allowing me to freely explore my academic passions. I am also grateful for the guidance I received from Professor Jolynn Dellinger and Justin Sherman and empowered by Micalyn Struble for her constant encouragement and similar pursuit of this academic work. Additionally, a special thank you to all my loving family and friends (particularly Joy Fu and Samarth Lamba) who provided endless support during the completion of this project.

Lastly, I am humbled and forever thankful for the privacy advocates, investigative journalists, and civil liberties groups who have continued to trailblaze a future with greater consumer protections and privacy for all.
# Table of Contents

Acknowledgments 2  
Table of Contents 3  
Abstract 5  
Introduction 6  
  Defining the Privacy Issues Presented by Emerging mHealth Technologies and the Data Broker Industry 6  
  Research Questions 9  
Theoretical Framework: the Relationship between COVID-19, mHealth Apps, Data Brokers, and Personal Data 10  
  The COVID-19 Pandemic and mHealth Apps 10  
  The Limitations of HIPAA and Current Privacy Laws 13  
  Data Brokers and the Data Broker Ecosystem 16  
  Data Brokers & mHealth Data 19  
Methods 21  
  Research Design Overview 21  
Empirical Findings: a Deep Dive into the Transactional Processes, Mental Health Data for Sale, and Privacy Policies 28  
  Quasi-Experimental Approach Findings: A Review of the Transactions & Communications with Data Brokers 28  
    Initial Contact 29  
    Response Rates and Types of Responses 32  
    Modes of Communication 33  
    Controls for Determining Clients 35  
    Accuracy of Data Elements 39  
    Non-disclosure Agreements 40  
    Mentioned Privacy? 40  
    Data Use: Standard Protocol, Limitations, and Delivery 42  
    Data Elements for Sale 46  
Case Study I Findings: An Analysis of Company A’s Dataset 47  
  Inquiring About the Dataset 47  
  Payment 48  
  Accessing the Data 49
Abstract

The purpose of this thesis was to make more transparent the opaque data broker industry and its processes for selling and exchanging mental health data about depressed and anxious individuals. The research is critical as more depressed and anxious individuals utilize personal devices and software-based health tracking applications (which are not protected by HIPAA), often unknowingly putting their sensitive mental health data at risk. By initiating sales inquiries with the data brokers, purchasing a dataset, and evaluating the privacy policies of ten firms, this thesis unveiled the data broker industry and the exchange of mental health data in the data economy. Furthermore, the findings highlight the industry’s lack of business best practices, particularly concerning issues of privacy. The unregulated and black-box nature of the data broker industry, the exchange of sensitive mental health data, and the lack of clear consumer privacy protections necessitate a comprehensive federal privacy law or, at the very least, an expansion of HIPAA’s privacy protections.
Introduction

On average, nearly 1 in 5 U.S. adults suffer from a mental illness each year, and the number of people affected by depression and anxiety only increased during the pandemic ("Mental Illness"; Stieg, 2021). 42% of people surveyed by the U.S. Census Bureau reported symptoms of depression and anxiety at the height of the COVID-19 pandemic, indicating an 11% increase from 2019 (Abbott, 2021). Individuals with mental illnesses face obstacles in obtaining proper care, such as the social stigma that accompanies mental disorders or the high costs for receiving therapy ("Stigma and Discrimination"; Tresla et al., 2020). Due to such factors, nearly 50% of individuals with a mental illness do not seek treatment or receive proper care ("Stigma and Discrimination"). Notably, historically-oppressed and impoverished communities are most negatively affected by these barriers while also being the most at-risk of developing mental disorders (The State Of Mental Health In America; Tresla et al., 2020).

The surge in depression and anxiety and the limitations on attending in-person therapy sessions during the pandemic led to a parallel shift towards telehealth and mental health care apps (mHealth apps), increasing mHealth app downloads by 200% between 2019 and 2020 (Evenstad, “COVID-19: Digital Health Trends”; Martinez-Martin et al., 2020). These mHealth apps became "effective in making therapy more accessible, efficient, and portable," reducing certain barriers to receiving therapy (like transportation costs) ("ADAA Reviewed Mental Health Apps"). There are currently around 10,000 to 20,000 total available apps, most of which were very recently developed ("ADAA Reviewed Mental Health Apps"). Additionally, marginalized communities became primary users of...
mHealth apps due to their historical mistrust of healthcare systems and the general affordability of the apps. For example, 86% of surveyed LatinX patients stated an “interest in utilizing a health app,” and LatinX smartphone users, in general, were 20% more likely to use a health app than white individuals (Schueller et al., 2019).

While many mHealth apps have expanded access to mental health resources, particularly for vulnerable populations, these apps have also collected and sold sensitive mental health information for use by other companies and entities (Martinez-Martin et. al, 2020). Since mHealth apps are not covered by the Health Insurance Portability and Accountability Act (HIPAA), private companies are not legally obligated to keep their users’ data confidential. Other emerging technologies, such as wearables or social media platforms, are also not covered by HIPAA (Marbury, 2020). Consequently, mHealth apps, wearables, social media platforms, etc. (collectively referred to as “emerging mHealth technologies”) do engage in misleading privacy and data collection practices and often share users’ data with third parties without their consent (Huckvale et. al., 2015).

According to the Federal Trade Commission, data brokers are third-party entities “that collect consumers’ personal information and resell or share that information with others” (Ramirez et al., 2014). Currently, the data broker industry remains largely unregulated, allowing the industry’s players to sell billions of sensitive data records for various purposes, such as targeted advertising (Ramirez et al., 2014). Furthermore, information regarding the number and types of data brokers, their privacy and data collection policies, as well as the categories of data that they sell remains largely unknown
The impact of data brokers on the exchange and use of mHealth data is even less transparent, as indicated by the lack of academic literature and the largely anecdotal evidence presented across various lay press sources.

The lack of regulatory oversight over the relationship between emerging mHealth technologies and the data broker industry puts the individual privacy of depressed and anxious individuals, as well as those in their social spheres, at-risk (Martin, 2020). Privacy International defines privacy as a “fundamental right” that is “essential to autonomy and the protection of human dignity” (Marbury, 2020). Privacy enables consumers to establish boundaries and limit access to their physical and digital spaces, giving individuals the “ability to assert our rights” when dealing with power imbalances (“What is Privacy”, 2017). When considering health data, privacy becomes even more critical since health information can include large quantities of personally identifiable data that is “sensitive” or “potentially embarrassing” (Nass SJ et. al., 2009). Patient confidentiality is also critical for maintaining “trust between patients and medical professionals”; notably, physicians who regard patient privacy obligations more seriously are more likely to have patients who report their symptoms honestly, resulting in better care (“Patient Confidentiality in Healthcare”, 2020).

As the use of emerging mHealth technologies continues to expand and the data industry remains largely unregulated, there is a critical need to understand how our personal and sensitive mHealth data is being traded and sold by data brokers. Furthermore, it becomes imperative to “ensure accountability” and implement mechanisms
to mitigate the “unauthorized or unpredictable use of mental health data” to provide consumer protections for personal privacy (Martinez-Martin et. al, 2020). Ultimately, the purpose of this thesis is to investigate the circulation and use of mHealth data within the data broker ecosystem, discuss the related privacy implications, and consider regulatory actions that can be taken to protect our society’s most vulnerable populations.

Research Questions

- What do general sales inquiries and transactions with data broker firms entail?
- What is the impact of data brokers on the exchange and use of data about individuals suffering from depression and/or anxiety?
- To what extent do data brokers consider privacy as an industry best practice?
Theoretical Framework: the Relationship between COVID-19, mHealth Apps, Data Brokers, and Personal Data

The COVID-19 Pandemic and mHealth Apps

The most common mental disorders in the U.S. include anxiety disorders and major depressive disorder, with around 40 million American adults suffering from anxiety disorders and 16.2 million American adults suffering from major depressive disorder ("3 Most Common Mental Health Disorders in America", 2021; Koskie, 2020; "Anxiety Disorders"). While major depressive disorder is the most salient type of depression in the U.S., other types of depression include persistent depressive disorder (1.5% U.S. adults affected), bipolar disorder (2.8% U.S. adults affected), seasonal depression (5% of the U.S. population affected), postpartum depression (80% of new mothers affected), and psychotic depression (1 in 13 people worldwide experience a psychotic episode) (Koskie, 2020). Anxiety disorders can also be categorized into subtypes, including but not limited to, generalized anxiety disorder, social anxiety disorder, panic disorder, and phobias ("Anxiety Disorders"). For this thesis, all the aforementioned types of depression and anxiety were considered during the data collection process.

The COVID-19 pandemic led to greater social and psychological burdens, including “social isolation, widespread unemployment, worries over contracting the virus, insomnia, social media exposure, and the rising death toll” - all of which contributed to a surge in mental health issues (Martinez-Martin et. al, 2020). Acute stress, anxiety, depression, and
suicidality increased during the pandemic, as social distancing measures contributed to “cognitive decline, substance abuse, and other mental health problems” (“Double Jeopardy”). Historically marginalized and vulnerable communities experienced the most mental stress, placing more undue burdens on their lives (“Double Jeopardy”). These mental health inequities resulted in “devastating health, social, and economic consequences” (Ruth and Starks, 2021).

As the rate of mental health disorders increased, the use of digital mental health resources, such as mHealth apps, became more common practice during the pandemic (Torous et. al, 2018). Today, therapy sessions are being held over Zoom and more people are turning to mHealth apps for their diagnosis and care. mHealth apps have become quite popular because they often include a wide array of capabilities: they connect individuals with therapists, facilitate guided meditations, provide “mood-tracking diaries”, and walk-through “cognitive behavioral therapy exercises” (“What are Mental Health Apps”). As mHealth apps continue to redefine mental health care delivery, many psychologists and scholars have praised these digital tools for their efficiency and broader accessibility (Blumenfield and Levin-Scherz, 2020). However, others have noted the need to evaluate the security and privacy mechanisms of these apps, noting that many mHealth apps do not provide basic data protections and sell their users’ data to third parties (Galvin and DeMuro, 2020). For example, a recent study found that 33 of the 36 mHealth apps surveyed sold insights on people’s digital behavior to “advertisers or data analytics companies,” as well as shared sensitive information, such as “health diary entries,
self-reports about substance use, and usernames” (Martinez-Martin et. al, 2020). In a
different study, researchers found that of the 79 apps the United Kingdom National Health
Service certified to be “clinically safe and trustworthy,” 89% transferred information to third
parties and 66% did not encrypt the data before sharing it (Iglesias-Posadilla et al., 2017;
Torous et al., 2018, Treskes, 2016; Huckvale et al., 2015). Many other studies have shown
that mHealth apps, in particular, do not use secure connections to transport the data or fail
to “protect the integrity” of the data (Singh et al., 2016; Müthing et al., 2017). Emerging
mHealth technologies, particularly mHealth apps, do not only have scant privacy practices
in place, but the platforms are increasingly becoming the target of cyberattacks, making
them vulnerable to security breaches. For example, one study found that 91% of health
apps failed one or more cryptographic tests, allowing cybercriminals to easily expose,
tamper, or steal patient data (“Cryptographic Vulnerabilities”). Another group found that
79% of the sampled health apps shared user data, and many of the companies who
owned the apps could aggregate and re-identify this user data (Grundy et. al, 2019).

