Duke SANFORD | CYBER POLICY PROGRAM

Data Brokers and the Sale of Americans' Mental Health Data

The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy

By Joanne Kim
February 2023

—

Overview: This report includes findings from a two-month-long study of data brokers and data on U.S. individuals' mental health conditions. The report aims to make more transparent the 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 (many of which are not protected by the Health Insurance Portability and Accountability Act), often unknowingly putting their sensitive mental health data at risk. This report finds that the industry appears to lack a set of best practices for handling individuals' mental health data, particularly in the areas of privacy and buyer vetting. It finds that there are data brokers which advertise and are willing and able to sell data concerning Americans' highly sensitive mental health information. It concludes by arguing that the largely unregulated and black-box nature of the data broker industry, its buying and selling of sensitive mental health data, and the lack of clear consumer privacy protections in the U.S. necessitate a comprehensive federal privacy law or, at the very least, an expansion of HIPAA's privacy protections alongside bans on the sale of mental health data on the open market.

Author: Joanne Kim was a researcher at Duke University's Technology Policy Lab.

Publication Note: The author's views are her own. These views do not necessarily represent those of the Duke University Sanford School of Public Policy or Duke University. This report was adapted from the author's senior undergraduate thesis at Duke University and was produced intellectually independently as part of the Duke Cyber Policy Program data brokerage research team.[1]

Overview of Problem and Current Legal/Policy Gaps

Problem: One in five U.S. adults suffers from a mental illness each year, and the number of Americans affected by depression and anxiety only increased during the COVID-19 pandemic.[2] 42% of people surveyed by the U.S. Census Bureau reported symptoms of depression and anxiety at the height of the pandemic, indicating an 11% increase from 2019.[3] This made a bad situation worse. Individuals with mental illnesses already faced obstacles in obtaining proper care, such as financial constraints and social stigmas.[4] Additionally, historically oppressed and impoverished communities are most negatively affected by these barriers—and are also at the highest risk of developing mental disorders.[5]

The surge in depression and anxiety and the limitations on accessing in-person therapy sessions during the pandemic led to a shift towards telehealth and software application alternatives (mHealth apps), increasing mHealth app downloads by 200% between 2019 and 2020.[6] mHealth apps became "effective in making therapy more accessible, efficient, and portable," reducing certain barriers to receiving therapy, like transportation costs.[7] Marginalized communities became primary users of mHealth 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.[8]

While many mHealth apps have expanded access to care, particularly for vulnerable populations, these apps have also sold sensitive mental health information for use by other companies and entities.[9] Since most mHealth apps are not covered by the Health Insurance Portability and Accountability Act (HIPAA), which only applies to certain covered health entities, many private companies operating mHealth apps are not legally obligated by HIPAA to keep their users' data confidential. Companies that produce and maintain other technologies that might collect health data, such as wearables or social media platforms, are also often not covered by HIPAA.[10] Consequently, mHealth apps, wearables, social media platforms, and many other technology companies (collectively referred to as "emerging mHealth technologies") can most often legally share, license, and sell users' health data (in addition to other data) to third parties without users' knowledge or consent.[11]

Current Legal and Policy Gaps: Data brokers, according to one of multiple definitions, are companies "that collect consumers' personal information and resell or share that information with others."[12] The industry remains largely unregulated for most sales of data, allowing brokers to sell billions of sensitive data records for various purposes, ranging from profiling individuals for targeted advertising to secretly assessing individuals' health costs.[13] Information regarding the number and types of data brokers, their privacy and data collection policies, as well as the categories of data they sell remains largely unknown, even as investigative reporting and civil society research have provided more insight into data brokers' sale of data to law enforcement, health insurance providers, and even criminal scammers, among others.[14] The impact of data brokers on the exchange and use of mHealth data is even less transparent, as indicated by the lack of literature and the largely anecdotal evidence in the public domain.

The lack of regulation of the data broker industry's direct and indirect relationships with emerging mHealth technology companies threatens the individual privacy of depressed and anxious individuals, as well as the privacy of those in their social spheres.[15] Privacy International defines privacy as a "fundamental right" that is "essential to autonomy and the protection of human dignity."[16] Privacy enables consumers to establish boundaries and limit access to their physical and digital spaces, giving individuals the "ability to assert [their] rights" when dealing with power imbalances.[17] When considering mental health data, protecting individual privacy becomes even more critical as health information can include large quantities of personally identifiable data that is highly "sensitive" or "potentially embarrassing."[18] The information could also be used for economic exploitation or racial profiling, among other harms.[19]

Previous research by Duke University has identified data brokers advertising highly sensitive data on hundreds of millions of Americans, including their sensitive demographic information, political preferences and beliefs, and whereabouts and real-time GPS locations, as well as data on students, first responders, government employees, and current and former members of the U.S. military.[20] This report builds on that work to better understand data brokers' selling of Americans' highly sensitive mental health information. Such research is additionally imperative in the health space, in order to "ensure accountability" to implement mechanisms that mitigate the "unauthorized or unpredictable use of mental health data" and provide comprehensive consumer protections for personal privacy.[21]

Policy Response: Ultimately, the purpose of this research 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 society's most vulnerable populations. The nation is in dire need of a comprehensive federal privacy law, and this report recommends that the federal government should also consider generally banning the sale of mental health data on the open market.

