Research Ethics in Behavior Analysis

From Laboratory to Clinic and Classroom

Edited by

David J. Cox
Institute for Behavioral Studies, Endicott College, Beverly, MA, United States

Noor Y. Syed
SUNY Empire State College and the Center for Autism Advocacy: Research, Education, and Supports, Saratoga Springs, NY, United States; Anderson Center International, Staatsburg, NY, United States; Endicott College, Beverly, MA, United States

Matthew T. Brodhead
Department of Counseling, Educational Psychology, and Special Education, Michigan State University, East Lansing, MI, United States

Shawn P. Quigley
Melmark, Andover, MA, United States; Berwyn, PA, United States; Charlotte, NC, United States

ACADEMIC PRESS
An imprint of Elsevier

Chapter 9

Data handling: ethical principles, guidelines, and recommended practices

Brent A. Kaplan, Shawn P. Gilroy, W. Brady DeHart, Jeremiah M. Brown and Mikahil N. Koffarnus

Department of Family and Community Medicine, University of Kentucky College of Medicine, Lexington, KY, United States; Department of Psychology, Louisiana State University, Baton Rouge, LA, United States; Optum Labs, Eden Prairie, MN, United States; Department of Human Nutrition, Foods, and Exercise, Fralin Biomedical Research Institute at VTC, Virginia Tech, Blacksburg, VA, United States

Data are encountered at nearly every step of the clinical and research process. The organization and management of data are important starting from the initial recruitment of a potential participant through dissemination of the research. Managing data may often be overlooked during the day-to-day activities of a productive laboratory or clinic, and established practices that "just work" might be assumed to preclude the need for direct training on data management and data handling. However, consistent and ethical systems for data handling are of the utmost importance given the many areas where data may be compromised, altered, or otherwise mishandled (e.g., human error). Specific trainings and established guidelines for the management of data help to protect the anonymity and confidentiality of individuals, reduce the chance of human error, minimize bias and conflicts of interest, and enhance reproducibility efforts, including dissemination and research synthesis endeavors (e.g., meta-analyses; Haidich, 2010; Lipsey & Wilson, 2001). With the ever-increasing quantity of research studies being conducted and disseminated, data need to be more quickly collected, organized, and synthesized. If data are not properly managed, the available data may be incomplete or contain bias, and this may limit the generality of subsequent analyses.

The overarching goal of this chapter is to provide considerations, practices, and suggestions to enhance the integrity of data handling throughout behavior analytic research and, to a lesser degree, clinical process. We (the authors of this chapter) have worked in various settings and have managed various types

Research Ethics in Behavior Analysis. https://doi.org/10.1016/B978-0-323-90969-3.00006-2

of data in behavior analysis, including applied clinical work, nonhuman animal laboratory, human experiential research, crowdsourced survey data, and archival medical record data. To accomplish our overarching goal, we first discuss various types of data commonly encountered in behavior analysis ranging from highly sensitive data (e.g., personally identifiable) to anonymized data. We then discuss the various stages of the data handling process, from initial data collection and storage, to data validation, to data analysis, and ultimately to dissemination. At each stage, we will discuss aspects of data handling across the different domains relevant to behavior analysts including applied clinical work, human and nonhuman experimental work, largely anonymized or survey data (i.e., data gathered is initially anonymized, not after the fact that can be done with clinical and experimental data), and archival data (e.g., medical records).

A theme we will weave throughout the discussion of this chapter is the integration of open-source tools in the process of data handling. Open-source software is software where the underlying code is distributed and can be modified for free or with at least some recognition of the developers. Many open-source tools are free to use, which can reduce barriers to sharing, inspecting changes/revisions, and reproducibility. Because open-source tools typically rely on code that is open for inspection, the code can be "vetted" to determine the extent to which the code does what it was intended to do (for example, open-source encryption software may be routinely audited to ensure vulnerabilities are fixed). We believe that understanding and utilizing open-source tools wherever possible and appropriate enhance the ethical handling of data, including protecting confidentiality, promoting transparency and responsibility (e.g., using audit trails), and aid in reproducibility.

Legal and regulatory landscape of data handling

There are a number of legal considerations of which behavior analysis should be aware with regards to data handling. The General Data Protection Regulation (GDPR; European Parliament and Council, 2016) is a regulation of the European Union that addresses the handling of personal data of individuals, giving European Union citizens rights over their personal data and limiting the actions of companies using personal data for marketing. The California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA) are laws in the US state of California that regulate businesses' ability to sell or share customer's personal data without consent (for more information on the GDPR and the CPRA, see Mueller et al., 2022, pp. 267–288). Relatedly, cybersecurity standards have been outlined by the National Institute of Standards and Technology (NIST; a nonregulatory agency of the U.S. Department of Commerce) to protect digital data. The Health Insurance Portability and Accountability Act (HIPAA; United States, 1996) stipulates how personally identifiable information should be protected from involuntary disclosure; the

Health Information Technology for Economic and Clinical Health Act (HITECH; United States, 2009) regulates certain aspects of privacy and data security, including health information exchange standards. Additionally, the Family Education Rights and Privacy Act (FERPA; United States, 1974) specifies the rights of parents to access, amend, and disclose certain components of their child's educational records. While an in-depth review of these laws and regulations is beyond the scope of this chapter, it should be noted that researchers and clinicians would likely benefit from collaborating with individuals with knowledge of relevant legal or cybersecurity concerns.

Types of data in behavior analysis

As mentioned earlier, data exist in various forms. Data can include things such as video recordings, products of behavior monitoring systems (both physical and digital including physiological measurements), electronic surveys, and more. Data range on a continuum from highly personal and identifiable to completely anonymized. Data containing sensitive or identifiable information need to be stored confidentially (e.g., video recordings), especially those including protected health information (PHI), whereas other data require minimal safeguards (e.g., anonymized survey responses). As a result, we will discuss data handling along a continuum ranging from highly sensitive data to not very sensitive data (e.g., anonymized, archival, nonhuman data). Here, we will briefly discuss and highlight some of the different types of data encountered in various behavior analytic settings.

Highly sensitive data

Research and practice in behavior analysis emphasize the role of information and data. Various dimensions of data and aspects of the environment are recorded and interpreted to inform ecologically based research and treatment. Broadly, this follows a sequence of participant characterization (i.e., identifying individual, diagnostic profile, demographics), treatment development (e.g., defining behavior, evaluating functional relations), and treatment evaluation (i.e., determining if treatment meaningfully influences behavior). Each of these steps is driven by different forms of information that must be organized and maintained to either inform treatment or answer specific research questions.

