PART 3

PROTECTING PRIVACY AND CONFIDENTIALITY

QUESTION #18

What Is Meant by "Privacy" and "Confidentiality," and Is There a Difference?

Privacy and confidentiality are two critical concepts that all researchers must address when designing and implementing research. Often assumed to have the same meaning, privacy and confidentiality are, in fact, two discrete but related concepts. An easy way to distinguish the two is to think of privacy as protecting individuals and confidentiality as protecting information—or data—that people share with researchers.

Privacy can be defined as having control over oneself—that is, people can choose when to share information about themselves and with whom. During recruitment, you can protect the privacy of prospective participants by implementing procedures that do not disclose information to others that would identify prospective participants as being part of a specific group, engaging in a specific behavior, or having a specific health condition. During data collection, you can reduce the likelihood of a violation of privacy by implementing procedures that allow participants to share their information with researchers where others cannot hear or see them.

When participants privately share their information with researchers, they expect that their information will remain confidential—that is, they expect that only the research team and other authorized individuals will have access to their data. In a practical sense, confidentiality refers to the specific steps researchers implement to keep information about participants unknown to others, to the extent possible. Federal research regulations require researchers to establish procedures to protect the confidentiality of information that is individually identifiable (meaning, the participant can be identified directly by the researcher or through identifiers that are linked with the data). However, researchers often implement the same confidentiality procedures for all types of data, as they are good standard research practices.

More questions? See #21, #22, and #23.

QUESTION #19

What Makes Data De-Identified?

Datasets that have been stripped of all personal identifiers are considered to be de-identified. The federal research regulations do not list specific personal identifiers. Instead, they loosely define identifiable to mean that "the identity of the subject is or may readily be ascertained by the investigator or associated with the information" (45 C.F.R. § 46.102(e)(5)). Although a universal list of personal identifiers does not exist, the 18 identifiers listed in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (such as participant name and date of birth) are reasonable identifiers that researchers should consider removing from their datasets when de-identifying them (USHHS, 2015a, 2015b).

The premise of de-identifying datasets is that by removing all personal identifiers, the participants' identities likely cannot be determined by those who see the data. Even after datasets are de-identified, however, a slight risk remains that participants can be re-identified, if someone had the interest in and means to do so. Therefore, researchers and ethicists debate the extent to which data can truly be de-identified.

Typically researchers must de-identify their datasets when they plan to share them with researchers outside the original study team (such as for secondary data analysis), when the data are to be made publicly available, or when they prepare data for long-term storage. Adequately de-identifying datasets may take considerable effort, depending on the type of information collected. Numerous procedures exist for removing or masking identifiers in *quantitative* datasets. For example, a specific process is required for removing all HIPAA identifiers from quantitative datasets in research that must follow the HIPAA Privacy Rule (USHHS, 2015a, 2015b).

Processes for de-identifying *qualitative* data are not as straightforward. Overall, it is very difficult to de-identify qualitative data. Researchers typically modify easily-identifiable data in interview transcripts. For example, proper names said by the participant, such as "my friend Bob," are removed and replaced with a general description ("my friend") or a pseudonym. However, that step alone likely does not make qualitative data de-identified. Larger segments, including very specific or unusual experiences, may need to be redacted from transcripts to preserve participants'

identities. Social and behavioral scientists must therefore be mindful of the quality of their data—both quantitative and qualitative—if a large amount of stripping must be done to de-identify them, and whether the necessary context will still remain to allow for valid interpretations to be made by others.

When de-identifying data for sharing or storage, the master list linking personal identifiers to the study data does not necessarily have to be destroyed. Institutional review boards often allow the original researcher to maintain the master list that links the participants' names to their identification numbers, but that list must be stored securely and not shared.

More questions? See #18, #20, and #24.

QUESTION #20

What Makes Data Anonymous?