Research Findings

Overview:

- Some data brokers are marketing highly sensitive data on individuals' mental health conditions on the open market, with seemingly minimal vetting of customers and seemingly few controls on the use of purchased data.
- 26 of the 37 contacted data brokers responded to inquiries about mental health data, and 11 firms were ultimately willing and able to sell the requested mental health data.
- Whether this data will be deidentified or aggregated is also often unclear, and many of the studied data brokers at least seem to imply that they have the capabilities to provide identifiable data.
- The 10 most engaged data brokers asked about the purpose of the purchase and the intended use cases for the data; however, after receiving that information (verbally or in writing) from the author, those companies did not appear to have additional controls for client management, and there was no indication in emails and phone calls that they had conducted separate background checks to corroborate the author's (non-deceptive) statements.
- The 10 most engaged brokers advertised highly sensitive mental health data on Americans including data on those with depression, attention disorder, insomnia, anxiety, ADHD, and bipolar disorder as well as data on ethnicity, age, gender, zip code, religion, children in the home, marital status, net worth, credit score, date of birth, and single parent status.
- Pricing for mental health information varied: one data broker charged $275 for 5,000 aggregated counts of Americans' mental health records, while other firms charged upwards of $75,000 or $100,000 a year for subscription/licensing access to data that included information on individuals' mental health conditions.
- One company that the author was in contact with depicted their firm as an advertising tech firm. The sales representative offered to ask their manager about coordinating a data deal on information from organizations they advertise for on behalf of the author.
- Data broker 1 emphasized that the requested data on individuals' mental health conditions was "extremely restricted" and that their team would need more information on intended use cases—yet continued to send a sample of aggregated, deidentified data counts.
- After data broker 1 confirmed that the author was not part of a marketing entity, the sales representative said that as long as the author did not contact the individuals in the dataset, the author could use the data freely.
- Data broker 2 implied they may have fully identified patient data, but said they were unable to share this individual-level data due to HIPAA compliance concerns. Instead, the sales representative offered to aggregate the data of interest in a deidentified form.
- Data broker 4 was the most willing to sell data on depressed and anxious individuals at the author's budget price of $2,500 and stated no apparent, restrictive data-use limitations post-purchase.

- Data broker 4 advertised highly sensitive mental health data to the author, including names and postal addresses of individuals with depression, bipolar disorder, anxiety issues, panic disorder, cancer, PTSD, OCD, and personality disorder, as well as individuals who have had strokes and data on those people's races and ethnicities.
- Two data brokers, data broker 6 and data broker 9, mentioned nondisclosure agreements (NDAs) in their communications, and data broker 9 indicated that signing an NDA was a prerequisite for obtaining access to information on the data it sells.
- Data broker 8 often made unsolicited calls to the author's personal cell. If the author was delayed in responding to an email from data broker 8, the frequency of calls seemed to increase.
- Some brokers imposed data use limitations on the possible sale of people's mental health information, ranging from "single-use" (which usually pertains to mailing purposes) to "multi-use" (which means the dataset is available for one year after purchase) based on the firm and the product purchased.
- Based on an evaluation of privacy policies, data brokers seem collectively less willing to provide access and disclosure to their customers and users about the collection or correction of personal data.

Research Methodology: This research had two main parts: Phase I, contacting data brokers and initiating sales inquiry processes; and Phase II, examining the privacy policies of the 10 most engaged data brokers.

The author initially contacted and approached 34 data brokers based on results from Google searches, such as "healthcare data providers," "mental health data brokers," "health information for sale," "mental health data for sale," and "data brokers who sell mental health data." The author was later referred to three additional firms during the process. To request information on the data that each broker held on individuals, the author submitted a contact form on the broker's website or emailed the company directly, identifying herself as a researcher. Overall, the author emailed five data brokers, was automatically referred to an additional firm via email, and filled out a contact form for 31 firms. The message that the author 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."*

10 data brokers were included in the study's primary cohort based on the following criteria: the author had at least one direct call or virtual meeting with a sales representative from the firm, the firm advertised the relevant data of interest, and the author successfully completed a discussion about the firms' relevant data services and products. Notably, the data brokers in the study upheld different obligations on confidentiality, limiting the author's ability to discuss each broker's individual processes by name. Consequently, throughout the report, the firms of interest have been numbered as "data broker 1", "data broker 2," etc. to maintain anonymity. All identifiable information about the data broker sales representatives—

including their roles, names, email addresses, phone numbers, pronouns, etc.—have also been redacted and removed to preserve their privacy.

The author also drafted a "Data Elements Wish List," which served as a guide for requesting data during the various sales inquiries (shown in Appendix A). For example, the author asked the firms for datasets that included specific ailments (like depression and anxiety) and common antidepressants (like Zoloft, Lexapro, and Prozac).

The author collected three types of data during interactions with the data brokers: call/communication observational field notes, email attachments, and email messages/conversations. The author only downloaded and analyzed the email attachments and exchanges for the 10 most engaging firms, as they provided the most documents and substantive information in their messages. Attachments included data dictionaries, product brochures, and pricing information. The author's observational and transcription notes included information on pricing, the alleged accuracy of the firms' products, the availability of certain datasets and data elements, the limitations of using the data post-purchase, and other characteristics.

In accordance with Duke University's Institutional Review Board (IRB) process, the author received IRB review and exemption of her project before initiating any contact with data brokers, and the author kept the university's IRB informed of the research throughout the project. The author managed the data collection process from a Duke University-provided device, which was configured in accordance with the Duke Office of Information Technology's security recommendations, including strong password protection, up-to-date general and security updates, and active antivirus technology. The author was the only one that had access to the datasets acquired during the course of this research, and the author took all research actions without any deliberate deception, in accordance with Duke University's IRB standards.