Data related to the characterization of participants are particularly sensitive because this information most closely associates individuals with their recorded data. For example, individual identification numbers (e.g., social security number) and dates of birth are (largely) unique and easily cross-referenced (cf., year of birth). Furthermore, certain forms of information (if made publicly available) have the potential to expose participants to negative social consequences or outright discrimination. For example, individuals may

be treated differently by their coworkers or community if it were known that they participated in mental health treatment (Sickel et al., 2014). It is for this reason that such highly sensitive data are protected (i.e., PHI) under the US HIPAA (https://www.hhs.gov/hipaa/) when related to healthcare. This act requires healthcare providers to avoid disclosing any identifiable information related to healthcare to any third party without explicit permission, including information that could be reasonably used to identify a patient. Although not always required for research studies outside healthcare settings, the information sharing systems developed for HIPAA compliance are often useful for managing participant information flows in research. For example, HIPAA compliance refers, but is not limited, to ensuring the confidentiality, integrity, and availability of PHI, identification and protection against reasonably anticipated threats to security, and protection against unauthorized disclosure. If dealing with PHI, researchers and clinicians should familiarize themselves with current guidelines regarding HIPAA (https://www.hhs.gov/hipaa/for-professionals/index.html) and in the context of schools and student records, the Family Educational Rights and Privacy Act (FERPA; https://studentprivacy.ed.gov/faq/what-ferpa).

Data collected for the purpose of treatment development and evaluation can be highly sensitive as well. These data often consist of interviews, open-ended surveys designed to support the definition of targeted behavior (along with respective deficits) and direct observations of behavior in context. For example, clinical interviews and reviews of family histories reveal a wealth of clinically relevant, but sensitive, information (e.g., history of medical, behavioral disorders). Once behavior is well-characterized and defined, behavior can be sampled to explore possible functional relations (e.g., structural analysis) as well as determine baseline levels (e.g., rate, duration) as they pertain to treatment. This may be extended to experimental manipulations as well (e.g., functional analysis), toward the same ends. Further, the preferences and perspectives of participants and their families may also be captured to support the social validity of programmed supports.

The specific types of data collected during treatment development and evaluation can vary significantly between cases. For example, the types of behavior being recorded could range from disruptive child behavior, to substance use, and to nonsuicidal self-injury. In such a situation, data may be collected using a range of formats (e.g., pencil/paper behavior data sheets, video-recorded interviews, electronic data monitoring) because there is considerable variability in the types of behavior observed and how those are tracked. Researchers should consider how each of these mediums of data collection is handled. For example, after data are collected on paper/pencil sheets, they should be placed in a binder and not left unattended. The binder with data sheets should be moved to the secure location (e.g., locked filing cabinet) when data collection is complete. Likewise, for electronic data entry, the researcher should ensure the browser is closed and the computer is locked

(or shut down) after collection is complete. Task analyses or a checklist for the steps required when completing a session may aid in ensuring data are properly transmitted and stored.

Several types of data are collected in human research studies. For example, these data may consist of those considered to be PHI (e.g., name, physical address, meeting appointment dates and times) while others, in combination, could be used to potentially identify someone (e.g., sex, age, zip code). These types of data are distinct from others that do not have as much of a possibility of identifying individuals (e.g., deidentified data; questionnaires, response times). According to the U.S. Department of Health and Human Services (The Office for Civil Rights (OCR) & Malin, 2012), the following pieces of information are considered identifiers of or related (relatives, employers) to the individual: names; geographic subdivisions smaller than a state; elements of a date (except year); telephone numbers; vehicle identification numbers (including license plates); fax numbers; device identifiers and serial numbers; email addresses; URLs that are explicitly associated with an individual; social security numbers; internet protocol (IP) addresses; medical record numbers; finger and voice prints; health plan beneficiary numbers; full-face photographs and comparable images (e.g., images that contain any unique characteristic related to the individual that could lead to identification such as a tattoo, a unique setting, a birthmark; Nettrour et al., 2018); account numbers; other unique identifying number (e.g., a subject identification number); or certificate and license numbers. In essence, these data can be directly used to identify the individual. As such, they are highly sensitive and should be handled and stored with the highest precautions. In later sections, we will consider certain safeguards and data handling techniques to maximize security and protection. For now, we mention that, depending on the scope of the human experiential studies, researchers may collect a relatively large or small amount of PHI.

Potentially sensitive data

Additional data gathered in the clinic or research lab may be considered potentially sensitive but, in and of themselves, are not sensitive. Certain types of demographic information may enable others to identify an individual when combined with other data. For example, when conducting research among diverse populations, uncommon cultural, linguistic, ethnic, racial, gender, ability, or other demographic characteristics may be identifying if few members of the community share those characteristics. While there may not be a general test to determine whether any of these demographic characters may be considered identifying information, the researcher or clinician should rely on their expert domain knowledge to assess whether there is a likely chance information may be identifiable. In instances where there may be a chance certain demographics are identifiable, researchers or clinicians should rely on a board such as an Institutional Review Board to determine whether to classify the information as identifiable.

Similarly, 5-digit ZIP codes may have a relatively small number of residents and—combined with other information such as a person's sex and age—their identity could be ascertained. A common strategy when handling with ZIP codes in public reporting of results or other contexts where divulging of PHI is unwarranted is to omit the final two numbers of the ZIP code, retaining only the first three. This provides a broader, but still useful, characterization of location without increasing the likelihood of being personally identifiable. In sparsely populated areas, even the first three digits are too specific, so the US Department of Health and Human Services maintains a list of 3-digit ZIP codes that contain a small enough number of individuals that they could be considered identifying information (The Office for Civil Rights (OCR) & Malin, 2012). For example, as of the 2000 census data, there are currently 17 three-digit restricted ZIP codes that contain 20,000 or fewer persons. Some of these ZIP codes include 036 (Bellows Falls, VT), 692 (Valentine, NE), 878 (Socorro, NM), and 063 (New London, CT). Researchers wanting to deidentify datasets including zip codes should consult the most recent census data available to determine whether the first three digits of a ZIP code may be restricted. In order to avoid ZIP code ambiguity completely, researchers and clinicians should determine whether they need to collect the information in the first place. Some studies or projects may not need ZIP codes and explicitly not collecting these data may be preferred.

Archival data are commonly collected for other purposes (e.g., healthcare, quality, and organizational improvement, treatment efficacy) and are retrospectively gathered and analyzed from various sources after the initial reason for the data collection is completed. Sometimes the archival data may be from previous research in the laboratory, or they can be from publicly available repositories (e.g., data.gov, census.gov, Panel Study of Income Dynamics), or they may be available upon request for specific purposes or analyses. For data collected previously in, for example, a laboratory, researchers and clinicians should ensure they have the appropriate permissions to access the data if the data contain identifiable information or otherwise have someone who has existing permissions deidentify the data. Archival data can exist in many forms including surveys, medical records, behavioral metrics such as web browser activity, and patient data logs. Archival data often include identifying information and PHI and steps should be taken to ensure PHI, if included, are protected. If the research can be accomplished without the PHI, then this may be the best approach as the data then may be classified as not very sensitive. Ultimately, researchers and clinicians working with existing and archival data sources should work directly with the custodians of those data to ensure privacy, if applicable, is maintained (e.g., understanding the scope of access to the data, if data can only be accessed on the server on which it is stored).