Given the lack of data privacy protections, hackers who target these mHealth apps
can access a multitude of sensitive information, including but not limited to, names,
addresses, dates of birth, Social Security numbers, allergies, medications, vital signs data,
pathology reports, test results, X-rays images, and even full medical records (“100% of
Tested”). After accessing the data, many hackers circulate and sell it in various
ecosystems, such as black markets or to other data brokers (“UCLA Health Hacked”).

These findings are highly concerning since many individuals consider mHealth
data to be more intimate and personal than other forms of data, and the exposure of one's mHealth data could have consequences on an individual's overall livelihood and well-being (Aitken et. al, 2016).

The Limitations of HIPAA and Current Privacy Laws

In 1996, the U.S. passed HIPAA to “improve the portability and accountability of health insurance coverage,” as well as combat waste, fraud, and abuse in the health insurance and healthcare delivery sectors (“HIPAA History”). Although not originally intended to protect patient privacy, the federal law was eventually amended to include the HIPAA Privacy and Security Rules in 2003 and 2005, respectively, due to a growing concern over the confidentiality of medical data (“HIPAA History”). The HIPAA Privacy Rule defines protected health information (PHI) as “any information held by a covered entity which concerns health status, the provision of healthcare, or payment for healthcare that can be linked to an individual.” The Privacy Rule also included language on how PHI should be disclosed, provided appropriate mechanisms for obtaining consent, and gave patients the “right to withhold information” (“HIPAA History”). Considering the original intentions for passing HIPAA, the law has several limitations when concerning individual privacy protections. For example, it fails to provide a private right of action (Terry and Nicholas, 2016) and it does not cover all medical records or parties that hold health information (Proskauer, 2016; Armontrout et al. 2016; Prochaska et al. 2017). HIPAA protects information stored in medical records, conversations between healthcare
professionals and patients, information about an individual in a health insurer's computer system, and an individual's medical billing data (“HIPAA for Individuals”). Institutions that are obligated to uphold HIPAA compliance include health plans, most health care providers, health care clearinghouses, and business associates (“HIPAA for Individuals”). While these entities are obligated to ensure patient confidentiality and privacy, the limited nature of HIPAA enables covered entities to sell and trade de-identified data; once all personal identifiers have been removed, the data is no longer protected by the law (“HIPAA for Individuals”). This is due to the Privacy Rule which treats personal data as “not individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding systems” (“45 CFR Parts 160 and 164”). Additionally, under current statutory provisions, covered entities that sell de-identified medical data are not obligated to obtain consent from patients (Wetsman, 2021).

HIPAA releases all other entities which are not a “covered entity” or “business associate” from its legal obligations. Categorized as “non-covered entities” (NCEs), these organizations can include but are not limited to, life insurers, private companies, applications, personal health record vendors, wearables, etc (“HIPAA Privacy Rules for Non-Covered Entities”, 2021). Other emerging NCEs, such as “peer health communities, online health management tools, and websites used for research” are also multiplying with little regulatory oversight (“Examining Oversight”). Given HIPAA’s limitations, NCEs can freely sell, exchange, and use the data they collect.

Aside from HIPAA, other federal laws, while equally limited, have been utilized as
an attempt to fill the regulatory gaps resulting from the lack of comprehensive federal privacy law. The Affordable Care Act (ACA), the Fair Credit Reporting Act (FCRA), and Gramm-Bliley Leach Act all provide some level of protection for individual privacy (Ramirez et al., 2014). For example, FCRA protects consumer data from consumer reporting agencies and considers whether the data will be utilized for credit, employment, insurance, housing, and similar eligibility determinations. However, it does not cover “the sale of consumer data for marketing and other purposes” (Ramirez et al., 2014). While laws, such as FCRA, are not directly related to the health industry, the statute still provides some guidelines for NCEs and what they may be able to do with the data they collect. Recently, California, Colorado, and Vermont enacted state privacy laws as a means of providing additional protection to consumers (Burt, 2018). While the state laws are significant for increasing data privacy protections, the laws do not fully address the regulatory void within the data economy. These legal limitations have subsequently left entire industries and the general exchange of sensitive mHealth data largely unregulated, harming the most vulnerable populations.

Data Brokers and the Data Broker Ecosystem

Data brokers are a specific type of NCE that deals extensively with the exchange, flow, and use of health data. These firms store information about an individual’s health interests based on purchasing habits, Internet browser histories, and more. For example, Datalogix (now part of Oracle) classifies people as “allergy sufferers” and “dieters.” Acxiom, another data broker, also sells data on which individuals have searched for
certain "ailments or prescriptions" ("Oracle Advertising"). According to the FTC, data brokers collect this information “for purposes such as marketing products, verifying an individual’s identity, or detecting fraud” (Ramirez et al., 2014). While most of the data broker ecosystem came into existence in the early 1980s, data brokerage firms did not produce considerable revenue until the 1990s (Stevens). Data brokers generate profits by sourcing and aggregating data and then selling certain categories of users to third parties (Wlosik, 2019). Most of these purchases are made through Internet transactions, with a very limited number of these exchanges occurring in-person (Stevens). The $200 billion industry has not stopped growing, as the world continues to generate large amounts of data (Ramirez et al., 2014). Consequently, as more data is generated, data brokers can aggregate and sell more useful insights and collections to their clients.

The FTC identified three different categories of data brokers: entities subject to the Fair Credit Reporting Act (FCRA); entities that maintain data for marketing purposes; non-FCRA covered entities that maintain data for non-marketing purposes that fall outside of the FCRA (i.e. detect fraud or locate people) (Ramirez et al., 2014). The firms can be sorted into even more specific categories, including advertising and marketing firms (collect data to create targeted ads); fraud detection data brokers (collect data to detect fraud and usually serve financial institutions); risk-mitigation data brokers (collect data to determine certain consumer offers); and people-search sites (collect data to sell data to others) (Wlosik, 2019).

To understand the industry, it is important to consider the categorization of players in the data economy, which includes first-party, second-party, and third-party data
collectors. First-party data collectors include any entities with which consumers directly share a relationship and intentionally interact. They typically gather customer feedback data, website traffic and patterns, and other survey data (Bernazzani, 2021).

Second-party data collectors include organizations that enter mutually beneficial data use agreements with first-party data collectors (Bernazzani, 2021). More simply put, it’s the secondhand use of firsthand data. Finally, third-party data collectors, such as data brokers, are entities that do not share a direct link or relationship with the customer and often sell aggregated data (Bernazzani, 2021). Notably, data brokers work closely with first-party data collectors, as most of their business exchanges deal with either selling to or buying from first-party data collectors. To illustrate this point, for example, the average person uses at least nine different apps every day and about 70% of those apps share their data with third parties (Stephen et al., 2018).

Data brokers can collect, sell, and exchange an extensive list of data records, including but not limited to: names, addresses, and contact information, demographics (race, age, occupation, and education level), life-event triggers (i.e. getting married/divorced, buying a home, sending a kid to college), hobbies, purchasing habits, salary and pay stub information, type of vehicles or property owned, public voting records and political involvement, health-related information, occupation, and income (Beckett, 2021; “Life-event Triggers”; Wlosik, 2019). The firms collect this data from several sources, including other data brokers, store loyalty card programs, hospitals, employers, government records (i.e. Florida Department of Motor Vehicles), publicly available
information, private sector companies, social media platforms, and web history (Ramirez et al., 2014). While most of the information that data brokers collect is anonymized, the aggregation and categorization of millions of patient files may make it possible to re-identify those files, posing serious privacy and security risks (Tanner, 2021).

Well-established data brokers can benefit consumers by preventing fraud, improving product offerings, and enabling identity verification (“The Data Brokers Quietly Buying and Selling Your Personal Information”). In terms of healthcare, some utilize their data collections and insights to drive better health delivery and outcomes for everyone. However, while some data brokers may have good intentions and motivations, the lack of regulation of the industry has also led to malpractice and has ultimately hurt consumers. To further explain, due to a lack of transparency (Ramirez et al., 2014) and information regarding the industry, data brokers may sustain opaque secret scoring practices that lead to discriminatory levels of service or product offerings, undermine consumer trust in the marketplace, increase the vulnerability of sensitive data, draw potentially sensitive or incorrect inferences from the aggregated data, increase digital redlining (the use of digital tools to perpetuate inequities), etc. (“The Data Brokers Quietly Buying and Selling Your Personal Information”).

**Data Brokers & mHealth Data**

When specifically examining the impact of data brokers on the use and exchange of mHealth data records, there is a lack of academic literature and quantitative evidence.
Still, many press sources and reports have noted that data brokers buy and sell health data, and more specifically, mHealth information (Tanner, 2021). Firms have begun specializing in gathering information from hospitals' and doctors' records, prescription and insurance claims, and laboratory tests ("How Data Brokers Make Money Off Your Medical Records"). Notably, many doctors, nurses, and patients are unaware that electronic health record information or lab test results may be anonymized and sold, while others are pressured to sell their patient and medical records (Farr, 2019). Much of the public is left even more in the dark, as the public remains unaware of the data broker industry or its implications on privacy.

The dominant player in the medical-data-trading industry was formerly IMS Health, as the firm recorded $2.6 billion in revenue in 2014 ("How Data Brokers Make Money Off Your Medical Records"). IMS automatically received data from the computerized records held by pharmacies, insurance companies, and other medical organizations—including federal and many state health departments. Furthermore, 75% of all retail pharmacies in the U.S. had sent some portion of their electronic records to IMS. The data that was collected and sorted by IMS was then purchased by entities such as Pfizer, which spent $12 million annually to buy health data from a variety of sources. Recently, however, Quintiles acquired IMS Health, and now, the merged entity is known as IQVIA. As of 2019, IQVIA made a revenue of $11.11 billion and currently manages more than a billion non-identified medical records ("About"). One of the firm’s products is marketed as providing information from “more than 550 suppliers including nearly 100 pharmaceutical
manufacturers" to develop key healthcare insights ("Sales Information"). IQVIA also sells a whole host of databases, including "Consumer Illness Ailment Data," "Medical Mailing List," and "Hospital Data List," among other lists and collections ("Physicians Emails & Data"). As aforementioned, IQVIA is just one of the major health data broker firms. Other big data brokers in the healthcare space, specifically, include Acxiom Healthcare, Experian Health, Optum, and EpsilonHealth - all of which have billions of data points on millions of patients and individuals. Furthermore, in the U.S. alone, there are at least 4,000 other data brokers, and the extent to which each firm deals with or sells health information is unknown (Rafter). This raises concerns over the impact of the data broker ecosystem on the exchange, flow, and use of mental health data.