Phase I: Contacting Brokers and Initiating Sales Inquiries

Initial Contact: The five data brokers the author directly emailed did not respond. Two other firms made referrals to other data brokers via email. Notably, one data broker referred the author to data broker 10 of their own volition, with the belief that data broker 10 would more effectively cater to the author's data needs.[22] By contrast, the author had to ask a different firm for alternative options after the data broker identified itself as more of a services firm than a data provider or data aggregator.[23]

Another data broker automatically referred the author to data broker 1 after receiving the sales inquiry through their website. The data broker had categorized the author's inquiry as a small or mid-size business opportunity and revealed that the firm had sustained a working relationship with data broker 1 for over 20 years. The fact of the referral suggests that some data brokers have close relationships with other information providers and that some firms may research or profile their potential clients.

Each contact form typically required the following: first and last name, email, phone number, company name, position/role/title, country, and a message (in response to such questions as "How can we help you?", "Which products and/or services are you most interested in?", and "Questions/Comments"). Several brokers were concerned with ensuring that the author was not a robot, and others included a clause about agreeing to their privacy policies.

Data Broker Responses: Of the 37 contacted data brokers, 26 firms responded, resulting in a response rate of approximately 70%. Of the 26 respondent firms, eight firms sent automated messages and 18 firms (including the top most engaged firms) had a sales representative directly reach out. Overall, 11 data brokers were willing and supposedly able to sell the relevant mental health data the author was looking for, and only 10 of those 11 brokers sustained their communications with the author.

The 10 most engaged data brokers used several different modes of communication, including calls made via Zoom, Microsoft Teams, cell phones, and other conference dial-in options. All sales representatives used email to some degree. Notably, data broker 8 often made unsolicited calls to the author's personal cell. If the author was delayed in responding to an email from data broker 8, the frequency of calls seemed to increase. All broker representatives who used video-enabled telecommunication tools did not reveal their faces. This deviated from standard sales call norms where sales representatives are usually expected and/or incentivized to have their cameras on. Of all the contacted firms, a sales representative from one firm was the only individual to turn their camera on. Notably, the sales representative clarified that the company was an advertising tech firm and not necessarily in the business of selling data; however, the sales representative was willing to ask their manager about potentially arranging a data deal with the information they do acquire from the organizations they advertise for (such as hospitals and pharmaceutical companies). They were also willing to recommend other data providers, if needed.[24]

Controls for Determining Clients: All 10 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. In the email exchanges and calls, there was typically no indication that a separate background check had been conducted to confirm the author's (non-deceptive) statements about this work.

The sales representative from data broker 6 did claim to have done research on the author before the call, but the representative did not seem to have any knowledge about the author's work.[25] The sales representative from data broker 1 mentioned that the requested data had become extremely restricted, and also emphasized that their team would consequently want to know the specifics of the author's work (such as how the data would be used).[26] However, the sales representative continued to send aggregated, deidentified samples on relevant data counts even before the author sent a specific description about the intended use of the data.

Still, a few email exchanges indicate that some firms may have controls for determining which clients to work with. In the case of one data broker, the sales representative stated that their team could not assist the author with her sales inquiry.[27] The author sent a follow-up in response, inquiring about the broker's reason for refusing to engage with the author and also requested recommendations for other firms to contact, but was not provided with any additional information or recommendations about other data providers.

The Cost of Buying Americans' Mental Health Data: When comparing the 10 firms' pricing models, the cost differed substantially based on the company, product, or service. For example, one firm charged $275 per 1,000 individuals' aggregated 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.

Another firm's pricing structure was also based on volume. The cost-per-record was $0.20 per record for a total of 10,000 aggregated records and a minimum expenditure of $2,000. Further, the firm's cost-per-record also decreased as the volume of requested records increased. For example, for 435,780 records, the cost per record was $0.06.

Other firms charged higher rates, with one broker quoting $20,000 for their annual license. The firm later updated their quote and offered their data license at a discounted price of $15,000. Another firm's products ranged from $75,000 to $100,000, while a different broker charged higher prices based on the service and product, quoting $30,000 for a product that provided counts on how many times a specific medication had been filled in a specific area and upwards of $100,000 for additional demographic data. Notably, some sales representatives were willing to adjust the price to some degree.

One firm charged rental fees for some of their data. A dataset that included all mental health and health professionals with verified addresses would cost around $50 per thousand records with a 5,000 minimum rental. The firm 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, the broker also required a prepayment for the data since it was the author's first purchase with the company.[28] Some firms also presented the option of licensing all or just a portion of the relevant data. One data broker stated that purchasing a report without licensing the data would cost a little below the 6-figure range. The typical report cost range was anywhere from $1,000 - $10,000, depending on the complexity of the analysis.[29] Additionally, one firm offered a promotional code on their website which would give customers a percentage off their data purchase.[30] The broker also offered information on their "Additional Fees," which included privacy protection. While it was not clear what this entailed, it cost an additional $110.

Notably, the data brokers did not always provide a full, transparent explanation about all their data products and services, making it difficult to understand whether a firm was offering data that was supposedly "deidentified" (e.g., without a name listed), more identifiable, or some degree of both types of data. Still, some firms clearly advertised data already directly linked to individuals, as they offered individual names, addresses, and various forms of contact information (such as phone numbers and emails) in a dataset.

Accuracy of Data Elements: The reported accuracy of each firm's data also varied. While not all 10 firms provided information about the accuracy of their products, some provided insight into the accuracy of their data (as shown in Table 1). In general, it was not possible to determine how each firm had calculated or determined their accuracy score, and it was even more challenging to identify whether there was an industry standard for what is considered a satisfactory accuracy score, and if so, its standard method of calculation.