Not very sensitive data

Finally, various other types of data collected can be considered less sensitive (i.e., less than what is necessary to identify individuals). We make the distinction here that these data are still sensitive and thus, should be handled with care regardless. These data, however, do not necessarily warrant the multiple safeguards that more sensitive information require. For example, nonhuman animal data do not require the same protections as human participant data. Additionally, data gathered in human experimental studies wherein that data are comprised of response times or choice responses or infusions of a drug are often deidentified, and as such, cannot be used to link back to the participant without identifying data. Not very sensitive data may include response times, responses to behavioral tasks, and other data (e.g., comments, general ratings of behavior). Although these data may not allow someone to directly identify individuals, they should be handled with care similar to other data, and researchers should ensure that these data are not located in the same place as PHI (i.e., digitally or otherwise), as this might allow one to link the responses.

Deidentification removes all identifying information that can be possibly linked to a given individual. Some of these identifying pieces of information have been discussed already (e.g., name, age, medical record numbers). There are two ways in which deidentifying data can occur. They include the Safe Harbor method and Expert Determination approach (Kayaalp, 2018, pp. 1044–1050). To meet the Safe Harbor method, all identifiers must be removed from the dataset. For the Expert Determination approach, certain statistical or scientific principles are applied to the data (e.g., adding random amounts to values but retaining the underlying distribution) that make it unlikely the data can be linked back to any specific individual.

Crowdsourcing is increasingly being used for collecting behavioral data (Chandler et al., 2019; Peer et al., 2017; Strickland & Stoops, 2019). Crowdsourcing services such as Amazon Mechanical Turk, Prolific, and Qualtrics Panels provide researchers additional or alternative avenues for collecting data for certain types of research questions. The type of data collected from these sites are often anonymous but need not be depending on the types of questions being asked. Researchers should be aware of the data being collected (e.g., IP address, other identifiable information) and take proper precautions and steps to ensure participants are being fully informed of the data being collected as well as proper mechanisms by which to transmit and store the data, which will be overviewed in later sections. For example, these mechanisms may include encrypted transmission, user-defined role access through a secured database, and how long any PHI will be stored. These safeguards should be in place to minimize risk and harm to the participant and protect confidential information.

Data collection

One of the core aspects of data handling is data collection. Planning for the data collection strategy (e.g., electronic, physical) occurs at the very beginning of the clinical or research process, and data collection begins the moment anything related to the potential participant or client is obtained and stored. We will highlight and discuss several themes and key aspects to increase the likelihood that data will be collected with high integrity, free of bias, and such that confidentiality is maintained. These considerations include proper training of research personnel, clear definitions related to what and how data are to be collected, and tools to support the ethical collection of complete data.

Training and Clear Definitions. First, research personnel should have sufficient and adequate training in the handling of data. Those who are collecting the data should have an intimate understanding of the different types of data being collected. For example, research staff should be able to identify data considered PHI compared to deidentified or anonymized data that cannot be directly linked to the participant. Valuable training resources include hands-on experiential training from others in the laboratory and courses provided in services such as the Collaborative Institutional Training Initiative (also known as CITI Program; https://about.citiprogram.org/en/homepage/). For example, the CITI program has courses entitled "Mobile Apps and Human Subjects Research" and "Good Laboratory Practice," which cover aspects of data collection and PHI. Whereas individuals who are not associated with a specific institution can sign up for these trainings as an independent learner and pay for courses, for individuals seeking alternatives to fee-based training, the Office for Human Research Protections offers free human research protection training (https://www.hhs.gov/ohrp/education-and-outreach/online-education/human-research-protection-training/index.html). Data not considered PHI should still be handled with care. With the increased availability of tools to find individuals online and the proliferation of social media and other public repositories of data, it is increasingly possible to use combinations of seemingly arbitrary data to identify a participant. Researchers should consider the volume and combination of data collected as a potential identifier when determining the extent to which different types of research data are confidential.

Proper training of research personnel can also reduce the likelihood of bias. This bias may manifest itself in the form of screening and/or deciding whether a potential participant meets inclusion criteria for the study. The need for clear and definable inclusion and exclusion criteria is important here, as well as for research personnel to identify whether someone meets those criteria. For example, in a study looking at a population of cigarette smokers, an inadequate criterion for inclusion may be "Currently smokes cigarettes." This criterion may be inadequate by virtue of being overly nonspecific because "currently" is not well defined and cigarette smoking does not correspond with a quantitative level. Consider the example of an individual who smokes a single cigarette

1 day before volunteering for the research study. Would this behavior count for inclusion and importantly, would this individual's data be reflective of other "current cigarette smokers"? A better inclusion criterion may be "Smokes at least five cigarettes per day on average during the past 30 days." Here, there is a clear window of time as well as frequency of behavior.

Tools can also minimize the risk of human error and bias. These tools may be in paper format, or they may be used electronically. One such tool is the use of a tracker (physical or electronic) along with physical (paper) research materials. A tracker can most easily be thought of as similar to a task analysis or checklist that breaks down every step the researcher must take to conduct the session according to the procedure defined by the IRB approved protocol or the behavioral support plan. In other words, the tracker tries to maximize the extent to which procedures are implemented as they are intended (e.g., treatment fidelity). In our studies, we have used trackers successfully to ensure enrollment logs are updated and data are checked for completeness after a participant completes a task. For example, a tracker created for the purposes of an informed consent session might have the following components: (a) date, (b) participant ID, (c) research assistant consenting the participant, (d) checkboxes for individual inclusionary/exclusionary criteria, (e) checkboxes associated with discussing the consent form with the participant, answering any questions, and obtaining participant and researcher signatures, and (f) a checkbox associated with payment delivered to the participant. Whereas this type of tracker is no replacement for proper training in effectively executing the protocol as written and approved, utilizing a tracker will help increase the likelihood all steps of the research protocol are implemented with high fidelity and adherence.

The degree to which trackers support the goals of high fidelity, close adherence, and minimizing errors and bias can be enhanced by integrating several safeguards. Two safeguards include validating data entry fields (among electronic trackers) and requiring double entry for important values or other values that may be easily transposed or misread. For data validation, the electronic tracker should be programmed and tailored to the expected data type. For example, if an entry field should include a date, then with most electronic software (e.g., Qualtrics Research Suite), a rule can be implemented whereby an error will display and the page will not advance until the entry field contains text in a date format (see Fig. 9.1 for an illustrative example). Double entry requires that an important value (e.g., a participant's assigned condition) is entered in twice and that both fields contain the exact same characters. Depending on the software this could be accomplished on the same page or on separate pages (e.g., on an initial intake page and a few pages later so as to identify the error quickly). Although double entry like this would be preferred with two different researchers or clinicians independently entering data in to the fields, this is not necessary and sometimes not feasible. This safeguard is highly useful for ensuring the correct value is implemented. For