**Methods**

**Research Design Overview**

The research questions were addressed by implementing a three-part methodology: the Quasi-Experimental Approach, Case Study I, and Case Study II. For the Quasi-Experimental Approach, I approached data brokers as a customer to inquire about the availability of mental health data, or more specifically, information related to depression and anxiety. Following the completion of the Quasi-Experimental Approach, I purchased a dataset from Company A\(^1\) for Case Study I. Finally, for Case Study II, I evaluated and compared the privacy policies of the “top ten most engaging” data brokers

\(^1\) Company A was kept confidential for legal compliance purposes.
(referred to as the “Top Ten”). These ten data brokers (as shown in Table 1) were included in the cohort based on the following criteria: had at least one direct call or virtual meeting with a sales representative and also initiated the data discussion phase. The data discussion phase includes having an active conversation about the relevant data elements provided by each broker, including but not limited to, the types of ailments or medication data available for purchase.

Table 1. Top 10 Most Engaging Data Brokers

<table>
<thead>
<tr>
<th>Data Broker</th>
<th>Description of Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alesco Data</td>
<td>Marketing Company: “We provide people-based audience targeting solutions coupled with cutting-edge Machine Learning algorithms…” (“Alesco Data - Big Data + Machine Learning”)</td>
</tr>
<tr>
<td>Avalere Health</td>
<td>Consulting Company: “Avalere is a healthcare consulting firm dedicated to improving healthcare.” (“Avalere Health: Leading Healthcare Consulting Firm”)</td>
</tr>
<tr>
<td>Company A</td>
<td>Marketing Company: Acquire new customers with highly targeted email, postal, and phone leads</td>
</tr>
<tr>
<td>Definitive Healthcare</td>
<td>Data Analytics Company: “We’re transforming billions of data points into meaningful answers that create paths to commercial success.” (“Healthcare Analytics &amp; Provider Data</td>
</tr>
<tr>
<td>IQVIA</td>
<td>Data Analytics Company: “IQVIA helps you drive healthcare forward by creating intelligent connections with an expansive portfolio of capabilities and technologies, unparalleled data, and global healthcare expertise.” (“IQVIA - Powering Healthcare with Connected Intelligence”)</td>
</tr>
<tr>
<td>HealthVerity</td>
<td>Data Exchange Company: “We have built a high-governance, privacy-compliant way for our partners to connect and exchange real-world data across the broadest ecosystem…” (“About us</td>
</tr>
<tr>
<td>HealthWise Data</td>
<td>Data Analytics Company: “Provider of unique data insights for the healthcare industry.” (“HealthWise Data”)</td>
</tr>
</tbody>
</table>
MedicoReach  Marketing Company: “With our accurate and comprehensive healthcare email list and medical email lists, we make sure that your messages reach the right prospect to give you better response rate.” (“Healthcare Email List - Medical Email List | MedicoReach”)


Redi-Data  Marketing Company: “Redi-Data is a leading provider of postal and email lists, data services, and direct marketing solutions. With our comprehensive databases and innovative technologies, we help our customers connect with their target audience to increase their business opportunities.” (“Email & Mailing Lists of Consumer, Business, and Healthcare | Direct Marketing & List Services”)

All three sections of the methodology were necessary for understanding the impact of the data broker ecosystem on the exchange and use of mental health data on depressed and anxious individuals.

**Quasi-Experimental Approach**

I initially contacted and approached 37 data brokers (shown in Appendix A) as a customer to better understand and make more transparent the transactions run by data broker companies. I initially identified this group of firms by googling specific search terms, including “healthcare data providers,” “mental health data brokers,” “health information for sale,” “mental health data for sale,” and “data brokers who sell mental health data.” From there, I looked through the top five search results for each search term and compiled a list of 34 data brokers. I was later referred to three other firms during the process. To submit a request for product information, I had to either submit a contact form on the broker’s website or email the company directly. Appendix A also includes an
indication of whether I emailed a firm, used a contact form, or was referred. Overall, I initially emailed five data brokers, was automatically referred to one via email and filled out a contact form for the other 31 firms. The message I submitted varied slightly based on the firm, but resembled the following:

“We are interested in learning about your data offerings. Specifically, we are hoping to look into any health and/or mental health data you may have available for purchase or use.”

To contact the data brokers, I used an email account that is owned by the Triangle Privacy Research Hub (TPRH), a non-profit research organization run by Duke faculty. I also utilized a Duke University-issued Chromebook throughout the data collection process. The device was password protected and up to date on security, anti-virus, and general updates. The device was also encryption enabled and had no other accounts with Administrator-level access. All these steps were taken in accordance with the Duke Office of Information Technology’s (OIT) recommendations. Only I have access to my datasets and all actions were taken without any deliberate deception, in accordance with Duke University’s IRB standards.

Finally, I created a “Data Elements Wish List” (shown in Appendix B), which served as a guide for requesting data during the various sales inquiries. For example, I asked the firms about datasets that included specific ailments (like depression and anxiety) and medications (like Zoloft, Lexapro, and Prozac). I also hoped to find demographic data about individuals who either identified as depressed or anxious or were related to these
individuals as healthcare providers, family members, or caregivers.

Quasi-Experimental Approach Data Questions:

- What do the exchanges and communications with the data brokers look like?
- What mental health data about depressed or anxious individuals or their providers are available for purchase?
- How much do datasets cost? And, how much does it differ between firms?
- How is the data transferred after a purchase? And, to what extent can I use, manipulate, and exchange the data after purchase?
- What internal controls, if any, do the firms have in place when deciding which clients to work with?

Data & Data Collection

I collected three types of data during my interactions with the data brokers: call/communication observational field notes, email attachments, and email messages/conversations. I only downloaded and analyzed the email attachments and exchanges for the Top Ten, as they provided the most documents and substantive information in their messages. For call/communication observational field notes, I typed important information about the company’s products, services, or operations, as well as any other interesting observations that resulted from the calls into a Microsoft Word document for each firm. Each document was labeled “[Title of the Firm] Communication Transcript/Notes” and each call was also labeled with the appropriate date.
Observational or transcription notes from the calls specifically included information on pricing, the accuracy of the data, the availability of certain datasets and data elements, the limitations of using the data post-purchase, etc. The notes also included dates and a short description of any email exchanges that took place. In terms of email attachments, I downloaded and kept any documents that were sent to me via email from the Top Ten. This included data dictionaries, product details, and pricing information. Similarly, I also downloaded my email exchanges with the Top Ten in the form of PDF documents for further analysis.

**Case Study I: Purchased Company A’s Dataset**

I purchased a dataset from Company A for data confirmation and validation purposes. To elaborate, I wanted to ensure that the data that was marketed for sale was delivered in the final product. Other general observations about the data were also made. During the sales inquiry process, Company A was most willing to sell data on depressed and anxious individuals; offer the data at my budget price of $2,500, and had no restrictive data-use limitations post-purchase. Company A also offered a more streamlined and quick payment and purchase process, which made the firm more interesting for this study.

**Data & Data Collection**

I analyzed the purchased dataset from Company A by cross-checking the data elements that I had intended to purchase to the information that was provided to me in the
Case Study II: Examining the Privacy Policies of the Top Ten

For Case Study II, I examined the Top Ten’s privacy policies to compare whether each firm had considered and implemented privacy best practices. I read through each privacy policy to map the common categories and clauses found in the policies and wrote down general observations about the policies in a spreadsheet.

Data & Data Collection

I labeled each privacy policy from the Top Ten by the appropriate date that it was downloaded, and then noted the last time a policy had been updated before reading through each document (shown in Appendix C). Common categories or clauses were aggregated and eventually compiled into a spreadsheet. Even if the language was vague, any mention of a common clause or category was still marked for each data broker. For example, if a firm was vague about their data security practices but still mentioned “data security” or “security” in their privacy policy, they earned an “x” in the spreadsheet, indicating that the privacy policy either touched on or contained some information about this category.
Empirical Findings: a Deep Dive into the Transactional Processes, Mental Health Data for Sale, and Privacy Policies

Quasi-Experimental Approach Findings: A Review of the Transactions & Communications with Data Brokers

All identifiable information about the sales representatives, including their role, name, email address, phone number, pronouns, etc. have been redacted and removed to preserve their privacy.

Initial Contact
Reached Out Via Email

The five data brokers I emailed include Adstra Health, Exactis, Ogilvy, Towerdata, and TriNetX. The email addresses I reached out to can be found in Appendix A. Interestingly, none of the data brokers I directly emailed responded.

Referrals

During the contact process, Intelius and Veeva recommended Definitive Healthcare and Redi Data, respectively. (Intelius had also recommended Experian, but I had already submitted a request to that firm.) Notably, the sales representative from Veeva voluntarily referred me to Redi Data (as seen in Figure 1), as they wrote, “I believe Redi Data would be a better fit for your needs” (Sales Representative at Veeva, Personal Correspondence, 9.27.21). In contrast, I had to prompt the Intelius sales representative for alternative options (as seen in Figure 2).
Figure 1. Veeva’s referral to Redi Data

Figure 2. Intelius’ referral to Definitive Healthcare and Experian

Acxiom Healthcare automatically referred me to Alesco Data after receiving my sales inquiry through their website. Figure 3 below shows the email notification regarding the automatic referral. The email indicates that Acxiom had categorized TPRH as a “small” or “mid-size” business opportunity and reveals that Alesco Data and Acxiom have been working together for nearly 20 years. The referral suggests that data brokers have close relationships with other information providers and that firms may research or
pre-categorize their potential customers.

*Figure 3. Automatic Referral to Alesco Data*

Hi Development,

Thank you for your interest in Axiom. Axiom has transitioned its small and mid-size business opportunities to our partner Alesco Data. This partnership means your business will continue to have access to Axiom's data and services through easy-to-use technology and expert consultative support.