Data are anonymous when they are not linked to any participant identifiers. In other words, the identity of a participant cannot be determined through his or her data. If the data are truly anonymous, even the study team cannot determine participants' identities. Researchers often choose to collect data anonymously for studies on stigmatized or illegal behaviors. Then, if unauthorized persons gain access to the data—or if the data were purposefully shared with other researchers for secondary analyses—participants' identities could not be detected because identifying information was never collected or known by the researchers at all. Importantly, data do not need to be anonymous to be considered ethical; employing secure procedures for limiting a confidentiality breach of identifiable data is ethically sufficient. Only in certain situations where extra protections are needed is it preferable to collect data anonymously. However, some researchers—regardless of whether the research topic is sensitive or not—choose to collect data anonymously for a study because they do not need participants' identifiers to answer their research questions.

If you want to collect data anonymously, you must consider several factors. First, your study design matters. Collecting anonymous data is likely an unrealistic option for research that requires data to be linked from multiple interactions with the same participant, such as in longitudinal research. In these situations, researchers should keep a master list linking participant names to their participant identification numbers so they can ensure that data are collected from the same participant at each time point. Participants can therefore be identified by anyone who has access to this list. Researchers who want their data to be anonymous should consider employing a study design that has only a single interaction with participants, such as a one-time interview or survey.

Second, consider how you are going to collect data. Collecting anonymous data is not possible when you (or another member of the research team) meet in person with a participant to conduct an interview or survey, for example. By conducting a face-to-face interview, you know what the person looks like and therefore can identify him or her, even if you do not know the participant's name or have any other identifiers. Depending on

the topic of the study, being identified as a participant (even without any disclosure of information discussed) could be potentially stigmatizing. For similar reasons, if a researcher wants to collect data anonymously, participants can neither be video recorded nor have their pictures taken because they can be visually identified. Data from audio-recorded interviews are also not considered anonymous because participants' voices are unique, like fingerprints, and therefore considered identifiable.

Third, consider the kinds of data to be collected. For data to be anonymous, you cannot collect *any* information that can identify a participant. This information includes, for example, participant names, email addresses, and date of birth. Eighteen specific identifiers are listed in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (USHHS, 2015b). The Family Educational Rights and Privacy Act Regulations (FERPA) also provide a list of direct and indirect identifiers (USDE, 2017). Some identifiers may be unique to a particular study participant, such as a description of a tattoo, and when known, could identify the participant. A combination of identifiers when viewed together, such as ethnicity, sexual orientation, and age, could also reveal the identity of a study participant in some situations, especially when research is conducted in small towns or communities.

Data that were originally collected with personal identifiers can become anonymous data, in theory, if all personal identifiers are removed from the data *and any documents linking identities and data are destroyed*.

Ultimately, it may be difficult to collect data that are truly anonymous. Often researchers want to know identifying information to provide context to the data or to maintain long-term contact with participants. If you need to collect participant identifiers but are concerned about the negative implications of others potentially discovering the identities of participants in your research study, use strict procedures to protect the confidentiality of study data and consider obtaining verbal consent, so that participants' names are not linked to the study though their signature on a consent form.

More questions? See #19, #23, and #25.

QUESTION #21

When Is Information (or Behavior) Considered Private Rather Than Public, and How Can Private Information Be Used in Research?

When information is provided in a private place, such as when a patient tells his doctor that he is feeling depressed during a medical checkup, the patient has a reasonable expectation that the information will remain private and will not be used for other purposes, such as for research. The federal research regulations state that "[p]rivate information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public (e.g., a medical record)" (45 C.F.R. § 46.102(e)(4)). When information is provided in a public space—especially nonsensitive information, such as when a person comments on an online news story—people's expectations of privacy are typically much lower.

Other situations, however, are not as straightforward, particularly when people share information they perceive to be private in public places. For instance, is it ethical for researchers to listen to and use for research purposes a conversation between two friends at a coffee shop, without their consent? Even though the individuals are communicating in a public space, they may believe that their conversation is private and have a perception that this information will remain private—that is, no one else is listening and documenting what they are saying. As a researcher, a good question to ask yourself is: Would these individuals feel that their privacy has been violated if they learned a researcher was using what they said for research purposes?