Table 1: Company-Reported Accuracy of Some of Data Brokers' Data

Company	**Broker's Data Accuracy Claim**
Data Broker 3	93% deliverability rate for emails
Data Broker 4	Postal data – 95%; Email data – 90%; Phone data – 85%
Data Broker 6	10x more accurate than competitors
Data Broker 7	Ailment predictions were proven 77%-85% accurate when compared against pharma data, which is uniquely strong for non-HIPAA-restricted data
Data Broker 9	Accuracy would depend on customer's data inputs

NDAs: Data brokers 6 and 9 mentioned nondisclosure agreements during the sales inquiry process. Data broker 6 mentioned that signing the NDA would provide full access to their data exchange solution at no cost. Data broker 9 stated that signing the NDA would give a potential customer access to their data attributes' guide and help them to understand what data inputs the customer would have to provide to utilize the firm's solutions.

Sales Representatives' Mentions of Privacy: Out of the 10 data brokers, only the sales representative from data broker 1 explicitly mentioned privacy concerns. The sales representative mentioned that their products would not produce outputs for specific medications due to HIPAA and privacy interests. They also stated that all of the broker's data is aggregated based on households by address as another means of protecting individual privacy. Interestingly, one of the requirements for purchasing data from data broker 1 was to have a privacy policy posted on the purchasing organization's website. According to the sales representative, data broker 1 is strict about their clients having a privacy policy. In fact, at that time, data broker 1 was in a lawsuit against a client who failed to provide a privacy policy before purchasing data.

Three other firms mentioned HIPAA during the calls. Data broker 2's sales representative mentioned that while they do have fully identified patient data, they were unable to share individual data due to HIPAA compliance concerns. Instead, the sales representative offered to aggregate the data of interest. It was unclear how and to what degree the data would be aggregated. Data broker 6 also briefly mentioned that their marketplace, or platform for exchanging data, processed 120 HIPAA-compliant data sources. According to the sales representative, deidentified data never reached the firm. According to the firm's website users of the marketplace can, however, source, de-identify, and link patient data in real-time with 10 times greater accuracy than competing platforms.[31]

Data broker 7 seemed less privacy oriented. The sales representative mentioned that because their data was not considered HIPAA-protected, it was easier to leverage and go around the "red tape."[32] The sales representative also acknowledged that the firm could help organizations utilize consumer data to effectively understand the patient as a consumer.

Broker Controls on the Use of Mental Health Data: Some brokers appeared to have controls in place around the sale, licensing, or transacting of mental health data, while others did not appear to have controls in place or at least did not enforce the controls that may have existed. Table 2 summarizes these findings.

Table 2: Apparent Data Broker Controls on Transacting in Mental Health Data

Company	Broker's Apparent Controls
Data broker 1	• Stated it typically requires sample mailing piece • Asked about how the data will be used for research, the end goal, and any information that would provide insight into the project—information which would then be

	shared with the compliance department
Data broker 2	• Inquired about the purpose of the author's work
Data broker 3	• Inquired about the purpose of the author's work
Data broker 4	• Required sample mailing piece; the piece was reviewed and edited based on guidelines from the sales representative • Inquired about the purpose of the author's work; specifically asked about the author's target audience, method of contact (postal, email, or phone), location (country, state, county, city, zip code, and radius), and budget
Data broker 5	• Inquired about the purpose of the author's work
Data broker 6	• Inquired about the purpose of the author's work • Author could sign NDA to gain view access to the platform-based solution at no cost
Data broker 7	• Inquired about the purpose of the author's work
Data broker 8	• Inquired about the purpose of the author's work
Data broker 9	• Inquired about the purpose of the author's work • Author could sign NDA to gain more information about the firm's products
Data broker 10	• Inquired about the purpose of the author's work • Required prepayment before receiving the data since it was the author's first time working with the broker • Required sample mail piece

Note: The information listed above reflects apparent controls in place, based on what was communicated, enforced, or seemingly enforced in an observable way.

Sample Mailing Piece: Some of the 10 firms either asked for a sample mailing piece or inquired whether the author would be directly reaching out to the individuals listed in the datasets. Specifically, data brokers 1, 10, and 4 required a sample mail piece. The request for sample mail pieces suggests that the firms have some controls in place before selling their data.

Data broker 4's sales representative informed the author that they could not release the dataset until a sample mail piece had been approved, even though the author's intended use for the data was for research purposes only. Interestingly, however, the sales representative provided payment options before the mail piece had been completely reviewed. Then, the author was asked by data broker 4 to edit the first mail piece. According to the sales representative, the sample was denied because it included verbiage on ailments.[33] The language used in both sample mail pieces that the author submitted 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, data broker 4 had three controls in place:[34]

1. A sample mail piece is required for approval
2. Data broker 4 could only send names and postal lists, not emails
3. If needed, emails could be sent through their email deployment system

Notably, while some of the firms required a sample mailing piece, there did not seem to be additional controls to ensure that the author (or, in a general case, the data buyer) would be using the language they submitted or that they would act based on the intentions stated in the sample letter. The lack of follow-up and vetting controls in place suggests that malicious actors or clients could use the data in unstated ways or easily lie about their intentions.