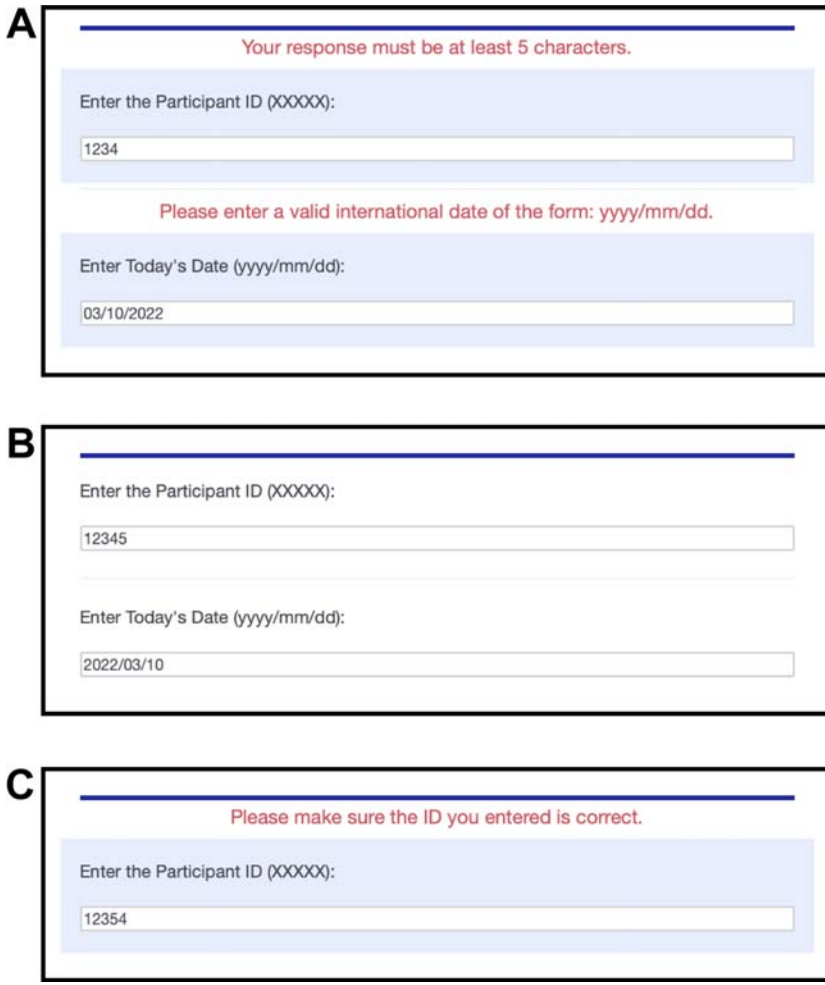

FIGURE 9.1 An illustration of content validation in a tracker. **Panel A** shows the validation error codes when an invalid response is provided. In the first case, the ID is not long enough and doesn't meet the five character requirements. In the second case, the date is entered incorrectly and any date other than in the format "yyyy/mm/dd" will not be accepted. **Panel B** shows correct responses that do not trigger validation errors. **Panel C** shows double entry validation. Notice the transposition of "5" and "4" at the end. The required text is "12,345."

example, a participant's identification code may be prone to transposition errors (e.g., correct: 12,345; transposed: 12,354; see Fig. 9.1 for an example), and so requiring entry twice should minimize such errors. We note, however, this is no guarantee and that those collecting data should understand where and when errors might be more likely to occur. An excellent way to determine where errors might occur is to have someone who is not familiar with the study

follow and complete the steps (e.g., complete the tracker, complete the study questions, conduct a "mock" session). When at all possible, extensive testing should occur prior to conducting actual sessions.

Paper data collection

Due to limitations in using physical paper to collect data, safe handling techniques are relatively limited in scope compared to electronic data collection. Clear, understandable instructions should be provided, as well as a sufficient range of possible answers. To decrease the likelihood of missing responses, instructions should remind participants to answer all questions and each question should be distinct as to not be easily missed or skipped. All paper data associated with a given participant should be kept together. If paper data contain highly sensitive data, they should be stored in accordance with those data (see Storage section below).

Electronic data collection

Many types of data are collected electronically to streamline the process of entering and saving information. Furthermore, not all types of studies require physical, in-person interaction. Data can be collected electronically regardless of whether the research occurs in the laboratory (e.g., linked to specific apparatus), in a clinical or educational setting (e.g., via direct observation), or electronically over the Internet (e.g., via survey software). These collection practices carry their own set of considerations and safeguards compared to traditional in-person research. Likewise, electronic data collection can range from highly sensitive to not very sensitive data.

Researchers often strive to minimize the proportion of missing responses during data collection, as missing responses can present difficulties when analyzing the data. However, in practice, requiring participants to answer every question may not be ethical as individuals may not wish to share certain types of information (e.g., trauma, legal, drug use history). In the case of electronic survey tools, this type of situation requires thought and planning. For instance, rather than making questions optional to answer, researchers often utilize "forced" response questions (whereby the participant must select some response before advancing) while including a specific "Do not wish to answer" option. This setup is preferred over optional answers as a non (or missing) answer does not distinguish whether the participant simply skipped the question or whether the participant was not comfortable answering that question. A high proportion of "Do not wish to answer" responses may be indicative of a poorly worded or unnecessarily intrusive question, and the researcher may wish to investigate the question.

Missing data may also occur in applied and educational settings. A therapist or teacher may be working with a client or student and may inadvertently

fail to record a behavior within an interval. Alternatively, school teachers may incorrectly enter an unrealistic achievement scores into a database such that the value is flagged and removed when validating the data. These cases are often outside the control of the researcher analyzing the data and statistical techniques (e.g., imputation) may be required to provide a "best guess" of what the data *ought* to be. When possible, frequent intermittent data checks should occur so that identifying missing or implausible data can lead to fixes or other solutions specific to the situation at hand.

Another feature of data collection that can enhance data integrity and collection is using attending questions. These questions may be used to evaluate whether the respondent is attending to the relevant stimuli and responding accordingly. For example, attending questions may be used after a vignette outlining a hypothetical scenario or for ensuring that the participant understands the operational definition of a term. Participants may be provided feedback on the spot; if they answer incorrectly, they may be told to reread the section and asked another similar attending question related to the content until comprehension is achieved. We have used attending-like questions in our studies in the form of a "consent teachback" (Talevski et al., 2020). After reading an informed consent form, potential participants will complete an electronic survey where portions of the consent forms are displayed. Potential participants then answer multiple choice questions related to that portion of the consent form. If answered incorrectly, potential participants are provided feedback on the wrongly answered questions and are provided additional chances to answer correctly (e.g., programmed instruction).

Data handling and integrity can be maximized when using electronic survey software to collect data. Several features of these surveys can be implemented to ensure compliance, and we have discussed these features previously in the context of an electronic tracker. The first feature is data type validation whereby entries must match a predefined data type. Relevant to collecting data from participants, researchers may restrict an entry to numbers only. This would avoid a situation in which a participant might be inclined to respond with a range (e.g., 2−4) and would be forced to provide a specific number (e.g., 3). The second feature is one of double entry. Researchers may want to reduce any ambiguity or error in the participant's response by requiring them to enter the response identically in two separate places.