You’re in good hands — Alesco Data has been a trusted Axiom partner for over 20 years and can recommend a solution that’s right for your business.

An Alesco Data representative will reach out to you shortly. If you have any immediate questions or need assistance, please don’t hesitate to contact them at 888-477-0232, leads@alesco-data.com or visit their site at www.alesco-data.com.

Sincerely,

The Axiom Team

**Filled Out Contact Form**

Each contact form usually asked for the following information: first and last name, email, phone number, company name, position/role/title, country, message (including, but not limited to “How can we help you?”, “Which products and/or services are you most interested in?”, and “Questions/Comments”). Screenshots of some of the Top Ten’s contact forms can be found in Appendix D.

Some contact forms were also unique in their request for information from the contact form. For example, several data brokers were concerned with ensuring that I was not a robot: Adfire Health required the completion of a captcha that had a simple math problem (as seen in Figure 4 below); Epsilon, Equifax, and Redi Data had an “I am not a robot” box that I had to check off. Avalere Health and Epsilon also had a clause about agreeing to each firm’s privacy policy.
Response Rates and Types of Responses

Of the 37 data brokers I contacted, 26 firms responded, resulting in approximately a 70% response rate as seen in Figure 5. Of the 26 firms, eight firms sent automated messages and 18 firms (including the Top Ten) had a sales representative directly reach out as seen in Figure 6. Of the 18 firms, 11 data brokers were willing to sell the relevant mental health data I was looking for.
**Modes of Communication**

Several different modes of communication were employed by the Top Ten brokers during each sales inquiry process. Calls were made via Zoom, Microsoft Teams, cell, and other conference dial-in options. Appendix E shows which mode(s) of communication were
utilized by each of the Top Ten firms. All sales representatives used email to some degree.

There were a few other observations made regarding communication methods. First, Medico Reach often made unsolicited calls to my cell number. As seen in Figure 7 below, illustrates how Medico Reach continued to make several calls.

*Figure 7. Some of Medico Reach’s Unsolicited Calls*

![Recents](image)

The number and frequency of calls seemed to increase if there was a delay in responding to Medico Reach’s email messages (as seen in Figure 8).

*Figure 8. Email Exchange about Communication with Medico Reach*
Another interesting point of observation was that all the data brokers which used video-enabled telecommunication tools did not reveal their faces. This deviated from standard sales calls where sales representatives are usually expected and/or incentivized to have their cameras on. Notably, PulsePoint's sales representative was the only individual to turn their camera on during the call.

Controls for Determining Clients

All Top Ten data brokers inquired about the purpose of the project and the intended use case for the data. However, while the sales representatives confirmed (either verbally or in writing) the purpose of the project, there did not seem to be additional controls for client management. To elaborate, at least in the email exchanges and calls, there was no indication that a separate background check had been conducted to confirm that what I had revealed about TPRH and my project was true.

The sales representative from HealthVerity did claim to have “done research on TPRH before the call,” but the representative did not have any knowledge about the organization (Sales Representative at HealthVerity, Personal Correspondence, 11.9.21). Relatedly, the sales representative from Alesco Data mentioned that the requested data had become “extremely restricted,” and also emphasized that her team would consequently want to know “exactly what you’re [TPRH] doing with the data and/or how you’re utilizing it” (Sales Representative at Alesco Data, Personal Correspondence, 11.10.21). However, the sales representative continued to send data samples and run counts even without a specific description.
Still, the email exchange with AccuData and Acxiom’s referral to Alesco Data (as explained above) indicates that some firms may have controls for determining which clients to work with. In the case of AccuData, the sales representative stated that their team had determined they were “not able to help you [TPRH] with our services” as seen in Figure 9 (Sales Representative at AccuData, Personal Correspondence, 10.4.21). Even after sending a follow-up to the email two days later, the sales representative did not provide any additional information or recommendations to other data providers.

Figure 9. Email Exchange with AccuData’s Sales Representative

<table>
<thead>
<tr>
<th>Re: Targeted Data Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Hoffman <a href="mailto:development@triangleprivacyhub.org">development@triangleprivacyhub.org</a></td>
</tr>
<tr>
<td>Wed 10/6/2021 12:08 AM</td>
</tr>
<tr>
<td>To: [redacted]</td>
</tr>
<tr>
<td>Hi [redacted],</td>
</tr>
<tr>
<td>Could you possibly elaborate? And, do you have any suggestions for other companies or data providers that we could contact based on our stated interests?</td>
</tr>
<tr>
<td>Thank you.</td>
</tr>
<tr>
<td>From: [redacted]</td>
</tr>
<tr>
<td>Sent: Monday, October 4, 2021 2:47 PM</td>
</tr>
<tr>
<td>To: David Hoffman <a href="mailto:development@triangleprivacyhub.org">development@triangleprivacyhub.org</a></td>
</tr>
<tr>
<td>Subject: RE: Targeted Data Solutions</td>
</tr>
<tr>
<td>HI David;</td>
</tr>
<tr>
<td>Thank you for the follow up. After reviewing your website my team has determined that we are not able to help you with our services.</td>
</tr>
</tbody>
</table>

Pricing

Overall, when comparing the Top Ten firms’ prices for the data, the price range differed substantially based on the company, product or service.
For example, Company A charges $275 per thousand records, with a minimum order of 5,000 records for “Ailment Contacts.” From there, the price per thousand records decreased, based on the number of records purchased (as seen in Figure 10).

Figure 10. Company A’s Pricing Information for Ailment Contacts

<table>
<thead>
<tr>
<th>Ailment Contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,000 records (minimum order) = $275 per thousand records</td>
</tr>
<tr>
<td>15k records or more = $260 per thousand records</td>
</tr>
<tr>
<td>25k records or more = $245 per thousand records</td>
</tr>
<tr>
<td>50k records or more = $230 per thousand records</td>
</tr>
<tr>
<td>100k records or more = $215 per thousand records</td>
</tr>
</tbody>
</table>

Medico Reach’s pricing structure was also based on volume. The cost-per-record was $0.20 per record for a total of 10,000 records and a minimum expenditure of $2,000. Similar to that of Company A, Medico Reach’s cost-per-record also decreased as the volume of requested records increased (as seen in Figure 11).

Figure 11. Medico Reach’s Pricing Information

<table>
<thead>
<tr>
<th>PRICING STRUCTURE BASED ON VOLUME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Records</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>10,000</td>
</tr>
<tr>
<td>50,000</td>
</tr>
<tr>
<td>100,000</td>
</tr>
<tr>
<td>200,000</td>
</tr>
<tr>
<td>435,780</td>
</tr>
</tbody>
</table>

Other Top Ten firms charged higher rates, with HealthWise Data quoting $20K for their annual license and HealthVerity’s sales representative estimating most of their data products to range from $75K to $100K. IQVIA also charged a higher price based on the service and product (as seen in Figure 12), quoting around $30K for their product,
Xponent®, and upwards of $100K for additional demographic data.

*Figure 12. IQVIA’s Pricing Information by Relevant Product*

Hey Joanne,

Thanks again for taking some time to meet and discuss your request this morning. Looking forward to working together.

Attached to this email you’ll find the following:

- Xponent overview- provides a view of scripts and units filled, down to the ZIP-code level (likely $30k or so depending on specifics)
- Prescriber Profiler overview- allows you to see how many scripts a physician is writing for, and all of the contact information for that physician (likely $15k-20k depending on specifics)
- Medical Claims Data overview (likely $75k+, not including any demographic info)

Let me know if you have any questions. Happy to set up any further discussions that you’d like.

Still, other Top Ten firms, like Redi Data, charged rental fees for some of their data. For a dataset that included all mental health and health professions from Redi Data’s State License Database, it would cost around $50 per thousand records with a 5,000 minimum rental. Redi Data also charged $25 to deliver the list as an Excel spreadsheet. In total, to rent all 15,378 records for a one-time mailing use would cost $793.90. Notably, Redi Data also required a prepayment for the data since “this would be your [TPRH] first order with us” (Sales Representative at Redi Data, Personal Correspondence, 11.12.21).

Some firms, such as HealthVerity and Optum, also presented the option of licensing all or just a portion of the relevant data. Optum’s sales representative wrote that, “If you are looking for a data report without licensing the data, those start just under the 6-Figure range. The typical report cost range is from $1,000 - $10,000, depending on the
complexity of your analysis” (Optum Marketing at Optum, Personal Correspondence, 10.19.21). Additionally, Company A offered a promotional code “DATA15” which would give customers 15% off their data purchase (Sales Representative at Company A, Personal Correspondence, 10.6.21).

In addition to their pricing structure, Company A also offered information on their “Additional Fees.” Interestingly, one of their additional fees included privacy protection. While it was not clear what this entailed, it cost an additional $110 to acquire.

**Accuracy of Data Elements**

The reported accuracy of each firm’s data also varied. While not all the Top Ten firms provided information about the accuracy of their products, Table 2 provides insight into the accuracy of the data from some of the brokers. In general, it was difficult to determine how each firm had calculated or determined their accuracy score, and even more challenging to identify an industry standard for what is considered to be a satisfactory accuracy score.

**Table 2. Reported Accuracy of Some of Top Ten’s Products**

<table>
<thead>
<tr>
<th>Company</th>
<th>Reported Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>HealthWise Data</td>
<td>Ailment predictions were proven 77%-85% accurate when compared against pharma data, which is uniquely strong for non-HIPAA- restricted data</td>
</tr>
<tr>
<td>HealthVerity</td>
<td>10x more accurate than competitors</td>
</tr>
<tr>
<td>LexisNexis</td>
<td>Accuracy would depend on customer’s inputs</td>
</tr>
<tr>
<td>Company A</td>
<td>Postal - 95% Email - 90% Phone - 85%</td>
</tr>
</tbody>
</table>
Non-disclosure Agreements

Two of the Top Ten firms mentioned non-disclosure agreements (NDAs) during the sales inquiry process. HealthVerity mentioned that signing an NDA would provide full access to their data exchange solution at no cost. LexisNexis barred much more information behind an NDA. A potential customer would have to sign an NDA to gain access to LexisNexis’ data attributes’ guide and to understand what data inputs the customer would have to provide to utilize the firm’s solutions.

Mentioned Privacy?

Out of the Top Ten firms, only the sales representative from Alesco Data directly mentioned “privacy” concerns. The sales representative mentioned that their products would not produce outputs for specific medications due to “HIPAA and privacy interests.” Additionally, all Alesco’s data is aggregated based on households as another means of protecting individual privacy. Interestingly, one of the requirements for purchasing data from Alesco was to have a privacy policy posted on the TPRH website. According to the sales representative, Alesco was “strict” about ensuring that their customers had a privacy policy. Alesco was in a lawsuit against a client who failed to provide a privacy policy before purchasing the data. Alesco Data is concerned with individual privacy to some degree.