In examining these situations through the lens of the federal research regulations, conversations between patients and their doctors are considered private, and researchers are not allowed to use that information as data for their study, unless they obtain the patient's consent or do not

collect any identifiable information about the patient. Similarly, a conversation between two people at a coffee shop would be considered private information, because the people would reasonably expect that their conversation would not be recorded, made public, or included in a research study without their knowledge. However, researchers could still observe and analyze this information, as long as they do not link any identifiable information with the conversation data. Comments made online in a public forum and public speeches would be considered public information and generally can be used for research purposes without obtaining informed consent.

More questions? See #60, #70, and #72.

QUESTION #22

What Can I Do to Protect Participants' Privacy During Data Collection and Reporting?

During data collection, you must limit the possibility that others will see participants taking part in research activities or will hear the information that they are sharing in an interview. As a general rule, all research interactions should be conducted in a private location where the conversation cannot be seen or heard by others. However, it is often desirable to hold study interviews in a neutral location, such as a public library. This is acceptable as long as privacy is maintained. While infrequent, participants may want someone else such as a spouse or other family member present during their interview. From a privacy perspective, this is allowed if it is requested by the participant. However, you will need to consider other factors: How sensitive are the questions being asked? Will the other person share information discussed in the interview with others? Is it possible that the participant may be less truthful in front of this other person? If the answer to any of these questions is "yes," then it may be best not to include this person in your research.

When reporting your research findings, data must be presented in a way that prevents individual participants from being identified. This is especially important in qualitative research where participant quotes are typically provided to illustrate an aspect of the data. While it is ethically acceptable to include participants' demographics with the quotes, such as the participant's gender and age, you must evaluate whether the context provided in the quote combined with demographic data could potentially identify the participant. Additionally, it may be difficult to conceal participants' identities when the research is conducted in small towns or communities. In some situations, it may also be difficult to conceal the geographic location of a study population when the authors' affiliations are included with the manuscript or are easily searchable online. You should consider these factors when deciding what information about participants—and which parts of quotes—to include when describing the findings from qualitative research.

More questions? See #56, #59, and #72.

QUESTION #23

What Can I Do to Protect the Confidentiality of Information Collected?

Researchers must implement procedures to limit the likelihood that people outside the study team can gain access to information shared by participants during research. Confidentiality of participant data can never be fully protected or guaranteed, as unanticipated breaches can occur due to human error. A study computer may be stolen, data might be stored on servers that are not secure or password protected, field notes may be lost, or transcripts might be left on the data analyst's desk at the end of the day. These errors allow unauthorized people to gain access to and read participant information.

Fortunately, there are several basic steps that can be taken to substantially limit the likelihood of a breach in confidentiality:

- Use participant identification numbers instead of participant names on all hard copy and electronic study documents, including surveys, field notes, photographs, and audio and video recordings.
- Collect only those personal identifiers that are absolutely necessary. If risk of a breach of data would pose significant harm to participants, consider collecting no personal identifiers.
- Breaches of confidentiality may happen during transport. After collecting data in the field, return to the study office immediately with any completed questionnaires, field notes, and recording devices to appropriately log and store them. If possible, upload the audio recordings to a secure location in the cloud prior to leaving the data collection site. Erase interviews from recording devices as soon as audio files are stored on a secure server.
- Password protect all electronic document files, and store them on secure servers or password-protected computers.
- Store all hard copy research records—such as handwritten interview notes and printed transcripts—in locked cabinets.

- Avoid storing any research records on portable USB flash drives. If such storage is temporarily necessary, the records should be copied to a secure server as soon as possible and deleted from the less secure temporary storage devices.
- Limit access to study files to essential study staff.
- Keep signed consent forms and other documents that include participant names, such as master participant lists and contact information sheets, separate from documents containing participant data. Hard copies of these files should be stored in separate, locked cabinets. Electronic files should be kept in separate electronic folders and have different passwords; for example, the same password should not be used for interview transcripts and the master participant list.
- If re-identification will not be necessary, destroy all documents that would allow for the re-identification of participants, such as the master list of participant identifiers, as soon as possible after research is complete.

More questions? See #25, #88, and #99.

QUESTION #24

When Must I Share Study Data—and Participant Names—With Individuals Outside of the Research Team?