In addition to tools that organize the collection of data, various types of software exist to perform the actual measurement of behavioral data. Historical methods of measuring behavior, in both research and practice, relied on manual data recording (e.g., long-form descriptions, completing data sheets) or physical products (e.g., tape audio/video recordings). These methods have good utility; however, there is considerable administration necessary when using these materials and may be inefficient and limit analytic options. For example, these methods require human interaction to score, evaluate for interrater reliability, and physically save the data for long-term archiving and

backup. Each of these steps is expensive in terms of staffing and resources, and this is a barrier to reliably implementing good data management processes. Multiple tools have been developed to address such barriers, largely concerning data collection (e.g., BDataPro—Bullock et al., 2017; DataTracker—https://github.com/miyamot0/DataTracker3). These electronic data collection tools have been designed to minimize the burden of proactive data management practices. For instance, the programs listed here automate the process of calculating interrater agreement and related indicators of data quality. Further, programs such as DataTracker store both human- and computer-readable data as an added layer of redundancy and layer of data validation. Having multiple layers of data allows for automated checks for the consistency of data records and auditing.

Various open-source tools and programming languages exist and can be used to automate different types of data collection. For example, we have used Integrated Development Environments (IDE) such as Visual Studio (https://visualstudio.microsoft.com) and modern programming languages such as Python (https://www.python.org) to collect different types of behavioral data. However, there are many alternatives for IDEs (e.g., Visual Studio Code) and languages (e.g., C#, F#, Go). Of course, the researcher will need to weigh the costs in terms of time, money, and training to implement such systems within their laboratory or clinical work. For instance, these tools have the potential to run highly specialized tasks and streamline data collection (i.e., limiting the potential for human error) but may require considerable expertise to effectively design and manage these systems.

Data storage

Information and data can be collected using different platforms and various platforms differ in how they are secured. We will separate storage methods by whether storage is physical (i.e., secured location) or electronic (i.e., stored via remote or local media). Sensitive data such as PHI should be collected and stored in locations where only approved study personnel can access the information. Physically, this may be in a locked cabinet in a locked office (i.e., double-locked and secured location). Only authorized personnel should have access to these files, and these files should be kept in secured storage except for when the data are actively being used (e.g., data validation, data analysis). Electronically, data should be stored using services that are not accessible to third parties and that offer robust encryption and rights management capabilities. Encryption means that the data are transmitted and stored in such a way as to hide the actual information and are only accessible by individuals who have the rights and methods to unhide the information. For particularly sensitive information, data collection tools such as RedCap (Harris et al., 2019) can be installed on local servers and provide additional security compared to “cloud”-based services that store data on the servers of private

companies (e.g., https://www.dropbox.com/). Nonetheless, data stored on cloud-based servers should still be protected via passwords and only authorized users should have access to the data. Two-factor authentication methods (e.g., use of a one-time code via an authenticator application) should be used whenever possible to reduce the likelihood of unauthorized personnel gaining access to the service. When data are locally stored on a computer, users should ensure the computer is password protected and locked whenever unattended, even for a short duration. When in doubt, researchers and clinicians should discuss storage options with their team and any governing body (e.g., Institutional Review Board) to ensure compliance.

Services used to collect and store the information should allow the ability to specify rights management, meaning users can be assigned roles and their roles dictate the extent to which they can interact with the data (see Fig. 9.2). For example, the principal investigator or primary project lead may have the right to read and edit all data, whereas a research assistant may only be able to read a subset (e.g., no sensitive information) of the data. Furthermore, raw data collected should be saved in a "read-only" manner whenever possible. Read-only refers to the permissions of a file whereby once saved, the data cannot be changed. These raw data files may be duplicated so that analyses can be conducted. Saving and storing raw data in this state greatly increase the chances that the original data do not get tampered with or altered. Analyses and data manipulations can be conducted on copies of the data, but a read-only copy of the original data ensures that an unaltered version is always available. This does not guarantee subsequent manipulations of the data will be free from error or other issues, but does allow reanalysis and external validation of the data if needed.

Retrospective archival data researchers often have little control over how data were collected. That is, these researchers typically do not have input on how the data were originally collected and can only deal with the data now. For example, patient data collected in the hospital may have billing codes instead of diagnoses codes. This may be due to other individuals having originally collected data or the data were collected for different purposes. When possible, researchers should collaborate with the teams collecting the data to encourage systems that maximize both the practical utility and the research potential of the data. To extend the example, billing codes may not provide the necessary specificity a research question necessitates (e.g., a billing code related to "high blood pressure" may be insufficient when actual values are needed). Nonetheless, researchers should work with the entities or other researchers who originally collected the data to understand the scope of the data available.

Depending on how securely the data are stored, researchers may not have direct access to the data. Frequently, organizations employ data analysts/extractors who are skilled in extracting data from archival sources. There are positives and negatives to such situations. On the one hand, this can cause delays in acquiring the data, and miscommunications between the research

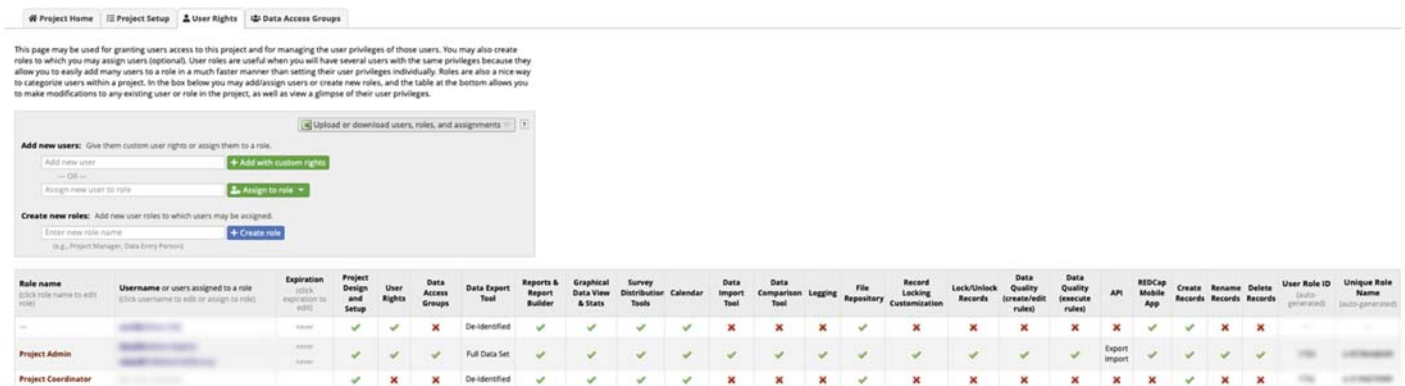

FIGURE 9.2 An example of user permissions in the RedCap interface. Different roles can be assigned and those roles can have prepopulated access to different permissions (e.g., renaming records, lock/unlock records). Notice how different users can have access to the full data set or the deidentified data.

team and the data analysts may lead to the incorrect type or level of data being attained. On the other hand, relying on another individual to extract the data can help increase HIPAA compliance and can allow access to complex database systems. The research analyst can deidentify the data and remove sensitive information to be in line with Safe Harbor guidelines discussed above (e.g., recode ages over 90, mask zip codes). However, despite this extra step of deidentification, the research team is under the same expectations to safeguard delivered datasets as other data types.