While the other Top Ten firms did not directly say the word “privacy,” three other firms mentioned HIPAA during their calls. Avalere Health’s sales representative
mentioned that while they do have fully-identified patient data, they were unable to share the individual data due to HIPAA compliance concerns. Instead, the sales representative offered to aggregate the data of interest. Similar to that of Avalere, HealthVerity briefly mentioned that their “marketplace,” or platform for exchanging data processed 120 HIPAA-compliant data sources. HealthVerity’s sales representative also mentioned that de-identified data never reached the firm which contradicts the information on their website that reads “source, de-identify, and link patient data in real-time with 10x greater accuracy than industry alternatives” (Sales Representative at HealthVerity, Personal Correspondence, 11.9.21).

Unlike the other firms, HealthWise Data seemed less privacy-orientated. The sales representative mentioned that their data was not “HIPAA-protected” so it is “easy to leverage” and go around the “red tape” (Sales Representative at HealthWise Data, Personal Correspondence, 10.22.21). The sales representative also acknowledged that HIPAA-restricted data could help industries better understand the patient from a clinical perspective; however, HealthWise could help organizations utilize consumer data to effectively understand the patient as a consumer.
Data Use: Standard Protocol, Limitations, and Delivery

Sample Mailing Piece

Several Top Ten Firms either asked for a sample mailing piece and inquired whether TRPH would be directly reaching out to the individuals listed in the datasets. Alesco Data, Redi Data, and Company A required a sample mail piece. The request for sample mail pieces suggests that the firms have some controls in place for the use of their datasets.

Company A’s sales representative informed me that they could not release the dataset until a sample mail piece had been approved, yet provided payment options before the mail piece had been reviewed. I was asked to edit the first mail piece I had sent. According to the sales representative, the “sample was denied” because it included verbiage on ailments (Sales Representative at Company A, Personal Correspondence, 10.18.21). The language I utilized in both sample mail pieces was as follows:

**Mail Piece 1:** Dear [name], We are reaching out to you because someone at this residential address may have or has been formerly diagnosed with depression and/or anxiety. Please contact us if anyone in your household would be interested in being connected with community resources.

**Mail Piece 2:** Dear [name], We are reaching out to your household to see if anyone in your family would be interested in being connected with mental health community resources.

Due to the “sensitive” nature of the “Ailments and Medications” list, Company A
had three controls in place (Sales Representative at Company A, Personal Correspondence, 10.11.21). The following guidelines were provided (Sales Representative at Company A, Personal Correspondence, 10.11.21):

1. “Sample Mail Piece is required for approval to release list to you.”
2. “We can only turn over to you the name and postal list, no emails.”
3. “Emails will need to be sent through our Email Deployment and we will provide you with delivery, open and click reporting (pricing listed below).”

Data Use Limitations and Guidelines

The Top Ten firms also had different approaches concerning to what extent and how the data could be used after purchase. For example, Alesco Data allows customers to purchase the data either for “single-use” (which usually pertains to mailing purposes) or “multi-use” (which means the dataset is available for one year after purchase). After Alesco Data confirmed that TPRH was not a marketing entity, the sales representative said that as long as TPRH did not contact the individuals in the dataset, “you could do what you wanted with the data.” This indicated that once I gained access and downloaded the dataset, the extent to which I could use the data for research purposes was unlimited (Sales Representative at Alesco Data, Personal Correspondence, 10.5.21). Company A had even fewer restrictions on how the data could be used after purchase.

The data was automatically considered multi-use. The sales representative mentioned that Company A even provided access to an online portal so that I could “always access the data if the file gets corrupted” (Sales Representative at Company
A, Personal Correspondence, 11.1.21). HealthWise Data’s sales representative also mentioned that there were on using the data for “internal analytics purposes”; however, unlike Alesco Data and Company A, HealthWise Data’s sales representative specifically stated that TPRH would not be allowed to resell the data (Sales Representative at HealthWise Data, Personal Correspondence, 11.3.21). Avalere Health mentioned an interesting limitation: the sales representative stated that there were different restrictions “depending on the data.” The sales representative also advised against conducting an analysis at the brand name level (i.e., Zoloft, Lexapro, and Prozac) and publishing it without Avalere Health’s guidance and review. The sales representative mentioned that the firm does work for pharmaceutical companies, so having Avalere Health’s brand attached to any unreviewed publications could potentially “rub our clients the wrong way” (Sales Representative at Avalere Health, Personal Correspondence, 10.8.21). In general, the sales representative repetitively mentioned that Avalere Health is “committed to objectivity and data-driven outcomes from research” (Sales Representative at Avalere Health, Personal Correspondence, 10.8.21).

**Aggregation, Analytics, and Deidentified Data**

Some of the firms mentioned data aggregation, de-identified data, or additional data analytics services and capabilities. For example, Optum mentioned that their U.S.-based claims and clinical data were de-identified and could be used for research purposes (Optum Marketing at Optum, Personal Correspondence, 10.19.21). As
aforementioned, HealthVerity also included language on de-identified data, although it was difficult to understand whether the firm did or did not work with de-identified data. Still, the sales representative did mention that the firm was not a data aggregator but more of a “broker of the data.” They also mentioned assigning HealthVerity ID numbers to their data, allowing all the elements to be interoperable and linkable (Sales Representative at HealthVerity, Personal Correspondence, 11.9.21). Avalere Health also mentioned linkability and aggregation. The firm can link external datasets to medical data. For example, the sales representative mentioned being able to link social determinants of health data to claims data. The firm also offered analytical services on top of curating the datasets, which would allow customers to look for specific outputs (Sales Representative at Avalere Health, Personal Correspondence, 10.8.21).

**Formal Agreements**

Some of the firms also mentioned data use agreements or data licenses. For example, Avalere Health is under a data use agreement because it is a research-focused firm that produces public-facing research (Sales Representative at Avalere Health, Personal Correspondence, 10.8.21). HealthWise Data and Optum specifically required customers to sign a data license. Optum also mentioned that the data would not be licensed directly to consultants or individuals representing a client (Optum Marketing at Optum, Personal Correspondence, 10.19.21). HealthVerity’s sales representative informed me that if a third party needed access to the information, then they would also need to sign a data use agreement (Sales
Delivery of Data

Some of the Top Ten firms also mentioned various ways of receiving the data upon purchase. HealthWise Data usually sends their data via an FTP file but was willing to accommodate to any format I needed. Avalere Health stated that the delivery of the data would be based on the different restrictions tied to my dataset of interest.

Others, such as Alesco Data and Company A, mentioned providing the data in the form of a spreadsheet, while Definitive Healthcare would provide the data as a SaaS tool in the cloud.

Data Elements for Sale

Throughout the sales inquiry process, I was offered a lot of information on what kinds of data the firms could offer. Appendix G offers the full list of data elements advertised by each of the Top Ten firms. In general, non-medical data elements that were frequently advertised included home market value, credit score, homeowner status, marital status, ethnicity/race, net worth, names, addresses, email/phone number, food insecurity, transportation, and purchasing habits. Medical data elements that were of interest and were available included mental health facilities, anxiety, depression, PTSD, bipolar disorder, ability to pay for medical expenses, caregivers, annual exams, and lab data. Notably, many of the brokers worked with other institutions or entities to derive complete insights or to procure the data they did not have in-house. While various firms had different levels of how specific or granular the data could be, most were able to provide data at
least by household, indicating that the firms could offer a lot of data about individuals.

**Case Study I Findings: An Analysis of Company A’s Dataset**

*Inquiring About the Dataset*

Several observations were made during the sales process with Company A, which began on September 23, 2021 and ended on November 18, 2021. There was a large delay in receiving information from the sales representative at certain points of the sales process, which prompted following up for information regarding the dataset of interest. The sales representative offered samples of counts based on my stated budget of $2,500. The counts included basic traits, including geography, target (or population of interest), the expected outputs, data inclusions (i.e., name and postal address), and the quantity (“x” number of contacts).

Interestingly, at some point, the sales representative tried to reach out with a “deal” that they would “able to work out” for me. They mentioned they could double the data for an additional $500 and believed that it would be “beneficial” to my research (Sales Representative at Company A, Personal Correspondence, 10.28.21).

*Payment*

Company A offered several options for completing the payment, including by credit card, PayPal, and a wire transfer. For the pay by credit card option, Company A’s sales representative wrote that I could provide payment to her via phone or fill out the Credit Card Authorization Form. For the PayPal option, I was provided with a link to process the payment and a request for the payment confirmation. Finally, for the wire transfer, the sales representative also attached instructions for the Wire Transfer.
For my purchase, I used PayPal to process the payment. The dataset cost $2,500. After sending a screenshot of the money transfer to the sales representative, I received a purchase order to complete. No other agreements or documents were needed to complete the process. The sales representative informed me that once the purchase order was signed, a Client Service rep would email me with the log-in information to download the list. Furthermore, the sales representative stated that the list would be ready within 48 hours, indicating a quick turnaround time.

**Accessing the Data**

After signing the purchase order on November 3, 2021, I received the login information and link for the dataset on November 5, 2021. The email included instructions for accessing Company A's portal, as well as a “short video tutorial” for new users. A brochure called “How to Best Utilize Your List” was also linked to the email. The informational brochure included recommendations for optimizing the purchased dataset, such as “we recommend executing any marketing campaign within 30 days to utilize the most up-to-date data available.” The brochure also recommended, “working with a third-party vendor” to help deploy email campaigns and purchasing an updated list three to six months after receiving the first dataset. Initially, the password for the account was also my email (development@triangleprivacyresearch.org), which was not a secure means of guarding the data from being accessed by non-account holders. After creating a new password, I was able to access an “Order Summary & Download Page” which included the order date, total spent on the dataset, date fulfilled, and a “download”
button. Interestingly, by clicking the “download” button, I received a message that read “Terms and Conditions for PO 1138441 agreed from IP [redacted] on November 3, 2021” indicating that the firm had collected my IP address data.

The Dataset

I was able to download the dataset as an excel file multiple times. The spreadsheet contained counts for individuals ages 18-24 and individuals ages 65+. Both tabs included other elements, such as Race/Ethnicity ALL with zip codes from the Durham, North Carolina, area. Furthermore, counts for the ailments depression, bipolar, anxiety, panic disorder, cancer, stroke, PTSD, OCD, and personality disorder were also included, along with numbers for alcohol (which I assume indicates alcohol abuse) and the anti-depressants I was interested in (Zoloft, Lexapro, and Prozac). There was another column titled “deduped” which I assume meant verified counts.