Data Use Limitations and Guidelines: The 10 firms also had different approaches concerning to what extent and how the data could be used after purchase. For example, data broker 1 allows customers to purchase the data either for single-use (which usually pertains to mailing campaign purposes) or multi-use (which means the dataset is available for one year or some contracted time after purchase). After data broker 1 confirmed that the author was not part of a marketing entity, the sales representative said that as long as the author did not contact the individuals in the dataset, the author could use the data freely.[35]

Data broker 4's products of interest were automatically considered multi-use. The sales representative mentioned that data broker 4 even provided access to an online portal so that the author could always redownload the data if the original file was corrupted.[36] Data broker 7's sales representative also mentioned that the data could be used for internal analytics purposes; however, unlike data brokers 1 and 4, data broker 7's sales representative specifically stated that the author would not be allowed to resell the data.[37] Data broker 2 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 the firm's guidance and review. The sales representative mentioned that the firm has partnerships with pharmaceutical companies, so having data broker 2's brand attached to any unreviewed publications could potentially "rub our clients the wrong way."[38] In general, the sales representative repeatedly mentioned that data broker 2 is committed to objectivity and data-driven outcomes from research.[39]

Aggregation, Analytics, and Deidentified Data: Some of the firms mentioned data aggregation, deidentified data, or additional data analytics services and capabilities. For example, one firm stated that their U.S.-based claims and clinical data were "deidentified" and could be used for research purposes. As aforementioned, data broker 6 also included language on deidentified data, although it remained unclear whether the firm did or did not directly work with data they would describe as "deidentified." Notably, the sales representative mentioned that the firm was not a data aggregator but more of a broker of the data. They also discussed the firm's process of assigning ID numbers to their data, allowing all the elements to be interoperable and linkable.[40] Data broker 2 also mentioned link-ability and aggregation. The firm can supposedly 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 in addition to curating the datasets, which would allow customers to look for specific outputs.[41]

Formal Agreements: Some of the firms mentioned data use agreements or data licenses. For example, data broker 2 conducts business under data use agreements because it often produces public-facing research.[42] Data broker 7 specifically requires customers to sign a data license. Data broker 6's sales representative stated that if a third party needed access to the information, then they would also need to sign a data use agreement.[43]

Delivery of Data: Some of the 10 firms also mentioned various ways of delivering the data upon purchase. Data broker 7 usually sends its data via a File Transfer Protocol (FTP) but was willing to accommodate any format. Data broker 2 stated that the delivery of the data would be based on the different restrictions tied to the dataset of interest. Others, such as data brokers 1 and 4, mentioned providing the data in the form of a spreadsheet, while data broker 3 would provide the data as a Software-as-a-Service (SaaS) tool in the cloud.

Data Elements Advertised During Sales Process: Appendix B offers the full list of data elements advertised by each of the 10 firms. In general, the data brokers advertised non-medical data elements on individuals including home market value, credit score, homeowner status, marital status, ethnicity/race, net worth, name, address, profession, and email/phone number, as well as information on food insecurity, transportation, and detailed purchasing habits. The data brokers also advertised medical data on individuals, including information on mental health facilities, anxiety, depression, PTSD, bipolar disorder, ability to pay for medical expenses, caregivers, annual exams, and biometric lab data.

For example, data broker 1 claimed to have the largest sources of consumer-reported health data. The firm also advertised that households with ailments are more likely to be interested in targeted offers about medical needs. During the sales process, the sales representative briefly mentioned that data about children would require more processing since the firm was strict about who could access this kind of data. Data broker 4 directed the author to lists titled "Anxiety Sufferers" and "Consumers with Clinical Depression in the United States." Among other mental health data elements, data broker 7 also advertised having data on those in the military (which could be sorted by Air Force, Marines, etc.) and assimilation codes. Assimilated individuals were noted as English-speaking while "unassimilated" individuals were marked as speaking their native language only. Data broker 7 also advertised that the firm could provide granular data at the individual level. Aside from mental health data, data broker 9 also provided brochures that advertised global watch lists, phone records, and Social Security Number records.

Phase II: Examining the Privacy Policies of the 10 Most Engaged Data Brokers

The author examined the privacy policies of the 10 most engaged data brokers to compare whether each firm had considered and implemented comprehensive privacy policies. This involved reading through each available privacy policy to map the common categories and clauses found in them; noting general observations about the policies; labeling each privacy policy from the 10 firms by the appropriate date that it was downloaded; and then noting the last time a policy had been updated before reading through each document. 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.

Several common elements emerged from sorting through the privacy policies (shown in Appendix C). All 10 firms included language about their data security protocols and explicitly mentioned collecting and retaining personal data. Nine of the 10 firms also collect traffic data (like IP addresses). For example, data broker 8 explicitly states that they collect URL referral information and clickstream data, IP addresses, the variant of Operating System and browser used, the page accessed and clicks registered from the user, and the duration for which a user viewed an item. Data broker 8's privacy policy also stated that the firm does not collect generic information about users.[44] The firm then wrote that their database is sourced from reliable and accurate sources, such as annual reports/U.S. Securities and Exchange Commission (SEC) filings, government records, B2B directories, newspaper subscription offers, transactional data, community postings, phone surveys, sign-up data from email campaigns, and more.[45]

Seven of the 10 firms also disclose data to third parties about an individual visiting their website, while only two of the firms (data brokers 5 and 7) require third parties to adhere to their privacy standards. This means that the firms collect data about individuals who look through their website and can share that information with other parties. Further, other than data brokers 5 and 7, the other firms do not require other third parties that they work with to necessarily adhere to the same privacy standards that they uphold (meaning, other third parties could have fewer protections in place for the data that they acquire about site visitors).

Seven 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, without including many details. Additionally, only two firms (data brokers 7 and 8) 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 data brokers collect, share, and aggregate data about individuals, they seem collectively less willing to provide access and disclosure to their customers and users about the collection or correction of their data. For example, only two firms (data brokers 1 and 7) included information about a formal customer complaint process, and just four of the 10 firms give individuals access to their personal information. While six 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).

General Observations about the Privacy Policies: All 10 data brokers' privacy policies included 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 data broker's privacy policy because it was hidden on a less-accessible page of the broker's website (like a FAQ page or unlinked privacy policy page).