Some data may contain or be derived from highly sensitive information and medical records. These data are especially sensitive and must be protected. Failure to properly protect medical records could violate established HIPAA laws and expose the research team and sponsoring facility to legal action including fines and lawsuits. Archival data are best stored on local servers that are user restricted or password-protected (see Storage section above). This ensures that only approved researchers can access the data. Archival data should never be stored on personal devices or personal cloud storage sites and should never be shared via email.

A final note about data storage and optimal conditions regards the concept of redundancy. Redundancy of data can help in the event something catastrophic should affect the data. In general, best practice suggests a "3-2-1" rule. Keep at least three copies or versions of the data; one copy serving as the primary copy and two as backups. These copies are distributed across at least two different locations, with at least one of the locations being "off-site" (for example, secure and encrypted data could be stored on a server or literally at another physical building). This rule, of course, means that each version or copy of the data (especially highly sensitive data) at each location would need to be protected with similar safeguards. While this rule may not be feasible in all laboratories and clinics, researchers should understand and be aware of the general concept and apply it accordingly in their settings. Further, given we live in the era of big data, many database vendors offer solutions that meet the "3-2-1" rule at a reasonable cost.

Data validation

Data validation refers to efforts to characterize the available data and ensure that the data record reflects what the instrument was designed to measure. During the research process, there are several opportunities to validate data: data collection, post data collection, and post data analyses. Data validation is a vitally important step to ensure the data collected are complete, error-free, and unaltered. Here, we will discuss considerations for validating data, including auditing of data for integrity, completeness, and replicability.

Modern data storage resources are inexpensive and plentiful, allowing for a full data auditing trail to be maintained. An easy and automatic way to increase the integrity of data auditing capabilities is to use versioning software

that stores old versions of data files indefinitely. Common office software such as Microsoft Office 365 (https://www.office.com) now supports versioning to automatically save old copies of files, and services such as GitHub (https://github.com) and the Open Science Framework (OSF; https://www.osf.io) support similar functionality for any type of file including research data files. If an error is made in a data cleaning or analysis step that compromises the integrity of the data, a copy of the original data or earlier versions of data files allow for auditing analysis steps and recovering uncompromised data. Whether or not versioning software is being used, naming conventions should be thoughtful. For example, a file's name may include the date it was downloaded (e.g., "20220310"; March 10th, 2022) and an indicator of whether the file contains original, raw data or manipulated data (e.g., ResearchStudy-20220310-raw.csv).

GitHub is a public platform supporting the Git version control language and supports the open science initiative (Gilroy & Kaplan, 2019). Version control software such as Git allow for detailed record keeping of changes in documents (working best with text-based files such as .txt, .csv, or other text-based coding files, e.g., .R) and allow for collaboration. Fig. 9.3 is an example from a GitHub repository (a project composed of many different files) maintained by the first author (Kaplan) showing collaboration with the second author (Gilroy). The second author copied the repository and made changes that he felt would be useful for the repository (in this specific case, an R package). The second author then made a "pull request" so that the first author could integrate the changes. A full description of Git and GitHub is beyond the scope of this chapter, but interested readers are directed to Gilroy and Kaplan (2019) for an introductory tutorial.

Collaborative research is the norm, and most research studies now involve individuals with varying needs to access sensitive data. Additionally, research collaborations are often complex across multiple studies such that each specific collaborator may have unique access requirements for components of the data. Data storage and sharing tools such as Microsoft Sharepoint or RedCap allow for sophisticated permission systems to control these data access hierarchies (see Fig. 9.2). The planning stages of any research project that includes highly sensitive data should include data permission charts and strategies to limit the dissemination of such data as much as possible. Data should be stored in locations that allow for access permissions to be dynamically modified as necessary and should not be shared with mechanisms such as email that do not allow for access permissions to be modified.

Archived data, because it is commonly collected for other purposes, may contain two levels of information: individual raw data and individual summary data. For example, a healthcare dataset may include both lab and vital values (collected at different time points) and diagnosis codes derived from those labs and vitals. As another example, behavior analysts may be analyzing previously collected data obtained from a clinic related to the effects of certain

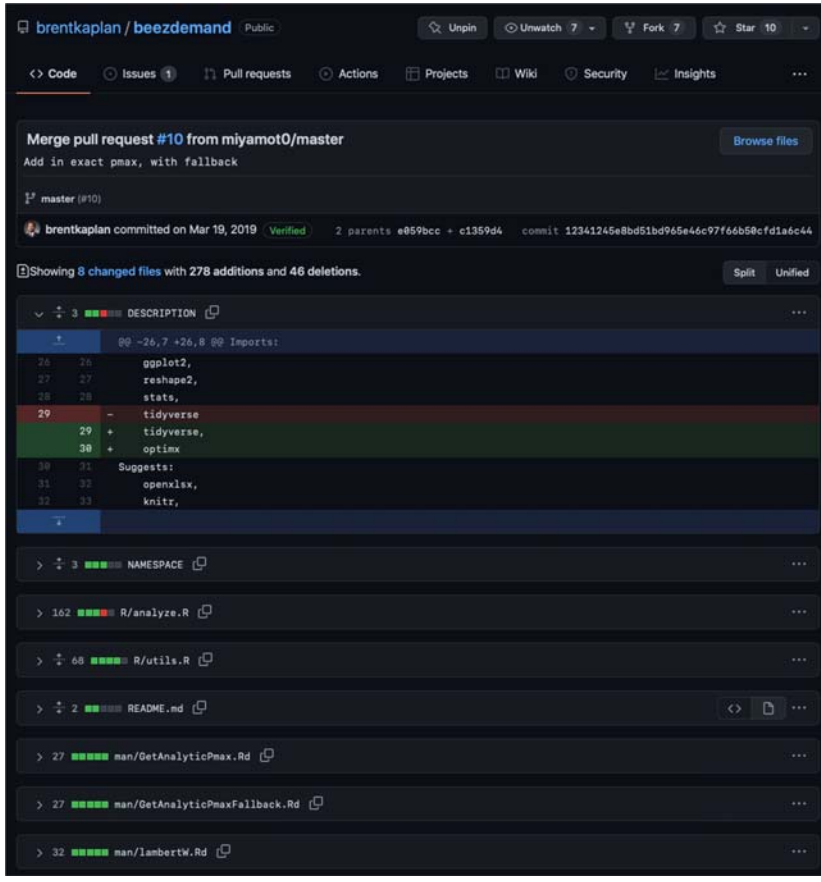

FIGURE 9.3 Snapshot from GitHub showing collaboration and version control benefits. The second author made a pull request wherein he made additions and deletions that would be useful for the repository. The first author accepted these changes and incorporated them (i.e., merge pull request) into the repository. Notice how the version control language documents each addition and subtraction from each of the different files.

interventions. Though the summary data can be useful because others have created those summary values, the researchers may consider re-creating their summary values based on the raw data available. In this sense, the researcher gains additional confidence in the validity of the summary data and can be more comfortable moving forward analyzing that summary data.