The dataset largely differed from the order I had discussed with the sales representative. Originally, I was promised a range of 7,000 to 9,000 contacts, as well as outputs such as race, ethnicity, income, marital status, presence of children, gender, name, and postal address. However, as described, the delivered dataset did not include any specific outputs concerning these elements. Furthermore, in total, the dataset only included 1,113 “deduped” counts (457 for the 18-24 and 656 for 65+). I contacted the sales representative concerning the discrepancies between the dataset that was offered and the dataset I received, specifically noting that names and addresses were not included. The sales representative did not respond to my email.
Case Study II Findings: A Comparison of Ten Data Broker Privacy Policies

Common Elements

Several common elements emerged from sorting through the Top Ten’s privacy policies, as seen in Appendix F. All ten firms included language about their data security protocols and explicitly mentioned collecting and retaining personal data. 9 of the 10 firms also collect traffic data. 7 of the 10 firms also disclose data to third parties, while only two of those firms (HealthWise Data and IQVIA) require the third parties to adhere to their privacy standards. Consequently, the data that is shared may not be protected or granted the same level of privacy protection as offered by the data brokers. Two of the firms (Health Verity and Medico Reach) included clauses on not being able to control what third parties did with their data at all. Furthermore, only two firms (IQVIA and Avalere Health) included a consent process that would enable individuals to offer consent for any data use cases unrelated to the original purpose.

7 of the 10 firms also mentioned aggregating personal data with other sources, without a clear indication of how such aggregated datasets would be used and, in most cases, did not include many details. Additionally, only two firms (Health Wise Data and Medico Reach) included language on what kinds of data were utilized to develop their products. Notably, some firms also seem to utilize personal data for non-service or Non-product-related activities (such as providing information to employers).

While the firms collect, share, and aggregate data about individuals, the brokers seem less collectively willing to provide access and disclosure to their customers and
users. For example, only two firms (Health Wise Data and Alesco Data) included information about a formal customer complaint process and just four of the ten firms give individuals access to their personal information. While 6 data brokers do provide opt-out options, it was unclear how effective these processes would be.

Finally, most of the privacy policies included additional and explicit protections for minors and individuals in certain regions or countries (based on the relevant privacy rules in a location). Alarmingly, IQVIA mentioned potentially receiving data that is not anonymized and de-identified and Health Verity mentioned potentially selling de-identified data - both could greatly compromise individual privacy.

General Observations about the Privacy Policies

All ten of the privacy policies included a lot of vague language or clauses that did not necessarily provide enough information about each firm’s data collection, storage, and aggregation practices. Furthermore, in some cases, it was rather difficult to find a firm’s privacy policy because it was hidden in a less accessible page of the firm’s website (like a FAQ or unlinked privacy policy page).

Conclusion

By pursuing this work utilizing a three-part methodology, unique and critical insights about the data broker industry, the processes for exchanging mental health data, and the firms’ privacy policies was derived to implicate several findings. Overall, the research illustrates an inconsistent and unregulated ecosystem that moves, sells, and
aggregates sensitive mental health data in large quantities with minimal and vague privacy protections in place. Considered in its entirety, these findings emphasize the critical need for a comprehensive federal privacy law that could provide data privacy protection to all Americans.

**Primary Findings: Data Brokers and the Impact on the Privacy of Depressed and Anxious Individuals**

*Main findings from the Quasi-Experimental Approach*

Several critical insights resulted from intersecting with the data brokers as a customer. In general, the overall response rate was relatively high at 70%. However, only 11 of the 37 data brokers that were contacted sold the relevant data of interest, highlighting that some firms were inappropriately labeled or categorized within online databases.

Interestingly, some of the firms directed me to other data brokers, indicating that an opaque network of partnerships may be in place. These “partnerships” or strong relationships also seem to exist between data brokers and certain industries. For example, one of the firms was concerned with maintaining a positive relationship with the pharmaceutical industry.

Some of the firms were unwilling to pursue a sale inquiry process with me, either due to my stated scope and purpose or because they had looked into TPRH. This indicates that the firms may have some controls in place for identifying and pursuing
clients. Notably, however, some firms, such as PulsePoint and Equifax, were willing to ask their colleagues if a “data deal” could be arranged for my unique research needs.

The most common means of initially reaching the firms included filling out a contact form as found on their websites. Generally, the forms asked for basic information, such as name, email address, and reason for contact. While the initial method for contacting the firms was quite standard, the preferred mode of communication thereafter was quite broker-dependent. Zoom, Microsoft teams, calls, etc. were all used in various ways throughout the data collection process to keep in contact with the firms.

Generally, all the Top Ten firms inquired about the intent and purpose of my work; some even asked for sample mail pieces or required signing NDAs before moving forward. However, while these requests were made, many of the firms were also willing to continue the sales inquiry process and share about their services and products without a complete understanding of my intentions. While the controls for determining clients remain unclear, some of the firms did have stated guidelines and limitations in place for how the data could be used following a purchase. For example, data use agreements and data licenses were brought up during the data collection process in relation to sharing the data with third parties or utilizing the data in general.

The pricing and accuracy of the marketed datasets varied based on volume requested, specificity of the data, and the pricing model the firm maintained. The prices for data for the Top Ten firms ranged from $2,500 to $100,000. The pricing varied based on
how much data was requested or available and how specific I wanted the dataset to be. More sensitive data, such as ailments and medications, typically equated to higher prices. Furthermore, there was no standardized payment model across the Top Ten firms. Some offered subscriptions, others rentals, and still others differentiated their products as “single-use” or “multi-use.” These various models all played a role in determining the final price of the product. The lowest stated accuracy between the Top Ten firms was 77%. It was difficult to understand how these accuracy measures were determined by the various firms, especially since some companies reported their accuracy in relation to their competitors. Furthermore, while the reported accuracy was high, it remains undetermined whether these measures are correct.

Some of the most critical findings from the quasi-experimental approach dealt with the mention of (or lack of) privacy during the sales inquiry process and the discussion of de-identified data, data aggregation, and data analytics. In general, many of the firms did not directly mention privacy; however, some of the Top Ten firms were concerned with HIPAA compliance while others were focused on getting around HIPAA's “red tape. There was a spectrum regarding how much the Top Ten firms seemed to care about privacy and upholding best practices. Most seemed more concerned with pushing the sales call forward, even if I was asking for what they deemed to be “sensitive” data. In terms of de-identified data, there was a clear risk of re-identification presented by the Top Ten firms, particularly those that specifically mentioned data aggregation or data linkage practices. Even if the firms did not aggregate the data, many offered extensive data
analytics services which enabled the companies to derive unique and detailed insights about certain populations. Furthermore, my personal information seemed to be at-risk as some of the firms tracked my IP address or found my number. My browser and search engines were also altered after pursuing this work. To elaborate, a search in Google often displayed results about purchasing data or bad links.

Main findings from Case Study I

The purchased dataset did not deliver the promised elements or provide the number of contacts the sales representative from Company A had mentioned. This indicates that there could be some discrepancies not just in the delivery of the data but in how accurate and complete the datasets may be. In this case, a rather small dataset was requested, yet even so, several outputs and quantity errors were noted.

In general, the process of acquiring a dataset was quite simple. After identifying the desired data elements, the sales representative followed up with a purchase order and payment options. Per submitting the payment, the data was delivered within 48 hours via a portal set up by Company A. The simple process, as well as Company A's willingness to accommodate the dataset based on the stated budget, implies that accessing and purchasing this data can be quite easy for the general public. Notably, not only was the process simple, the portal was secured only by a password, which did not have to be complex. This called into question how secure the data was.

Main Findings from Case Study II

By evaluating the privacy policies of the Top Ten firms, it became clear that certain
business practices were more standardized across the industry than others. For example, all firms included language on their security protocols and all collected personal data. Except for one firm, all the other brokers also collected traffic data. Most of the companies also seemed to share information with third parties and had minimal to no controls in place for holding the third parties to their privacy standards. This is cause for concern, especially since most of the brokers did not offer comprehensive data privacy rights to their customers. To explain, only one firm had internal monitoring controls in place to evaluate their ability to uphold the privacy policy, and in general, only four firms gave customers access to their information. Most concerning is that eight of the brokers do aggregate data, but how and to what extent the aggregated data is used was rarely disclosed in the privacy policies.

**Limitations**

While pursuing this project, some limitations impacted the overall research. While I received funding for my research, I still faced financial constraints. Many data brokers with more specific and comprehensive datasets sold the information at a price much greater than $2,500. Consequently, I was only able to purchase one dataset with a limited number of counts. To elaborate, I was unable to acquire additional counts (as offered by the sales representative) from Company A due to my strict budget. There were also clear time constraints that affected my interactions with the data brokers. My response rates and ability to follow-up were largely affected by external commitments. Furthermore, considering the timeline for the thesis, I had to halt my communications with the brokers,
despite potentially being able to pursue other discussions.

Another limitation that arose during the data collection process was my use of an email address for TPRH. In certain cases, such as with AccuData, it is highly likely that approaching the firm as a research entity focused on privacy may have deterred them from pursuing a sale with me. Notably, even if I had not approached the firms as TPRH, contacting them for data intended for research purposes also posed an unforeseen limitation. Some of the firms, such as Equifax, were unwilling and unable to sell data because I would be using it for research rather than marketing and/or sales purposes. Finally, there was also confusion about my real name and identity, since the name associated with the inbox was “David Hoffman,” which made some interactions with a few of the firms difficult to navigate.

Policy Implications

Based on the primary findings, sensitive mental health data and other personal data are being used, circulated, and sold by data brokers. This comes as a great concern, especially since the firms are either unaware or loosely concerned about providing privacy protections and HIPAA standards. A lack of regulation and the opaque nature of the data broker industry has also allowed the firms to maintain inconsistent business practices, ranging from data quality, accuracy, de-identification, and data aggregation processes, to data procurement and storage considerations. These inconsistent practices combined with vague privacy policies point to a critical need for greater consumer protections, particularly of sensitive data, such as mental health data.
By failing to provide data privacy protections for mental health data, many vulnerable populations are at a greater risk of experiencing discrimination, social isolation, and health complications. As aforementioned, mental health disorders carry a stigma that keeps many people from seeking the appropriate care. For example, in a study of 90,000 people, 60% of the surveyed participants reported that they feared receiving care due to related stigmas (Clark, 2021). Others in the survey were also concerned with how knowledge of their mental disorders may affect employment or social life (Clark, 2021). Additionally, depressed and/or anxious individuals (the population of interest in this thesis) are more prone to commit suicide (“Does depression increase the risk for suicide?”, 2014; Hitti, 2005). Consequently, failing to protect their personal and mental health data may have detrimental and even lethal consequences.