Analysis of Policy Implications for the United States

This research highlights a largely unregulated data brokerage ecosystem that sells sensitive mental health data in large quantities, with either vague or entirely nonexistent privacy protections. It varies, and sometimes it is simply unclear, whether a broker will advertise data that is already clearly and explicitly linked to an individual, or whether it will advertise data that has some level of obscurity in place to hide the identity of specific people. Considered in its entirety, these findings emphasize the critical need for a comprehensive federal privacy law that could provide data privacy protections to all Americans. These findings also underscore the need for shorter-term Congressional and state-level bans on the sale of sensitive mental health data on the open market.

Data brokers are collecting, aggregating, analyzing, circulating, and selling sensitive mental health data on individuals. This comes as a great concern, especially since the firms seem either unaware of or loosely concerned about providing comprehensive privacy protections. The lack of regulation in the U.S. and the opaque nature of the data broker industry has allowed data brokers that sell Americans' mental health data to maintain inconsistent business practices concerning data quality, accuracy, deidentification, data aggregation, data procurement, and data storage. These inconsistent practices, combined with vague privacy policies, point to a critical need for greater consumer protections, particularly for sensitive data, such as mental health data.

The unregulated collection, aggregation, sharing, and sale of data on individuals' mental health conditions puts vulnerable populations at greater risk of discrimination, social isolation, and health complications. Health insurance providers—which already buy individuals' race, education level, net worth, marital status, and other data without their knowledge or full consent to predict healthcare costs—could buy mental health data to discriminately charge individuals for care or discriminately target vulnerable populations with advertisements.[46] Scammers could purchase mental health data from data brokers to exploit and steal from individuals living with mental health conditions, as scammers have done to steal from payday loan applicants.[47]

Mental health disorders carry a stigma that inhibits many individuals from seeking 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.[48] Others in the survey were also concerned with how information about their mental disorders may affect future employment or their social life.[49] Depressed and/or anxious individuals in particular are more prone to committing suicide, making it even more critical to ensure that they do not face greater challenges in receiving care.[50,51] Privacy abuses are already a problem and a huge source of potential for harm for those at risk of self-harm or suicide—as a *Politico* story exposing Crisis Text Line, a suicide hotline, showed that the hotline was using people's data on self-harm, emotional abuse, and suicidal thoughts to "create and market customer service software."[52] Although the nonprofit claims that sharing the data will make customer support more "human, empathetic, and scalable," it seems dangerous for any institution to hold such sensitive and deeply personal without any regulatory controls in place.[53] 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, a comprehensive federal privacy law is long overdue. While states have begun to pass privacy laws, they only cover a specific jurisdiction.[54] 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 of the collection of their data, 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, as well as have more controls in place for client management.[55] Congress should continue to introduce and push bills, such as the Information Transparency and Personal Data Control Act (which was introduced as a piece of comprehensive privacy legislation in March of 2021), the American Data Privacy and Protection Act (which was introduced in June of 2022), and the Health and Location Data Protection Act.[56] In addition, Congress should consider providing protections for whistleblowers and researchers so that confidentiality obligations in data broker terms and conditions are not used to shield the industry from needed transparency and oversight. Appropriate enforcement and monitoring mechanisms, such as giving the Federal Trade Commission (FTC) more regulatory authority on these issues, should also 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 would be to extend HIPAA's protections. 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 regulations.[57]

Overall, sensitive mental health and personal data are being sold and exchanged in a data economy that is opaque and unregulated. To protect vulnerable populations and preserve Americans' privacy, regulatory actions must be taken to either advance a comprehensive federal privacy law or extend HIPAA to encompass the ever-evolving data economy and eruption of mHealth technologies.

Appendix A. Data Elements Wish List

General Asks:

- Target Audience
    - B2C, focused on mental health patients who have depression or anxiety
    - Would also like information on mental health care providers or physicians
- Contact information
- 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: $2,500

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 and presence of children in the household
    - Gender
- Top 3 most popular antidepressants: sertraline (Zoloft), escitalopram (Lexapro), and fluoxetine (Prozac)
- Data on recent visits to a 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; and personality disorder

Appendix B. Available Data Elements as Discussed by 10 Firms

*Not all 10 firms offered all the elements listed here; this is an aggregated list of elements that were offered and advertised by one, some, or all the firms

Non-Medical/Non-Healthcare Data Elements	**Medical/Healthcare-Related Data Elements**
• First/last name • Date of birth/exact age • Income • Gender • Marital status • Single-parent status • Ethnicity/race • Residential addresses • City name, state, and zip code • Home market value • Credit score • Homeowner status • Number of people in the household • Neighborhood characteristics • Net worth • Data about children • Driver's licenses • Social Security Number • Active government workers (state) • Active living Jews • Wealthy seniors nearing retirement • Social determinants of health data (anything that happens to a patient outside the hospital) • Grocery source (what's in the shopping cart) • Consumer data (such as retail purchases and other interests) • Religion • Language • Profession/professional licenses • Education/degrees • Pets in the household • Exercise habits • Assimilation codes • Bankruptcy • Criminal records	• Mental health data available: depression, attention disorder, insomnia, anxiety, ADHD, treatments (medication for ADHD/ADD), antidepressants, and bipolar disorder • Likelihood of having depression percentiles/scores • Likelihood of having anxiety percentiles/score • Other ailments data (such as diabetes, allergies, Alzheimer's, heart problems, bladder control difficulties, frequent headaches, high blood pressure, etc.) • Health plans • Medical events • Clinical facilities (such as, surgery centers, imaging centers, health clinics, long term facilities, hospitals/IDNs, and connected care organizations) • Lab data (genomic data)/biomarkers • Data from patient/disease registries, wearables/connected devices, and patient support/management programs • Physician profiles and physician groups (general practice, addiction specialists, etc.) • Data on the technologies being used inside facilities (i.e., when they installed an EHR, software use, etc.) • Average costs per procedure, practice locations, physician specialties, licensing, and roles • B2C (individual) or B2B data concerning health or mental health data • Specific medication uses and prescriptions