Data analysis and dissemination

The act of analyzing data involves numerous different ethical considerations, especially when statistical methods are applied. We will focus broadly on

considerations related to conflicts of interest among those overseeing the data analysis, the use of preregistration prior to conducting the research, and utilizing software that minimizes errors and enhances replicability.

First, anyone who is handling data and oversees carrying out the primary duties related to data analysis should ensure they are free from conflicts of interest. Conflicts of interest are situations in which financial or personal considerations may interfere with or otherwise influence the decisions made regarding or interpretations of the data (Brody, 2011). Even perceived conflicts of interest may inadvertently inject bias in how decisions are made or influence the extent to which the consumer of the analytics trusts the purported results. To deter these potential sources of bias, the individuals tasked with analyzing and validating the data must be free from any conflicts of interest. Many colleges and universities have departments or offices (e.g., Office of Research Integrity and Compliance) that specialize in overseeing and advising potential conflicts of interest, as well as developing management plans for such instances.

Another consideration to minimize ethical complications related to data analysis is to preregister the study (https://www.cos.io/initiatives/prereg). Preregistration can fall on a continuum from precommitting an analytical plan to a time-stamped site (e.g., GitHub) to writing various aspects of the research process including methods, data collection, and data analytic plans and getting the preregistered report reviewed by peers. By preregistering the study and obtaining expert feedback prior to analyzing the data, sources of bias may be identified beforehand, or sources of bias may be minimized or eliminated completely by adhering to the prespecified research plan (Simmons et al., 2021).

Finally, using certain types of software that produce a reproducible record will enhance the ability to verify the data were analyzed in ethical, appropriate, and scientifically defensible ways. For example, open-source statistical programs (or software that support specific statistical packages) such as R (Jamovi is a graphical user interface for easier use of R; see also RStudio, an IDE (RStudio Team, 2020)), SAS, and Python (see SciPy package) provide the ability to document the entire data analytic strategy from the unaltered, raw data to the final products (statistical results, graphs, tables). Different software packages, tools, and tutorials are being created by behavior analysts for behavior analysts (e.g., Gilroy & Kaplan, 2019; Gilroy et al., 2018, 2017; Kaplan et al., 2019), making use of these open-source tools more accessible and appealing.

Considerations in Archival Datasets. Archival datasets can present unique analytic challenges. Multiple steps can be necessary to organize the data for statistical analyses because data were collected from different sources, stored in different locations, and compiled by an analyst who is not part of the research team. Indeed, many archival data researchers report that the bulk of their interaction with the dataset is spent on data "cleaning" or "normalizing."

Data cleaning refers to screening, identifying, and correcting missing or implausible values, and ensuring the data are valid (e.g., data fall within a certain range for a given response), accurate, complete, consistent, and uniform as possible. Data normalizing refers to bringing the variables of interest in proportion to one another so that different variables can be compared (e.g., data can be centered such that the mean is zero and the standard deviation is one). Because archival data often require these complex preparations, proficiency in software that records these analytic steps is essential. These programs often allow for data to be combined and manipulated without altering the original datasets. These programs also document the steps taken and allow for analyses audits (e.g., the ability to trace the steps of the analyses starting from the raw data all the way to the final result) and replications.

Analyzing archival data can require a flexibility that may not be necessary in personally collected datasets, meaning the tools to analyzing these types of data will be largely dependent on the types of data available (as mentioned before, one may have billing codes instead of diagnoses codes). Miscoded values, institutional labeling differences, and missing data are common challenges that the research team may face. Fortunately, archival datasets can be very large (on the order of thousands of entries), which allows the researcher to remove individual data based on a priori exclusionary criteria. However, such large datasets present additional challenges. Because of the possible large sample sizes, obtaining statistical significance (e.g., $P < .05$) but not clinical significance (e.g., a "real-world" outcome that would encourage implantation or intervention; effect size) is a common possibility. Therefore, the calculating and reporting of effect size (e.g., how large the change or effect is, not simply the probability of observing the effect as large or larger if the null hypothesis is correct) are essential to provide the information necessary to evaluate the strength of your findings.

Sharing data with other researchers and the public can accelerate the pace of scientific discoveries by allowing individuals with diverse skills to access valuable datasets and by providing an opportunity for research conclusions to be verified by others. Many funding agencies and journals encourage or require data to be shared publicly or upon request when a manuscript is published in the scientific literature. Typically, an Informed Consent Form wherein the potential participant is informed of their rights, benefits, costs, and other aspects of the study will outline under what conditions participant's data will be used. Potential participants are typically informed as to what happens to their data if they withdraw. Furthermore, Institutional Review Boards (IRB) or other governing boards require researchers to follow guidelines requiring all data to be stored for a specific period of time, after which identifying information should be destroyed. This regulation is outlined in the US Department of Health and Human Services' Code of Federal Regulations Title 45 Part 46 (45CFR46), also known as the "Common Rule." Simultaneously meeting all these requirements and best practices can be greatly facilitated by avoiding the

mixing of highly sensitive data with other study data during the data collection phase, which makes it much easier to create a deidentified dataset.

Depending on the type of information requested and transmitted, the parties may need to engage in a Data Use Agreement (see Appendix for a template Data Use Agreement provided by the National Institutes on Health; https://www.niaid.nih.gov/research/sample-data-sharing-plan). These Data Use Agreements broadly specify the scope of use of the data, who will access the data, safeguards in place to protect the data, etc. However, in cases where only deidentified data are shared, once this dataset is created sharing raw data with other researchers is quick and easy, and the IRB required data management steps years after a study concludes are also straightforward to complete. However, we stress that researchers should ensure the deidentified dataset is truly considered deidentified. In academic institutions, there is often someone in the Office of Research Integrity who may help assist and verify the data are truly deidentified. As we discussed previously, the combination of certain demographic information (e.g., certain full zip codes, specific ages) may be linked together to identify the individual. When sharing datasets that may contain these types of linking information, researchers may use strategies such as binning values (e.g., instead of age 32, this dataset could be coded as between 30 and 34) or perturbing values (e.g., instead of age 32, this dataset could be coded as 34) whereby the qualities of the data (e.g., variance) are maintained.

Finally, researchers and clinicians should be aware of data ownership when considering disseminating results from a study. In academic institutions, research funding is typically awarded to the institution, and so it is the institution that has responsibilities of overseeing activities related to the research, including ensuring the researcher is maintaining records and proper storage. Although institutions differ in their ownership policies, researchers should not assume that they may take the data as they wish when they transfer institutions. Ownership should be clearly articulated prior to data collection, and this is often conveyed either by the funding institution or company or through discussions with the fundee.

Conclusions

In this chapter, we have overviewed different links in the research process—all of which require considerations of proper and ethical data management. We have overviewed different types of data behavior analysts might encounter across a range of settings and disciplines, ranging from clinical work to large, archival datasets. We have discussed considerations in the data collection phase for minimizing errors, and ensuring bias is minimized among those who handle the data. We then discussed considerations and recommendations for ensuring data are analyzed with integrity and that data are valid, such as recognizing potential conflicts of interest and using replicable software. Finally, we discussed how

dissemination efforts can meet the goals of accelerating scientific discoveries while ensuring participant protection. While integrating all the aforementioned recommendations may not be viable for researchers, we encourage researchers to consider the gaps in their research process with respect to data handling. Ensuring policies and procedures are in place for maintaining confidentiality, minimizing errors and bias, and enhancing reproducibility should be a priority in any research laboratory and clinic, and these procedures should be audited and evaluated as frequently as necessary.