Given these findings and the significance of protecting mental health data, comprehensive federal privacy law is long overdue. While states have begun to pass privacy laws, they are lacking in nature and only cover a specific jurisdiction (Klosowski, 2021). By enacting a comprehensive federal privacy law, all Americans will be granted baseline privacy protections and data brokers will be held more accountable for their business practices. Such a law should include provisions that allow consumers to opt-out, gain access to their information, and correct any discrepancies. Furthermore, data brokers should be obligated to be more transparent about their use and exchange of data (Fazlioglu, 2021). Notably, Congress should continue to introduce and push for the enactment of bills, such as the Information Transparency and Personal Data Control Act,
which was introduced as the first comprehensive privacy legislation in March of 2021 (Fazlioglu, 2021). Additionally, appropriate enforcement and monitoring mechanisms, such as giving the FTC more regulatory authorities on these issues, should be considered.

While a comprehensive federal privacy law would be most ideal, there are still many barriers to enacting such a significant law. Another alternative policy recommendation that could serve as a supplement in the interim would be to expand HIPAA. Some scholars have suggested expanding the definition of covered entities or business associates to include some emerging mHealth technologies, such as apps. Furthermore, the FTC should be responsible for handling the enforcement of these new conduct codes (Butler, 2017).

Overall, sensitive and intimate mental health and personal data are being sold and exchanged in a data economy that is opaque and unregulated. To protect many vulnerable populations, legal actions must be taken to either move forward a comprehensive federal privacy law or expand HIPAA to adapt to the ever-evolving data economy and eruption of mHealth technologies.

**Further Research**

Future research is necessary to make the data broker industry more transparent and to raise more awareness about the use and exchange of some of our most vulnerable personal data. Additional research should aim to overcome the stated limitations by acquiring a larger research budget, or by approaching the data brokers via a different email address and as a different entity entirely. Other work could include measuring and
evaluating the impact of interacting with data brokers on my privacy. To elaborate, it would be interesting to understand why my search engines and browsers were impacted after starting this work, as well as take a deeper dive into understanding how my personal information was stored or collected by the firms. For example, as aforementioned in the data analysis section, Company A did store my IP address. Additionally, evaluating any personal information that was collected for accuracy could provide additional insights.

Another vein of research that could prove important to consumers would be to evaluate the various opt-out processes as offered by the firms. By conducting such an analysis, it would be possible to document and make more transparent the process for opting-out as well as test whether or not the stated processes are effective. Such research could provide critical insights on whether data broker firms uphold the various components of their privacy policy in practice and serve as an opportunity to provide additional recommendations on expanding consumer protections and rights in the data economy.

Acquiring more datasets is another stream of future work that could lead to additional insights into the data broker industry. While I was able to purchase a dataset, it was only from Company A, which is notably a smaller firm in the market. Additionally, the data purchased for this thesis did not include all the promised elements. It would be interesting to purchase multiple datasets to determine the quality of the datasets provided by other firms and also evaluate to what extent and how the delivery of the data is secured. Furthermore, the research could also explore whether firms have controls in place to flag clients who purchase datasets from different data brokers within a short
period. This could help better understand whether the firms consider the potential risks and harms of data aggregation.

Finally, as of current, a limited number of scholars have explored and reported on the business practices of data brokers by directly engaging with them. I initiated this approach alongside fellow Sanford Honors student Micalyn Struble, and we encourage others to adopt this approach to discern more about the data broker industry in the future.
References


https://www.achn.net/about-access/whats-new/health-resources/3-most-common-mental-health-disorders-in-america/.


HIPAA Journal. “100% of Tested MHealth Apps Vulnerable to API Attacks,” February 16, 2021.


https://doi.org/10.1038/d41586-021-00175-z.


https://www.nami.org/About-Mental-Illness/Mental-Health-Conditions/Anxiety-Disorders.

Armontrout, James, John Torous, Matthew Fisher, Eric Drogin, and Thomas Gutheil.

https://doi.org/10.1007/s11920-016-0726-x.


https://blog.hubspot.com/service/first-party-data.


g

https://thehill.com/opinion/technology/402775-states-are-leading-the-way-on-data-


p

privacy.


crease-in-health-app-downloads-research-shows.


https://www.orcha.co.uk/media/1746/covid_report_jan_2021_final-version.pdf

“Cryptographic Vulnerabilities, Data Leakage and Other Security Breaches in

https://www.securitymagazine.com/articles/93524-cryptographic-vulnerabilities-dat
a-leakage-and-other-security-breaches-in-healthcare-apps.

Federal Trade Commission. “Data Brokers: A Call For Transparency and
Accountability: A Report of the Federal Trade Commission (May 2014),” May 27,
2014.

eral-trade-commission-may-2014.

“Data Brokers Track Everywhere You Go, but Their Days May Be Numbered |

https://techcrunch.com/2020/07/09/data-brokers-tracking/?guccounter=1&guce_ref
errer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAALTpTiQ


Farr, Christina. “Hospital Execs Say They Are Getting Flooded with Requests for Your
Health Data." CNBC, December 18, 2019.


https://www.psychiatrictimes.com/view/mental-health-disparities-among-black-ame


“IQVIA Home Solutions Sticky Hovers.” Accessed December 3,


“Mental Health Problems and Social Media Exposure during COVID-19 Outbreak.”


Riley, Charles. “Insurance Giant Anthem Hit by Massive Data Breach.”

CNNMoney, February 4, 2015.


https://www.cnbc.com/2021/10/10/depression-increased-during-covid-pandemic-how-to-feel-better-cope.html.


Tanner, Adam. “How Data Brokers Make Money Off Your Medical Records.”


https://doi.org/10.1038/scientificamerican0216-26.


“That Mental Health App Might Share Your Data without Telling You - The Verge.”


“The Data Brokers Quietly Buying and Selling Your Personal Information.”


https://doi.org/10.1136/eb-2018-102891.


https://doi.org/10.1080/17434440.2016.1218277.


# Appendix

## Appendix A. List of Contacted Data Brokers

<table>
<thead>
<tr>
<th>Data Broker</th>
<th>Method of Initial Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(emailed v. contact form v. referral)</td>
</tr>
<tr>
<td>AccuData</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Acxiom Healthcare</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Adfire Health</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Adstra Health</td>
<td>Emailed (<a href="mailto:connect@adstradata.com">connect@adstradata.com</a>)</td>
</tr>
<tr>
<td>Alesco Data</td>
<td>Automatically referred</td>
</tr>
<tr>
<td>Allscripts</td>
<td>Contact Form</td>
</tr>
<tr>
<td>AnalyticsIQ</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Avalere Health</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Cerner</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Company A</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Crossix (Veeva)</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Definitive Healthcare</td>
<td>Contact Form &amp; Referred</td>
</tr>
<tr>
<td>Epsilon</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Equifax</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Exactis</td>
<td>Emailed (<a href="mailto:Success@Exactis.com">Success@Exactis.com</a>)</td>
</tr>
<tr>
<td>Experian Health</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Health Catalyst</td>
<td>Contact Form</td>
</tr>
<tr>
<td>HealthVerity</td>
<td>Contact Form</td>
</tr>
<tr>
<td>HealthWise Data</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Company</td>
<td>Contact Details</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>IHS Markit</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Imprivata PatientSecure</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Intelius</td>
<td>Contact Form</td>
</tr>
<tr>
<td>IQVIA</td>
<td>Contact Form</td>
</tr>
<tr>
<td>LexisNexis</td>
<td>Contact Form</td>
</tr>
<tr>
<td>MedicoReach</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Ogilvy</td>
<td>Emailed (<a href="mailto:nick.cavarra@ogilvy.com">nick.cavarra@ogilvy.com</a>)</td>
</tr>
<tr>
<td>Optum</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Oracle Health</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Pulsepoint</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Sisense</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Towerdata</td>
<td>Emailed (<a href="mailto:sales@towerdata.com">sales@towerdata.com</a>)</td>
</tr>
<tr>
<td>TransUnion</td>
<td>Contact Form</td>
</tr>
<tr>
<td>TriNetX</td>
<td>Emailed (<a href="mailto:join@trinetx.com">join@trinetx.com</a>)</td>
</tr>
<tr>
<td>Truven Health</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Veradigm</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Verisk</td>
<td>Contact Form</td>
</tr>
<tr>
<td>Redi Data</td>
<td>Contact Form &amp; Referred</td>
</tr>
</tbody>
</table>
Appendix B. Data Elements Wish List

- **Target Audience**
  - B2C, focused on mental health patients who have depression or anxiety
  - Would also like information on mental health care providers or physicians
- **Channel of Contact I am looking for:**
  - Postal/email/phone
- **Location of Interest:**
  - United States
  - Would like a subset of data to come specifically from Durham, NC
- **Budget for the Campaign →** $2500

**Other Specifications**

- **Ailment 1: Depression (and by its subtypes if possible)**
  - Major Depression
  - Persistent Depressive Disorder
  - Bipolar Disorder
  - Psychotic Depression
- **Ailment 2: Anxiety (and by its subtypes if possible)**
  - Generalized anxiety disorder
  - Panic disorder
  - Social anxiety disorder
  - Agoraphobia
  - Separation anxiety
- **Demographics**
  - Age (would like to focus more on younger populations, 18-24 and below 18; other age groups of interest → 40+, 65+)
  - Race/Ethnicity
  - Income
  - Marital status + children
  - Gender
- **Top 3 most popular antidepressants:** sertraline (Zoloft), escitalopram (Lexapro), and fluoxetine (Prozac)
- **Recently visited/has been in mental health facility or institution**
- **Therapy status - are they receiving a form of therapy?**
- **Related healthcare professionals, caregivers, or family members**
- **Other Ailments/Comorbidities:** Cancer; Stroke; Acute coronary syndrome; Drug and/or alcohol dependence; Post-Traumatic Stress Disorder; Obsessive-Compulsive Disorder; Personality Disorder
## Appendix C. Top Ten’s Privacy Policy Download Date & Version

<table>
<thead>
<tr>
<th>Data Broker Firm</th>
<th>Privacy Policy Download Date &amp; Last Updated (if provided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alesco Data</td>
<td>Download Date: 10/24/21&lt;br&gt;Last Updated: Not Provided</td>
</tr>
<tr>
<td>Avalere Health</td>
<td>Download Date: 10/24/21&lt;br&gt;Last Updated: 2/23/21</td>
</tr>
<tr>
<td>Company A</td>
<td>Download Date: 10/24/21&lt;br&gt;Last Updated: 6/17/21</td>
</tr>
<tr>
<td>Definitive Healthcare</td>
<td>Download Date: 11/18/21&lt;br&gt;Last Updated: 4/1/21</td>
</tr>
<tr>
<td>IQVIA</td>
<td>Download Date: 11/18/21&lt;br&gt;Last Updated: 7/22/21</td>
</tr>
<tr>
<td>HealthVerity</td>
<td>Download Date: 11/18/21&lt;br&gt;Last Updated: 10/6/20</td>
</tr>
<tr>
<td>HealthWise Data</td>
<td>Download Date: 10/24/21&lt;br&gt;Last Updated: 8/23/19</td>
</tr>
<tr>
<td>MedicoReach</td>
<td>Download Date: 11/18/21&lt;br&gt;Last Updated: 12/1/2020</td>
</tr>
<tr>
<td>LexisNexis</td>
<td>Download Date: 11/18/21&lt;br&gt;Last Updated: 8/18/21</td>
</tr>
<tr>
<td>Redi-Data</td>
<td>Download Date: 11/18/21&lt;br&gt;Version/Last Updated: Not Provided</td>
</tr>
</tbody>
</table>
Appendix D. Some of the Top Ten’s Contact Forms

Alesco Data’s contact form

Contact our team

TPRH Development

Analyst Triangle Privacy Research Hub

development@triangleprivacyhub.org Phone

Reason For Contacting Us*

Get in touch with sales

How can we help you?