While uncommon in the social and behavioral sciences, there are circumstances in which people outside of the study team may need to look at your research data. For example, an institutional review board (IRB) may need to review study data to investigate participant complaints of mistreatment or accusations of data falsification/fabrication. Funding agencies are also typically allowed to view study data if deemed necessary. To inform prospective participants that people outside of the study team may have access to their data, researchers often disclose, during the informed consent process, all the possible groups that may have access to research data. This way, participants are fully informed of who may see their information when they make their decision about research participation.

You may be required to share research data—and participant names—with individuals outside of the research team when it is necessary to ensure the safety of the participant or others. State laws vary, but most require certain individuals, referred to as "mandatory reporters," such as medical doctors, nurses, teachers, guidance counselors, and social workers, to share any disclosure of child or elder abuse or neglect with authorities. In these situations, a breach in confidentiality and privacy may be necessary to protect participant safety or the safety of others.

If it is possible that participants may disclose information about their own safety or the safety of others during the course of your research, you should investigate your state laws, engage your local IRB, and contact legal counsel at your institution. You'll need to determine if you are considered a mandatory reporter and to ensure that your research follows all applicable laws and professional ethical requirements. When conducting research on a topic that could lead to the disclosure of

reportable information, prospective participants must be informed during the consent process that any information they share about potential harm to themselves or others will be reported to the authorities, and that this could result in legal action against them.

More questions? See #25, #93, and #94.

QUESTION #25

What Is a Certificate of Confidentiality?

A Certificate of Confidentiality (CoC) protects researchers from forced disclosure of participant data—including participants' identities—to local, state, and federal authorities. Researchers who have a CoC can refuse to provide information about participants when solicited by the authorities. Without a CoC in place, researchers may be legally required to disclose participants' data and identities, for example, if their data are subpoenaed for a civil or criminal case.

Regardless of who is conducting or funding the research, CoCs are issued by the National Institutes of Health (NIH, 2016) and other Department of Health and Human Services (HHS) agencies to the researcher's institution. NIH automatically issues CoCs for NIH-funded research that collects or uses identifiable, sensitive information. NIH describes "identifiable, sensitive information" to mean "information about an individual that is gathered or used during the course of biomedical, behavioral, clinical, or other research, where the following may occur: an individual is identified; or for which there is at least a very small risk, that some combination of the information, a request for the information, and other available data sources could be used to deduce the identity of an individual" (301(d) Public Health Service Act (42 U.S.C 241)). Researchers with other sources of funding may request a CoC from the NIH when they collect identifiable, sensitive information as part of health-related research.

Importantly, having a CoC in place does not mean that researchers may never share information about participants with authorities. A researcher may need to contact authorities when a participant discloses plans to harm someone, for example. A CoC prevents only involuntary and forced disclosure of research information by researchers.

More questions? See #12, #23, and #98.

QUESTION #26

What Privacy Laws Must I Follow?

If you plan to conduct research in U.S. public schools, you should become familiar with the Federal Education Rights and Privacy Act (FERPA). This federal law "affords parents the right to have access to their children's education records, the right to seek to have the records amended, and the right to have some control over the disclosure of personally identifiable information from the education records" (USDE, 2017). It outlines rules regarding informed consent that must be followed when accessing student education records for research purposes. With some exceptions, parental consent is required for researchers to access records and collect identifiable information of students who are under the age of 18 (or the age of majority in the specific state). Parental consent may not be required for collecting nonidentifiable data from student records.

If you plan to conduct research in medical settings or with patients, you should become familiar with the Health Insurance Portability and Accountability Act (HIPAA; USDHHS, 2015a, 2015b). The purpose of the HIPAA Privacy Rule is to protect information in individuals' medical records (electronic or other) as well as other personal health information provided by individuals when engaged in or paying for health care. Often information that was originally collected for health care purposes can be used to answer social and behavioral research questions. For example, if you want to analyze existing medical records data to examine the association between heart disease and mental health, or to identify patients with heart disease to participate in interviews on nutrition, you will need to follow the HIPAA Privacy Rule. If you plan to conduct research with patients or by using medical records, check with your institution to determine how to comply with the HIPAA Privacy Rule.

More questions? See #33, #52, and #53.