Appendix

Example Plan addressing Key Elements for a Data Sharing Plan under NIH Extramural Support

(For questions, contact the NIH Office of Extramural Research (OER), Email Sharing@nih.gov)

Example Data Sharing Plan for FOA-XX-XXXX

What data that will be shared:

I will share phenotypic data associated with the collected samples by depositing these data at _____, which is an NIH-funded repository. Genotype data will be shared by depositing these data at _____. Additional data documentation and deidentified data will be deposited for sharing along with phenotypic data, which includes demographics, family history of XXXXXX disease, and diagnosis, consistent with applicable laws and regulations. I will comply with the NIH GWAS Policy and the funding IC's existing policies on sharing data on XXXXXX disease genetics to include secondary analysis of data resulting from a genome-wide association study through the repository. Meta-analysis data and associated phenotypic data, along with data content, format, and organization, will be available at _____. Submitted data will confirm with relevant data and terminology standards.

Who will have access to the data:

I agree that data will be deposited and made available through _____, which is an NIH-funded repository, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes such as finding genes that contribute to process of XXXXXX. I agree that the names and Institutions of persons either given or denied access to the data, and the bases for such decisions will be summarized in the annual progress report. Meta-analysis data and associated phenotypic data, along with data content, format, and organization, will be made available to investigators through _____.

Where will the data be available:

I agree to deposit and maintain the phenotypic data and secondary analysis of data (if any) at _____, which is an NIH-funded repository and

that the repository has data access policies and procedures consistent with NIH data sharing policies.

When will the data be shared:

I agree to deposit genetic outcome data into _____ repository as soon as possible but no later than within 1 year of the completion of the funded project period for the parent award or upon acceptance of the data for publication, or public disclosure of a submitted patent application, whichever is earlier.

How will researchers locate and access the data:

I agree that I will identify where the data will be available and how to access the data in any publications and presentations that I author or coauthor about these data, as well as acknowledge the repository and funding source in any publications and presentations. As I will be using _____, which is an NIH-funded repository, this repository has policies and procedures in place that will provide data access to qualified researchers, fully consistent with NIH data sharing policies and applicable laws and regulations.

Rev. 20100831

References

Brody, H. (2011). Clarifying conflict of interest. *The American Journal of Bioethics, 11*(1), 23–28. https://doi.org/10.1080/15265161.2010.534530

Bullock, C. E., Fisher, W. W., & Hagopian, L. P. (2017). Description and validation of a computerized behavioral data program: "bDataPro.". *The Behavior Analyst, 40*(1), 275–285. https://doi.org/10.1007/s40614-016-0079-0

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond mechanical turk. *Behavior Research Methods, 51*, 2022–2038. https://doi.org/10.3758/s13428-019-01273-7

European Parliament and Council. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri%BCCELEX:02016R0679-20160504.

Gilroy, S. P., Franck, C. T., & Hantula, D. A. (2017). The discounting model selector: Statistical software for delay discounting applications. *Journal of the Experimental Analysis of Behavior, 107*(3), 388–401. https://doi.org/10.1002/jeab.257

Gilroy, S. P., & Kaplan, B. A. (2019). Furthering open science in behavior analysis: An introduction and tutorial for using GitHub in research. *Perspectives on Behavior Science, 42*(3), 565–581. https://doi.org/10.1007/s40614-019-00202-5

Gilroy, S. P., Kaplan, B. A., Reed, D. D., Koffarnus, M. N., & Hantula, D. A. (2018). The demand curve analyzer: Behavioral economic software for applied research. *Journal of the Experimental Analysis of Behavior, 110*(3), 553–568. https://doi.org/10.1002/jeab.479

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14*(Suppl. 1), 29–37.

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium:

Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Kaplan, B. A., Gilroy, S. P., Reed, D. D., Koffarnus, M. N., & Hursh, S. R. (2019). The R package beezdemand: Behavioral economic easy demand. *Perspectives on Behavior Science, 42*(1), 163−180. https://doi.org/10.1007/s40614-018-00187-7

Kayaalp, M. (2018). Modes of de-identification. In *, 2017. AMIA annual symposium proceedings*. American Medical Informatics Association.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. In *Applied social research methods series*. SAGE Publications.

Mueller, S., Taylor, C. R., & Mueller, B. (2022). Managing change related to consumer privacy laws: Targeting and personal data use in a more regulated environment. In M. Karmasin, S. Diehl, & I. Koinig (Eds.), *Media and change management: Creating a path for new content formats, business models, consumer roles, and business responsibility*. Springer International Publishing. https://doi.org/10.1007/978-3-030-86680-8_15

Nettrour, J. F., Burch, M. B., & Bal, B. S. (2018). Patients, pictures, and privacy: Managing clinical photographs in the smartphone era. *Arthroplasty Today, 5*(1), 57−60. https://doi.org/10.1016/j.artd.2018.10.001

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153−163. https://doi.org/10.1016/j.jesp.2017.01.006

RStudio Team. (2020). *RStudio*. Boston, MA: Integrated Development for R. RStudio, PBC. URL http://www.rstudio.com/.

Sickel, A. E., Seacat, J. D., & Nabors, N. A. (2014). Mental health stigma update: A review of consequences. *Advances in Mental Health, 12*(3), 202−215. https://doi.org/10.1080/18374905.2014.11081898

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration: Why and how. *Journal of Consumer Psychology, 31*(1), 151−162. https://doi.org/10.1002/jcpy.1208

Strickland, J. C., & Stoops, W. W. (2019). The use of crowdsourcing in addiction science research: Amazon mechanical turk. *Experimental and Clinical Psychopharmacology, 27*, 1−18. https://doi.org/10.1037/pha0000235

Talevski, J., Wong Shee, A., Rasmussen, B., Kemp, G., & Beauchamp, A. (2020). Teach-back: A systematic review of implementation and impacts. *PLOS ONE, 15*(4), e0231350. https://doi.org/10.1371/journal.pone.0231350

The Office for Civil Rights (OCR), & Malin, B. (2012). *Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule*. Health Information Privacy.

United States. (1974). *Family Education Rights and Privacy Act (FERPA)*, 20 U.S.C. § 1232g; 34 CFR Part 99.

United States. (1996). *The Health Insurance Portability and Accountability Act (HIPAA)*. Washington, DC: U.S. Department of Labor, Employee Benefits Security Administration.

United States. (2009). *Health Information Technology for Economic and Clinical Health Act. TITLE XIII—Health Information Technology. American Recovery and Reinvestment Act of 2009* (ARRA) (Pub.L. 111-5). Retrieved from https://www.congress.gov/bill/111th-congress/house-bill/1.