We are interested in learning about your data offerings. Specifically, we are hoping to look into any health and/or mental health data you may have available for purchase or use.

Contact Us

Definitive Healthcare’s contact form
Contact us

If you have a question about the HealthVerity technology platform, or if you'd just like to say hello.

1818 Market Street
Suite 700
Philadelphia, PA 19103

info@healthverity.com
(267) 262-6776

Development
development@triangleprivacyhub.org

Triangle Privacy Research Hub
We are interested in learning about your data offerings. Specifically, we are hoping to look into any health and/or mental health data you may have available for purchase or use.

Send
## Appendix E. Top Ten’s Modes of Communication

<table>
<thead>
<tr>
<th>Mode of Communication</th>
<th>Data Broker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>HealthWise Data, HealthVerity, and Definitive Healthcare</td>
</tr>
<tr>
<td>Microsoft Teams</td>
<td>Avalere Health, IQVIA, Redi Data, and LexisNexis</td>
</tr>
<tr>
<td>Cell/Google Voice</td>
<td>Company A and MedicoReach</td>
</tr>
<tr>
<td>Alternative Conference Dial-In</td>
<td>Alesco Data</td>
</tr>
</tbody>
</table>
### Appendix F. The Common Elements Found in the Top Ten’s Privacy Policies

<table>
<thead>
<tr>
<th>Common Elements</th>
<th>Number of Firms Who Included the Common Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Google Analytics</td>
<td>2</td>
</tr>
<tr>
<td>Mentions Internal Monitoring/Evaluation Processes</td>
<td>1</td>
</tr>
<tr>
<td>Disclosures about Changing the Policy</td>
<td>8</td>
</tr>
<tr>
<td>Mentions Accuracy of the Data</td>
<td>2</td>
</tr>
<tr>
<td>Includes a Customer Complaint Process</td>
<td>2</td>
</tr>
<tr>
<td>Gives Customer Access to their Personal Information</td>
<td>4</td>
</tr>
<tr>
<td>Includes an Acceptable Use Policy (i.e., ethical and moral use of the data)</td>
<td>2</td>
</tr>
<tr>
<td>Describes the Data Used to Develop their Products</td>
<td>2</td>
</tr>
<tr>
<td>Intends to be Explicit about Collecting Personal Data</td>
<td>1</td>
</tr>
<tr>
<td>Collect Data through Automated Technologies</td>
<td>1</td>
</tr>
<tr>
<td>Collect Personal Data</td>
<td>10</td>
</tr>
<tr>
<td>Collect Traffic Data (i.e., IP addresses or device information)</td>
<td>9</td>
</tr>
<tr>
<td>May Aggregate Personal Data with Other Sources</td>
<td>7</td>
</tr>
<tr>
<td>Mentions Regional and/or International Privacy Laws</td>
<td>8</td>
</tr>
<tr>
<td>Uses Personal Data to Provide a Service</td>
<td>6</td>
</tr>
<tr>
<td>Uses Personal Data for Non-Service Activities (i.e., advertising)</td>
<td>4</td>
</tr>
<tr>
<td>Feature Description</td>
<td>Count</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Individuals can Offer Consent for Data Use</td>
<td>2</td>
</tr>
<tr>
<td>Cases Unrelated to the Original Purpose</td>
<td></td>
</tr>
<tr>
<td>Discloses Data to Third Parties</td>
<td>8</td>
</tr>
<tr>
<td>Does Not Have any Explicit Control Over Third Parties</td>
<td>2</td>
</tr>
<tr>
<td>Third Parties Must Adheres to the Same Privacy Standards</td>
<td>2</td>
</tr>
<tr>
<td>Enables Targeted Advertisements</td>
<td>4</td>
</tr>
<tr>
<td>Includes Opt-Out Processes</td>
<td>6</td>
</tr>
<tr>
<td>May Work with Data that is Not Anonymized or De-identified</td>
<td>1</td>
</tr>
<tr>
<td>May Sell Deidentified Data</td>
<td>1</td>
</tr>
<tr>
<td>Includes a Policy Protecting Minors</td>
<td>6</td>
</tr>
<tr>
<td>Includes a Data Retention Clause for Storing Personal Data</td>
<td>4</td>
</tr>
<tr>
<td>Provides Language on Consumer Rights and Choices</td>
<td>5</td>
</tr>
<tr>
<td>Mentions Data Security Protocol</td>
<td>10</td>
</tr>
</tbody>
</table>
## Appendix G. Top Ten’s Data Elements

<table>
<thead>
<tr>
<th>Data Broker</th>
<th>Data Elements</th>
</tr>
</thead>
</table>
| Alesco Data               | - Home market value, credit score, homeowner status, date of birth, income, age, interests (timeshare), marital status, number of people in the household, neighborhood characteristics, ethnicity, net worth, etc.  
                          | - Ailments and data about children  
                          | - Mental health data available: depression, attention disorder, insomnia, anxiety, ADHD, treatments (medication for ADHD/ADD), antidepressants, and bipolar disorder  
                          | - Provided “Sample Records - Ailments” spreadsheet which had real last names, first names, residential addresses, city name, state, and zip code                                                                 |
| Avalere Health            | - Medical claim for every patient  
                          | - 300 health plans  
                          | - Medicare/Medicaid claims data  
                          | - Age bands  
                          | - 62 billion medical events, 336 million unique patients, 580,000 clinical facilities, and 50 states + DC & Puerto Rico                                                                                           |
| Definitive Healthcare     | - 2 million physician profiles, physician groups (general practice, addiction specialists, etc.), clinics, and imaging centers  
                          | - Track over 200 technologies being used inside facilities (i.e., when they installed an EHR, installed across all sites, software use, etc.)  
                          | - Work with 4 largest clearinghouses (blue cross/blue shield/Aetna)  
                          | - Average cost per procedure                                                                                                                   |
| Company A                                                                 | • Can provide B2C (individual) or B2B data concerning health or mental health data  
|                                                                          | • Names, addresses, emails, and phone numbers  
|                                                                          | • Can filter by illness, facility, medication, etc.  
|                                                                          | • Has over 10,000 filters including household income, age, gender, children in the home, address, and specific medications  
|                                                                          | • Anxiety Sufferers List  
|                                                                          | • Consumers with Clinical Depression in the United States List  
| IQVIA                                                                    | • Prescriber information (name, PI number, DEA number, age, email address, telephone number, allowed to contact or not, etc.)  
|                                                                          | • Age, gender, and zip code  
|                                                                          | • Subnational Database - monthly breakdown of how many scripts of each medication were filled in the zip code  
|                                                                          | • 5.1 billion medical and Rx claims representing 312 million patients (95% of US population) (medical claims data brochure - 11.12.21)  
| HealthVerity                                                             | • 75 different data providers and partnerships  
|                                                                          | • Medical claims sources  
|                                                                          | • Pharmacy sources  
|                                                                          | • EMR (electronic medical records)  
|                                                                          | • Hospital data (400+ health systems and hospitals + procedures tied to billing code)  
|                                                                          | • Lab data (genomic data)  
|                                                                          | • Imaging and biospecimen source  
|                                                                          | • Consumer data  
|                                                                          | • Social Determinants of Health - anything that happens to a patient outside the hospital  
|                                                                          | o 2,000 different attributes (demographics)  
|                                                                          | o 3 large sources  
|                                                                          | • Grocery source (see what's in the shopping court)  
|                                                                          | • Mortality data  

| HealthWise Data  | Consumer Data (base consumer file consisting of 250 million US adults with contact information and social determinants of health attributes) included some data elements, such as:  
  ● Religion  
  ● Language  
  ● Self-Improvement  
  ● Single Parent  
  ● Exact Age  
  ● Presence of Children  
  ● Retail Purchases  

HealthWise360 (proprietary individual-level health and wellness attributes of full 250 million universe) included some data, such as:  
  ● Likely to engage in risky health behaviors  
  ● Likelihood of having depression centiles  
  ● Likelihood of having depression score  
  ● Likelihood of having anxiety centiles  
  ● Likelihood of having anxiety score  
  ● Likelihood of living in a high crime area  
  ● Likelihood of being food insecure or living in a food desert  
  ● Likelihood of having access to transportation  

pH Segmentation (overview of their “pH personas for health” segmentation system)  

WELLth Segmentation (overview of their “wellth” ability to pay segmentation system)  
  ● “Millennials ages 18-42 years, with low ability to pay for medical expenses”  
  ● “Gen X and Young Boomers ages 43-64 years, with a medium ability to pay for medical expenses”  

| SDOH+8 Data | (overview of their SDOH+8 bundle)  

| Medico Reach |  
  ● Has information on pharmacists and B2B data  

| LexisNexis |  
  ● 442 clinically validated attributes  
  ● Not health record-based; just based on attributes  


| Redi Data | • Provides up to four scores -- readmission risk score, motivation score, medication adherence score, and total cost risk score (how much would they cost to the healthcare system over a few months)  
| | • Can provide counts for physicians who have prescribed sertraline/Zoloft, escitalopram/Lexapro, or fluoxetine/Prozac in the past quarter |