• Property tax assessments • Voter registration • Aircraft/watercraft registration • Accident reports • Global watch lists • College attendance records • 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 *Among many other available datasets and elements	• Prescriber information (name, PI number, DEA number, age, email address, telephone number, allowed to contact or not, etc.) • Data on the frequency of how many scripts of each medication were filled in a specific zip code or area • EMR data (electronic medical records) • General hospital systems data (number of procedures performed, departmental data, etc.) • Mortality data • Overall health score • Risky health behavior data • BMI estimate • Ability to pay for healthcare sorted in segments (such as 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) • Readmission risk scores • Medication adherence scores • Total cost risk scores (how much would they cost to the healthcare system over a few months) • Physicians who have prescribed sertraline/Zoloft, escitalopram/Lexapro, or fluoxetine/Prozac in a specific period • Data on abortion clinics *Among many other available datasets and elements

Appendix C. The Common Elements Found in the Top 10's Privacy Policies

Common Elements	Number of Firms Who Included the Common Element
Use Google Analytics	2
Mentions Internal Monitoring/Evaluation Processes	1
Disclosures about Changing the Policy	8
Mentions Accuracy of the Data	2
Includes a Customer Complaint Process	2
Gives Customer Access to their Personal Information	4
Includes an Acceptable Use Policy (i.e., ethical and moral use of the data)	2
Describes the Data Used to Develop their Products	2
Intends to be Explicit about Collecting Personal Data	1
Collect Data through Automated Technologies	1
Collect Personal Data	10
Collect Traffic Data (i.e., IP addresses or device information)	9
May Aggregate Personal Data with Other Sources	7
Mentions Regional and/or International Privacy Laws	8
Uses Personal Data to Provide a Service	6
Uses Personal Data for Non-Service Activities (i.e., advertising)	4
Individuals can Offer Consent for Data Use Cases Unrelated to the Original Purpose	2
Discloses Data to Third Parties	8
Does Not Have any Explicit Control Over Third Parties	2
Third Parties Must Adheres to the Same Privacy Standards	2
Enables Targeted Advertisements	4
Includes Opt-Out Processes	6
May Work with Data that is Not Anonymized or De-identified	1
May Sell Deidentified Data	1

Includes a Policy Protecting Minors	6
Includes a Data Retention Clause for Storing Personal Data	4
Provides Language on Consumer Rights and Choices	5
Mentions Data Security Protocol	10

Endnotes

[1] For complete disclosure of funders of the Duke University Sanford Cyber Policy Program, see the External Relationships section of https://scholars.duke.edu/person/David.Hoffman. No specific direct funding was provided for this report, and no funders reviewed any of this research before publication or had any editorial control over the report's substantive content.

[2] National Institute of Mental Health (NIMH). "Mental Illness." Accessed December 3, 2021. https://www.nimh.nih.gov/health/statistics/mental-illness; Stieg, Cory. "Depression Increased during Covid Pandemic: How to Feel Better, Cope," CNBC, October 10, 2021. https://www.cnbc.com/2021/10/10/depression-increased-during-covid-pandemic-how-to-feel-better-cope.html.

[3] Abbott, Alison. "COVID's Mental-Health Toll: How Scientists Are Tracking a Surge in Depression." *Nature* 590, no. 7845 (February 3, 2021): 194–95. https://doi.org/10.1038/d41586-021-00175-z.

[4] "Stigma and Discrimination." Psychiatry.org, accessed December 3, 2021. https://www.psychiatry.org/patients-families/stigma-and-discrimination

[5] Mental Health America. "The State of Mental Health in America." MHANational.org, accessed December 3, 2021. https://mhanational.org/issues/state-mental-health-america.

[6] Martinez-Martin, Nicole, Ishan Dasgupta, Adrian Carter, Jennifer A Chandler, Philipp Kellmeyer, Karola Kreitmair, Anthony Weiss, and Laura Y Cabrera. "Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities." JMIR Mental Health 7, no. 12 (December 22, 2020): e23776. https://doi.org/10.2196/23776.

[7] Anxiety & Depression Association of America. "ADAA Reviewed Mental Health Apps." ADAA.org, accessed December 3, 2021. https://adaa.org/finding-help/mobile-apps.

[8] Schueller, S. M., J. F. Hunter, C. Figueroa, and A. Aguilera. "Use of Digital Mental Health for Marginalized and Underserved Populations." *Current Treatment Options in Psychiatry* 6, no. 3 (September 15, 2019): 243–55. https://doi.org/10.1007/s40501-019-00181-z.

[9] "Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities."

[10] Marbury, Donna. "3 Reasons Why Wearables Bring New Complications for HIPAA Compliance." *Health Tech Magazine*, September 23, 2020. https://healthtechmagazine.net/article/2020/09/3-reasons-why-wearables-bring-new-complications-hipaa-compliance.

[11] Huckvale, Kit, José Tomás Prieto, Myra Tilney, Pierre-Jean Benghozi, and Josip Car. "Unaddressed Privacy Risks in Accredited Health and Wellness Apps: A Cross-Sectional Systematic Assessment." *BMC Medicine* 13 (September 7, 2015): 214. https://doi.org/10.1186/s12916-015-0444-y.

[12] Ramirez, Edith. "California Legislature Passes Nation's Second 'Data Broker Registration' Law," 2014. https://www.troutman.com/insights/california-legislature-passes-nations-second-dat a-broker-registration-law.html.

[13] Ibid.

[14] Grauer, Yael. "What Are 'Data Brokers,' and Why Are They Scooping Up Information About You?," *VICE*, March 27, 2018. https://www.vice.com/en/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection.

[15] "Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities."

[16] "3 Reasons Why Wearables Bring New Complications for HIPAA Compliance."

[17] Privacy International. "What Is Privacy?," PrivacyInternational.org, October 23, 2017. http://privacyinternational.org/explainer/56/what-privacy

[18] Nass, Sharyl J., Laura A. Levit, Lawrence O. Gostin, and Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press (US), 2009. https://www.ncbi.nlm.nih.gov/books/NBK9579/.

[19] Sherman, Justin. "Data Brokers are a Threat to Democracy," *WIRED*, April 13, 2021. https://www.wired.com/story/opinion-data-brokers-are-a-threat-to-democracy/

[20] Sherman, Justin. *Data Brokers and Sensitive Data on U.S. Individuals*, Duke University Sanford School of Public Policy, August 2021. https://sites.sanford.duke.edu/techpolicy/wp-content/uploads/sites/17/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf
[21] "Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities."
[22] Data broker, Personal Correspondence, 9.27.21.
[23] Data broker, Personal Correspondence, 10.18.21.
[24] Data broker, Personal Correspondence, 11.12.21.
[25] Sales Representative at data broker 6, Personal Correspondence, 11.9.21.
[26] Sales Representative at data broker 1, Personal Correspondence, 11.10.21.
[27] Data broker, Personal Correspondence, 10.4.21.
[28] Sales Representative at data broker 10, Personal Correspondence, 11.12.21.
[29] Data broker, Personal Correspondence, 10.19.21.
[30] Sales Representative at data broker 4, Personal Correspondence, 10.6.21.
[31] Sales Representative at data broker 6, Personal Correspondence, 11.9.21.
[32] Sales Representative at data broker 7, Personal Correspondence, 10.22.21.
[33] Sales Representative at data broker 4, Personal Correspondence, 10.18.21.
[34] Sales Representative at data broker 4, Personal Correspondence, 10.11.21.
[35] Sales Representative at data broker 1, Personal Correspondence, 10.5.21.
[36] Sales Representative at data broker 4, Personal Correspondence, 11.1.21.
[37] Sales Representative at data broker 7, Personal Correspondence, 11.3.21.
[38] Sales Representative at data broker 2, Personal Correspondence, 10.8.21.
[39] Sales Representative at data broker 2, Personal Correspondence, 10.8.21.
[40] Sales Representative at data broker 6, Personal Correspondence, 11.9.21.
[41] Sales Representative at data broker 2, Personal Correspondence, 10.8.21.
[42] Sales Representative at data broker 2, Personal Correspondence, 10.8.21.
[43] Sales Representative at data broker 6, Personal Correspondence, 11.9.21.
[44] Data broker 8. "Privacy Policy." Accessed April 14, 2022. [link redacted].
[45] Ibid.
[46] Marshall Allen, "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates," NPR, July 17, 2018, https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.
[47] U.S. Federal Trade Commission, "FTC Charges Data Brokers with Helping Scammers Take More Than $7 Million from Consumers' Accounts," FTC.gov, August 12, 2015, https://www.ftc.gov/news-events/news/press-releases/2015/08/ftc-charges-data-brokers-helping-scammer-take-more-7-million-consumers-accounts.
[48] Clark, Maria. "30 Statistics on Mental Health Stigma Infographic — Etactics." Etactics | Revenue Cycle Software. Accessed December 3, 2021. https://etactics.com/blog/statistics-on-mental-health-stigma.
[49] Ibid.
[50] Division (DCD), Digital Communications. "Does Depression Increase the Risk for Suicide?" Text. HHS.gov, 2014. https://www.hhs.gov/answers/mental-health-and-substance-abuse/does-depression-increase-risk-of-suicide/index.html.
[51] Hitti, Miranda. "Study Links Anxiety, Nervousness to Suicide." WebMD, August 10, 2005. https://www.webmd.com/anxiety-panic/news/20050810/study-links-anxiety-nervous ness-to-suicide.
[52] Levine, Alexandra S. "Suicide hotline shares data with for-profit spinoff, raising ethical questions." Politico, January 28, 2022. https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617.
[53] Ibid.
[54] Klosowski, Thorin. "The State of Consumer Data Privacy Laws in the US (And Why It Matters)." Wirecutter: Reviews for the Real World (blog), September 6, 2021. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/.
[55] Fazlioglu, Müge. "The First but Not Last Comprehensive US Privacy Bill of 2021," March 17, 2021. https://iapp.org/news/a/the-first-but-not-the-last-comprehensive-u-s-federal-privacy- bill-of-2021/.
[56] Ibid; Warren, Elizabeth. "Warren, Wyden, Murray, Whitehouse, Sanders Introduce Legislation to Ban Data Brokers from Selling Americans' Location and Health Data." June 15, 2022.

https://www.warren.senate.gov/newsroom/press-releases/warren-wyden-murray-whitehouse-sanders-introduce-legislation-to-ban-data-brokers-from-selling-americans-location-and-health-data; Duball, Joseph. "US House committee showcases federal privacy momentum, opportunity." IAPP, June 15, 2022. https://iapp.org/news/a/us-house-committee-unpacks-federal-privacy-momentum-opportunity/.

[57] Butler, Mary. "Is HIPAA Outdated? While Coverage Gaps and Growing Breaches Raise Industry Concern, Others Argue HIPAA Is Still Effective." Journal of AHIMA 88, no. 4 (April 2017): 14-17, 52.